

Neural and perceptual correlates of individual gaze behavior in complex scenes

Synopsis zur kumulativen Dissertation

Zur Erlangung des Doktorgrades der Naturwissenschaften

– *Dr. rer. nat.* –

Vorgelegt von

M. Sc., Diana Kollenda

Betreuer und Erstgutachter: Prof. Dr. Benjamin de Haas

Zweitbetreuerin: Prof. Dr. Jutta Billino

Giessen, Mai 2026

I collected some inspiring quotes related to my studies:

"You and I have more or less the same brain maps [...]. But universality and uniqueness are close companions throughout our biological world. For instance, fingerprints are universal to humans – we all have them and yet the idiosyncrasies of individual fingerprints make each one unique. The same is true for brain maps, except that their idiosyncrasies matter. They affect a person's or a creature's capacity to perceive, attend, remember, recognize, and react to the wider world."

– Schwarzlose, R. (2021). *Brainscapes: The Warped, Wondrous Maps Written in Your Brain-and how They Guide You*. Mariner Books.

"Gaze behavior is the thread used to weave the objects, actions, and events of a scene into the fabric of a story."

and

"Scene understanding is ultimately an interpretation, it is whatever a person tells you it is — the specific story they have attached to a given collection of pixels."

– Zelinsky, G. J. (2013). Understanding scene understanding. *Frontiers in psychology*, 4, 954.

Danksagung

Mit genau diesem Teil habe ich die Ausarbeitung meiner Thesis angefangen und sie auch beendet. Zwischendurch habe ich immer mal wieder dran gedacht, weil mir bewusst wurde, wie viele Menschen mich unterstützt haben, etwa meine Freund:innen und die großartigen Wissenschaftler:innen die ich unterwegs kennenlernen durfte. Da eine vollständige Liste den Rahmen sprengen würde, möchte ich mich an dieser Stelle auf jene konzentrieren, die dieses Projekt ganz besonders geprägt haben.

Ben, danke, dass Du mir die Freiheit gegeben hast, meine vielfältigen Interessen zu verfolgen, und mir von Anfang an dein Vertrauen geschenkt hast. Ich erinnere mich noch sehr lebhaft an unser erstes Treffen, als ich noch im Master war: Du hattest gerade deinen ERC Starting Grant erhalten und warst auf der Suche nach Doktorand:innen. Nie zuvor hatte mich jemand so mit seiner Begeisterung für Wissenschaft und seiner offenen Unterstützung beeindruckt – und dass, obwohl ich damals noch eine völlig fremde war. Ich hatte mir damals vorausschauend notiert: „Ich glaube Ben wäre wirklich ein cooler Chef, von dem man viel lernen könnte.“ Und das sollte sich bewahrheiten. Im Laufe der Jahre hast Du es geschafft, eine harmonische und zugleich unglaublich unterhaltsame Arbeits- und Lernatmosphäre zu schaffen, immer offen für neue Herausforderungen; Plank-Challenges eingeschlossen.

Petra, Marcel und Max, danke, dass ihr diese PhD-Zeit zu etwas so Schöнем und Unvergesslichem gemacht habt. Ich hätte mir nie ein so vertrautes, unkompliziertes und zugleich produktives Arbeitsumfeld erträumt. Ich werde euch und all die gemeinsamen Momente in bester Erinnerung behalten und bin mir sicher, dass euch noch viele spannende Abenteuer bevorstehen. Ich hoffe sehr auf eine baldige Reunion mit allen von euch. Großer Dank geht auch an **Sophie und Malin** für durchweg emotionalen Support.

Meine Familie, ihr wart immer eine große Unterstützung und nach Dortmund kommen fühlte sich oft an wie ein kurzer Erholungsurlaub. Ihr seid wunderbare Menschen und ich liebe euch von ganzen Herzen. **Papa**, was ich sehr an dir schätze, ist deine Gelassenheit. In Momenten, in denen ich ungeduldig wurde oder zweifelte, hat mir deine Perspektive geholfen, kurz innezuhalten. **Mama**, deine endlose Neugier und Weltoffenheit sind ansteckend und in dieser Hinsicht bist du meine Muse. **Mike**, mein Bruderherz, du tust wofür dein Herz brennt und das ist unglaublich inspirierend.

Und am allerwichtigsten: **Sascha**, du bist mein Herz und mein Halt. Du hast immer an mich geglaubt, mich unterstützt, und ganz ohne Worte ein Gespür dafür bewiesen, wie es mir gerade wirklich geht. Du hast mich mit deinem klugen Kopf inspiriert, aus meiner Komfortzone gelockt, zum Lachen gebracht, zum Sport motiviert, und damit für die so wichtige Balance gesorgt.

Summary

Scene perception is not a passive registration of visual input but an active, selective process through which observers explore and make sense of complex visual environments. Interestingly, people differ systematically in where they look when viewing the same scene. For example, some people fixate faces twice as much as others, while others fixate text much more than others do. This thesis investigates how such systematic individual differences during scene viewing relate to neural organization in the ventral stream and to perceptual consequences. Across two studies, we show that individual differences in the tendency to prioritize faces and text are systematically linked to category-selective brain representations and how observers describe scenes.

The first study examines how individual fixation preferences for faces and text relate to category-selective responses in the ventral stream. We show that individual tendencies to fixate faces and text are related to representational distinctiveness of distributed responses for these categories. This link is especially evident for first fixation tendencies, which also correlate with the size of face- and word-selective regions. Additionally, the distinctiveness of category-specific neural responses is linked to performance in corresponding behavioral tasks. These associations suggest that active vision is linked to category-selectivity in the ventral stream and preferentially targets object categories that are distinctively represented in the individual brain.

The second study examines whether such gaze biases translate into systematic differences in perception. We find that observers who look more similar at scene elements also tend to produce more similar descriptions, especially in their noun usage. This link generalizes across independent scene sets. Moreover, the tendency to dwell on people or text correlates with how often these categories are referenced in descriptions. We observed anti-correlations in the prioritization of people versus text in gaze behavior as well as in descriptions. These observations support a trait-like coupling between where individuals look and how they describe scenes.

Taken together, our findings outline a systematic pattern of covariation between active vision, individual neural organization, behavioral task performance, and perceptual interpretations of scenes. Idiosyncratic category-selective representations in the ventral stream are associated with stable gaze priorities for socially and culturally relevant categories such as faces and text. These priorities appear to be related to the perceptual information that is ultimately sampled and included in descriptions. This thesis therefore proposes that treating individual variability as meaningful signal, rather than noise, is crucial for a more comprehensive understanding of human visual processing and perception.

Zusammenfassung

Die Wahrnehmung von Szenen ist keine passive Registrierung visueller Reize, sondern ein aktiver, selektiver Prozess, durch den Beobachter komplexe visuelle Umgebungen erkunden und interpretieren. Interessanterweise unterscheiden sich Personen systematisch darin, wohin sie blicken, wenn sie dieselbe Szene betrachten. So fixieren manche Personen beispielsweise Gesichter doppelt so häufig wie andere, während wiederum andere deutlich häufiger auf Text schauen. Diese Dissertation untersucht, wie solche systematischen individuellen Unterschiede während der Szenenbetrachtung mit der neuronalen Organisation im ventralen visuellen Verarbeitungspfad und mit perzeptuellen Konsequenzen zusammenhängen. In zwei Studien zeigen wir, dass individuelle Unterschiede in der Tendenz, Gesichter und Text zu fixieren, systematisch mit kategorieselektiven Hirnrepräsentationen sowie mit der Art und Weise zusammenhängen, wie Versuchspersonen Szenen beschreiben.

Die erste Studie untersucht den Zusammenhang zwischen individuellen Fixationspräferenzen für Gesichter und Text und kategorieselektiven Aktivitäten im ventralen Pfad. Wir zeigen, dass die individuellen Tendenzen, Gesichter und Text zu fixieren, mit der Distinktheit verteilter neuronaler Aktivitätsmuster für diese Kategorien zusammenhängen. Dieser Zusammenhang ist besonders ausgeprägt für Erstfixationen, die zudem mit der Größe gesichts- und wortselektiver Areale korrelieren. Darüber hinaus hängt die Distinktheit kategoriespezifischer neuronaler Aktivitätsmuster mit der Leistung in entsprechenden Verhaltensaufgaben zusammen. Diese Befunde legen nahe, dass aktives Sehen mit der Selektivität für bestimmte Kategorien im ventralen Pfad verknüpft ist und bevorzugt auf Objektkategorien gerichtet ist, die im individuellen Gehirn besonders distinkt repräsentiert sind.

Die zweite Studie untersucht, ob solche systematischen Unterschiede in Blickpräferenzen mit Unterschieden in der perzeptuellen Wahrnehmung einhergehen. Wir zeigen, dass Versuchspersonen mit ähnlichen Blickmustern auch dazu tendieren, ähnlichere Beschreibungen zu produzieren, insbesondere in ihrer Substantivverwendung. Dieser Zusammenhang generalisiert über unabhängige Szenen hinweg. Darüber hinaus korreliert die Tendenz, sich auf Personen oder Text zu konzentrieren, mit der Häufigkeit, mit der diese Kategorien in Beschreibungen erwähnt werden. Zudem zeigen sich negative Korrelationen in der Priorisierung von Personen versus Text sowohl im Blickverhalten als auch in den Beschreibungen. Diese Befunde sprechen für eine stabile Kopplung zwischen Blickverhalten und Szenenbeschreibung.

Zusammenfassend zeigen die Ergebnisse eine systematische Kovariation zwischen aktivem Sehen, individueller neuronaler Organisation, Verhaltensleistung und perzeptuellen Interpretationen von Szenen. Idiosynkratische kategorieselektive Repräsentationen im ventralen Pfad gehen mit stabilen Blickprioritäten für sozial und kulturell relevante Kategorien wie Gesichter und Text einher, und diese scheinen mit der perzeptuellen Information zusammenzuhängen, die in Beschreibungen integriert wird. Diese Dissertation plädiert daher dafür, individuelle Variabilität nicht als Rauschen, sondern als bedeutungsvolles Signal zu betrachten, um ein umfassenderes Verständnis der menschlichen visuellen Verarbeitung und Wahrnehmung zu ermöglichen.

Contents

I. Synopsis.....	1
1. Introduction	1
1.1. Gaze allocation in complex scenes	2
1.2 Neural and genetic correlates of gaze and behavior.....	3
1.3 Perceptual and behavioral correlates of gaze	5
1.4 Rationale of the present studies	6
2. Study 1: Active vision is linked to category selectivity in the individual brain.....	7
3. Study 2: Individual gaze predicts individual scene descriptions	9
4. Discussion.....	11
4.1 Linking active vision, neural category-selectivity, and perception	11
4.2. Broader aspects and implications	15
4.3. Limitations and further directions.....	17
4.4 Conclusion	20
5 References.....	21
II. Publications.....	38
Study 1: Active vision is linked to category selectivity in the individual brain	39
Study 2: Individual gaze predicts individual scene descriptions	95
3. List of all publications	112
4. Selbstständigkeitserklärung	113

I. Synopsis

1. Introduction

Imagine sitting on a train and watching a scene unfold on the platform outside: a woman stands beside a man and a child. They look tense. You may conclude this ‘family is worried about a delayed train.’ But then you notice something you had not seen before: a name tag on the woman’s coat that reads “Conductor.” Suddenly, the meaning shifts to ‘a conductor has stopped a father and son for riding without tickets.’ This example stresses two things: first, the prevalence of faces and text in our everyday life and second, that in just a few seconds, the perception of the same scene changes as the eyes move from one detail to another. That is, visual understanding of scenes can be rapidly informed by global contextual information (Torralba et al., 2006), but is further refined through successive gaze shifts that continuously integrate newly acquired details into a more elaborated scene representation (cf. Henderson & Castelhana, 2005).

Although the human visual field spans roughly 180 degrees, shifting fixation is essential because the uncrowded, high-acuity vision needed for detailed object recognition is concentrated in the fovea, the center of gaze (see Figure 1b for an illustration of this, where only the face of the woman is clearly recognizable). The fovea spans only about 1.7 degrees of visual angle, and visual information declines sharply toward the periphery (Rosenholtz, 2016; Strasburger et al., 2011; Watson, 2014). This gradient reflects the nonuniform distribution of cone photoreceptors and retinal ganglion cells, whose densities decrease progressively away from the fovea (Curcio et al., 1990; Curcio & Allen, 1990). Foveal-peripheral asymmetries are further evident in cortical magnification in the primary visual cortex, where foveal input occupies a disproportionately large share of cortical area relative to both peripheral input and the fovea’s tiny retinal surface area (Daniel & Whitteridge, 1961; Wandell et al., 2007).



Figure 1. (a) Scene generated by ChatGPT 5.2 based on the following prompt: “Create an image depicting the following scene unfolding outside a train window: a woman stands beside a man and a child. They look tense. The woman is wearing a name tag on the coat that reads ‘Conductor’, from which one concludes that she is a conductor that has stopped a father and son for riding without tickets.” (b) Illustration of foveal high-acuity vision versus blurred periphery (Gaussian-convolved to simulate reduced acuity). This manipulation is not intended as a biologically precise model of human vision but serves illustrative purposes only.

The disparity between foveal and peripheral acuity represents an evolutionary compromise, balancing head size and energy demands against the needs for both fine-detail processing and broad spatial navigation and

detection (Vater et al., 2022). This resource allocation is highly efficient, because our eyes can move quickly and perform about three gaze shifts per second that bring different parts of a scene to the fovea. However, this dynamic interplay between fixations and eye movements creates a complex challenge: the visual system must constantly decide *where* and *what* to look at next based on information from the lower-resolution periphery. Understanding how these moment-to-moment fixation decisions are informed has long been, at least since Buswell (1935) and Yarbus (1967), a central question in vision science.

1.1. Gaze allocation in complex scenes

In the search for an answer to what explains gaze allocation, early research has focused primarily on commonalities in fixation location and timing across different observers, seeking image features that reliably attract attention (i.e., are salient and thus stand out relative to neighboring image content; Parkhurst et al., 2002). This led to the development of influential saliency models (cf. Itti et al., 1998; Harel et al., 2006), which were further inspired by the early visual system and assume that gaze is automatically captured by low-level image features. Within this framework, fixation allocation is driven by regions with high local contrast in color, intensity, or orientation (Itti and Koch, 2001).

However, low-level saliency is only one piece of the puzzle for explaining gaze in complex, semantically rich scenes. Observers are generally drawn to objects in scenes (Einhäuser et al., 2008b). Moreover, local informativeness (i.e., the judged isolated meaning of scene regions) can guide gaze as early as the first fixation (Henderson et al., 2018; Henderson & Hayes, 2017; Peacock et al., 2019). In particular, high-level semantic categories such as faces and text, which carry high behavioral and informational relevance in complex scenes, are relatively strong gaze attractors. They account for gaze allocation more effectively than low-level features alone (Cerf et al., 2009; Wang & Pomplun, 2012) and basic object properties such as size and eccentricity (Xu et al., 2014).

Additionally, top-down cognitive factors, such as task demands, goals, emotions, expectations and prior knowledge, strongly bias gaze toward behaviorally relevant regions (Buswell, 1935; Einhäuser et al., 2008; Yarbus, 1967). For example, when instructed to search for a target object, semantic priors about typical object locations (e.g., pillows on bed) can guide gaze, even when the target (e.g., pillow) is not visually salient (Henderson et al., 2009) or entirely absent (Võ et al., 2019). During free viewing, when no explicit task is assigned, as well as in a paradigm combining free viewing with subsequent scene descriptions, recent work by Murlidaran & Eckstein (2025) suggests that observers prioritize scene elements that support the goal of scene understanding. In sum, the aforementioned factors account for a substantial proportion of variance in gaze behavior across observers.

However, these factors typically treat stable between-observer differences as unexplained noise, even though incorporating such individual differences has been shown to improve gaze prediction beyond generic, group-based low- and high-level salience models (Li & Chen, 2018). A growing body of research shows that healthy adult observers exhibit systematic, trait-like gaze tendencies that govern how frequently and when they

sample semantically meaningful elements of complex scenes. In particular, de Haas et al. (2019) reported highly consistent and temporally stable tendencies to fixate specific semantic object categories such as faces, text, implied motion, taste, touched objects, and emotional content (see also Linka & de Haas, 2020). Interestingly, although gaze patterns becomes more similar by adulthood (Linka et al., 2025), stable and systematic individual differences in gaze toward these semantic categories persist (de Haas et al., 2019; for faces, see Guy et al., 2019).

Particularly, the tendency to direct first fixations after image onset toward a face or text as well as the overall proportion of fixation durations on each category (i.e., cumulative dwell time), vary substantially between participants and constitute robust individual-difference measures (de Haas et al., 2019; Linka & de Haas, 2020). Some observers tend to (first) fixate faces up to twice as much as others across images, whereas other observers show a stronger tendency to (first) fixate text. Thus, although faces, and to a lesser extent, text, are strong attractors on average (Cerf et al., 2009; Xu et al., 2014), their salience is not uniform across observers. Importantly, these tendencies vary along a continuum (from weaker to stronger gaze tendencies for faces and text, respectively) and are defined relative to the distribution within the tested sample. Interestingly, despite faces being fixated more frequently overall, individual tendencies to fixate faces and text tend to be anti-correlated (de Haas et al., 2019), suggesting a competitive relationship for attention. Because both categories require high foveal resolution, observers may prioritize one over the other, especially under limited viewing time. Accordingly, in the opening example, some observers may preferentially look (first) at the faces of the father, son, or conductor, whereas others may instead prioritize the name tag.

Together, the outlined literature suggests that spontaneous gaze allocation in complex scenes is driven by low-level, semantic, and contextual image features, providing a robust predictive baseline. However, such image-based approaches overlook the systematic inter-individual variability in gaze behavior. Crucially, studying these idiosyncratic gaze tendencies is essential to better understand which mechanisms mediate an observer's particular view of the world. An important, yet unresolved question in this context is whether idiosyncratic gaze tendencies are linked to robust differences in the way the visual system represents behaviorally relevant categories such as faces and text.

1.2 Neural and genetic correlates of gaze and behavior

A potential neural correlate for stable gaze tendencies towards faces and text (i.e., words) lies in category-selective visual cortex, particularly the ventral temporal cortex (VTC) (Haxby et al., 2001; Kanwisher, 2010; Peelen & Downing, 2005). Within VTC, faces and pseudowords (hereafter 'word') are among the strongest drivers of distributed neural responses (Haxby et al., 2001) and of highly category-selective regions such as the fusiform face area (FFA) and visual word form area (VWFA), respectively (Dehaene & Cohen, 2011; Grill-Spector & Weiner, 2014; Kanwisher & Yovel, 2006). Consistent with the high-acuity demands of both categories, these regions show a foveal bias, preferring stimuli typically processed around the center of the visual field (Hasson et al., 2002; Levy et al., 2001; Malach et al., 2002). Moreover, faces and words also differ

in hemispheric lateralization: face-selectivity is typically right-dominant, whereas word-selectivity is left-dominant (e.g., Blauch et al., 2025). This may reflect their respective connections to broader lateralized networks: language in the left hemisphere (Saygin et al., 2016) and social cognition in the right (Powell et al., 2018).

The functional importance of VTC is especially evident in acquired visual agnosia, where category recognition is impaired despite intact basic vision (Farah, 2004). Notable subtypes of this condition include prosopagnosia (face recognition deficits) and pure alexia (word recognition deficits; cf. Kleinschmidt & Cohen, 2006). Furthermore, causal evidence for this functional specialization is provided by intracranial stimulation of a site within the right FFA, which acutely distorts face perception (Parvizi et al., 2012). Moreover, individual differences in the size of the right FFA and left VWFA correlate with face recognition performance and reading skill, respectively (Furl et al., 2011; Golarai et al., 2010; Mitchell et al., 2025). This suggests that structural variation in category-selective VTC tracks behavioral abilities.

Furthermore, the categories of faces and text have markedly different evolutionary and developmental histories. Face perception has been shaped over long evolutionary timescales (Parr, 2011) and shows selectivity already in infancy (Deen et al., 2017; Kosakowski et al., 2022) and even in congenitally blind individuals responding to sounds representing faces (e.g. laughing, whistling etc.; van den Hurk et al., 2017). Twin studies further indicate that both face-selective neural activity (Abbasi et al., 2020; Polk et al., 2007) and the tendency to look at social content such as faces are partly influenced by genetic factors (Constantino et al., 2017; Kennedy et al., 2017). In contrast, reading is a recent cultural invention (Roser & Ortiz-Ospina, 2018), and word-selective regions emerge with the onset of literacy, as children learn to map visual symbols to spoken language (Dehaene et al., 2010; Dehaene-Lambertz et al., 2018).

Despite these different origins, visual experience appears critical for the expansion and refinement of both face- and word-selective regions (Golarai et al., 2010; Kubota et al., 2024). For instance, depriving macaques (i.e. non-human primates) of early face exposure prevents the emergence of face-selective patches and typical face-directed gaze behavior (Arcaro et al., 2017). Moreover, macaques that naturally lack face-selective regions showed reduced gaze toward faces (Vinken & Vogels, 2019). Importantly, longitudinal work in humans shows that distinctiveness of multivariate representations for faces and words in VTC develops in a protracted manner across childhood and adolescence and predicts performance in face recognition and speed-reading tasks, respectively (Nordt et al., 2023). In parallel, cross-sectional work has shown that the tendencies to fixate faces and text each change gradually across the first two decades of life (Linka et al., 2023, 2025).

Recent evidence further shows that similarity in individual gaze tendencies toward faces and text during free viewing of movie clips predicts similarity of neural representations in category-selective cortex (Borovska & de Haas, 2024). Together, these findings suggest a tight link between gaze behavior, category-selective representations, and perceptual abilities. Yet this putative link has not been directly tested, and it remains an

open question whether neural representations, measured under controlled conditions of central fixation of the same visual input, differ between individuals with divergent gaze tendencies. A further important question is whether such individual gaze tendencies are related to perceptual abilities and to how observers interpret complex scenes.

1.3 Perceptual and behavioral correlates of gaze

Empirical evidence indicates that the tendency to fixate on text and faces (i.e. their salience) is associated with behavioral performance. Text salience during free viewing (Linka et al., 2025) as well as reading skill (e.g., Nordt et al., 2023) has been shown to increase across the first two decades of life following the onset of formal reading instruction. That is, learning to read appears to raise the informativeness of written text for children and adolescents not only in formal school settings but also in more naturalistic contexts, including tasks such as image viewing. Furthermore, face salience has likewise been reported to develop in a protracted manner (Linka et al., 2025) as do face recognitions skills (Germine et al., 2011; Nordt et al., 2023). Importantly, the tendency to first fixate on faces was directly linked to face recognition ability in student samples (de Haas et al., 2019). Consistent with this association, gaze behavior toward faces varies systematically across groups defined based on their face recognition performance. Individuals with developmental prosopagnosia, characterized by severe impairments in face recognition, allocate a smaller proportion of dwell time to faces compared to control participants (Bobak et al., 2017). In contrast, so-called super-recognizers, who exhibit exceptional face recognition abilities, show increased proportions of both first fixations and dwell time directed toward faces relative to controls (Linka et al., 2022). Notably, individual differences in face recognition performance have also been found to predict how well observers understand an unfolding story during natural viewing of the same movie (Noad & Andrews, 2025). Consequently, differences in gaze behavior may lead observers to extract different information from the same visual scenes.

Brief scene descriptions offer a useful tool for capturing such differences in perception. Although observers can report rich object- and scene-level information after a single glance (Fei-Fei et al., 2007), generating a brief description requires prioritization among multiple potentially reportable aspects of a scene. As a result, descriptions can vary substantially (cf. Chen et al., 2015; Jas & Parikh, 2015). Importantly, this variability is not random: what observers prioritize in their description can serve as a proxy for what they consider as important (cf. Berg et al., 2012). In addition, scene descriptions provide a behavioral index of scene understanding (cf. Murlidaran & Eckstein, 2025).

There is substantial evidence that the spatial distribution of gaze is closely linked to subsequent scene descriptions. For example, observers are more likely to fixate on objects they subsequently describe, particularly foreground and animate objects (e.g., Liang et al., 2018; Yun et al., 2013). Moreover, Coco & Keller (2012) demonstrated that greater similarity in scan-paths (defined as the temporal order of fixations on objects) is associated with greater similarity in scene descriptions. They further showed that a participant's description of a given scene can be identified from a pool of candidate descriptions solely based on their scan-

path. However, if not only gaze behavior (de Haas et al., 2019; Linka & de Haas, 2020) but also the link between gaze and descriptions is trait-like, gaze behavior should predict scene descriptions not only for the same scene but also new ones. This hypothesis has yet to be tested. Moreover, it remains unclear whether observers systematically differ in the perceptual importance they assign to specific scene elements. Specifically, it is an open question whether established individual differences in gaze toward faces and text (de Haas et al., 2019; Linka & de Haas, 2020) are associated with systematic differences in scene descriptions.

1.4 Rationale of the present studies

Building on the work that has established connections between gaze, neural patterns, and behavioral performance, typically averaging across observers, the main aim of this thesis is to shift the focus to the *individual* observer. Specifically, here we systematically investigate neural (Study 1) and perceptual correlates (Study 2) of *individual* gaze behavior in complex scenes.

To test individual gaze tendencies, both studies recorded participants' eye movements while freely viewing complex scenes for three seconds each. Each scene contained numerous objects spanning diverse categories such as people, faces, text, animals, and touched objects. For every image, pixel-level object masks with category labels (obtained from Xu et al. (2014) and for people from Broda & de Haas (2022a) were used to quantify gaze allocation across objects and categories (see Figure 2 for examples). More precisely, semantic labels and associated pixel coordinates served as the basis for calculating the proportions of first fixations after image onset (Study 1) and dwell times (i.e. cumulative fixation duration), either on individual objects (Study 2) or on semantic categories (Study 1 and Study 2) relative to other objects or categories.



Figure 2. (a) Example scene from the OSIE stimulus set (Xu et al., 2014) used in both studies, which participant freely explored for three seconds. (b) Same scene with overlaid object masks indicating labeled objects. Object masks 3, 6, 8 and 14 were labeled as belonging to the category 'face'; masks 2, 5, 7 and 13 included heads and were labeled as 'people', and mask 1 was labeled as 'text'.

Hereinafter, the specific background, research questions, methods, and key findings of the two studies that are the core of this dissertation are summarized and briefly discussed. This is followed by an integrative discussion of their implications in the final chapter. Please refer to the original manuscripts for detailed descriptions of the methods.

2. Study 1: Active vision is linked to category selectivity in the individual brain

By Kollenda, Akbari, Broda & de Haas (2026, accepted for publication in *Nature Human Behaviour*)

Previous vision research often focused on group averages, but recent studies show stable individual differences in gaze behavior, such as tendencies to look preferentially at faces or text (de Haas et al., 2019; Guy et al., 2019; Linka & de Haas, 2020). These tendencies are partly linked to genetics (Constantino et al., 2017; Kennedy et al., 2017) and recognition abilities (de Haas et al., 2019; Linka et al., 2022), but the underlying mechanisms remain unclear. This study examines whether these individual gaze tendencies are associated with category-selective representations in the VTC. Specifically, Study 1 addresses the following questions:

- Are gaze tendencies toward faces and text during free viewing linked to corresponding representational distinctiveness in the individual VTC during central fixation of faces and words?
- Is representational distinctiveness in adults related to performance in face recognition and speed-reading tasks?
- Are these links also observed for the extent of functional selective activation in VTC for faces and words?

The study used a multi-method approach including three separate experimental sessions using eye tracking, functional neuroimaging (fMRI) and behavioral tasks. In total, N = 102 adult participants took part in the eye tracking session. N = 61 of these also took part in the neuroimaging session, N = 69 of these in the Cambridge Face Memory task (CFMT; N = 34 overlapping with the MRI sample) and N = 64 took part in the speed-reading test (SLRT-2; N = 25 overlapping with the MRI sample). In the eye tracking session, participants viewed 700 complex scenes in a free viewing paradigm (see above), allowing us to quantify the proportion of first fixations and dwell time for faces and text, respectively. In a separate fMRI localizer session, participants were asked to maintain central fixation while viewing centrally presented category-specific stimuli (faces, pseudowords etc.), ensuring that all participants received identical visual input. This allowed us to derive two measures of category-selectivity within left and right lateral VTC: First, we analyzed representational distinctiveness based on multivariate response patterns (cf. Nordt et al., 2023; see Figure 3). A higher value indicates that the distributed response pattern for a given category (e.g., faces) is more consistent across runs and more dissimilar from other categories (e.g., words, bodies, houses, cars, and limbs). Second, we analyzed the individual extent of face- and word-selective activity in group-defined functional regions (cf. Julian et al., 2012) corresponding to the right fusiform face area (FFA) and left visual word form area (VWFA), respectively.

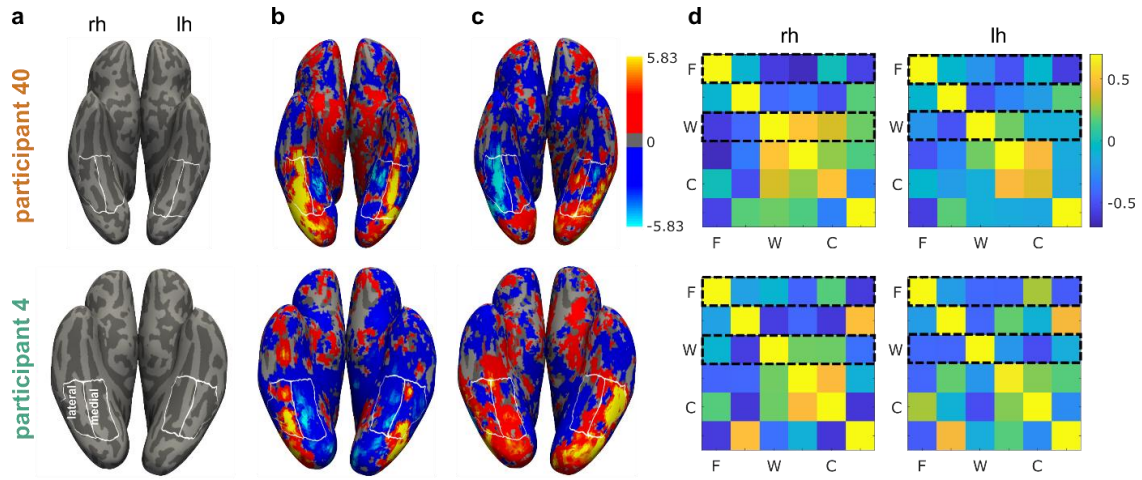


Figure 3. Data from two example participants (ID40, top; ID4, bottom) are shown across panels. (a) Individual anatomical regions of interest (ROIs) in lateral and medial ventral temporal cortex (VTC; both hemispheres; cf. Nordt et al., 2023). (b-c) t-contrast maps for faces and words, respectively, with colors corresponding to the scale bar on the right. (d) Representation similarity matrices based on category-selective activity patterns in lateral VTC across odd versus even localizer runs (per hemisphere). Correlation values are indicated by the color bar on the right. The respective stimulus categories are faces (F), bodies, words (W), houses, cars (C), and limbs. Dashed rows mark entries used to compute representational distinctiveness (within-category similarity minus mean between-category similarity) for faces and words.

We replicated the robust and substantial variation in gaze tendencies toward faces and text between participants reported previously (de Haas et al., 2019; Linka & de Haas, 2020). We also replicated the strong anti-correlation between these tendencies when considering all images (de Haas et al., 2019). However, this relationship became noticeably weaker when we restricted the analysis to images that did not contain faces and text simultaneously. We further replicated the previously reported association between the tendency to first fixate faces and better face recognition performance (de Haas et al., 2019). Crucially, individuals with higher face salience showed higher face distinctiveness in the right lateral VTC (see Figure 4) and the proportion of first fixations on faces was associated with a larger size of right FFA. An analogous pattern was observed for text salience and word distinctiveness in the left lateral VTC and size of the VWFA. These links were category-specific (e.g., gaze tendencies toward faces were not linked to word distinctiveness). Furthermore, higher face distinctiveness and larger size of right FFA were associated with better face recognition performance, whereas word distinctiveness was associated with better pseudoword speed-reading performance. Correlations were generally stronger for first fixations after image onset.

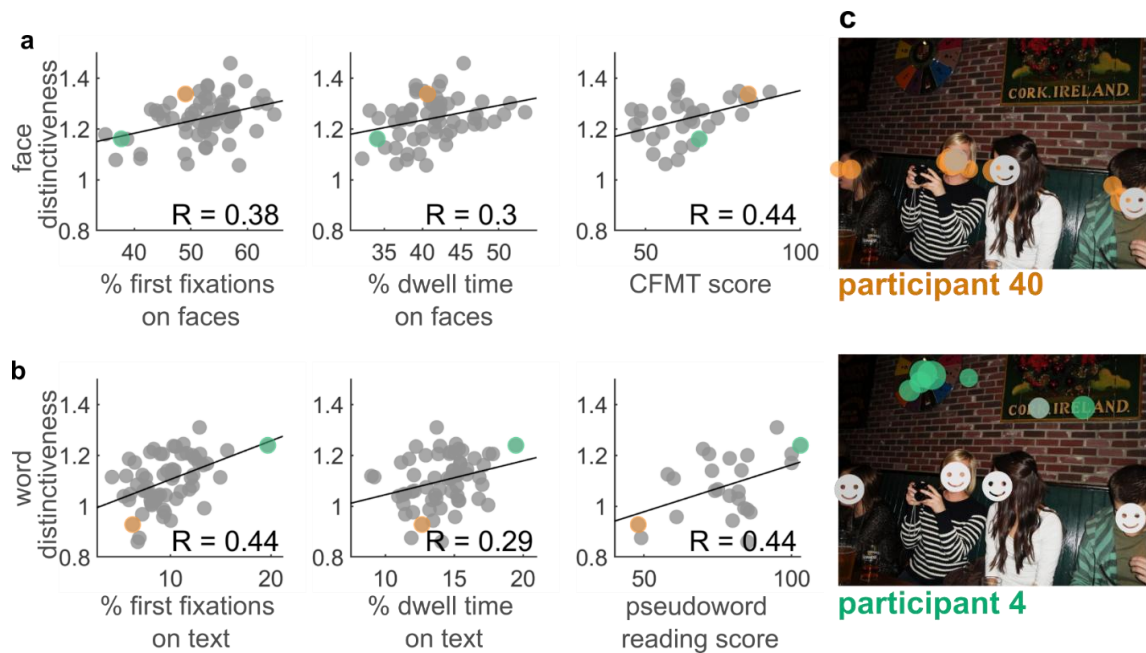


Figure 4. (a) Greater face distinctiveness in the right lateral ventral temporal cortex (VTC) was associated with a higher tendency to first fixate on faces, higher dwell times on faces, and better Cambridge Face Memory Test (CFMT) performance ($N = 61$, 61 , and 34 , respectively). (b) Similarly, greater word distinctiveness in left lateral VTC was linked to a higher tendency to first fixate on text, higher dwell time on text, and higher scores in pseudoword speed reading ($N = 61$, 61 , and 25 , respectively). (c) Example participants (ID40, orange; ID4, green) are highlighted across panels and figures. Fixation maps show their gaze during free viewing, with marker size reflecting the duration and the first fixation depicted in lighter color.

Together, these findings suggest that moment-to-moment gaze behavior systematically targets image content that the individual ventral stream represents more distinctively. This may imply a role of VTC in at least partially informing gaze by signaling the relevance and representational precision of depicted objects (cf. Discussion, section 4.1 – Implications of Study 1).

3. Study 2: Individual gaze predicts individual scene descriptions

By Kollenda, Reher & de Haas (2025, *Scientific Reports*)

Study 1 showed that, although humans share a general functional neural layout, the category-selectivity patterns differ idiosyncratically between individuals and systematically covary with gaze tendencies. However, it remains unclear whether the systematic gaze patterns observed in Study 1 and prior work (Broda & de Haas, 2022a; de Haas et al., 2019; Guy et al., 2019; Linka & de Haas, 2020) lead to systematically different perceptual consequences. Study 2 therefore asks whether individual gaze biases translate into systematic, predictable differences in how observers perceive and communicate visual content. Specifically, we address the following questions in Study 2:

- Are gaze tendencies toward people and text elements during free viewing linked to how often these categories are referenced in subsequent scene descriptions?
- Is there a general link between gaze tendencies and scene descriptions that is not restricted to a-priori categories?

- Does the link between gaze and description similarity generalize to new sets of images?

The study used a within-subject design and tested $N = 30$ participants. Participants viewed 100 complex scenes in a free viewing paradigm (see above) and were asked to provide a brief description of each scene immediately following its presentation. We analyzed complete descriptions and specific word classes (nouns, verbs, adjectives) using semantic sentence (sBERT; Reimers & Gurevych, 2019) and word embeddings (fastText; Grave et al., 2018; Mikolov et al., 2018), respectively. We then compared the similarity of these between participants, yielding Description Dissimilarity Matrices (DDMs). Similarly, we analyzed the similarity of dwell time distributions across all object pixel masks in all scenes to derive a Gaze Dissimilarity Matrix (GDM). Data were split into odd versus even scenes to assess the consistency of description and gaze similarity patterns and to perform decoding analyses testing whether an observer's identity could be predicted across independent image sets. We further analyzed the split-half reliability and proportions of dwell time on people and text elements and compared these measures to the reliability and frequency of corresponding references in descriptions.

Both gaze and description patterns consistently differed between individuals. Gaze similarity significantly correlated with description similarity, with the strongest association for nouns, which mapped more directly to fixated objects than verbs and adjectives. We successfully decoded observer identity across independent image sets, such that we could identify which participant produced a set of descriptions based on gaze data for a *different* image set. Finally, observers who spent more time looking at text elements referenced them more frequently, and those who spent more time looking at people referenced them more often. Notably, the tendencies to look at people versus text elements and to reference them were each anti-correlated.

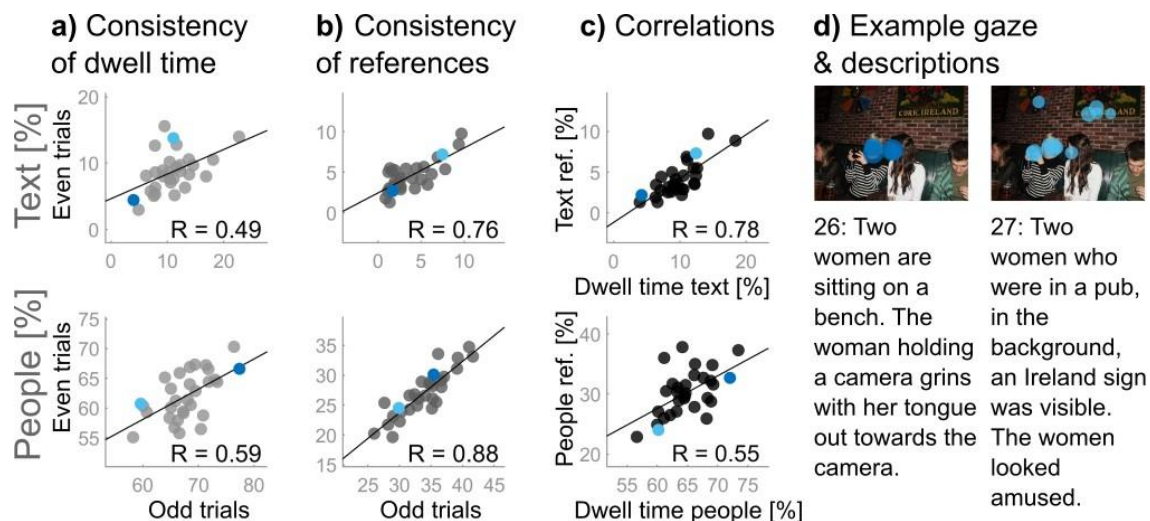


Figure 5. (a) Split-half correlations show consistent individual tendencies of dwell time distribution between odd and even trials for text ($p = .001$) and people ($p < .001$). (b) Similarly, individuals showed reliable consistency in how often they reference these elements in their descriptions across halves for text and people (both $p < .001$). (c) Dwell times on text and people correlate with how frequently participants reference these elements in their descriptions (text: $p < .001$; people: $p = .002$). (d) Example participants (ID 26, dark blue, ID 27, light blue) with corresponding fixation patterns and verbal descriptions, translated from German to English.

Together, these findings demonstrate that scene descriptions are linked to individual gaze. Where observers allocate their gaze in scene predicts the perceptual impression they form and communicate, and these patterns are stable across images, supporting trait-like differences in both gaze and descriptions behavior as well as in their link.

4. Discussion

This thesis investigated the neural (Study 1) and perceptual correlates (Study 2) of idiosyncratic gaze behavior in complex scenes. Our findings demonstrate that inter-individual differences in gaze behavior are not random. Instead, they reflect the functional organization of the individual ventral temporal cortex (VTC) and predict how observers interpret and communicate what they see. Thus, in line with a growing shift in the field of vision science (Charest & Kriegeskorte, 2015; de Haas et al., 2019; Finn et al., 2020; Wang et al., 2024), we argue that embracing idiosyncrasy as meaningful signal is essential for a more comprehensive account of where people look and, more broadly, how human visual perception operates.

4.1 Linking active vision, neural category-selectivity, and perception

While Study 1 focused on the neural correlates of individual differences in human gaze behavior, Study 2 focused on their perceptual consequences. Together, the results point to a systematic covariance between gaze behavior, category-selective representations in the ventral temporal cortex (VTC), perceptual abilities, and perceptual interpretation of scenes. Study 1 demonstrated that moment-to-moment gaze preferentially samples scene elements (e.g., faces or text) that are distinctively represented in an individual's ventral stream. This suggests that the VTC may play an active role in informing gaze and allocation of foveal resources. Study 2 extends this account to its perceptual and communicative consequences: what is fixated is more likely to be perceived and preferentially encoded in brief scene descriptions. Although we did not directly test the link between neural representations and perceptual consequences, converging evidence suggests such a relationship at least for naturalistic movie stimuli and subsequent interpretations (Han & Bonner, 2026). Taken together, our findings support a three-fold link between individual differences in functional neural architecture, moment-to-moment gaze decisions, and subjective scene understanding in adults.

Notably, however, the neural and perceptual correlates appear to manifest at different temporal dynamics of gaze, with the relative weighting of bottom-up and top-down influences shifting as scene exploration proceeds (Schütt et al., 2019). First fixations are often thought to be driven automatically by salient bottom-up image properties, including low-level features (e.g., Anderson et al., 2015; Parkhurst et al., 2002) and high-level semantic and 'meaning'-related properties (e.g., Foulsham & Underwood, 2011; Henderson et al., 2018; Henderson & Hayes, 2017). Dwell time, on the other hand, might reflect more deliberate, top-down cognitive processes that accumulate over longer viewing periods. In Study 1, we observed that the link between individual gaze tendencies and category-selective neural properties in the VTC was stronger for first fixations than for dwell time. This suggests that the functional layout of the individual visual system finds its most direct

expression in the earliest, relatively automatic moments of scene exploration. Conversely, scene descriptions, which reflect a more deliberate, cognitive act, correlated strongly with object dwell time distributions. This suggests that the perceptual content ultimately communicated in scene descriptions is shaped across the course of visual exploration. Crucially, however, these two processes should not be treated as strictly dissociable. For example, Murlidaran & Eckstein (2025), demonstrated that objects that are relevant for scene understanding were fixated not only more frequently but also earlier than irrelevant ones. Together, this implies that bottom-up and top-down influences do not operate in strictly sequential manner but interact throughout the exploration of scenes. They jointly steer gaze toward content that is both salient for the individual and meaningful for scene understanding.

Implications of Study 1. Prior work has shown that low- and high-level image properties guide gaze from the very first fixation (Anderson et al., 2015; Foulsham & Underwood, 2011; Henderson et al., 2018; Henderson & Hayes, 2017; Parkhurst et al., 2002). Our findings in Study 1 extend this work by showing that the salience of a stimulus category differs systematically between observers from the earliest moments of scene exploration (see also de Haas et al., 2019; Linka & de Haas, 2020). That is, although faces and text elements embedded in scenes are strong gaze attractors, they were not prioritized equally by all participants. This dissociation stresses a distinction between *universal* salience, which reflects average gaze tendencies toward certain image features, and *individual* salience, which reflects observer-specific preferences (cf. Xu et al., 2019).

A previous longitudinal study has shown that representational distinctiveness increases in a protracted manner across childhood and adolescence alongside increasing behavioral skills (Nordt et al., 2023). We replicated that face and word distinctiveness is linked to behavioral performance in face recognition and reading speed tests in the adult visual system. Furthermore, we found that representational distinctiveness is linked to the individual salience of the corresponding category. Interestingly, these findings resonate with predictive processing frameworks, which conceptualize perception as an active inference process: the brain continuously generates predictions ('best guesses') about incoming sensory input, with gaze serving as a mechanism for sampling information that is most likely to reduce prediction errors (Friston et al., 2012; Pezzulo et al., 2024). This can be achieved in two ways: by refining predictions and by sampling what matches predictions best (Parr & Friston, 2019). On this view, the distinctive category-selectivity of the VTC may function as a form of learned prior, shaping not only how visual input is processed once fixated, but also which regions are selected for detailed foveal inspection in the first place. This selection process may be preceded by a covert, pre-saccadic analysis of meaning across the visual field (Mack, 2003). Rapidly extracted global scene gist (Torralba et al., 2006) and scene context (Tasliyurt-Celebi et al., 2024), may aid this process by signaling where potentially meaningful and informative regions are likely to be located. Interestingly, a recent preprint shows that individual internal models (i.e. visual expectations) can predict idiosyncratic scene exploration, particularly when participants engage in strategic exploration under increased peripheral uncertainty (Engeser et al., 2026). These findings suggest that presumably both stimulus-driven scene properties and observer-specific expectations may jointly influence where observers look in complex scenes.

Furthermore, understanding how peripheral and foveal information processing interacts to support anticipatory guidance remains an interesting question in vision science. For instance, Stewart et al. (2020) propose an integrated, bidirectional exchange (but see Zhaoping (2024) who argues for strictly separated functions). That is, peripheral input may bias foveal processing toward relevant visual features and foveal feedback in turn sharpens peripheral sensitivity by establishing learned expectations. Over time, the visual system could link relatively impoverished peripheral representations with richer foveal ones through a learning history enabled by periphery-fovea linking eye movements (Gegenfurtner, 2016). As a result, coarse peripheral information processing could later function as an ‘intelligent header’ guiding the detailed inspection needed for precise identification (Thorpe et al., 2001). Consistent with this account, Sheinberg & Logothetis (2001) observed that neurons in the temporal cortex of macaques exhibit selective activation shortly before saccades to behaviorally relevant target objects, with activity increasing again upon foveal fixation. Preliminary intracranial evidence suggests that a similar mechanism may exist in humans, with pre-saccadic category-selective responses observed in the VTC (Alreja et al., 2025). Notably, however, Xiao et al. (2024) reported only modest evidence for feature-selective activity that anticipates upcoming input or integrates information across fixations during natural image viewing in macaques. Furthermore, Fan et al. (2016) and Williams et al. (2008) reported that behaviorally relevant *peripheral* target objects, which are typically selected for an upcoming saccade target, are represented in foveal retinotopic cortex. Their findings further suggest that foveal cortex can be recruited for tasks requiring high spatial detail and high-level category information, even when the stimulus lies well outside the receptive field of any foveal neuron. This might be triggered via feedback from temporal areas (Gilbert & Li, 2013; Rockland & Van Hoesen, 1994). Although speculative, such a mechanism could help generate expectations about soon-to-be foveated input (see also Wolfe & Whitney, 2014, 2015). Together, these findings support the view that gaze is not only reactive to visual input, but also anticipatory and guided by neural priors about what is most informative to sample. Taken together with our findings from Study 1, these observations suggest that temporal areas such as the VTC may support gaze guidance in two complimentary ways: by signaling behavioral relevance to classic visuomotor structures and by providing foveal feedback that anticipates incoming information, thereby facilitating post-saccadic target processing. Further work is needed to clarify the precise nature and timing of these potential VTC influences. Specifically, category-selective responses in VTC should be contrasted, ideally on the individual level, for peripheral stimuli before a saccade and when only attended covertly, relative to those during central fixation as in Study 1.

Implications of Study 2. Prior work has already suggested that where people look is linked to what they later recall (Einhäuser et al., 2008b; Pertzov et al., 2009) and mention when describing the same scenes (Coco & Keller, 2012; Liang et al., 2018; Yun et al., 2013). In Study 2, we extend these findings by showing that individual gaze tendencies (i.e. the distribution of dwell time across objects within scenes) are associated with descriptions tendencies, not only for the same set of scenes but also generalize to different sets. This may suggest a trait-like, bidirectional pattern: allocating more dwell time to certain inputs can narrow how people

interpret a scene and what they include in their descriptions and thus deem important. In turn, priorities are likely to guide where they preferentially direct their gaze and what they dwell on in subsequent fixations.

We further found that gaze tendencies toward people and text translated to the tendency to reference them in subsequent scene descriptions. In addition, the anti-correlation between gaze toward people and text was paralleled in the tendency to describe people versus text. This pattern indicates that the competitive face-text dynamic extends from gaze to verbal output, providing converging behavioral evidence for a trait-like linkage across modalities. Together, these findings bridge the gap between research on eye movements and subjective perception, supporting the validity of gaze tendency measures as predictors of stable perceptual priorities and individual differences.

Notably, the link between dwell time and verbal descriptions does not seem to be a simple “readout” of what objects observers looked at. Yun et al. (2013) showed that observers were more likely to have looked at the objects they described than to describe every object they had looked at. This asymmetry suggests that scene descriptions constitute a higher-level interpretation assembled from fixated objects, whose perceived relevance is reevaluated in the current context, rather than reflecting pure gaze history. Foveal sampling thus might be a necessary but not sufficient condition for description. In part, this could explain why the correlation between gaze and descriptions in Study 2 is high but does not account for the whole explainable variability in description similarity. It would be interesting to test in future studies whether individuals with relatively low spontaneous gaze tendencies toward faces or text would include these categories in their descriptions when fixation on them is experimentally enforced (e.g. Murlidaran & Eckstein's (2025) Figure 7a) or steered by narrative image sequences (e.g., (Berlot et al., 2026). Alternatively, they may still omit them because they considered them less important by default. Moreover, next to gaze tendencies, other cognitive traits such as working memory capacity and verbal fluency could additionally influence description behavior, and future studies should systematically compare their unique contributions to the variance in individual description similarity.

A further question is whether objects must be directly fixated to be incorporated into scene descriptions, and whether this requirement differs across stimulus classes such as text and (familiar) faces, both of which require fine-detail resolution for recognition. On the one hand, Murlidaran & Eckstein (2025) showed that fixation location systematically influences scene descriptions: when observers are required to fixate objects that are irrelevant for global scene understanding, their descriptions diverged more strongly from a gold-standard description than when the fixation is directed on a relevant object. On the other hand, Fei-Fei et al. (2007) demonstrated that participants could provide rich object- and scene-level descriptions already at very brief (e.g., 67 ms) presentation times. Furthermore, a recent preprint, using a gaze-contingent paradigm to selectively mask peripheral or foveal input, indicates that peripheral vision loss reduces the production of informative descriptions more strongly than central vision loss. This was particularly the case for scenes depicting social interactions (Johnson et al., 2025). Together, these findings underscore the importance of both foveal and peripheral vision for the rapid extraction of scene semantics. Future studies should

systematically compare the influence of central or foveal processing on references to text and people in descriptions, relative to baseline individual tendencies during unimpaired scene viewing.

Moreover, individual scene viewing may not only influence scene descriptions shortly after scene presentation but potentially also event comprehension of movies and long-term memory. With regard to event comprehension, a study by Noad & Andrews (2025) demonstrated that individual differences in face recognition performance, which has been linked to different gaze tendencies (Bobak et al., 2017; de Haas et al., 2019; Linka et al., 2022), predicted narrative understanding as measured with a written recall of the movie and specific questions about events. Interestingly, in the context of long-term memory, image memorability has been proposed as an image specific property that reliably predicts the likelihood that an image will be remembered across observers (Goetschalckx et al., 2018; Isola et al., 2011). This memorability has been partially linked to the presence of semantic categories such as people and text (Bainbridge, 2019). In both our studies, participants showed systematic individual differences in their tendency to fixate on these categories during scene exploration. In Study 1, we further demonstrated that the tendency to first fixate on faces is associated with better face recognition performance (see also de Haas et al., 2019). Building on these findings, future research could examine whether such individual gaze tendencies modulate memory performance for scenes depicting people, text, or both.

4.2. Broader aspects and implications

The developmental origins of individual gaze tendencies. The correlational findings in Study 1 suggest that, by adulthood, gaze tendencies toward faces and text align with the precision of corresponding category-selective representations and size of functional regions of the mature VTC. Study 2 further demonstrated that gaze priorities correspond to perceptual interpretations. A key open question, however, is how the link between gaze priorities and VTC category-selectivity emerges over development. One plausible account is that this link arises through a multicausal process involving both biological predispositions and experience-dependent plasticity. For instance, early visual development may be guided by biological priors, including heritable patterns of spontaneous gaze (Constantino et al., 2017; Kennedy et al., 2017), early cortical microstructural maturation and myelination (Natu et al., 2021), pre-existing connectivity patterns (Saygin et al., 2016), and early category selectivity (Deen et al., 2017; Kosakowski et al., 2022). These factors likely interact with experience as the visual system adapts to the statistical and social regularities of the individual environment (Powell et al., 2018). Evidence from deprivation and variation in early experience supports this interactionist view. For example, macaques raised without exposure to faces fail to develop face-selective patches and showed reduced spontaneous gaze tendencies toward faces (Arcaro et al., 2017). In less extreme cases, both macaque and human infants who experience less responsive face-to-face interaction later exhibit reduced spontaneous gaze toward novel faces (Dettmer et al., 2016; Field, 1992). At the same time, concurrent multimodal social input may further scaffold this developmental trajectory: gentle touch has been shown to enhance face-identity learning in infants (Della Longa et al., 2019), and infant-directed speech captures and

sustains attention (Fernald, 1985). In school-age children and teens, faces and text become more frequent in the visual environment as well as communicatively relevant. These changes in the ‘visual diet’ co-occur with increasing face and word-selective representations in VTC (Kubota et al., 2024; Nordt et al., 2021). Throughout this developmental trajectory, reward and social feedback may consolidate gaze habits that have proven perceptually and behaviorally successful. In adults, gaze patterns are more homogenous than those of children (Linka et al., 2025) and category-selectivity is stable (Peelen & Downing, 2005). This suggests that individual gaze behavior, developing expertise, and cortical specialization may enter a period of mutual stabilization. Definitive evidence for the developmental trajectory will require longitudinal data combined with mono- and dizygotic twin studies to better tease apart genetic and environmental contributions.

Applications across disciplines. The findings of this thesis extend beyond vision science and contribute to a growing literature showing that eye movement behavior is relevant across disciplines, including biology, medicine, computer science, and media. Within clinical and social domains, atypical face-directed gaze has been linked to several conditions. For example, altered fixation patterns (e.g., reduced fixation time in the eye region) have been reported in autism spectrum disorder (Constantino et al., 2017; Wang et al., 2015), and gaze to social content predicts aspects of social competence (Klin et al., 2002). Furthermore, in developmental prosopagnosia, altered gaze strategies may reflect or contribute to impairments in face perception (Bobak et al., 2017; Lowes et al., 2025). Altered gaze pattern have also been linked to social anxiety, depression, schizophrenia, and other clinical traits (e.g., Armstrong & Olatunji, 2012; Hayes & Henderson, 2018; H. Li et al., 2023; Molitor et al., 2015). These findings raise the possibility that gaze behavior could serve as a diagnostic or prognostic marker (cf. Jiang & Zhao, 2017), especially when combined with standardized tests.

Cross-sectional developmental work further suggests that face recognition abilities continue to improve into adulthood, at least until around age 30 (Germine et al., 2011). This prolonged developmental window may, in principle, allow for targeted interventions aimed at improving face memory, face matching or perceptual discrimination (cf. DeGutis et al., 2014; White et al., 2014). Supporting this idea, individuals with developmental prosopagnosia who completed an intensive perceptual training (approximately 11 weeks of online training) showed persistent and generalized improvements in face discrimination (Corrow et al., 2019). However, observations from real-world settings with prolonged exposure to face-identification tasks (e.g. daily photo ID verification) show only limited benefits: Notaries with many years of experience (but no specific training) did not substantially outperform newer notaries or student samples in a face matching task (Papesh, 2018). Notably, interventions may not need to be restricted to face-specific training. Learning to read has been associated with improved object and face recognition performance (e.g., van Paridon et al., 2021). This observation aligns with the idea that literacy may refine visual processing more broadly (Behrmann & Plaut, 2013). However, this relationship does not appear to be obligatory, as face recognition deficits can occur despite intact word recognition (Susilo et al., 2015), and contradicts accounts proposing competitions between visual categories in VTC (see below). Future studies therefore should systematically test whether such

interventions could improve face recognition abilities, and potentially well-being in social situations (cf. Corrow et al., 2019; Lowes et al., 2025).

More broadly, individual differences in gaze tendencies may partially explain variation in cognitive abilities (for a similar approach considering scan patterns, cf. Hayes & Henderson, 2017) and educational outcomes. The present findings suggest that individuals differ in which visual content they treat as informative, and that these tendencies are related to skills such as face recognition and reading fluency. In educational contexts, it would therefore be valuable to examine whether stable gaze tendencies toward text, diagrams, faces, or objects predict reading development, STEM (science, technology, engineering, and mathematics) performance, or other academic outcomes. Such work could clarify how visual expertise and academic achievement co-develop and may eventually support more individualized learning materials. At the population level, integrating eye-tracking measures into large longitudinal representative cohorts, such as the UK Biobank or the German National Cohort, could provide important insights into the lifespan developmental of visual attention. The broad range of measures assessed in these databanks may in turn help clarify how visual attention relates to social behavior, well-being and health.

Finally, the finding that individual gaze patterns predict the perceptual impressions observers form and communicate in Study 2 has implications for education, media, and personalized communication. Specifically, we show that what individuals fixate on is systematically related to what they perceive as important and subsequently include in their descriptions. Building on this link, tailoring visual content to individual viewing tendencies could improve comprehension and engagement, although this remains to be tested directly. Recent work in generative AI points in a similar direction. For example, personalized persuasion (e.g., in the context of political appeals for pro-environmental actions) can be enhanced when individual characteristics such as purchase histories (Kumar & Kapoor, 2024) or personality traits (Matz et al., 2024) are taken into account. Provided that ethical guidelines are followed, stable gaze tendencies of the kind documented in this thesis may provide an additional basis for adaptive systems that tailor information to individual observers.

4.3. Limitations and further directions

Ecological validity and the free viewing paradigm. Following previous research investigating spontaneous gaze behavior (e.g., de Haas et al., 2019; Xu et al., 2014), our Studies 1 and 2 employed a free viewing paradigm in which observers view scenes for a brief duration without an explicit task. This approach seeks to approximate natural, ecologically valid gaze behavior, such as looking out a train window, within the constraints of a controlled laboratory setting. However, it could be questioned whether this paradigm invites participants to self-impose their own implicit task, such that individual gaze tendencies reflect differing task goals rather than trait-like differences. Speaking against this concern, findings from Murlidaran & Eckstein (2025) suggest that the primary goal during free viewing is scene understanding. Furthermore, faces and text remain robust attractors across varying task demands (Cerf et al., 2009). Nevertheless, future studies should systematically compare individual gaze tendencies across multiple task conditions. These should include

search tasks and real-world action tasks (also see below), which help to determine how robust individual differences are when observers are directed toward particular categories and attentional load varies.

To better understand how individuals interact with their visual environment, future research should adopt more naturalistic viewing (and description) conditions. This will further help to test whether gaze tendencies observed in the laboratory generalize to real-world contexts (cf. Risko et al., 2012). Some researchers have already pioneered this shift by utilizing dynamic visual content (Borovska & de Haas, 2024; Broda & de Haas, 2022b) and simulating real-life activities such as walking ('t Hart & Einhäuser, 2012) or everyday tasks like tea and sandwich-making (Hayhoe et al., 2003; Land et al., 1999). Mobile eye tracking promises a reliable tool for such endeavors. For example, Varela et al. (2023) combined mobile eye tracking with automated region-of-interest analysis and found consistent individual differences in fixations on people and faces in natural settings, despite overall reduced gaze toward unfamiliar passersby compared to image-based studies. Notably, fixations on people increased during direct interactions, suggesting that social norms and the potential for interactions, which are largely absent during media viewing, play an important role in shaping gaze behavior in live environments. A related key open question is whether the stable individual gaze tendencies toward social stimuli observed in laboratory settings (Broda & de Haas, 2022a, 2022b) extent robustly to real-world social contexts (cf. Peterson et al., 2016, for a related demonstration of generalization in within-face fixation patterns). Alternatively, they may be dynamically modulated, or even suppressed, by learned social norms. Cross-cultural differences in such norms may further influence these patterns and warrant systematic investigation.

It is also worth noting that artificially constraining viewing time to three seconds in our paradigm likely increased competition for attentional resources relative to real-world viewing, where observers can more freely choose how long to inspect a scene and its content. Relatedly, Study 1 showed that images containing both faces and text resulted in stronger anti-correlations of gaze tendencies toward these categories than images containing only one of them. Future work could systematically examine how viewing duration and the number of competing elements jointly influence gaze tendencies. Notably, the observed anti-correlation between face and text salience (Studies 1 and 2) is consistent with accounts proposing competition between visual categories in VTC (Rossion et al., 2004). Moreover, they align with the neural recycling hypothesis (Dehaene & Cohen, 2007), which posits a competition for limited cortical territory between categories such as faces and words. More recent longitudinal evidence, however, suggests that word-selective cortex may preferentially emerge from limb- and tool-selective regions rather than face-selective cortex (Kubota et al., 2024; Nordt et al., 2021). This is also consistent with developmental shifts in gaze away from limbs and toward text (Linka et al., 2023, 2025). Moreover, face- and word-selective regions are typically lateralized to opposite hemispheres, and their specialization may reflect the integration within broader networks (Powell et al., 2018; Saygin et al., 2016) rather than local competition. Interestingly, in this context, face-selective regions exhibit idiosyncratic patterns of whole-brain functional connectivity (Watson et al., 2026). Future research should examine whether these connectivity profiles relate systematically to individual differences in gaze behavior,

face recognition, reading ability, representational tuning in VTC (cf. Study 1), and verbal descriptions (cf. Study 2). This may help to clarify whether the observed anti-correlations between face and text processing reflect a direct competition of cortical resources or largely independent specialization within larger networks.

Sample diversity and generalizability. A further limitation concerns the diversity of the population sampled. Participants in the present studies were predominantly German university students. This sample was therefore obtained from a western, educated, industrialized, rich, and democratic (WEIRD) society (Henrich et al., 2010). Cultural differences in context-independent versus context-dependent perceptual styles have been reported, which may contribute to distinct gaze tendencies, perception, and memory (Chua et al., 2005; Nisbett & Miyamoto, 2005). Thus, the individual differences observed within our culturally homogenous sample may be even more pronounced in more diverse populations. Furthermore, we did not include age groups beyond young adults, even though visual exploration rates change across the lifespan (Linka et al., 2025). Beyond participant diversity, the stimuli employed in this study were also limited in scope (e.g., depicting primarily Western scenes, people and characters and mainly the visual environments of adults). Future work should therefore employ cross-cultural and lifespan designs, as well as more diverse stimulus sets, to build a more comprehensive and generalizable picture of the gaze tendencies reported in this thesis.

Focus on certain categories. The present thesis focuses on individual differences in gaze tendencies toward faces (extended to people more broadly in Study 2) and text. This focus was motivated by evidence showing that both semantic categories exhibit robust inter-individual variability and are each associated with distinct category-selective areas in the VTC. However, other ecologically meaningful categories may reveal similarly systematic individual patterns. A particularly promising candidate is the category of limbs (and touched objects), which reliably attracts gaze at systematically different levels across individuals (Broda & de Haas, 2022a; for touched objects see de Haas et al., 2019; Linka & de Haas, 2020) and shows clear category-selectivity within the VTC (Weiner & Grill-Spector, 2010). Importantly, this category follows an interesting developmental trajectory: limb salience (Linka et al., 2023, 2025), limb-selectivity (Nordt et al., 2021) and limb distinctiveness within the VTC (Nordt et al., 2023) all appear to decline throughout the first decade of life. A post-hoc analysis based on data from Study 1, however, found no link between individual gaze tendencies toward limbs and their corresponding representational distinctiveness. This suggests that behavioral relevance in adulthood may be critical for observing such relationships. To test this more directly, future studies could examine limb distinctiveness in individuals who use sign language, for whom limbs retain strong behavioral relevance. This approach would provide a unique opportunity to investigate how visual expertise shapes both gaze tendencies and neural representations of a category that otherwise undergoes developmental decline. Consistent with this idea, expertise has been linked to category-specific salience and selectivity patterns in the VTC (Gauthier et al., 2000; Gomez et al., 2019; James, 1890; Reeder et al., 2016; Stein et al., 2016). More broadly, extending our line of research to a wider range of semantic categories, such as numbers (Nordt et al., 2023), food, tools, animals (cf. Contier et al., 2024; Huth et al., 2012) and domains

of expertise (Gauthier et al., 2000; Gomez et al., 2019), would enrich our understanding how gaze tendencies and representational spaces covary across individuals.

4.4 Conclusion

Where we look is not merely a passive reflection of the current visual environment but appears to be an active reflection of who we are. This thesis demonstrated that individual differences in gaze behavior are neither arbitrary nor inconsequential. Study 1 showed that the tendency to fixate faces or text is linked to more distinctive and extended neural representations of those categories in the individual VTC, and to corresponding behavioral skills. This suggests that VTC may inform gaze by signaling the representational relevance of depicted content. Study 2 demonstrated that individual gaze patterns reliably predicted how observers described the same scenes, particularly in their use of nouns and their tendency to refer to people and text. Together, these findings reveal a systematic covariation of gaze tendencies, individual neural organization, behavioral performance, and perceptual interpretations of scenes. Hence, we argue that considering individual differences is crucial for a more comprehensive understanding of visual perception. More precisely, it enriches our understanding of how the visual system represents and makes sense of the visual environment. These results align with a broader shift in vision science away from the average observer toward the individual as a meaningful unit of analysis. Beyond the theoretical implications of the findings presented in this thesis, they raise a practical possibility: that eye movements may serve as a non-invasive, behavioral signature of individual differences in neural organization and perception.

5 References

- Abbasi, N., Duncan, J., & Rajimehr, R. (2020). Genetic influence is linked to cortical morphology in category-selective areas of visual cortex. *Nature Communications*, 11(1), 709.
<https://doi.org/10.1038/s41467-020-14610-8>
- Alreja, A., Boring, M., Ward, M., Richardson, M., Abel, T., & Ghuman, A. (2025). Intracranial Electroencephalography reveals the neural correlates of eye-movement guidance to objects. *Journal of Vision*, 25(9), 2905–2905. <https://doi.org/10.1167/jov.25.9.2905>
- Anderson, N. C., Ort, E., Kruijne, W., Meeter, M., & Donk, M. (2015). It depends on when you look at it: Salience influences eye movements in natural scene viewing and search early in time. *Journal of Vision*, 15(5), 9. <https://doi.org/10.1167/15.5.9>
- 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>
- Armstrong, T., & Olatunji, B. O. (2012). Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clinical Psychology Review*, 32(8), 704–723.
<https://doi.org/10.1016/j.cpr.2012.09.004>
- Bainbridge, W. A. (2019). Memorability: How what we see influences what we remember. In *Psychology of learning and motivation* (Bd. 70, S. 1–27). Elsevier.
<https://www.sciencedirect.com/science/article/pii/S0079742119300015>
- Behrmann, M., & Plaut, D. C. (2013). Distributed circuits, not circumscribed centers, mediate visual recognition. *Trends in Cognitive Sciences*, 17(5), 210–219. <https://doi.org/10.1016/j.tics.2013.03.007>
- Berg, A. C., Berg, T. L., Daume, H., Dodge, J., Goyal, A., Han, X., Mensch, A., Mitchell, M., Sood, A., & Stratos, K. (2012). Understanding and predicting importance in images. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 3562–3569.
<https://ieeexplore.ieee.org/abstract/document/6248100/>

- Berlot, E., Schmitt, L.-M., Huber-Huber, C., Peelen, M. V., & de Lange, F. P. (2026). Narrative context shifts gaze from visual to semantic salience. *Communications Psychology*, 4(1), 59.
<https://doi.org/10.1038/s44271-026-00426-7>
- Blauch, N. M., Plaut, D. C., Vin, R., & Behrmann, M. (2025). Individual variation in the functional lateralization of human ventral temporal cortex: Local competition and long-range coupling. *Imaging Neuroscience*, 3, imag_a_00488. https://doi.org/10.1162/imag_a_00488
- Bobak, A. K., Parris, B. A., Gregory, N. J., Bennetts, R. J., & Bate, S. (2017). Eye-Movement Strategies in Developmental Prosopagnosia and “Super” Face Recognition. *Quarterly Journal of Experimental Psychology*, 70(2), 201–217. <https://doi.org/10.1080/17470218.2016.1161059>
- Borovska, P., & de Haas, B. (2024). Individual gaze shapes diverging neural representations. *Proceedings of the National Academy of Sciences*, 121(36), e2405602121.
<https://doi.org/10.1073/pnas.2405602121>
- 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.22.12.9>
- Broda, M. D., & de Haas, B. (2022b). Individual fixation tendencies in person viewing generalize from images to videos. *I-Perception*, 13(6), 20416695221128844. <https://doi.org/10.1177/20416695221128844>
- Buswell, G. T. (1935). *How people look at pictures: A study of the psychology and perception in art* (S. 198). Univ. Chicago Press.
- Cerf, M., Frady, E. P., & Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of Vision*, 9(12), 10. <https://doi.org/10.1167/9.12.10>
- 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, X., Fang, H., Lin, T.-Y., Vedantam, R., Gupta, S., Dollar, P., & Zitnick, C. L. (2015). *Microsoft COCO Captions: Data Collection and Evaluation Server* (arXiv:1504.00325). arXiv.
<https://doi.org/10.48550/arXiv.1504.00325>

- Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences*, 102(35), 12629–12633.
<https://doi.org/10.1073/pnas.0506162102>
- Coco, M. I., & Keller, F. (2012). Scan Patterns Predict Sentence Production in the Cross-Modal Processing of Visual Scenes. *Cognitive Science*, 36(7), 1204–1223. <https://doi.org/10.1111/j.1551-6709.2012.01246.x>
- Constantino, J. N., Kennon-McGill, S., Weichselbaum, C., Marrus, N., Haider, A., Glowinski, A. L., Gillespie, S., Klaiman, C., Klin, A., & Jones, W. (2017). Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature*, 547(7663), 340–344. <https://doi.org/10.1038/nature22999>
- Contier, O., Baker, C. I., & Hebart, M. N. (2024). Distributed representations of behaviour-derived object dimensions in the human visual system. *Nature Human Behaviour*, 8(11), 2179–2193.
<https://doi.org/10.1038/s41562-024-01980-y>
- Corrow, S. L., Davies-Thompson, J., Fletcher, K., Hills, C., Corrow, J. C., & Barton, J. J. S. (2019). Training face perception in developmental prosopagnosia through perceptual learning. *Neuropsychologia*, 134, 107196. <https://doi.org/10.1016/j.neuropsychologia.2019.107196>
- Curcio, C. A., & Allen, K. A. (1990). Topography of ganglion cells in human retina. *Journal of Comparative Neurology*, 300(1), 5–25. <https://doi.org/10.1002/cne.903000103>
- Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *Journal of Comparative Neurology*, 292(4), 497–523. <https://doi.org/10.1002/cne.902920402>
- Daniel, P. M., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *The Journal of Physiology*, 159(2), 203–221.
<https://doi.org/10.1113/jphysiol.1961.sp006803>
- de Haas, B., Iakovidis, A. L., Schwarzkopf, D. S., & Gegenfurtner, K. R. (2019). Individual differences in visual salience vary along semantic dimensions. 116, 11687–11692.
<https://doi.org/10.1073/pnas.1820553116>

- Deen, B., Richardson, H., Dilks, D. D., Takahashi, A., Keil, B., Wald, L. L., Kanwisher, N., & Saxe, R. (2017). Organization of high-level visual cortex in human infants. *Nature Communications*, 8(1), 13995. <https://doi.org/10.1038/ncomms13995>
- DeGutis, J. M., Chiu, C., Grosso, M. E., & Cohan, S. (2014). Face processing improvements in prosopagnosia: Successes and failures over the last 50 years. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00561>
- Dehaene, S., & Cohen, L. (2007). Cultural Recycling of Cortical Maps. *Neuron*, 56(2), 384–398. <https://doi.org/10.1016/j.neuron.2007.10.004>
- Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, 15(6), 254–262. <https://doi.org/10.1016/j.tics.2011.04.003>
- Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., & Cohen, L. (2010). How Learning to Read Changes the Cortical Networks for Vision and Language. *Science*, 330(6009), 1359–1364. <https://doi.org/10.1126/science.1194140>
- Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLOS Biology*, 16(3), e2004103. <https://doi.org/10.1371/journal.pbio.2004103>
- Della Longa, L., Gliga, T., & Farroni, T. (2019). Tune to touch: Affective touch enhances learning of face identity in 4-month-old infants. *Developmental Cognitive Neuroscience, Social Touch: A new vista for developmental cognitive neuroscience?*, 35, 42–46. <https://doi.org/10.1016/j.dcn.2017.11.002>
- Dettmer, A. M., Kaburu, S. S. K., Simpson, E. A., Paukner, A., Sclafani, V., Byers, K. L., Murphy, A. M., Miller, M., Marquez, N., Miller, G. M., Suomi, S. J., & Ferrari, P. F. (2016). Neonatal face-to-face interactions promote later social behaviour in infant rhesus monkeys. *Nature Communications*, 7(1), 11940. <https://doi.org/10.1038/ncomms11940>
- Einh user, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, 8(2), 2. <https://doi.org/10.1167/8.2.2>

- 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>
- Engeser, M., Babaei, N., & Kaiser, D. (2026). *Can Individual Internal Models Predict Idiosyncratic Scene Exploration?* (S. 2026.04.01.715777). bioRxiv. <https://doi.org/10.64898/2026.04.01.715777>
- Fan, X., Wang, L., Shao, H., Kersten, D., & He, S. (2016). Temporally flexible feedback signal to foveal cortex for peripheral object recognition. *Proceedings of the National Academy of Sciences*, 113(41), 11627–11632. <https://doi.org/10.1073/pnas.1606137113>
- Farah, M. J. (2004). *Visual Agnosia, second edition*. MIT Press.
- Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, 7(1), 10. <https://doi.org/10.1167/7.1.10>
- Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, 8(2), 181–195. [https://doi.org/10.1016/S0163-6383\(85\)80005-9](https://doi.org/10.1016/S0163-6383(85)80005-9)
- Field, T. (1992). Infants of depressed mothers. *Development and Psychopathology*, 4(1), 49–66. <https://doi.org/10.1017/S0954579400005551>
- 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, 116828. <https://doi.org/10.1016/j.neuroimage.2020.116828>
- Foulsham, T., & Underwood, G. (2011). If Visual Saliency Predicts Search, Then Why? Evidence from Normal and Gaze-Contingent Search Tasks in Natural Scenes. *Cognitive Computation*, 3(1), 48–63. <https://doi.org/10.1007/s12559-010-9069-9>
- Friston, K., Adams, R., Perrinet, L., & Breakspear, M. (2012). Perceptions as Hypotheses: Saccades as Experiments. *Frontiers in Psychology*, 3. <https://doi.org/10.3389/fpsyg.2012.00151>
- Furl, N., Garrido, L., Dolan, R. J., Driver, J., & Duchaine, B. (2011). Fusiform Gyrus Face Selectivity Relates to Individual Differences in Facial Recognition Ability. *Journal of Cognitive Neuroscience*, 23(7), 1723–1740. <https://doi.org/10.1162/jocn.2010.21545>

- Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, 3(2), 191–197.
<https://doi.org/10.1038/72140>
- Gegenfurtner, K. R. (2016). The Interaction Between Vision and Eye Movements†. *Perception*, 45(12), 1333–1357. <https://doi.org/10.1177/0301006616657097>
- Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition*, 118(2), 201–210.
<https://doi.org/10.1016/j.cognition.2010.11.002>
- Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, 14(5), 350–363. <https://doi.org/10.1038/nrn3476>
- Goetschalckx, L., Moors, P., & Wagemans, J. (2018). Image memorability across longer time intervals. *Memory*, 26(5), 581–588. <https://doi.org/10.1080/09658211.2017.1383435>
- Golarai, G., Liberman, A., Yoon, J. M. D., & Grill-Spector, K. (2010). Differential development of the ventral visual cortex extends through adolescence. *Frontiers in Human Neuroscience*, 3.
<https://doi.org/10.3389/neuro.09.080.2009>
- Gomez, J., Barnett, M., & Grill-Spector, K. (2019). Extensive childhood experience with Pokémon suggests eccentricity drives organization of visual cortex. *Nature Human Behaviour*, 3(6), 611–624.
<https://doi.org/10.1038/s41562-019-0592-8>
- Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). Learning word vectors for 157 languages. *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*. <https://aclanthology.org/L18-1550.pdf>
- Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, 15(8), 536–548.
<https://doi.org/10.1038/nrn3747>

- Guy, N., Azulay, H., Kardosh, R., Weiss, Y., Hassin, R. R., Israel, S., & Pertzov, Y. (2019). A novel perceptual trait: Gaze predilection for faces during visual exploration. *Scientific Reports*, 9(1), 10714.
<https://doi.org/10.1038/s41598-019-47110-x>
- Han, C., & Bonner, M. F. (2026). High-dimensional structure underlying individual differences in naturalistic visual experience. *Current Biology*, 36(3), 723-733.e6. <https://doi.org/10.1016/j.cub.2025.12.039>
- Harel, J., Koch, C., & Perona, P. (2006). Graph-based visual saliency. *Advances in neural information processing systems*, 19.
https://proceedings.neurips.cc/paper_files/paper/2006/hash/4db0f8b0fc895da263fd77fc8aecabe4-Abstract.html
- 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](https://doi.org/10.1016/S0896-6273(02)00662-1)
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, 293(5539), 2425–2430. <https://doi.org/10.1126/science.1063736>
- Hayes, T. R., & Henderson, J. M. (2017). Scan patterns during real-world scene viewing predict individual differences in cognitive capacity. *Journal of Vision*, 17(5), 23. <https://doi.org/10.1167/17.5.23>
- Hayes, T. R., & Henderson, J. M. (2018). Scan patterns during scene viewing predict individual differences in clinical traits in a normative sample. *PLOS ONE*, 13(5), e0196654.
<https://doi.org/10.1371/journal.pone.0196654>
- Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, 3(1), 6. <https://doi.org/10.1167/3.1.6>
- Henderson, J. M., & Castelano, M. S. (2005). Eye movements and visual memory for scenes. *Cognitive processes in eye guidance*, 213–235.

- Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behaviour*, 1(10), 743–747. <https://doi.org/10.1038/s41562-017-0208-0>
- Henderson, J. M., Hayes, T. R., Rehrig, G., & Ferreira, F. (2018). Meaning Guides Attention during Real-World Scene Description. *Scientific Reports*, 8(1), 13504. <https://doi.org/10.1038/s41598-018-31894-5>
- Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review*, 16(5), 850–856. <https://doi.org/10.3758/PBR.16.5.850>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- 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>
- Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011). What makes an image memorable? *CVPR 2011*, 145–152. <https://ieeexplore.ieee.org/abstract/document/5995721/>
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194–203. <https://doi.org/10.1038/35058500>
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11), 1254–1259. <https://doi.org/10.1109/34.730558>
- James, W. (1890). Attention. *The principles of psychology*, 1, 402–458.
- Jas, M., & Parikh, D. (2015). *Image Specificity*. 2727–2736. https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Jas_Image_Specificity_2015_CVPR_paper.html
- Jiang, M., & Zhao, Q. (2017). *Learning Visual Attention to Identify People With Autism Spectrum Disorder*. 3267–3276.

https://openaccess.thecvf.com/content_iccv_2017/html/Jiang_Learning_Visual_Attention_ICCV_2017_paper.html

Johnson, B. A., Soni, A. K., Murlidaran, S., Beyeler, M., & Eckstein, M. P. (2025). *Distinct Roles of Central and Peripheral Vision in Rapid Scene Understanding* (S. 2025.07.21.665750). bioRxiv.

<https://doi.org/10.1101/2025.07.21.665750>

Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage*, 60(4), 2357–2364.

<https://doi.org/10.1016/j.neuroimage.2012.02.055>

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*, 107(25), 11163–11170.

<https://doi.org/10.1073/pnas.1005062107>

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1476), 2109–2128.

<https://doi.org/10.1098/rstb.2006.1934>

Kennedy, D. P., D’Onofrio, B. M., Quinn, P. D., Bölte, S., Lichtenstein, P., & Falck-Ytter, T. (2017). Genetic Influence on Eye Movements to Complex Scenes at Short Timescales. *Current Biology*, 27(22), 3554–

3560.e3. <https://doi.org/10.1016/j.cub.2017.10.007>

Kleinschmidt, A., & Cohen, L. (2006). The neural bases of prosopagnosia and pure alexia: Recent insights from functional neuroimaging. *Current Opinion in Neurology*, 19(4), 386.

<https://doi.org/10.1097/01.wco.0000236619.89710.ee>

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual Fixation Patterns During Viewing of Naturalistic Social Situations as Predictors of Social Competence in Individuals With Autism. *Archives of General Psychiatry*, 59(9), 809–816. <https://doi.org/10.1001/archpsyc.59.9.809>

Kosakowski, H. L., Cohen, M. A., Takahashi, A., Keil, B., Kanwisher, N., & Saxe, R. (2022). Selective responses to faces, scenes, and bodies in the ventral visual pathway of infants. *Current Biology*, 32(2), 265–

274.e5. <https://doi.org/10.1016/j.cub.2021.10.064>

- Kubota, E., Grill-Spector, K., & Nordt, M. (2024). Rethinking cortical recycling in ventral temporal cortex. *Trends in Cognitive Sciences*, 28(1), 8–17. <https://doi.org/10.1016/j.tics.2023.09.006>
- Kumar, M., & Kapoor, A. (2024). *Generative AI and Personalized Video Advertisements* (SSRN Scholarly Paper Nr. 4614118). Social Science Research Network. <https://doi.org/10.2139/ssrn.4614118>
- Land, M., Mennie, N., & Rusted, J. (1999). The Roles of Vision and Eye Movements in the Control of Activities of Daily Living. *Perception*, 28(11), 1311–1328. <https://doi.org/10.1068/p2935>
- 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>
- Li, A., & Chen, Z. (2018). Personalized Visual Saliency: Individuality Affects Image Perception. *IEEE Access*, 6, 16099–16109. <https://doi.org/10.1109/ACCESS.2018.2800294>
- Li, H., Zhang, X., Yang, Y., & Xie, A. (2023). Abnormal eye movements in Parkinson’s disease: From experimental study to clinical application. *Parkinsonism & Related Disorders*, 115, 105791. <https://doi.org/10.1016/j.parkreldis.2023.105791>
- Liang, H., Jiang, M., Liang, R., & Zhao, Q. (2018). CapVis: Toward Better Understanding of Visual-Verbal Saliency Consistency. *ACM Trans. Intell. Syst. Technol.*, 10(1), 10:1-10:23. <https://doi.org/10.1145/3200767>
- 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), 13. <https://doi.org/10.1167/jov.20.9.13>
- Linka, M., Karimpur, H., & de Haas, B. (2025). Protracted development of gaze behaviour. *Nature Human Behaviour*, 9(9), 1887–1897. <https://doi.org/10.1038/s41562-025-02191-9>
- Linka, M., Sensoy, Ö., Karimpur, H., Schwarzer, G., & de Haas, B. (2023). Free viewing biases for complex scenes in preschoolers and adults. *Scientific Reports*, 13(1), 11803. <https://doi.org/10.1038/s41598-023-38854-8>

- Lowes, J., McGregor, L. M., Hancock, P. J. B., Duchaine, B., & Bobak, A. K. (2025). This condition impacts every aspect of my life: A survey to understand the experience of living with developmental prosopagnosia. *PLOS ONE*, 20(4), e0322469. <https://doi.org/10.1371/journal.pone.0322469>
- Mack, A. (2003). Inattentional Blindness: Looking Without Seeing. *Current Directions in Psychological Science*, 12(5), 180–184. <https://doi.org/10.1111/1467-8721.01256>
- Malach, R., Levy, I., & Hasson, U. (2002). The topography of high-order human object areas. *Trends in Cognitive Sciences*, 6(4), 176–184. [https://doi.org/10.1016/S1364-6613\(02\)01870-3](https://doi.org/10.1016/S1364-6613(02)01870-3)
- Matz, S. C., Teeny, J. D., Vaid, S. S., Peters, H., Harari, G. M., & Cerf, M. (2024). The potential of generative AI for personalized persuasion at scale. *Scientific Reports*, 14(1), 4692. <https://doi.org/10.1038/s41598-024-53755-0>
- Mikolov, T., Grave, E., Bojanowski, P., Puhersch, C., & Joulin, A. (2018). Advances in pre-training distributed word representations. *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*. <https://aclanthology.org/L18-1008.pdf>
- Mitchell, J. L., Yablonski, M., Stone, H. L., Jimenez, M., Takada, M. E., Tang, K. A., Tran, J. E., Chou, C., & Yeatman, J. D. (2025). The balance between stability and plasticity of the visual word form area in dyslexia. *Nature Communications*, 17(1), 368. <https://doi.org/10.1038/s41467-025-67054-3>
- Molitor, R. J., Ko, P. C., & Ally, B. A. (2015). Eye Movements in Alzheimer’s Disease. *Journal of Alzheimer’s Disease*, 44(1), 1–12. <https://doi.org/10.3233/JAD-141173>
- Murlidaran, S., & Eckstein, M. P. (2025). Eye movements during free viewing to maximize scene understanding. *Nature Communications*, 17(1), 940. <https://doi.org/10.1038/s41467-025-67673-w>
- Natu, V. S., Rosenke, M., Wu, H., Querdasi, F. R., Kular, H., Lopez-Alvarez, N., Grotheer, M., Berman, S., Mezer, A. A., & Grill-Spector, K. (2021). Infants’ cortex undergoes microstructural growth coupled with myelination during development. *Communications Biology*, 4(1), 1191. <https://doi.org/10.1038/s42003-021-02706-w>
- Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: Holistic versus analytic perception. *Trends in Cognitive Sciences*, 9(10), 467–473. <https://doi.org/10.1016/j.tics.2005.08.004>

- Noad, K. N., & Andrews, T. J. (2025). Differences in face recognition predict the understanding of events during natural viewing. *Perception*, 03010066251401485.
<https://doi.org/10.1177/03010066251401485>
- Nordt, M., Gomez, J., Natu, V., Jeska, B., Barnett, M., & Grill-Spector, K. (2019). Learning to Read Increases the Informativeness of Distributed Ventral Temporal Responses. *Cerebral Cortex*, 29(7), 3124–3139.
<https://doi.org/10.1093/cercor/bhy178>
- Nordt, M., Gomez, J., Natu, V. S., Rezai, A. A., Finzi, D., Kular, H., & Grill-Spector, K. (2021). Cortical recycling in high-level visual cortex during childhood development. *Nature Human Behaviour*, 5(12), 1686–1697. <https://doi.org/10.1038/s41562-021-01141-5>
- Nordt, M., Gomez, J., Natu, V. S., Rezai, A. A., Finzi, D., Kular, H., & Grill-Spector, K. (2023). Longitudinal development of category representations in ventral temporal cortex predicts word and face recognition. *Nature Communications*, 14(1), 8010. <https://doi.org/10.1038/s41467-023-43146-w>
- Papesh, M. H. (2018). Photo ID verification remains challenging despite years of practice. *Cognitive Research: Principles and Implications*, 3(1), 19. <https://doi.org/10.1186/s41235-018-0110-y>
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42(1), 107–123. [https://doi.org/10.1016/S0042-6989\(01\)00250-4](https://doi.org/10.1016/S0042-6989(01)00250-4)
- Parr, L. A. (2011). The evolution of face processing in primates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1571), 1764–1777. <https://doi.org/10.1098/rstb.2010.0358>
- Parr, T., & Friston, K. J. (2019). Generalised free energy and active inference. *Biological Cybernetics*, 113(5), 495–513. <https://doi.org/10.1007/s00422-019-00805-w>
- 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>
- Peacock, C. E., Hayes, T. R., & Henderson, J. M. (2019). Meaning guides attention during scene viewing, even when it is irrelevant. *Attention, Perception, & Psychophysics*, 81(1), 20–34.
<https://doi.org/10.3758/s13414-018-1607-7>

- Peelen, M. V., & Downing, P. E. (2005). Within-subject reproducibility of category-specific visual activation with functional MRI. *Human Brain Mapping, 25*(4), 402–408. <https://doi.org/10.1002/hbm.20116>
- Pertsov, Y., Avidan, G., & Zohary, E. (2009). Accumulation of visual information across multiple fixations. *Journal of Vision, 9*(10), 2. <https://doi.org/10.1167/9.10.2>
- 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), 12. <https://doi.org/10.1167/16.7.12>
- Pezzulo, G., Parr, T., & Friston, K. (2024). Active inference as a theory of sentient behavior. *Biological Psychology, 186*, 108741. <https://doi.org/10.1016/j.biopsycho.2023.108741>
- Polk, T. A., Park, J., Smith, M. R., & Park, D. C. (2007). Nature versus Nurture in Ventral Visual Cortex: A Functional Magnetic Resonance Imaging Study of Twins. *Journal of Neuroscience, 27*(51), 13921–13925. <https://doi.org/10.1523/JNEUROSCI.4001-07.2007>
- Powell, L. J., Kosakowski, H. L., & Saxe, R. (2018). Social Origins of Cortical Face Areas. *Trends in Cognitive Sciences, 22*(9), 752–763. <https://doi.org/10.1016/j.tics.2018.06.009>
- Reeder, R. R., Stein, T., & Peelen, M. V. (2016). Perceptual expertise improves category detection in natural scenes. *Psychonomic Bulletin & Review, 23*(1), 172–179. <https://doi.org/10.3758/s13423-015-0872-x>
- Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, 3982–3992. <https://aclanthology.org/D19-1410/>
- Risko, E. F., Laidlaw, K. E., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with real versus reel stimuli: Toward an empirical approach to concerns about ecological validity. *Frontiers in Human Neuroscience, 6*. <https://doi.org/10.3389/fnhum.2012.00143>
- Rockland, K. S., & Van Hoesen, G. W. (1994). Direct Temporal-Occipital Feedback Connections to Striate Cortex (V1) in the Macaque Monkey. *Cerebral Cortex, 4*(3), 300–313. <https://doi.org/10.1093/cercor/4.3.300>

- Rosenholtz, R. (2016). Capabilities and Limitations of Peripheral Vision. *Annual Review of Vision Science*, 2(Volume 2, 2016), 437–457. <https://doi.org/10.1146/annurev-vision-082114-035733>
- Roser, M., & Ortiz-Ospina, E. (2018). Literacy. *Our World in Data*. <https://ourworldindata.org/literacy>
- Rossion, B., Kung, C.-C., & Tarr, M. J. (2004). Visual expertise with nonface objects leads to competition with the early perceptual processing of faces in the human occipitotemporal cortex. *Proceedings of the National Academy of Sciences*, 101(40), 14521–14526. <https://doi.org/10.1073/pnas.0405613101>
- Saygin, Z. M., Osher, D. E., Norton, E. S., Youssoufian, D. A., Beach, S. D., Feather, J., Gaab, N., Gabrieli, J. D. E., & Kanwisher, N. (2016). Connectivity precedes function in the development of the visual word form area. *Nature Neuroscience*, 19(9), 1250–1255. <https://doi.org/10.1038/nn.4354>
- Schütt, H. H., Rothkegel, L. O. M., Trukenbrod, H. A., Engbert, R., & Wichmann, F. A. (2019). Disentangling bottom-up versus top-down and low-level versus high-level influences on eye movements over time. *Journal of Vision*, 19(3), 1. <https://doi.org/10.1167/19.3.1>
- Sheinberg, D. L., & Logothetis, N. K. (2001). Noticing Familiar Objects in Real World Scenes: The Role of Temporal Cortical Neurons in Natural Vision. *Journal of Neuroscience*, 21(4), 1340–1350. <https://doi.org/10.1523/JNEUROSCI.21-04-01340.2001>
- Stein, T., Reeder, R. R., & Peelen, M. V. (2016). Privileged access to awareness for faces and objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance*, 42(6), 788–798. <https://doi.org/10.1037/xhp0000188>
- Stewart, E. E. M., Valsecchi, M., & Schütz, A. C. (2020). A review of interactions between peripheral and foveal vision. *Journal of Vision*, 20(12), 2. <https://doi.org/10.1167/jov.20.12.2>
- Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, 11(5), 13. <https://doi.org/10.1167/11.5.13>
- Susilo, T., Wright, V., Tree, J. J., & Duchaine, B. (2015). Acquired prosopagnosia without word recognition deficits. *Cognitive Neuropsychology*, 32(6), 321–339. <https://doi.org/10.1080/02643294.2015.1081882>

- 't Hart, B. M., & Einhäuser, W. (2012). Mind the step: Complementary effects of an implicit task on eye and head movements in real-life gaze allocation. *Experimental Brain Research*, 223(2), 233–249.
<https://doi.org/10.1007/s00221-012-3254-x>
- Tasliyurt-Celebi, S., de Haas, B., Vö, M. L. H., & Dobs, K. (2024). Using CNNs to understand how bottom-up and top-down processes shape human face detection. *Cognitive Computational Neuroscience (CCN)*.
https://2024.ccneuro.org/pdf/155_Paper_authored_CCN_2024_Sule_Tasliyurt_Celebi_Authored.pdf
- Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & Bülthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *European Journal of Neuroscience*, 14(5), 869–876.
<https://doi.org/10.1046/j.0953-816x.2001.01717.x>
- Torrallba, A., Oliva, A., Castelhana, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113(4), 766–786. <https://doi.org/10.1037/0033-295X.113.4.766>
- van den Hurk, J., Van Baelen, M., & Op de Beeck, H. P. (2017). Development of visual category selectivity in ventral visual cortex does not require visual experience. *Proceedings of the National Academy of Sciences*, 114(22), E4501–E4510. <https://doi.org/10.1073/pnas.1612862114>
- van Paridon, J., Ostarek, M., Arunkumar, M., & Huettig, F. (2021). Does Neuronal Recycling Result in Destructive Competition? The Influence of Learning to Read on the Recognition of Faces. *Psychological Science*, 32(3), 459–465. <https://doi.org/10.1177/0956797620971652>
- Varela, V. P. L., Towler, A., Kemp, R. I., & White, D. (2023). Looking at faces in the wild. *Scientific Reports*, 13(1), 783. <https://doi.org/10.1038/s41598-022-25268-1>
- Vater, C., Wolfe, B., & Rosenholtz, R. (2022). Peripheral vision in real-world tasks: A systematic review. *Psychonomic Bulletin & Review*, 29(5), 1531–1557. <https://doi.org/10.3758/s13423-022-02117-w>
- Vinken, K., & Vogels, R. (2019). A behavioral face preference deficit in a monkey with an incomplete face patch system. *NeuroImage*, 189, 415–424. <https://doi.org/10.1016/j.neuroimage.2019.01.043>

- Võ, M. L.-H., Boettcher, S. E., & Draschkow, D. (2019). Reading scenes: How scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology, Attention & Perception*, 29, 205–210. <https://doi.org/10.1016/j.copsyc.2019.03.009>
- Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual Field Maps in Human Cortex. *Neuron*, 56(2), 366–383. <https://doi.org/10.1016/j.neuron.2007.10.012>
- Wang, G., Foxwell, M. J., Cichy, R. M., Pitcher, D., & Kaiser, D. (2024). Individual differences in internal models explain idiosyncrasies in scene perception. *Cognition*, 245, 105723. <https://doi.org/10.1016/j.cognition.2024.105723>
- Wang, H.-C., & Pomplun, M. (2012). The attraction of visual attention to texts in real-world scenes. *Journal of Vision*, 12(6), 26. <https://doi.org/10.1167/12.6.26>
- Wang, S., Jiang, M., Duchesne, X. M., Laugeson, E. A., Kennedy, D. P., Adolphs, R., & Zhao, Q. (2015). Atypical Visual Saliency in Autism Spectrum Disorder Quantified through Model-Based Eye Tracking. *Neuron*, 88(3), 604–616. <https://doi.org/10.1016/j.neuron.2015.09.042>
- Watson, A. B. (2014). A formula for human retinal ganglion cell receptive field density as a function of visual field location. *Journal of Vision*, 14(7), 15. <https://doi.org/10.1167/14.7.15>
- Watson, D. M., Noad, K. N., Quinn, B. P. A., & Andrews, T. J. (2026). Patterns of functional connectivity differentiate individuals and individual regions in face and scene selective networks. *Neuropsychologia*, 225, 109416. <https://doi.org/10.1016/j.neuropsychologia.2026.109416>
- Weiner, K. S., & Grill-Spector, K. (2010). Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *NeuroImage*, 52(4), 1559–1573. <https://doi.org/10.1016/j.neuroimage.2010.04.262>
- White, D., Kemp, R. I., Jenkins, R., & Burton, A. M. (2014). Feedback training for facial image comparison. *Psychonomic Bulletin & Review*, 21(1), 100–106. <https://doi.org/10.3758/s13423-013-0475-3>
- Williams, M. A., Baker, C. I., Op de Beeck, H. P., Mok Shim, W., Dang, S., Triantafyllou, C., & Kanwisher, N. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nature Neuroscience*, 11(12), 1439–1445. <https://doi.org/10.1038/nn.2218>

- Wolfe, B. A., & Whitney, D. (2014). Facilitating recognition of crowded faces with presaccadic attention. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00103>
- Wolfe, B. A., & Whitney, D. (2015). Saccadic remapping of object-selective information. *Attention, Perception, & Psychophysics*, 77(7), 2260–2269. <https://doi.org/10.3758/s13414-015-0944-z>
- 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*, 27(6), 1157–1166. <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*. 14. <https://doi.org/10.1167/14.1.28>
- Xu, Y., Gao, S., Wu, J., Li, N., & Yu, J. (2019). Personalized Saliency and Its Prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(12), 2975–2989. <https://doi.org/10.1109/TPAMI.2018.2866563>
- Yarbus, A. L. (1967). Eye Movements During Perception of Complex Objects. In A. L. Yarbus (Hrsg.), *Eye Movements and Vision* (S. 171–211). Springer US. https://doi.org/10.1007/978-1-4899-5379-7_8
- 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. <https://doi.org/10.3389/fpsyg.2013.00917>
- Zhaoping, L. (2024). Peripheral vision is mainly for looking rather than seeing. *Neuroscience Research, How to Study Subjective Experience*, 201, 18–26. <https://doi.org/10.1016/j.neures.2023.11.006>

II. Publications

Study 1: Active vision is linked to category selectivity in the individual brain

Kollenda, D., Akbari, E., Broda, M. D., & de Haas, B. (2026, April 16). Active vision is linked to category selectivity in the individual brain. *Psyarxiv*

Accepted for publication in *Nature Human Behaviour*

1 **Active vision is linked to category selectivity in the individual brain**

2 Diana Kollenda^{*1,2}, Elaheh Akbari^{1,2}, Maximilian D. Broda^{1,2} & Benjamin de Haas^{1,2}

3 1 Experimental Psychology, Justus-Liebig University Giessen, Germany

4 2 Center for Mind, Brain and Behaviour (CMBB), Marburg and Giessen, Germany

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20 *diana.kollenda@googlemail.com
21 Experimental Psychology, Justus-Liebig University,
22 Giessen, Germany

23

24

Abstract

25 Individuals reliably differ in how they look at complex visual scenes, with the most prominent
26 variation in their propensity to fixate faces and text. In 61 adults, we tested the hypothesis
27 that these differences in gaze are linked to representational properties of the individual visual
28 system. Eye-tracking captured each observer's characteristic gaze tendencies during
29 naturalistic scene viewing, and independent functional magnetic resonance imaging recorded
30 category-selective responses to faces, words, and other stimuli when participants were
31 instructed to fixate centrally. We find that the propensity to fixate faces or text goes along
32 with enhanced distinctiveness and enlarged functional regions of corresponding categorical
33 representations in the ventral stream. These in turn predicted performance on reading and
34 face recognition tasks. Thus, active vision appears linked to the precision of category-
35 selective encoding and corresponding neural resources in the individual brain.

36

Introduction

37 Vision is an active process that requires constant eye movements to explore different points
38 of interest with high acuity. Typically, this process is modelled based on group averages¹⁻³.
39 However, recent findings have documented robust differences between individuals viewing
40 the same complex scenes, most prominently in the propensity to fixate faces and text⁴⁻⁷.
41 These gaze preferences are linked to a tendency to prioritize corresponding elements when
42 describing a scene⁸ and, at least for faces, to category-specific recognition skills.^{5,9} Despite
43 the striking effect sizes of individual differences in gaze^{5,7} and their known genetic
44 component^{10,11}, the underlying biological mechanisms are unclear. Why do some individuals
45 fixate faces two to three times more often than other observers, even when viewing identical
46 stimuli? Why does the same hold for the tendency to fixate text?

47 Several lines of research suggest that face and text fixation preferences may be linked to
48 representational properties of the ventral stream. Just as individual differences in the
49 functional anatomy of early visual cortex predict low-level perceptual and psychophysical
50 biases^{12,13}, face recognition performance correlates with the degree of face selectivity and

51 the size of the right fusiform face area^{14,15}. Moreover, individual differences in the tendency to
52 fixate faces and text while watching movie clips predict inter-individual representational
53 divergence in inferior temporal cortex¹⁶. Generally, for adult readers, faces and text are both,
54 the most salient types of fixation targets during scene viewing^{5,17} and the most prominent
55 drivers of neural activity in distinct patches of the ventral stream¹⁸⁻²⁰. Crucially, recent data
56 from non-human primates show strongly reduced face fixation preferences in animals that
57 lack ventral face patches, either naturally²¹ or due to experimentally induced developmental
58 face deprivation²². Interestingly, the animals in these studies also showed an increased
59 tendency to fixate hands, which at least in Arcaro et al.'s. (2017)²² study went along with
60 increased neural selectivity for hands in the typical anatomical location of face patches. This
61 resonates with recent human studies which suggest a developmental push-pull relationship
62 between ventral selectivity for faces and hands²³. Further, recent developmental studies have
63 documented that the unfolding of text and face salience is protracted well into
64 adolescence^{24,25}, which parallels the development of corresponding ventral category
65 representations^{23,26-29}. Specifically, Nordt et al., (2023)²⁶ employed a measure of
66 representational distinctiveness, which captures both the reliability and specificity of
67 multivariate activation patterns evoked by a given stimulus category. Their data show that the
68 distinctiveness of word and face representations continuously increases into adolescence, in
69 parallel to reading and face recognition skills. This is strikingly similar to the findings of the
70 large-scale eye-tracking study by Linka et al. (2025)²⁵, who found that the proportion of
71 immediate saccades towards faces and text during scene viewing continuously increases
72 from age five to at least 16.

73 These parallel developmental trajectories raise the possibility of a systematic link between
74 active vision and representational spaces. Specifically, gaze tendencies may follow neural
75 development, actively prioritizing semantic categories which can be represented more
76 distinctively. Vice versa, developing gaze tendencies and changes in visual diet may shape
77 representational distinctiveness. These hypotheses are not mutually exclusive and share a
78 straightforward prediction: If there is a systematic link between gaze and representational

79 distinctiveness, large individual differences in gaze tendencies among adults should correlate
80 with corresponding category distinctiveness in the individual brain. To test these predictions,
81 we quantified the individual representational distinctiveness of faces and words²⁶ in the
82 ventral temporal cortex (VTC) of 61 adult observers for whom we also quantified
83 corresponding gaze tendencies in an independent free-viewing experiment^{5,7}.

84

85 **Results**

86 **Replication of stable individual gaze tendencies toward faces and text**

87 Gaze tendencies towards faces and text have been shown to be consistent across hundreds
88 of complex scenes as well as across time^{5,7}. This is true not only for the total duration spent
89 on a given semantic category (i.e. cumulative dwell time), but also for the tendency to target
90 a given category with the first saccade after image onset. Here, we replicate these findings in
91 102 participants. For the individual proportion of first fixations on faces, we found excellent
92 split-half reliability across odd and even images, $r(100) = .93$, $p < .001$, 95% CI [0.90, 0.95]
93 (maximum-minimum ratio of 2.04). The same was true for the individual proportion of
94 cumulative dwell time on faces, $r(100) = .95$, $p < .001$, 95% CI [0.92, 0.97] (max-min ratio of
95 1.66). Similarly, we found high split-half reliability for the individual proportion of first fixations
96 on text elements, $r(100) = .87$, $p < .001$, 95% CI [0.80, 0.91] (max-min ratio of 4.71), as well
97 as for the proportion of cumulative dwell time on text elements, $r(100) = .88$, $p < .001$, 95% CI
98 [0.84, 0.91] (max-min ratio of 2.16). In line with previous studies, there was a strong anti-
99 correlation between the individual tendency to fixate faces or text (first fixations: $r(100) = -.50$,
100 $p < .001$, 95% CI [-0.64, -0.34]; dwell time: $r(100) = -.52$, $p < .001$, 95% CI [-0.65, -0.36]).
101 However, these anticorrelations were weaker when excluding the 175/700 images containing
102 both faces and text, which put them in direct competition for attention (first fixations: $r(100) =$
103 $-.15$, $p = .136$, CI [-0.33, 0.05], dwell time: $r(100) = -.27$, $p = .006$, 95%, CI [-0.44, -0.09]).
104 Taken together, we could replicate large and consistent individual differences in the tendency

105 to fixate faces and text (Figure 1), which are strongly anticorrelated when brought into direct
106 competition for attention.

107

108 **Representational distinctiveness is linked to individual gaze tendencies**

109 In a separate session, we tested 61 of our eye-tracking participants in a standard functional
110 localizer experiment in a magnetic resonance imaging scanner (subsequently abbreviated as
111 fMRI session). This allowed us to identify specific patterns of neural activity associated with
112 different object categories in the ventral temporal cortex (VTC) of each individual. These
113 categories included faces and pseudowords/characters along with a number of control
114 categories (six categories in total: characters, headless bodies, faces, houses, cars and
115 limbs; cf. ³⁰). Following Nordt et al. (2023)²⁶, we determined the distinctiveness of the pattern
116 evoked by a given category as the subtraction of average between-category correlations
117 from within-category correlations (with the latter corresponding to pattern reliability, see
118 Figure 2 and methods). This measure of pattern distinctiveness has a theoretical range of -2
119 to 2, with higher values indicating more distinctive representations.

120

121 We calculated category distinctiveness separately for both hemispheres within individually
122 defined regions of interest of the lateral VTC, following the anatomical definition by Nordt et
123 al. (2023)²⁶ (see also ¹⁸). Lateral VTC is associated with a centre-biased recognition of form,
124 objects and categories such as faces³¹ and words/characters²⁹, while the medial part is
125 associated with a periphery bias and selectivity to places and scenes (cf. ^{30,32,33}). Notably,
126 there is a hemispheric lateralization, with character-selective regions predominantly in the left
127 hemisphere and face-selective regions primarily in the right³⁴. We replicated this lateralization
128 pattern in our sample, but found no association between the individual degree of
129 lateralization and gaze or perceptual performance (see Supplementary Methods and
130 Results).

131 The individual degree of representational distinctiveness for faces and words varied
132 considerably. Face distinctiveness in the right lateral VTC exhibited a max-min ratio of 1.38
133 across participants. Word distinctiveness in the left lateral VTC showed a max-min ratio of
134 1.52 across participants.

135 Next, we tested two primary a priori hypotheses: First, the individual tendency to fixate faces
136 is linked to face distinctiveness in the right lateral VTC and second, the individual tendency to
137 fixate text is linked to word distinctiveness in the left lateral VTC. We quantified category-
138 specific fixation tendencies as the proportions of first fixations and dwell time spent on
139 corresponding scene elements in the free-viewing experiment (and applied Bonferroni
140 corrected significance thresholds of $\alpha = .025$ to account for the use of these two measures).
141 Crucially, face distinctiveness in right lateral VTC significantly correlated with the tendency to
142 direct first fixations ($r(59) = .38, p = .003, 95\% \text{ CI } [0.15, 0.56]$) and dwell time ($r(59) = .30, p =$
143 $.017, 95\% \text{ CI } [0.13, 0.47]$) to faces in the independent free-viewing experiment (Figure 3a;
144 see also Table S1 for an overview of all correlations). Similarly, word distinctiveness in left
145 lateral VTC was correlated with the tendency to direct first fixations ($r(59) = .44, p < .001,$
146 $95\% \text{ CI } [0.25, 0.60]$), and dwell time ($r(59) = .29, p = .023, 95\% \text{ CI } [0.08, 0.49]$) on text
147 (Figure 3b).

148 Furthermore, we explored whether the observed relationships between neural and gaze
149 patterns are hemisphere-specific. Word distinctiveness in right lateral VTC was positively
150 correlated with the tendency to direct first fixations on text, $r(59) = .35, p = .006, 95\% \text{ CI}$
151 $[0.13, 0.55]$, but not dwell time, $r(59) = .14, p = .299, 95\% \text{ CI } [-0.05, 0.32]$. There were no
152 significant correlations between face distinctiveness in left lateral VTC and gaze tendencies
153 towards faces (first fixation and dwell time both $p(59) > .12$; see Table S1 for detailed
154 results). An exploratory whole-brain searchlight analysis replicated associations between
155 gaze tendencies and corresponding category distinctiveness in VTC. Moreover, we found
156 that the tendency to direct first fixations to text correlated with word distinctiveness in regions
157 of the extended reading network, especially in the left hemisphere (cf. ^{62,63}; see
158 Supplementary Methods and Results, including Figures S1 and S2).

159 Reassuringly, the observed correlations between representational distinctiveness and gaze
160 were category-specific. Cross-correlations between word distinctiveness in left lateral VTC
161 and gaze tendencies towards faces (first fixations and dwell time) were non-significant, both
162 $p(59) > .22$ (see Supplementary Table S1 for detailed results). Interestingly, face
163 distinctiveness in right lateral VTC correlated negatively with the tendency to direct first
164 fixations to text ($r(59) = -.36, p = .005, 95\% \text{ CI } [-0.55, -0.14]$), while dwell time on text did not
165 ($r(59) = -.24, p = .058, 95\% \text{ CI } [-0.43, -0.04], n.s.$).

166 Together, these findings show a robust and specific link between individual gaze tendencies
167 towards faces and text and the distinctiveness of corresponding representations in the
168 individual ventral stream.

169

170 **Representational distinctiveness is linked to behavioural performance**

171 Nordt et al. (2023)²⁶ reported that word distinctiveness in left and face distinctiveness in right
172 lateral VTC predicts reading and face recognition performance across children and
173 adolescents. To test whether this relationship extends to individual differences in our adult
174 sample, we invited participants back for additional testing and were able to do so in
175 subsamples. Specifically, we examined in 34 participants whether face distinctiveness in the
176 right lateral VTC correlates with face recognition ability measured with the Cambridge Face
177 Memory Test in the long format (CFMT)³⁵. In line with prior developmental findings²⁶, we
178 observed a positive correlation between face distinctiveness in right lateral VTC and the
179 CFMT score, $r(32) = .44$, one-tailed $p = .004, 95\% \text{ CI } [0.20, 0.64]$ (cf. Methods and Figure 3a,
180 right most plot). Furthermore, we investigated whether word distinctiveness in left lateral VTC
181 is linked to reading performance, measured with a speed-reading test for words and
182 pseudowords (SLRT-2)³⁶ in a subsample of 25 participants. Here, we observed a positive
183 correlation between word distinctiveness in left lateral VTC and the reading-speed scores for
184 pseudowords $r(23) = .44$, one-tailed $p = .012, 95\% \text{ CI } [0.03, 0.72]$ (cf. Methods and Figure
185 3b, right most plot). This also is in line with the previous study by Nordt et al. (2023)²⁶. There

186 was no significant correlation between word distinctiveness in left lateral VTC and reading-
187 speed scores for words, $r(23) = .17$, one-tailed $p = .204$, 95% CI [-0.30, 0.62].

188

189 **The individual sizes of rFFA and IVWFA are linked to gaze tendencies and behavioural** 190 **performance**

191 Our previous analyses focused on the multivariate measure of category distinctiveness within
192 anatomically defined ROIs, which develops in tandem to category-specific salience (cf. ^{25,26}).
193 However, many previous studies employed univariate measures of face and word-selectivity
194 (cf. ^{23,29,30}). Everything else staying equal, both types of measures are expected to scale with
195 the signal to noise ratio of category-specific responses. However, in principle they may also
196 vary independently. While distinctiveness probes the sharpness of (potentially distributed)
197 category-specific response patterns at the population level, the extent of category-specific
198 responses probes the amount of neural resources dedicated to processing a given category.
199 Here, we sought to test whether individual gaze tendencies and perceptual performance are
200 also linked to univariate category selectivity, specifically the individual sizes of the right
201 fusiform face (rFFA) and left visual word form area (IVWFA).

202 To this end, we quantified the participant-specific selective activation within group-level
203 parcels for the rFFA and IVWFA (cf. Julian et al., 2012⁷⁶; see methods for details). We
204 observed high split-half reliability (odd versus even localizer runs) for inter-individual
205 differences in the sizes of these functionally defined regions of interest (rFFA: $r(59) = .92$, $p <$
206 $.001$, 95% CI [0.88, 0.95]; IVWFA: $r(59) = .82$, $p < .001$, 95% CI [0.73, 0.89]). See
207 Supplementary Table S1 for an overview of all tested correlations with functional region of
208 interest sizes.

209 Individual sizes of rFFA correlated positively with the tendency to first fixate on faces ($r(59) =$
210 $.39$, $p = .002$, 95% CI [0.17, 0.58]), but not with the tendency of higher dwell times on faces
211 ($r(59) = .28$, $p = .031$, 95% CI [0.02, 0.52], n.s. at the Bonferroni corrected threshold of $\alpha =$
212 $.025$). Individual rFFA sizes were negatively correlated with gaze tendencies towards text

213 (first fixations: $r(59) = -.37$, $p = .004$, 95% CI [-0.56, -0.13]; dwell time: $r(59) = -.34$, $p = .007$,
 214 95% CI [-0.52, -0.13]). Furthermore, rFFA sizes correlated positively with face distinctiveness
 215 in the right lateral VTC ($r(59) = .55$, $p < .001$, 95% CI [0.36, 0.70]) and with behavioural
 216 performance on the CFMT ($N = 34$; $r(32) = .37$, one-tailed $p = .014$, 95% CI [0.09, 0.60]; cf.
 217 Methods and Figure 4b). The latter result aligns with previous findings by Furl et al. (2011)¹⁴
 218 and Elbich & Scherf (2017)¹⁵.

219 Individual sizes of IVWFA correlated positively with the tendency to first fixate text ($r(59) =$
 220 $.35$, $p = .006$, 95% CI [0.15, 0.54]), but not with dwell time on text ($r(59) = .26$, $p = .039$, 95%
 221 CI [0.06, 0.45], n.s. at the Bonferroni corrected threshold of $\alpha = .025$). No significant cross-
 222 correlations were observed for gaze tendencies towards faces (first fixations: $r(59) = .09$, $p =$
 223 $.514$, 95% CI [-0.14, 0.31]; dwell time: $r(59) = .01$, $p = .962$, 95% CI [-0.21, 0.21]). IVWFA
 224 sizes showed a strong positive correlation with word distinctiveness in the left lateral VTC
 225 ($r(59) = .75$, $p < .001$, 95% CI [0.63, 0.83]). However, the correlation between IVWFA sizes
 226 and reading-speed scores for pseudowords did not reach significance ($N = 25$; $r(23) = .32$,
 227 one-tailed $p = .060$, 95% CI [-0.05, 0.59]; cf. Methods, Figure 4c and Table S1), which may
 228 reflect the small sample size, especially in light of previous reports of significant correlations
 229 between IVWFA size and reading scores³⁷.

230 Overall, these results mirror the patterns observed between gaze tendencies, category
 231 distinctiveness and behaviour (cf. Figure 3 and Supplementary Table S1). They suggest that
 232 active vision is linked to both: the sharpness of neural tuning for and resources dedicated to
 233 processing a given category in the individual brain. Our results are also consistent with the
 234 previous developmental finding that increases in category selectivity go along with increases
 235 in category distinctiveness (cf. ²⁶). However, in contrast to distinctiveness, category-specific
 236 associations between areas sizes and dwell time as well as reading scores were not
 237 significant.

238

239 **No support for a link between gaze tendencies and elevated levels of attention during**
240 **the localizer task**

241 The link between gaze tendencies and category-selective activity in VTC could in principle be
242 mediated by category-specific levels of attention, generalizing between free-viewing and the
243 localizer task. Specifically, participants with a stronger tendency to fixate faces or words
244 during free-viewing may also have allocated higher levels of attention to the respective type
245 of stimulus when it was presented centrally and in isolation during the fMRI localizer task.
246 This hypothesis would align with the finding that experts direct more attention to their object
247 of expertise⁶⁴⁻⁶⁶, potentially leading to increased activation levels in VTC⁶⁷⁻⁷⁰.

248 To test this hypothesis, we conducted an additional psychophysics experiment in 60 naïve
249 adults who also completed an abbreviated, but sensitive version of the free-viewing
250 experiment (200 images; cf. Linka & de Haas, 2020⁷). Crucially, the new psychophysics
251 experiment mirrored our fMRI localizer design but introduced a secondary blob-detection
252 task to probe differential levels of attentional competition with faces or pseudowords (cf. ⁷¹,
253 and see Methods for details). This design enabled two key tests: First, to determine whether
254 there are systematic individual differences in the level of attentional competition afforded by
255 isolated, centrally presented faces and pseudowords. Second, to test whether these
256 differences in attention correlate with the tendency to fixate faces or words during free-
257 viewing. While a positive correlation would support the mediation account, the absence of a
258 detectable correlation would speak against attentional engagement mediating the observed
259 link between gaze tendencies and category-selective activity in VTC.

260 First, participants indeed showed reliable individual differences in secondary task
261 performance for face blocks ($M = 73\%$, $SD = 10.2\%$; split-half reliability $r(58) = .66$, $p < .001$,
262 95% CI [0.45, 0.80]), pseudoword blocks ($M = 85\%$, $SD = 17.2\%$; split-half reliability $r(58) =$
263 $.92$, $p < .001$, 95% CI [0.58, 0.97]), and the performance difference between them (face-block
264 advantage $M = 11.7\%$ $SD = 16.6\%$; split-half reliability $r(58) = .81$, $p < .001$, 95% CI [0.05,
265 0.93]). Second, however, we found no significant associations between this indicator of

category-specific differential levels of attention and gaze tendencies towards faces (first fixations: $r(58) = -.03$, $p = .814$, $BF_{10} = 0.17$, 95% CI [-0.20, 0.16]; dwell time: $r(58) = .08$, $p = .528$, $BF_{10} = 0.20$, 95% CI [-0.19, 0.32]) or text (first fixations: $r(58) = -.03$, $p = .844$, $BF_{10} = 0.16$, 95% CI [-0.19, 0.24]; dwell time: $r(58) = -.01$, $p = .951$, $BF_{10} = 0.16$, 95% CI [-0.14, 0.24]). This suggests that the relationship between gaze and category-selective activity in VTC is unlikely to be mediated by elevated levels of category-specific attention.

Immediate face salience is linked to face recognition performance

CFMT performance scores showed considerable variability, with a max-min ratio of 1.98 among the 34 fMRI participants ($M = 64.08$, $SD = 11.87$) and 2.33 for an extended sample of 69 participants ($M = 64.37$, $SD = 11.95$), including participants who took part in the free-viewing experiment and CFMT but not the fMRI session. In the extended sample, a significant positive correlation emerged between the proportion of first fixations on faces and CFMT scores, $r(67) = .39$, one-tailed $p < .001$, 95% CI [0.17, 0.58]. This relationship was not significant within the fMRI subsample ($r(32) = .23$, one-tailed $p = .094$, 95% CI [-0.07, 0.49], *n.s.*; cf. Figure 5a). In line with previous results⁵, we observed no significant correlation between dwell times on faces and CFMT performance in both the extended sample ($r(67) = .18$, one-tailed $p = .077$, 95% CI [-0.04, 0.38], *n.s.*) and the smaller fMRI sample ($r(32) = .16$, one-tailed $p = .174$, 95% CI [-0.17, 0.49], *n.s.*). These findings replicate those of de Haas et al. (2019)⁵, who reported an association between CFMT scores and the tendency to direct first fixations to faces, but not dwell times.

Performance on the speed-reading test showed considerable variability across the 25 fMRI participants and an extended sample of 65 participants for both words (subsample: $M = 118.40$, $SD = 16.17$, max-min ratio of 1.97; extended sample: $M = 111.35$, $SD = 18.07$, max-min ratio of 2.11) and pseudowords (subsample: $M = 77.92$, $SD = 14.57$, max-min ratio of 2.15; extended sample: $M = 74.68$, $SD = 16.36$, max-min ratio of 3.52). Word and pseudoword performance were positively correlated in the extended sample, $r(63) = .65$, p

293 <.001, 95% CI [0.50, 0.78]. Among the fMRI participants, we observed significant positive
 294 correlations between the proportion of first fixations on text in complex scenes and speed-
 295 reading test scores for both words ($r(23) = .50$, one-tailed $p = .004$, 95% CI [0.23, 0.74]) and
 296 pseudowords ($r(23) = .69$, one-tailed $p < .001$, 95% CI [0.43, 0.84]). However, in the extended
 297 sample this relationship vanished for words ($r(63) = -.04$, one-tailed $p = .386$, 95% CI [-0.31,
 298 0.24]) and pseudowords ($r(63) = .19$, one-tailed $p = .065$, 95% CI [-0.04, 0.42], *n.s.*, see also
 299 Figure 5b). No significant correlations were observed between the proportion of dwell time on
 300 text and both reading performances (see Table S1). While this renders the evidence
 301 regarding a potential relationship between gaze tendencies and reading performance
 302 inconclusive, very strong effect sizes seem unlikely, given the high internal consistency of
 303 both the gaze (see above) and performance measures (cf. ³⁶).

304 Taken together, our results show that category-specific differences in behavioural
 305 performance among adults are linked to individual representational distinctiveness (cf. ²⁶),
 306 and, at least for faces, also to immediate category-specific salience and the size of rFFA (cf.
 307 ^{5,7}). However, given these experiments were limited to small subsamples available for
 308 behavioural testing they should be interpreted with caution.

309 Discussion

310 Across functional magnetic resonance imaging and eye-tracking, our findings show
 311 that the individual tendency to fixate faces or text targets elements of a scene which can be
 312 represented more distinctively in the individual visual system. Moreover, a complementary
 313 analysis showed that the individual tendency to immediately fixate faces or text also goes
 314 along with larger corresponding regions of interest (rFFA and IVWFA, respectively).

315 This provides a link between individual biology and recently discovered large and
 316 systematic differences in gaze⁵. A previous study from our group demonstrated that
 317 individual gaze during movie-watching goes along with concurrent representational
 318 differences in the ventral stream. Specifically, the authors showed that the interindividual
 319 similarity of neural representations during movie watching is determined by the similarity of

gaze¹⁶. Here, we asked a different question: Do neural representations *as such* differ between people with different gaze tendencies, even under conditions of central fixation? If so, *how* do neural representations differ between people with different gaze tendencies? We found that centrally presented faces and pseudowords are represented more distinctively in individuals who tend to fixate these objects when they are embedded in complex scenes. Moreover, the brains of individuals prioritizing faces or text had larger corresponding regions of interest in the ventral stream (rFFA and IVWFA, respectively). This suggests that gaze prioritizes retinal input which the individual brain can process more distinctively.

Furthermore, we found that stronger face recognition and reading performance correlated with greater corresponding representational distinctiveness in the lateral ventral temporal cortex. These results extend previous findings on neurodevelopmental²⁶ and individual¹⁴ links between categorical tuning in the ventral stream and perceptual performance. We also replicated a correlation between the tendency to initially fixate faces and better face recognition performance⁵.

The links between gaze tendencies, neural selectivity and perceptual performance were overall stronger for first fixations, i.e. the tendency to immediately fixate corresponding elements in a scene. An intriguing distinction between first and subsequent fixations is that first fixations are often considered to be driven by automatic, bottom-up processes³⁸, which tend to decrease as viewers freely and more deliberately explore visual scenes³⁹. Interestingly, our results suggest that such bottom-up processes may be particularly aligned with category-selective processing in the individual VTC. Notably, Vinken & Vogels (2019)²¹ reported a pronounced reduction of face salience for early fixations in a monkey with lacking face patches. Furthermore, Sheinberg & Logothetis (2001)⁴⁰ found that neurons in the temporal cortex of monkeys exhibited selective activation shortly before saccades to behaviourally relevant (i.e., target) objects, and this activity raised again during fixation. Similarly, a preliminary report of intracranial recordings confirms pre-saccadic, category-specific VTC responses in humans⁴⁴. Thus, although VTC traditionally is not considered part of the oculomotor system, it appears tightly linked to the control of gaze, perhaps by

348 signalling object identity and relevance to downstream visuomotor networks (cf. ^{41,42}). Future
349 research is needed to clarify the nature and timing of this potential influence.

350 Importantly, our findings do not allow any direct conclusions regarding the
351 developmental interplay of category distinctiveness, gaze tendencies, and behavioural
352 performance. These relationships may be modulated by individual visual experience (cf. a
353 recent preprint by Yang et al. (2025)⁴⁵), expertise in visual object recognition, attentional
354 habits, and genetic predispositions which can shape perceptual biases and neural
355 architecture (also see below). Regardless of preceding developmental trajectories, our
356 findings reveal an empirical association between moment-to-moment gaze behaviour and the
357 overarching capability of more distinctive neural representation in the adult ventral stream.
358 This broadly echoes theories of active inference⁴⁶, which propose that individuals actively
359 seek out observations reducing uncertainty given their internal representations of the world.

360 Why do active vision and neural tuning in the ventral stream match each other? Our
361 main approach is correlational, but several additional observations lead us to an informed
362 speculation: Individual gaze and neural tuning in the ventral stream may mutually shape
363 each other across the first two decades of life, while adults match their moment-by-moment
364 gaze to stabilized tuning properties of the matured ventral stream.

365 The development of both, gaze tendencies²⁵ and neural tuning in the ventral
366 stream^{23,26,47} is protracted across the first two decades of life. Developing fixation habits likely
367 shape the individual foveal diet. Visual experience in turn can shape category-specific neural
368 responses in the ventral stream, as shown for reading acquisition (e.g. ^{27,29,48-50}), extensive
369 exposure to Pokémon characters^{51,52} or face-deprivation in non-human primates²². At the
370 same time, a lack of neural preferences goes along with a strongly reduced tendency to
371 spontaneously fixate faces^{21,22}, underscoring the possibility of a bidirectional link. Moreover,
372 spontaneous gaze shows surprisingly strong genetic components^{10,11}, which could partly be
373 mediated by heritable differences in neural tuning^{53,54}, including ventral stream preferences
374 for faces^{55,56}. Taken together, individual differences in gaze and neural tuning may mutually

reinforce each other across development. However, this remains to be tested in multimodal longitudinal studies (see below).

In adults, both the individual layout of the ventral stream⁵⁷ and individual differences in gaze⁵ appear stable. However, this stability seems more pronounced for neural tuning than gaze. While attention and task instructions can modulate ventral stream responses top-down, this does not lead to a reversal of category-specific preferences in areas like the FFA⁵⁸. Spontaneous gaze tendencies, on the other hand, can be overridden and reversed top-down: Even individuals with a strong tendency to look at faces manage to mostly look away from them if instructed to do so⁴³. This suggests that, at least for adults, the moment-by-moment plasticity of gaze exceeds that of neural tuning in the ventral stream. Furthermore, individual differences in adult neural response patterns do not depend on concurrent differences in gaze. The results of our localizer experiment documented such differences under conditions of central fixation. Taken together, adults appear to show persistent individual differences in their neural preferences, which gaze caters to second by second. Future studies should test the causal nature of this link.

More generally, further research should trace neural and gaze development longitudinally to disentangle their interplay. Doing this in mono- and dizygotic twins would allow to differentiate between genetic and environmental effects^{10,11}. Furthermore, studies utilizing perceptual training could manipulate the distinctiveness of individual representations to test their effect on gaze.

Theoretically, both gaze tendencies and neural preferences could be mediated by habitually elevated levels of covert and overt attention. If an individual has persistently elevated levels of attention for a given type of stimulus, this may boost corresponding neural responses and bias gaze in tandem. If true, this should also render the corresponding stimulus a more potent distractor in dual task scenarios. Our additional psychophysical attention experiment was designed to put this prediction to a well-powered test. However, we found no such relationship, rendering persistent individual differences in category-specific

402 levels of attention an unlikely mediating mechanism for the relationship between neural
403 tuning and gaze.

404 Notably, expertise may play a role for both, salience and cortical selectivity (cf. ^{51,52}).
405 For instance, greater car expertise is associated with better car detection in scenes⁶⁶,
406 increased bottom-up salience for cars⁷², and increased VTC selectivity⁶⁷. Given performance
407 differences on face recognition and reading tasks go along with corresponding category
408 selectivity and gaze tendencies, our results can be interpreted along these lines.
409 Developmentally, gaze tendencies, perceptual skills and neural tuning may reinforce each
410 other (see above). This could be interpreted as a spiral of expertise, achieving different levels
411 in different individuals.

412 Interestingly, however, expertise research found that widespread category-selective
413 activity for objects of expertise diminished to novice-like responses when stimulus
414 engagement was low⁷⁰. Future studies could test whether high category-specific engagement
415 tasks increase the size of the effects we observed here. This could be part of a broader
416 endeavour to test how well individual differences in gaze, perceptual skills and neural tuning
417 generalise between different types of tasks, including real-world behaviour. A first step in this
418 direction is the observation of elevated face salience in super recognisers, some of whom
419 are employed for real-world face recognition tasks by the police⁹.

420 Notably, the observed correlations are based on a relatively homogeneous sample of healthy
421 university students. Nevertheless, we observed striking individual variability: Gaze
422 tendencies varied substantially, with a 4.7-fold difference in the tendency to first fixate text
423 elements and a 2-fold difference for faces. Neural representations also showed marked
424 variation: word distinctiveness in the left lateral VTC differed by up to 1.5-fold, and face
425 distinctiveness in the right lateral VTC 1.3-fold. Similarly, category-specific recognition
426 performance varied across individuals, with a 2.3-fold range in face recognition ability and up
427 to 3.5-fold differences in reading speed. The degree of (co)variance in this homogenous
428 sample suggests that the observed effects are either relatively independent of age, level of

education, and general cognitive ability, or could be even more extreme in more diverse samples. Recent developmental findings suggest the latter, by documenting that gaze patterns in natural scenes tend to become less idiosyncratic and more canonical across childhood and adolescence²⁵. Larger and more diverse samples could probe the role of multiple third variables and potential moderators such as gender, personality traits or academic-domain performance. However, the evidence for gender and personality associations with gaze tendencies is limited. For example, some previous studies observed a visual preference for social stimuli in females⁷³⁻⁷⁵, while recent large-scale eye-tracking work has reported only very small gender differences in the opposite direction (Linka et al., 2025²⁵; $d \approx 0.2$). Likewise, prior work has found no reliable link between personality traits like extraversion and gaze tendencies (de Haas et al., 2019⁵). Given the observed associations with reading scores, it seems especially important to investigate the developmental interplay of schooling, gaze tendencies, neural development and educational attainment.

In summary, our key finding is that individual differences in the tendency to fixate faces or text are systematically linked to corresponding neural resources and representational distinctiveness in the ventral stream. Previous research has shown that both individual gaze and the functional layout of the VTC undergo protracted and experience-dependent development. Our results show that in adulthood, active vision and neural tuning in the ventral stream are matched to each other. Moment-by-moment saccadic choices systematically prioritize elements of a scene which can be represented more distinctively in the individual visual brain.

450

451 **Methods**

452 **Participants**

453 We recruited a total of 102 healthy participants (mean age = 24.64 years, $SD = 4.58$; 30
454 male, 69 female, 3 diverse) from the participant mailing list at Justus Liebig University
455 Giessen, Germany. All participants had normal or corrected-to-normal visual acuity and

456 reported no known visual deficits. Every participant completed the eye-tracking experiment
457 (~120 minutes). Additionally, a subset of participants (mean age = 25.59 years, *SD* = 5.13;
458 19 male and 42 female) took part in further tasks: 61 completed the functional magnetic
459 resonance imaging (fMRI) session (~90 minutes), 69 completed the Cambridge Face
460 Memory Test in the long format (CFMT; ~15 minutes, including 34 who also participated in
461 the fMRI session), and 65, with German as their mother tongue, completed the speed-
462 reading task (~5 minutes, including 25 who also participated in the fMRI session).
463 Participants received either course credit or monetary compensation (10 euros/hour) for their
464 participation. The study was approved by the local ethics committee of the Justus-Liebig-
465 University Giessen (LEK-FB06) and was conducted in accordance with the relevant
466 guidelines and regulations. All participants provided written informed consent.

467 **Procedure**

468 Every participant took part in an eye-tracking session, followed by an fMRI session and/or a
469 behavioural session. If participants took part in both the fMRI and behavioural sessions, the
470 behavioural tests were typically completed after the scanning, otherwise, the CFMT was
471 completed subsequently online using testable.org⁵⁹. Before the eye-tracking session, fMRI
472 session and behavioural tests, participants gave written informed consent and received task
473 instructions. During the fMRI session, participants performed additional scans not related to
474 the current study, including retinotopy, free viewing of 150 images and diffusion tensor
475 imaging. Each session began with three localizer runs, followed by three retinotopy runs, and
476 then another set of three localizer runs.

477 **Eye-tracking**

478 Stimuli were displayed and data collected using Psychtoolbox version 3.0.16⁶⁰ and MATLAB
479 (R2021b, MathWorks, Natick, MA) on a Windows 10 computer and 23.8-inch LG Ultra HD28
480 monitor at a resolution of 3840 × 2160 pixels. The display was placed at a distance of 55 cm
481 and stimuli were presented at a size of 2400 × 1800 pixels, spanning 34.3 × 25.7 degree
482 visual angle. Movements of the right eye were recorded at a frequency of 1 kHz using an

483 Eyelink 1000 (SR Research, Ottawa, Canada). Following the experimental design reported
484 by de Haas et al. (2019)⁵, participants run 700 trials in total, each showing one scene image
485 from the OSIE image set², which they could freely explore, and they were offered to take
486 short breaks after every 100 trials. Each 100-trial block started with a nine-point calibration,
487 followed by a validation. Calibration was accepted only if the maximum error was ≤ 1.2
488 degrees of visual angle (dva) and the average error was ≤ 0.6 dva. The observed mean
489 validation error was 0.37 dva ($SD = 0.11$ dva). Participants initiated each trial by fixating on a
490 central point while simultaneously pressing the space bar, which triggered the presentation of
491 a scene image lasting for three seconds. The entire experiment lasted approximately 90
492 minutes. Note, in connection with an additional attention experiment, we conducted another
493 eye-tracking experiment (see below, 'Attention experiment').

494 ***Measured variables and preprocessing***

495 In line with de Haas et al. (2019)⁵ and manufacturer's recommendations, we excluded
496 fixations with onset latencies or durations shorter than 100 ms, as well as fixations that could
497 not be assigned a label. This allowed us to restrict our analysis to object-directed fixations
498 occurring after stimulus onset. To account for foveal extent and eye-tracking accuracy, a
499 fixation was assigned to a labelled object pixel mask for which the distance between the
500 fixation coordinate and pixel mask was within a tolerance margin of ~ 0.5 degrees visual
501 angle (cf. ⁵). We then quantified the cumulative fixation duration (i.e. dwell time) for each
502 observer that fell on text elements or faces as the percentage of cumulative dwell time on
503 any labelled pixel masks. Similarly, we quantified the percentage of labelled first fixations
504 after image onset that fell on text elements or faces (cf. ⁵).

505 ***Magnetic resonance imaging***

506 We collected neuroimaging data at the Benders Institute of Neuroimaging (BION) at Justus-
507 Liebig University using a 3-Tesla Siemens Prisma scanner and a 64-channel head coil.

508 ***Structural imaging***

509 High-resolution anatomical images were obtained using a T1-weighted magnetization-
510 prepared rapid acquisition gradient-echo (MPRAGE) sequence with a Field of view (FoV) of
511 240 × 240 mm, an echo time (TE) of 3.53 ms and a repetition time (TR) of 1880 ms. Further
512 parameters for obtaining functional data were as follows: inversion time = 949 ms, in-plane
513 resolution = 0.94 mm × 0.94 mm, number of slices = 176, slice thickness = 0.94 mm, flip
514 angle (FA) = 8°. Magnetic field perturbations were accounted for by measuring a field map
515 with the following scan parameters: FoV = 220 × 220 mm, TE (1) = 10 ms, TE (2) = 12.46
516 ms, TR = 1000 ms, in-plane resolution = 2 × 2 mm, slice thickness = 3 mm, number of slices
517 = 40 (transversal), FA = 90°

518 ***Functional imaging***

519 The functional data was acquired with the same scanner and head coil as the structural
520 images. Functional images covered the whole brain and were obtained using a multiband
521 echo-planar imaging sequence (EPI) with the following scan parameters: TE = 30 ms; TR =
522 1500 ms; FoV = 220 × 220 mm, in-plane resolution = 2.2 mm × 2.2 mm, 46 slices
523 (descending) with a thickness of 2.2 mm and a distance factor of 20%, FA = 71°, in-plane
524 acceleration factor = 2 and slice-acceleration factor = 2. We acquired 1086 volumes of
525 functional data per participant (181 per run).

526 We used an MRI compatible eye camera (12M-I, Arrington Research Inc., Scottsdale, AZ) to
527 observe participants' compliance with task requirements online (e.g. to hold fixation and do
528 not move the eyes). Stimuli were projected onto a screen in the scanner bore, just behind the
529 head coil using an Epson EB-G5600 projector with a resolution of 1024 x 768 pixels and a
530 refresh rate of 60 Hz.

531 ***Localizer experiment***

532 During functional imaging, participants completed a total of six runs of a standard localizer
533 experiment, in which stimuli of six different categories were displayed from the fLoc
534 functional localizer package (faces, bodies, characters, houses, cars, limbs; cf. ³⁰). Category
535 stimuli were placed on a phase-scrambled background, which was generated from randomly

selected images³⁰. All stimuli were presented in greyscale and the total image size including the background spanned 28 dva square. Individual images were presented only once in a trial but could be repeated randomly across trials. Images of one category were presented in 6 second blocks and were presented at category-specific rates (faces = 3.3 Hz; bodies = 4.5 Hz; characters = 2 Hz; houses = 2.6 Hz; cars = 2 Hz; limbs = 2 Hz). This is following previous research, suggesting category-specific temporal processing and encoding differences³⁰. At the beginning and end of each run, a baseline block was added, showing a grey luminance screen. Each run consisted of six blocks of each category. These blocks were shown in random order and intermixed with rest blocks showing only a grey background. Category blocks contained 12 – 27 stimuli, whereas rest blocks contained no stimuli. Participants were instructed to view the images while fixating on a central dot throughout. In each category block, one target was randomly interspersed, consisting of only the phase-scrambled background (i.e. no object overlayed). The oddball task required participants to press a button within 1 s of the appearance of this target. Responses from two participants were excluded due to a button-box malfunction. For the remaining 59 participants, performance on the oddball task during scanning was high for both face (hit rate = 80%, SD = 15%) and word blocks (hit rate = 83%, SD = 13%). In a supplementary analysis, we found no correlations between performance in the oddball tasks and category-specific gaze tendencies or category distinctiveness (see Table S2).

Assessing perceptual performance

A subset of our participants completed one or both of the following behavioural tasks: the Cambridge-Face-Memory-Test in the long format (CFMT³⁵) and the German speed-reading task SLRT-2 (short for: Salzburger Lese- und Rechtschreibtests³⁶).

Assessing face recognition ability

We assessed participants' face recognition ability using the CFMT³⁵, a self-paced recognition task consisting of trials that include separate learning and test phases. In each learning phase, participants study the identity of a previously unfamiliar face from three different

viewpoints. In the test phase, a triplet of faces is shown, and participants are asked to identify the learned identity amongst distractor faces. The test becomes increasingly more difficult (unknown viewing angles, lighting and superimposed noise). The score is obtained by measuring the percentage of correct responses during the test phase.

Assessing reading ability

In a subset of native German speakers, we assessed their reading ability using the speed reading test SLRT-2³⁶. In this task, participants read one list of real words and one list of pseudowords as quickly and accurately as possible. The test has a time limit of one minute per list and scoring is based on the total number of (pseudo-) words read minus the number of pronunciation errors.

Attention experiment

60 naïve participants (mean age = 22.87 years, $SD = 3.25$; 8 male and 52 female) completed a shorter, but otherwise similar version of the free-viewing experiment reported above (using 200 instead of 700 images; cf. Linka & de Haas, 2020⁷). Again, calibration was accepted only if the maximum error was ≤ 1.2 degrees of visual angle (dva) and the average error was ≤ 0.6 dva. The observed mean validation error was 0.35 dva ($SD = 0.08$ dva). These participants also completed a dual-task based on the localizer experiment. The task lasted ~ 10 min and contained 6 practice trials and 84 experimental trials total (2 x 42, alternating face and pseudowords). The experimental trials were divided into three equal blocks (28 trials each) with short breaks in between. The stimuli (spanning 28 dva square) and oddball task were identical to the face and pseudoword blocks in the localizer experiment, except for the addition of a secondary task. For the primary task, participants used the left arrow key to report the oddball target of a phase-scrambled background (i.e. no object overlaid; same task as in the scanner). The new, secondary task ("blob detector") added a small (4.3 dva), mid-grey disk in one of the four image corners on 1-3 frames per trial. These secondary targets were reported using the right arrow key and considered a hit if they occurred within 1 s of onset. Per category, the blob counts followed a fixed distribution (3 trials with 1 blob

occurrence, 20 with 2, 19 with 3), yielding 200 secondary targets overall (100 during face trials, 100 during word trials). Image order and blob positions were identical for all participants (fixed assignment), and the minimum temporal spacing between any two targets (primary or secondary) was two primary stimuli (606 ms during face trials, 1000 ms during word trials). Timing and keypresses were logged continuously. For each participant, we computed category-specific hit rates (e.g., number of secondary targets in face blocks/hits) and the difference in hit rate between text and face blocks (text – face; that is, positive and negative values indicate an advantage during the face and text blocks, respectively). To assess split-half reliability, we computed this indicator and category-specific hit rates separately for the first and second halves of the experiment and correlated it across participants. Participants were compensated with course credit or payment (10 euros/hour). The study was conducted in accordance with the guidelines and regulations of the local ethics committee of Justus-Liebig-University Giessen (LEK-FB06), and all participants gave written informed consent prior to participation.

Data analysis

We conducted all statistical analyses and generated all figures using MATLAB (R2024b, MathWorks, Natick, MA) and SPM ([https:// www.fil.ion.ucl.ac.uk/spm/](https://www.fil.ion.ucl.ac.uk/spm/)). For visualizations and definition of regions of interest based on anatomical landmarks (cf. ²⁶; see also Figures 2 and 4), we processed each observer's anatomical scans using FreeSurfer (<http://surfer.nmr.mgh.harvard.edu/>) to delineate the grey-white matter and grey matter pial boundaries. These boundaries were then used to reconstruct inflated cortical surfaces on which we projected statistical maps obtained using SPM.

Definition of lateral and medial VTC regions

Following previous studies^{18,26}, we defined lateral and medial VTC regions based on anatomical landmarks on the native inflated cortical surfaces of each participant's hemisphere in FreeSurfer (<http://surfer.nmr.mgh.harvard.edu/>). The posterior boundary of both VTC regions of interests (ROIs) was set along the middle section of the posterior

617 transverse collateral sulcus (ptCoS), while the anterior boundary was aligned with the
618 anterior tip of the mid-fusiform sulcus (MFS). However, we positioned this boundary slightly
619 more anteriorly, as mFus-Faces is predominantly located in this region¹⁸. The lateral border
620 of the lateral VTC ROI was placed along the lateral edge of the inferior temporal gyrus (ITG),
621 while the medial and lateral VTC ROIs were separated by a boundary along the MFS. The
622 medial border of the medial VTC ROI was set along the medial edge of the collateral sulcus
623 (CoS). In line with reported effects in a previous study²⁶, we focused our analysis related to
624 category distinctiveness on the lateral VTC.

625 **Preprocessing of functional MRI data**

626 Functional data preprocessing was performed using SPM (<https://www.fil.ion.ucl.ac.uk/spm/>).
627 Functional images were realigned and unwarped using voxel displacement maps derived
628 from the field maps. Functional data were then co-registered to each participant's structural
629 scan and spatially smoothed with a 5 mm Gaussian kernel. Supplementary analyses
630 confirmed that light smoothing led to more robust estimates of category-specific patterns (cf.
631 Supplementary Information, 'Reliability of multivariate patterns').

632 **Extraction of neural category selectivity**

633 We specified a general linear model (GLM) with boxcar regressors for each stimulus
634 category and six nuisance regressors for estimated rigid motion parameters based on the
635 realignment. Category regressors were convolved with the canonical hemodynamic response
636 function as implemented in SPM. This allowed us to estimate the response amplitude for
637 each of the six conditions at the voxel level. Specifically, we computed voxel-wise t-statistics
638 for the contrast between a given category and the average of all others.

639 **Analysis of category distinctiveness**

640 For each category, multivoxel patterns (MVPs) were defined as vectors of the corresponding
641 t-values across all voxels within a given ROI (lateral VTC in the left or right hemisphere). For
642 each participant, we computed one representational similarity matrix (RSM⁶¹; per hemisphere

643 by correlating MVPs from one half of the localizer runs (odd runs: 1, 3, and 5) with those from
 644 the other half (even runs: 2, 4, and 6). This yielded a 6×6 RSM for each hemisphere, with
 645 within-category correlations represented on the diagonal and between-category correlations
 646 on the off-diagonal (see Figure 2d). These matrices were then used to calculate face and
 647 word distinctiveness. These two categories were selected based on consistent findings from
 648 prior research demonstrating robust and substantial individual differences in adults' gaze
 649 tendencies for faces and words^{5,7} as well as the protracted development and behavioural
 650 relevance of corresponding representational distinctiveness²⁶. The distinctiveness of a
 651 category is calculated as the within category minus the average between-category similarity
 652 of multivariate responses with the other five categories and the resulting values can
 653 theoretically range from -2 to 2.

654 ***Definition of functional ROIs and their individual size***

655 We defined functional regions of interest (fROIs) using the automated Group-constrained
 656 Subject-Specific (GSS) approach (cf. ⁷⁶⁻⁷⁸). This method provides unbiased, participant-
 657 specific fROIs and is highly suitable for our sample size (N = 61). First, individual 1st-Level
 658 GLMs were estimated in MNI space using unsmoothed data. We then used the `spm_ss`
 659 toolbox (https://www.nitrc.org/projects/spm_ss; ⁷⁸) to generate binary activation maps for the
 660 contrasts of faces > other categories and words > other categories, respectively, both
 661 thresholded at $p < .0001$ (uncorrected). These binary maps were overlaid across participants,
 662 spatially smoothed (FWHM = 6mm), and thresholded to retain only voxels with at least 10%
 663 overlap across the sample. Resulting group-level parcels were restricted to those within
 664 which at least 60% (36/61) of participants showed significant activation. We compared
 665 resulting group-level peaks of respective parcels to previous literature and focused on the
 666 right Fusiform Face Area (rFFA, MNI peak coordinates: +40, -47, -21; cf. ⁷⁶) and left Visual
 667 Word Form Area (lVWFA; MNI peak coordinates: -44, -66, -13; cf. ⁵⁰; see Figure 4a, for an
 668 illustration of fROI locations).

669 To quantify participant-specific activation, we performed a masked voxel-count analysis.
670 Each participant's t-contrast map (unsmoothed data; faces > other categories or words >
671 other categories) was masked by the respective group-level parcel (rFFA or IVWFA). Within
672 these constraints, we calculated the total number of significant voxels exceeding a threshold
673 of $p < .0001$ (uncorrected; see Figure 4a and b).

674 **Statistics**

675 We used Pearson correlations to test the split-half reliability (cf. ⁵) of individual gaze
676 tendencies (proportions of first fixations and dwell time on faces or text elements) and rFFA
677 or IVWFA sizes. Because two metrics were used to quantify gaze tendencies, significance
678 thresholds were Bonferroni-adjusted: we reported original P-values throughout, but only
679 values below $\alpha = .025$ ($.05/2$) are considered significant. Furthermore, we examined the
680 relationships between category distinctiveness, functional region of interest sizes, gaze
681 tendencies, and behavioural performance on the CFMT and speed-reading tests.

682 To assess the respective significance of the observed correlations, we first computed the
683 correlation coefficient between the two original vectors. Next, we generated a null distribution
684 by randomly shuffling one of the vectors and recalculating the correlation coefficient with the
685 unshuffled vector. This procedure was repeated 10,000 times to create a distribution of
686 correlation coefficients expected under the null hypothesis of no association. Finally, we
687 calculated the proportion of these randomly generated correlations that were as extreme or
688 more extreme than the observed correlation, providing an empirical estimate of the p-value.
689 By default, reported p-values are two-tailed unless otherwise specified in the text.

690 We tested two primary a priori hypotheses: First, individual face salience is linked to face
691 distinctiveness in the right lateral VTC and second, individual text salience is linked to word
692 distinctiveness in the left lateral VTC. We quantified category-specific individual salience as
693 the individual proportions of first fixations and dwell time spent on corresponding stimuli
694 during the free-viewing experiment. To maintain statistical rigor, we adjusted significance
695 thresholds for the use of these two metrics using the method of Bonferroni. We report original

696 P-values but only interpret values below $\alpha = .025$ (.05/2) as significant. Note, we signalled
697 significant results with asterisks in Table S1. We also adopted a conservative stance by
698 employing two-sided tests, despite directional predictions for positive associations. We also
699 tested the hypotheses that individual perceptual performance for reading and face
700 recognition are linked to respective individual fixation tendencies and to corresponding
701 representational distinctiveness. Here, we also corrected for the two salience measures
702 tested (where applicable), but employed one-sided tests. This decision was based on
703 previous findings supporting the directional hypotheses (cf. ^{14,15,26}) and the lower sample
704 sizes for perceptual performance tests. Significance levels for associations with speed
705 reading scores were additionally corrected for the two reading measures employed
706 (pseudowords and words).

707 Complementing our main analyses, we explored associations across hemispheres and
708 categories (e.g. face distinctiveness in left lateral VTC vs. face salience and face
709 distinctiveness in right lateral VTC vs. text salience). These served as specificity controls to
710 test if the observed relationships between neural and gaze patterns are hemisphere- and
711 category-specific. Following reviewer suggestions, we also explored associations with
712 IVWFA and rFFA sizes. For all these exploratory analyses, we applied the same corrections
713 as for the main analysis to account for the two metrics of salience (proportion of first fixations
714 and dwell time).

715 For the attention experiment, we additionally computed Bayesian Pearson correlations in
716 JASP (Version 0.96; JASP Team, 2026⁷⁹) using a Jeffreys-Zellner-Siow (JZS) prior. The
717 Bayes Factors (BF_{10}) were interpreted following Wetzels and Wagenmakers (2012)⁸⁰, where
718 values below 1 indicate evidence in favour of the null hypothesis.

719 **Sample size**

720 Our sample size aimed for good sensitivity (power > .85) for effect sizes of $r \geq .4$. This was
721 based on previously observed effect sizes for the relation between category-specific
722 distinctiveness and perceptual skills among children²⁶. Under our hypothesis, we expected

723 somewhat larger effect sizes for the relationship between representational distinctiveness
724 and gaze tendencies among adults, given our estimates of distinctiveness were based on
725 three times the number of localizer runs per data point (six instead of two; Nordt et al,
726 2023²⁶); the generally better signal-to-noise ratio of BOLD responses in adults (e.g.
727 Thomason et al., 2005⁸¹); and the excellent internal consistency of individual differences in
728 face- and word-fixation tendencies among adults^{5,82}.

729 **Data availability**

730 The anonymized data used to generate the results in this study and generate the main
731 figures are freely available from <https://osf.io/9fhxk/>. The raw (f)MRI data generated in this
732 study are protected by privacy laws and not available. However, raw or processed data may
733 be shared with accredited researchers upon request.

734 **Code availability**

735 MATLAB code and stimuli used to generate the figures and results are freely available on
736 OSF: <https://osf.io/9fhxk/>.

737 **Acknowledgements**

738 This work was supported by European Research Council Starting Grant 852885 INDIVIDUAL
739 and Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project
740 No.222641018–SFB/TRR 135 TP C9 (DK, EA, MDB, BdH), as well as under Germany's
741 Excellence Strategy (EXC 3066/1 “The Adaptive Mind”, Project No. 533717223; DK, MDB,
742 BdH). The funders had no role in study design, data collection and analysis, decision to
743 publish or preparation of the manuscript. We thank Xu et al. for making their stimuli, pixel
744 masks and metadata publicly available. We also would like to thank Diana Weissleder,
745 Zeynep C. Demirkan, Tugce Dalmis, Hannah R. Wilkening, Sophie J. Halstenberg and Karin
746 Schmidt for their help with data collection. MR imaging for this study was performed at the
747 Bender Institute of Neuroimaging at the Justus Liebig University Giessen, Germany.

748 **Author contributions**

DK, MDB and BdH conceptualized the experiment. DK and EA collected the data. DK analysed the data in consultation with BdH. DK drafted the manuscript. DK, EA, MDB and BdH revised the manuscript and approved its final version.

Competing interests

The authors declare that they have no competing interests.

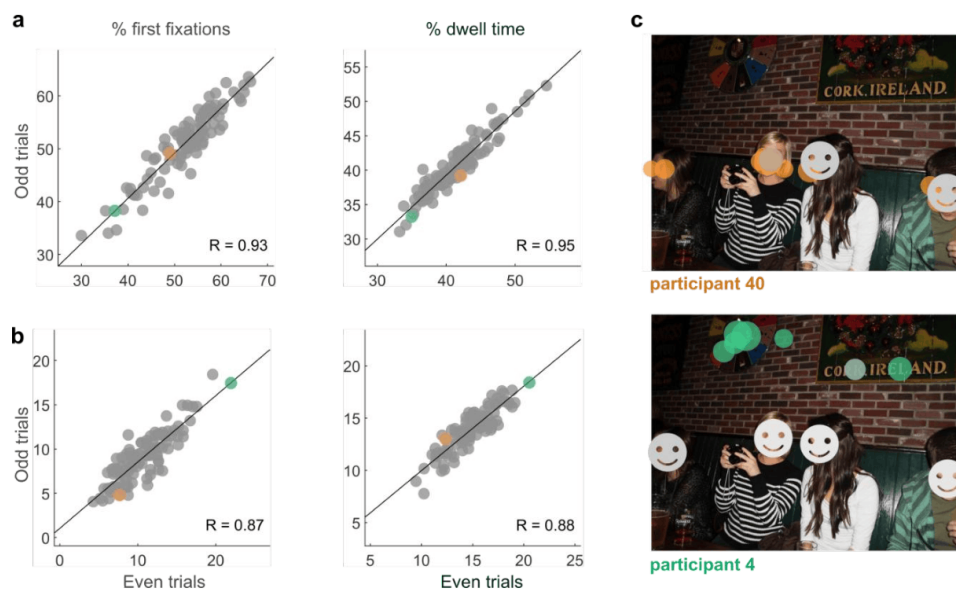


Figure 1: Consistent gaze tendencies towards faces and text in complex scenes.

Consistency of individual proportions of first fixations and dwell time on (a) faces and (b) text across odd and even images ($N = 102$, all $r > .87$, $p < .001$). Coloured dots highlight data from participant ID4 in green and participant ID40 in orange. Corresponding gaze patterns for an example scene are visualized in panel (c). Lighter colours represent first fixations, and darker shades later ones. Disk sizes scale according to fixation durations. Faces were visible in the experiment but are obscured in this reproduction to protect privacy. The original image is from Xu et al.'s (2014) OSIE dataset, published under the Massachusetts Institute of Technology license² and adapted under a Creative Commons license CC BY 4.0.

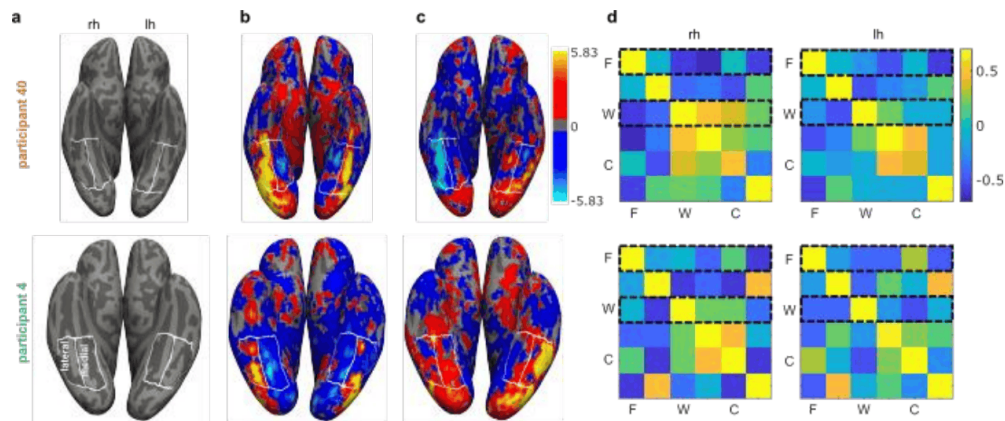


Figure 2: Multivariate pattern responses of individual participants in right and left lateral VTC.

All panels show data from example participants ID4 (bottom) and ID40 (top). (a) White outlines show individual anatomical regions of interests for the lateral and medial VTC, in the left (lh) and right hemisphere (rh), based on anatomical landmarks following Nordt et al. (2023)²⁶. (b) and (c) Overlays of *t*-contrasts for face- (b) and word- (c) specific activations according to the inset colour bar on the right. (d) Representational similarity matrices (RSMs) for six stimulus categories including faces (F), words (W) and cars (C) in the lateral VTC, showing pattern-correlations according to the inset colour bar on the right. Rows used to calculate the distinctiveness of face and word representations are highlighted with dashed lines (within category similarity in odd and even runs, i.e. on diagonal values, minus the mean of between category similarity in odd and even runs, i.e. off-diagonal values).

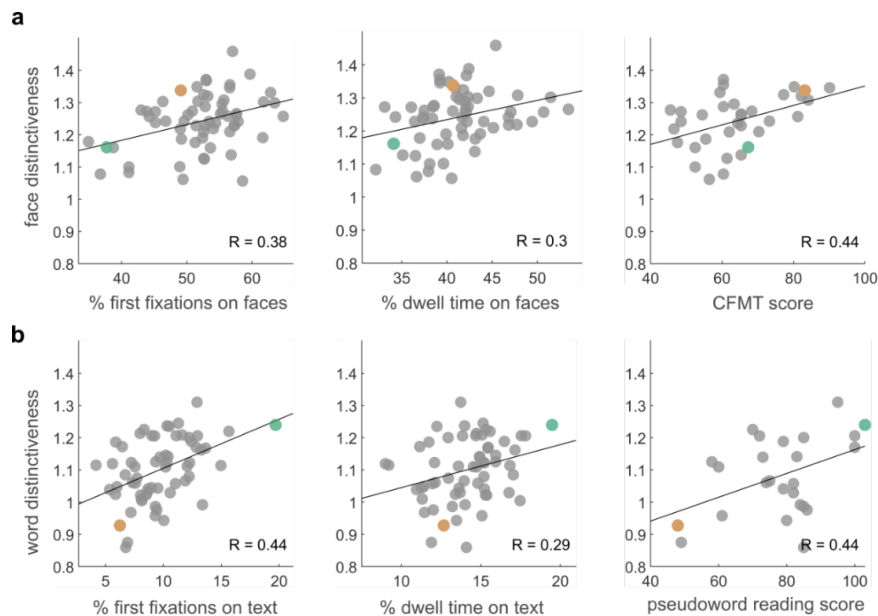


Figure 3: Category distinctiveness in right and left lateral VTC is linked to gaze tendencies and behavioural performance.

(a) Pearson correlations between face distinctiveness in right lateral VTC and the proportion of first fixations on faces ($N = 61$, $p = .003$), the proportion of dwell time on faces ($N = 61$, $p = .017$), and the Cambridge Face Memory Test (CFMT) score ($N = 34$, one-tailed $p = .004$), respectively. (b) Pearson correlations between word distinctiveness in left lateral VTC and the proportion of first fixations on text ($N = 61$, $p < .001$), the proportion of dwell time on text ($N = 61$, $p = .023$), and the pseudoword reading score ($N = 25$, one-tailed $p = .012$), respectively. Coloured dots highlight the same individuals across figures (participant ID4 in green, participant ID40 in orange).

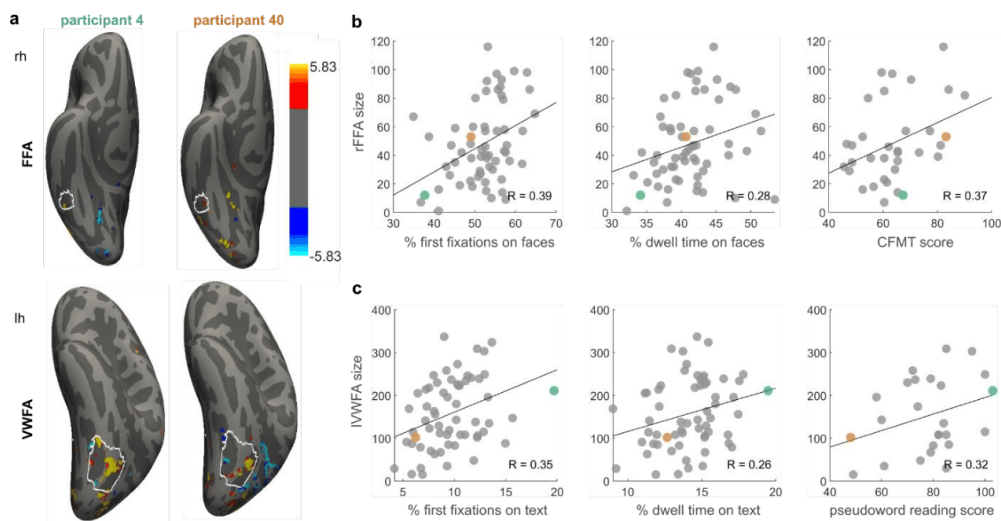


Figure 4: The relationship between the individual size of rFFA and IVWFA, gaze tendencies and behavioural performance.

(a) the top panel shows the outline of the group-level rFFA parcel in MNI space overlaid with individual t-contrasts for face-specific activations, ranging from $|t| > 4.28$ ($p < .0001$) to 5.83, as indicated by the colour bar to the right. Similarly, the bottom panel shows group-level parcels for the IVWFA overlaid with individual t-contrasts for word-specific activations. (b) Pearson correlations between the individual size of rFFA and the proportion of first fixations on faces ($N = 61$, $p = .002$), the proportion of dwell time on faces ($N = 61$, $p = .031$, n.s., Bonferroni-corrected), and the Cambridge Face Memory Test (CFMT) score ($N = 34$, one-tailed $p = .014$), respectively. (c) Pearson correlations between the individual size of IVWFA and the proportion of first fixations on text ($N = 61$, $p = .006$), the proportion of dwell time on text ($N = 61$, $p = .039$, n.s., Bonferroni-corrected), and the pseudoword reading score ($N = 25$, one-tailed $p = .060$, n.s.), respectively. Coloured dots highlight the same individuals across figures (participant ID4 in green, participant ID40 in orange).

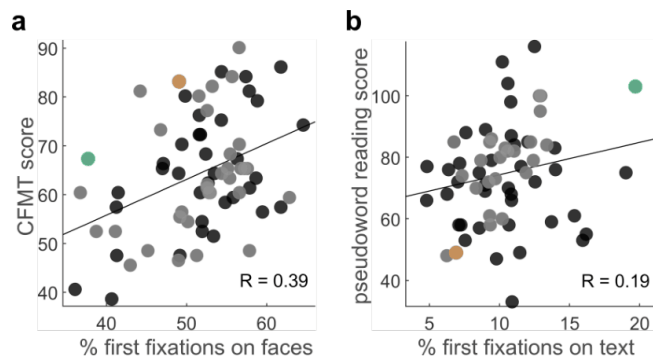


Figure 5: Proportions of first fixations on faces are related to face recognition, whereas first fixations on text are not significantly related to reading performance.

Pearson correlation between the proportion of (a) first fixations on faces and CFMT scores ($N = 69$, $p < .001$), as well as (b) first fixations on text and pseudoword reading scores ($N = 65$, $p = .065$, $n.s.$). In both plots, black dots represent participants from the eye-tracking experiment only, and light grey dots indicate those who also participated in the fMRI session. Coloured dots highlight the same individuals across figures (participant ID4 in green, participant ID40 in orange).

References

1. Itti, L., Koch, C. & Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 1254–1259 (2002).
2. Xu, J., Jiang, M., Wang, S., Kankanhalli, M. S. & Zhao, Q. Predicting human gaze beyond pixels. *Journal of vision* **14**; 10.1167/14.1.28 (2014).
3. Kümmerer, M., Bethge, M. & Wallis, T. S. A. DeepGaze III: Modeling free-viewing human scanpaths with deep learning. *Journal of vision* **22**, 7; 10.1167/jov.22.5.7 (2022).
4. Berlijn, A. M., Hildebrandt, L. K. & Gamer, M. Idiosyncratic viewing patterns of social scenes reflect individual preferences. *Journal of vision* **22**, 10; 10.1167/jov.22.13.10 (2022).
5. de Haas, B., Iakovidis, A. L., Schwarzkopf, D. S. & Gegenfurtner, K. R. Individual differences in visual salience vary along semantic dimensions. *Proceedings of the National*

- 832 *Academy of Sciences of the United States of America* **116**, 11687–11692;
833 10.1073/pnas.1820553116 (2019).
- 834 6. Guy, N. et al. A novel perceptual trait: gaze predilection for faces during visual
835 exploration. *Scientific reports* **9**, 10714; 10.1038/s41598-019-47110-x (2019).
- 836 7. Linka, M. & de Haas, B. OSIEshort: A small stimulus set can reliably estimate
837 individual differences in semantic salience. *Journal of vision* **20**, 13; 10.1167/jov.20.9.13
838 (2020).
- 839 8. Kollenda, D., Reher, A.-S. & de Haas, B. Individual gaze predicts individual scene
840 descriptions. *Scientific reports* **15**, 9443; 10.1038/s41598-025-94056-4 (2025).
- 841 9. Linka, M., Broda, M. D., Alsheimer, T., de Haas, B. & Ramon, M. Characteristic
842 fixation biases in Super-Recognizers. *Journal of vision* **22**, 17; 10.1167/jov.22.8.17 (2022).
- 843 10. Kennedy, D. P. et al. Genetic Influence on Eye Movements to Complex Scenes at
844 Short Timescales. *Current biology* : CB 27, 3554-3560.e3; 10.1016/j.cub.2017.10.007
845 (2017).
- 846 11. Constantino, J. N. et al. Infant viewing of social scenes is under genetic control and is
847 atypical in autism. *Nature* **547**, 340–344; 10.1038/nature22999 (2017).
- 848 12. de Haas, B., Kanai, R., Jalkanen, L. & Rees, G. Grey matter volume in early human
849 visual cortex predicts proneness to the sound-induced flash illusion. *Proceedings. Biological*
850 *sciences* **279**, 4955–4961; 10.1098/rspb.2012.2132 (2012).
- 851 13. Moutsiana, C. et al. Cortical idiosyncrasies predict the perception of object size.
852 *Nature communications* **7**, 12110; 10.1038/ncomms12110 (2016).
- 853 14. Furl, N., Garrido, L., Dolan, R. J., Driver, J. & Duchaine, B. Fusiform gyrus face
854 selectivity relates to individual differences in facial recognition ability. *Journal of cognitive*
855 *neuroscience*, 1723–1740 (2011).

- 856 15. Elbich, D. B. & Scherf, S. Beyond the FFA: Brain-behavior correspondences in face
857 recognition abilities. *NeuroImage* **147**, 409–422; 10.1016/j.neuroimage.2016.12.042 (2017).
- 858 16. Borovska, P. & de Haas, B. Individual gaze shapes diverging neural representations.
859 *Proceedings of the National Academy of Sciences of the United States of America* **121**,
860 e2405602121; 10.1073/pnas.2405602121 (2024).
- 861 17. Cerf, M., Frady, E. P. & Koch, C. Faces and text attract gaze independent of the task:
862 Experimental data and computer model. *Journal of vision* **9**, 10.1-15; 10.1167/9.12.10
863 (2009).
- 864 18. Weiner, K. S. & Grill-Spector, K. Neural representations of faces and limbs neighbor
865 in human high-level visual cortex: evidence for a new organization principle. *Psychological*
866 *research* **77**, 74–97; 10.1007/s00426-011-0392-x (2013).
- 867 19. Kanwisher, N. & Dilks, D. The Functional Organization of the Ventral Visual Pathway
868 in Humans. Chalupa, L. & Werner, J. (Eds.), *The New Visual Neurosciences* (2012).
- 869 20. Grill-Spector, K. & Weiner, K. S. The functional architecture of the ventral temporal
870 cortex and its role in categorization. *Nature reviews. Neuroscience* **15**, 536–548;
871 10.1038/nrn3747 (2014).
- 872 21. Vinken, K. & Vogels, R. A behavioral face preference deficit in a monkey with an
873 incomplete face patch system. *NeuroImage* **189**, 415–424;
874 10.1016/j.neuroimage.2019.01.043 (2019).
- 875 22. Arcaro, M. J., Schade, P. F., Vincent, J. L., Ponce, C. R. & Livingstone, M. S. Seeing
876 faces is necessary for face-domain formation. *Nature neuroscience* **20**, 1404–1412;
877 10.1038/hn.4635 (2017).
- 878 23. Nordt, M. et al. Cortical recycling in high-level visual cortex during childhood
879 development. *Nature human behaviour* **5**, 1686–1697; 10.1038/s41562-021-01141-5 (2021).

- 880 24. Linka, M., Sensoy, Ö., Karimpur, H., Schwarzer, G. & de Haas, B. Free viewing
881 biases for complex scenes in preschoolers and adults. *Scientific reports* **13**, 11803;
882 10.1038/s41598-023-38854-8 (2023).
- 883 25. Linka, M., Karimpur, H. & de Haas, B. Protracted development of gaze behaviour.
884 *Nature human behaviour*; 10.1038/s41562-025-02191-9 (2025).
- 885 26. Nordt, M. et al. Longitudinal development of category representations in ventral
886 temporal cortex predicts word and face recognition. *Nature communications* **14**, 8010;
887 10.1038/s41467-023-43146-w (2023).
- 888 27. Nordt, M. et al. Learning to Read Increases the Informativeness of Distributed Ventral
889 Temporal Responses. *Cerebral cortex* (New York, N.Y. : 1991) **29**, 3124–3139;
890 10.1093/cercor/bhy178 (2019).
- 891 28. Kubota, E., Grill-Spector, K. & Nordt, M. Rethinking cortical recycling in ventral
892 temporal cortex. *Trends in Cognitive Sciences* **28**, 8-17; 10.1016/j.tics.2023.09.006 (2023).
- 893 29. Dehaene-Lambertz, G., Monzalvo, K. & Dehaene, S. The emergence of the visual
894 word form: Longitudinal evolution of category-specific ventral visual areas during reading
895 acquisition. *PLoS biology* **16**, e2004103; 10.1371/journal.pbio.2004103 (2018).
- 896 30. Stigliani, A., Weiner, K. S. & Grill-Spector, K. Temporal Processing Capacity in High-
897 Level Visual Cortex Is Domain Specific. *The Journal of neuroscience : the official journal of*
898 *the Society for Neuroscience* **35**, 12412–12424; 10.1523/JNEUROSCI.4822-14.2015 (2015).
- 899 31. Grill-Spector, K., Weiner, K. S., Kay, K. & Gomez, J. The Functional Neuroanatomy of
900 Human Face Perception. *Annual review of vision science* **3**, 167–196; 10.1146/annurev-
901 vision-102016-061214 (2017).
- 902 32. Levy, I., Hasson, U., Avidan, G., Hendler, T. & Malach, R. Center–periphery
903 organization of human object areas. *Nature neuroscience* **4**, 533–539; 10.1038/87490
904 (2001).

- 905 33. Epstein, R. & Kanwisher, N. A cortical representation of the local visual environment.
906 *Nature* **392**, 598–601; 10.1038/33402 (1998).
- 907 34. Rossion, B., Joyce, C. A., Cottrell, G. W. & Tarr, M. J. Early lateralization and
908 orientation tuning for face, word, and object processing in the visual cortex. *NeuroImage* **20**,
909 1609–1624; 10.1016/j.neuroimage.2003.07.010 (2003).
- 910 35. Russell, R., Duchaine, B. & Nakayama, K. Super-recognizers: people with
911 extraordinary face recognition ability. *Psychonomic bulletin & review* **16**, 252–257;
912 10.3758/PBR.16.2.252 (2009).
- 913 36. Moll, K. & Landerl, K. SLRT-2: Lese-und Rechtschreibtest: Weiterentwicklung des
914 Salzburger Lese-und Rechtschreibtests (SLRT) (2010).
- 915 37. Mitchell, J. L. et al. The balance between stability and plasticity of the visual word
916 form area in dyslexia. *Nature communications* **17**, 368; 10.1038/s41467-025-67054-3 (2025).
- 917 38. Parkhurst, D., Law, K. & Niebur, E. Modeling the role of salience in the allocation of
918 overt visual attention. *Vision research* **42**, 107–123; 10.1016/S0042-6989(01)00250-4
919 (2002).
- 920 39. Einhäuser, W., Rutishauser, U. & Koch, C. Task-demands can immediately reverse
921 the effects of sensory-driven saliency in complex visual stimuli. *Journal of vision* **8**, 2.1-19;
922 10.1167/8.2.2 (2008).
- 923 40. Sheinberg, D. L. & Logothetis, N. K. Noticing Familiar Objects in Real World Scenes:
924 The Role of Temporal Cortical Neurons in Natural Vision. *Journal of neuroscience* **21**, 1340–
925 1350; 10.1523/JNEUROSCI.21-04-01340.2001 (2001).
- 926 41. Thompson, K. G. & Bichot, N. P. A visual salience map in the primate frontal eye field.
927 *Prog. Brain Res.* **147**, 251–262; 0.1016/S0079-6123(04)47019-8 (2005).
- 928 42. Bisley, J. W. & Goldberg, M. E. Attention, intention, and priority in the parietal lobe.
929 *Annual review of neuroscience* **33**, 1–21; 10.1146/annurev-neuro-060909-152823 (2010).

- 930 43. Broda, M. D., Borovska, P. & de Haas, B. Individual differences in face salience and
931 rapid face saccades. *Journal of vision* **24**, 16; 10.1167/jov.24.6.16 (2024).
- 932 44. Alreja, A., Boring, M., Ward, M., Richardson, M., Abel, T., & Ghuman, A. Intracranial
933 Electroencephalography reveals the neural correlates of eye-movement guidance to objects.
934 *Journal of Vision*, 25(9), 2905-2905; 10.1167/jov.25.9.2905 (2025).
- 935 45. Yang, J., Sepuri, T., Tan, A. W. M., Frank, M. C. & Long, B. Quantifying infants'
936 everyday experiences with objects in a large corpus of egocentric videos (2025).
- 937 46. Pezzulo, G., Parr, T. & Friston, K. Active inference as a theory of sentient behavior.
938 *Biological psychology* **186**, 108741; 10.1016/j.biopsycho.2023.108741 (2024).
- 939 47. Golarai, G. et al. Differential development of high-level visual cortex correlates with
940 category-specific recognition memory. *Nature neuroscience* **10**, 512–522; 10.1038/nn1865
941 (2007).
- 942 48. Feng, X., Monzalvo, K., Dehaene, S. & Dehaene-Lambertz, G. Evolution of reading
943 and face circuits during the first three years of reading acquisition. *NeuroImage* **259**, 119394;
944 10.1016/j.neuroimage.2022.119394 (2022).
- 945 49. Dehaene, S., Cohen, L., Morais, J. & Kolinsky, R. Illiterate to literate: behavioural and
946 cerebral changes induced by reading acquisition. *Nature reviews. Neuroscience* **16**, 234–
947 244; 10.1038/nrn3924 (2015).
- 948 50. Dehaene, S. et al. How Learning to Read Changes the Cortical Networks for Vision
949 and Language. *Science* **330**, 1359–1364; 10.1126/science.1194140 (2010).
- 950 51. Gomez, J., Barnett, M. & Grill-Spector, K. Extensive childhood experience with
951 Pokémon suggests eccentricity drives organization of visual cortex. *Nature human behaviour*
952 **3**, 611–624; 10.1038/s41562-019-0592-8 (2019).

953 52. James, T. W. & James, K. H. Expert individuation of objects increases activation in
954 the fusiform face area of children. *NeuroImage* **67**, 182–192;
955 10.1016/j.neuroimage.2012.11.007 (2013).

956 53. Alvarez, I. et al. Heritable functional architecture in human visual cortex. *NeuroImage*
957 **239**, 118286; 10.1016/j.neuroimage.2021.118286 (2021).

958 54. Benson, N. C. et al. Variability of the Surface Area of the V1, V2, and V3 Maps in a
959 Large Sample of Human Observers. *The Journal of neuroscience : the official journal of the*
960 *Society for Neuroscience* **42**, 8629–8646; 10.1523/JNEUROSCI.0690-21.2022 (2022).

961 55. Polk, T. A., Park, J., Smith, M. R. & Park, D. C. Nature versus nurture in ventral visual
962 cortex: a functional magnetic resonance imaging study of twins. *The Journal of neuroscience*
963 *: the official journal of the Society for Neuroscience* **27**, 13921–13925;
964 10.1523/JNEUROSCI.4001-07.2007 (2007).

965 56. Abbasi, N., Duncan, J. & Rajimehr, R. Genetic influence is linked to cortical
966 morphology in category-selective areas of visual cortex. *Nature communications* **11**, 709;
967 10.1038/s41467-020-14610-8 (2020).

968 57. Peelen, M. V. & Downing, P. E. Within-subject reproducibility of category-specific
969 visual activation with functional MRI. *Human brain mapping* **25**, 402–408;
970 10.1002/hbm.20116 (2005).

971 58. Çukur, T., Nishimoto, S., Huth, A. G. & Gallant, J. L. Attention during natural vision
972 warps semantic representation across the human brain. *Nature neuroscience* **16**, 763–770;
973 10.1038/nn.3381 (2013).

974 59. Rezlescu, C., Danaïla, I., Miron, A. & Amariei, C. More time for science: Using
975 Testable to create and share behavioral experiments faster, recruit better participants, and
976 engage students in hands-on research. *Progress in Brain Research*, 243–262 (2020).

977 60. Kleiner, M., Brainard, D. & Pelli, D. What's new in Psychtoolbox-3? (2007).

- 978 61. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis -
 979 connecting the branches of systems neuroscience. *Frontiers in systems neuroscience* **2**, 4;
 980 10.3389/neuro.06.004.2008 (2008).
- 981 62. Turker, S., Fumagalli, B., Kuhnke, P. & Hartwigsen, G. The 'reading' brain: Meta-
 982 analytic insight into functional activation during reading in adults. *Neurosci. Biobehav. Rev.*
 983 **173**, 106166 (2025).
- 984 63. Jobard, G., Crivello, F. & Tzourio-Mazoyer, N. Evaluation of the dual route theory of
 985 reading: a metanalysis of 35 neuroimaging studies. *NeuroImage* **20**, 693–712 (2003).
- 986 64. Hershler, O., & Hochstein, S. The importance of being expert: Top-down attentional
 987 control in visual search with photographs. *Attention, Perception, & Psychophysics*, **71**(7),
 988 1478-1486; 10.3758/APP.71.7.1478 (2009).
- 989 65. Golan, T., Bentin, S., DeGutis, J. M., Robertson, L. C., & Harel, A. Association and
 990 dissociation between detection and discrimination of objects of expertise: Evidence from
 991 visual search. *Attention, Perception, & Psychophysics*, **76**(2), 391-406; 10.3758/s13414-013-
 992 0562-6 (2014).
- 993 66. Reeder, R. R., Stein, T., & Peelen, M. V. Perceptual expertise improves category
 994 detection in natural scenes. *Psychonomic bulletin & review*, **23**(1), 172-179; 10.3758/s13423-
 995 015-0872-x (2016).
- 996 67. Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. Expertise for cars and
 997 birds recruits brain areas involved in face recognition. *Nature neuroscience*, **3**(2), 191-197;
 998 10.1038/72140 (2000).
- 999 68. Kanwisher, N. Domain specificity in face perception. *Nature neuroscience*, **3**(8), 759-
 1000 763; 10.1038/77664 (2000).
- 1001 69. McKone, E., Kanwisher, N., & Duchaine, B. C. Can generic expertise explain special
 1002 processing for faces? *Trends in cognitive sciences*, **11**(1), 8-15; 10.1016/j.tics.2006.11.002
 1003 (2007).

- 1004 70. Harel, A., Gilaie-Dotan, S., Malach, R., Bentin, s., Top-Down Engagement Modulates
1005 the Neural Expressions of Visual Expertise, *Cerebral Cortex*, 20(10), 2304–2318;
1006 10.1093/cercor/bhp316 (2010).
- 1007 71. Cartwright-Finch, U., & Lavie, N. The role of perceptual load in inattentional
1008 blindness. *Cognition*, 102(3), 321-340; 10.1016/j.cognition.2006.01.002 (2007).
- 1009 72. Stein, T., Reeder, R. R., & Peelen, M. V. Privileged access to awareness for faces
1010 and objects of expertise. *Journal of Experimental Psychology: Human Perception and*
1011 *Performance*, 42(6), 788–798; 10.1037/xhp0000188 (2016).
- 1012 73. Lutchmaya, S., & Baron-Cohen, S. Human sex differences in social and non-social
1013 looking preferences, at 12 months of age. *Infant Behavior and Development*, 25(3), 319-325;
1014 10.1016/S0163-6383(02)00095-4 (2002).
- 1015 74. Proverbio, A. M. Sex differences in social cognition: The case of face processing.
1016 *Journal of neuroscience research*, 95(1-2), 222-234; 10.1002/jnr.23817 (2017).
- 1017 75. Proverbio, A. M., Zani, A., & Adorni, R. Neural markers of a greater female
1018 responsiveness to social stimuli. *BMC neuroscience*, 9(1), 56; 10.1186/1471-2202-9-56
1019 (2008).
- 1020 76. Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. An algorithmic method for
1021 functionally defining regions of interest in the ventral visual pathway. *Neuroimage*, 60(4),
1022 2357-2364; 10.1016/j.neuroimage.2012.02.055 (2012).
- 1023 77. Fedorenko, E., Hsieh, P. J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher,
1024 N. New method for fMRI investigations of language: defining ROIs functionally in individual
1025 subjects. *Journal of neurophysiology*, 104(2), 1177-1194; 10.1152/jn.00032.2010 (2010).
- 1026 78. Nieto-Castañón, A., & Fedorenko, E. Subject-specific functional localizers increase
1027 sensitivity and functional resolution of multi-subject analyses. *Neuroimage*, 63(3), 1646-1669,
1028 10.1016/j.neuroimage.2012.06.065 (2012).

- 1029 79. JASP Team (2026). JASP (Version 0.96.0)[Windows].
- 1030 80. Wetzels, R., & Wagenmakers, E. J. A default Bayesian hypothesis test for
1031 correlations and partial correlations. *Psychonomic bulletin & review*, 19(6), 1057-1064,
1032 10.3758/s13423-012-0295-x (2012).
- 1033 81. Thomason, M. E., Burrows, B. E., Gabrieli, J. D., & Glover, G. H. Breath holding
1034 reveals differences in fMRI BOLD signal in children and adults. *Neuroimage*, 25(3), 824-837;
1035 10.1016/j.neuroimage.2004.12.026 (2005).
- 1036 82. de Haas, B. How to enhance the power to detect brain–behavior correlations with
1037 limited resources. *Frontiers in human neuroscience*, 12, 421; 10.3389/fnhum.2018.00421
1038 (2018).

Contents

Supplementary Materials	2
Calculation of fMRI laterality index of category-specific responses.....	2
Exploratory searchlight analysis	2
Supplementary Results.....	4
Detailed results reported in the main text.....	4
Reliability of multivariate patterns.....	8
No observed correlation between oddball task accuracy and gaze tendencies or category distinctiveness	8
No robust correlations between lateralization of category-specific responses and gaze tendencies or behavioural performance	9
Whole brain searchlight analysis of category distinctiveness and its relation to gaze tendencies	10
References.....	14

Supplementary Materials

Calculation of fMRI laterality index of category-specific responses

To test a potential relationship between fixation tendencies and the lateralization of category-specific responses, we followed the approach described in Blauch et al. (2025)¹. To identify visually responsive cortex, we first defined a stimulus-responsive mask. For each participant, we specified a first-level general linear model (GLM) contrast comparing the average of all six stimulus conditions (faces, bodies, words, houses, cars, limbs) to the blank condition (all stimuli > blank). The resulting t-map was thresholded at $p < .001$ (uncorrected, voxel-wise; extent threshold = 0 voxels) and binarized. The subsequent laterality analysis was restricted to the intersection of this stimulus-responsive mask and the anatomically defined ROIs: lateral and medial ventral temporal cortex (VTC) in the left and right hemispheres (cf. methods section 'Definition of lateral and medial VTC regions').

For each participant and each category-selective contrast (faces and words), we extracted the corresponding first-level t-map. Within the left and right stimulus-responsive VTC masks, we then computed the sum of positive t-values. This measure captures the overall category-selective response profile within the stimulus-responsive anatomical ROI, combining effect magnitude and spatial extent.

Hemispheric laterality was quantified with a laterality index (LI) defined as the relative difference of summed t-values between the left and right hemisphere. $LI > 0$ indicates left-lateralized activation for a given category, $LI < 0$ indicates right-lateralized activation, and $LI \sim 0$ indicates bilateral or weakly lateralized activation.

Exploratory searchlight analysis

For each participant, we computed whole-brain searchlight maps of neural distinctiveness separately for faces and words in native space. To this end, a spherical searchlight with a radius of 4 voxels (~8mm) was iteratively centred on each voxel of a given individual's grey-

matter-constrained image (SPM C1 image, thresholded at $>.02$). For each iteration, we computed category distinctiveness of responses intersecting with the sphere as described in Methods section 'Analysis of category distinctiveness'. We then projected the resulting distinctiveness values onto the corresponding centre voxel. This analysis was implemented using the CoSMoMVPA toolbox².

The resulting distinctiveness maps for word and face-representations were normalized to MNI space using the participant-specific forward deformation fields estimated from each participant's T1-weighted anatomical image in SPM 12 (Wellcome Trust Centre for Neuroimaging, London, UK) and restricted to cerebral cortex using the cortical mask of the Harvard-Oxford atlas (HOA; ^{3,4}).

At the group level, we first tested where category distinctiveness was reliably greater than zero across participants. For each stimulus category (words, faces), we entered the corresponding individual distinctiveness maps into a second-level one-sample t-test in SPM (random-effects GLM). This yielded whole brain statistical parametric maps identifying voxels in which mean distinctiveness values were significantly positive across participants. The resulting maps were thresholded with cluster-level FWE correction at $p < .05$, and significant clusters were saved as binary masks for subsequent analyses (word- and face-selective distinctiveness clusters, respectively).

We then tested whether inter-individual differences of representational distinctiveness within these clusters related to gaze tendencies. For this, we fitted second-level multiple regression models in SPM with the respective distinctiveness map as dependent images and the participant-specific gaze tendencies (first fixation or dwell time) as the covariate of interest. To restrict the analysis to regions that showed reliable category distinctiveness, we used the significant clusters from the corresponding group-sample t-test as an explicit mask in the regression model. We then computed t-contrasts testing for positive associations between category distinctiveness and gaze tendencies (first fixation and dwell time). Anatomical

regions associated with significant peak-level MNI coordinates were identified based on the Harvard-Oxford atlas (HOA; ^{3,4})

For visualizations, the group-level statistical t-maps were projected onto the Freesurfer fsaverage inflated cortical surface. This yielded a topography of cortical regions where local neural category distinctiveness covaried with individual differences in gaze tendencies.

Supplementary Results

Detailed results reported in the main text

Table S1

Overview of Results Reported in the Main Text

Test	Result	95% confidence interval	Two-tailed	Alpha
Result section: Representational distinctiveness is linked to individual differences in gaze				
Max-min ratio of face distinctiveness in the right lateral VTC	1.38			
Max-min ratio of word distinctiveness in the left lateral VTC	1.52			
Correlation between face distinctiveness in right lateral VTC and (1) first fixations and (2) dwell time on faces	(1) $r(59) = .38, p = .003^*$ (2) $r(59) = .30, p = .017^*$	(1) [0.1452, 0.5630] (2) [0.1251, 0.4653]	yes	.025
Correlation between word distinctiveness in left lateral VTC and (1) first fixations and (2) dwell time on text	(1) $r(59) = .44, p < .001^*$ (2) $r(59) = .29, p = .023^*$	(1) [0.2476, 0.6008] (2) [0.0824, 0.4853]	yes	.025
Correlation between word distinctiveness in right lateral VTC and (1) first fixations	(1) $r(59) = .35, p = .006^*$ (2) $r(59) = .14, p = .299$	(1) [0.1303, 0.5467] (2) [-0.0450, 0.3174]	yes	.025

and (2) dwell time on text				
Correlation between face distinctiveness in left lateral VTC and (1) first fixations and (2) dwell time on faces	(1) $r(59) = .20, p = .124$ (2) $r(59) = .05, p = .726$	(1) [-0.0376, 0.4242] (2) [-0.2003, 0.2902]	yes	.025
Cross-correlations between word distinctiveness in left lateral VTC and (1) first fixations and (2) dwell time on faces	(1) $r(59) = .16, p = .222$ (2) $r(59) = -.03, p = .818$	(1) [-0.0787, 0.3985] (2) [-0.2813, 0.2134]	yes	.025
Cross-correlations between face distinctiveness in right lateral VTC and (1) first fixations and (2) dwell time on text	(1) $r(59) = -.36, p = .005^*$ (2) $r(59) = -.24, p = .058$	(1) [-0.5503, -0.1372] (2) [-0.4271, -0.0379]	yes	.025
Result section: Representational distinctiveness is linked to behavioural performance				
correlation between face distinctiveness in right lateral VTC and the CFMT score	$r(32) = .44$, one-tailed $p = .004^*$	[0.2044, 0.6427]	no	.05
correlation between word distinctiveness in left lateral VTC and reading-speed scores for (1) pseudowords and (2) words	(1) $r(23) = .44$, one-tailed $p = .012^*$ (2) $r(23) = .17$, one-tailed $p = .204$	(1) [0.0311, 0.7218] (2) [-0.2953, 0.6170]	no	.025
Result section: The individual size of IVWFA and rFFA is linked to fixation tendencies and behavioural performance				
Correlation between rFFA cluster size and (1) first fixations and (2) dwell time on faces	(1) $r(59) = .39, p = .002^*$ (2) $r(59) = .28, p = .031$	(1) [0.1650, 0.5831] (2) [0.0175, 0.5230]	yes	.025
Cross-correlation between rFFA cluster size and (1) first fixations and (2) dwell time on text	(1) $r(59) = -.37, p = .004^*$ (2) $r(59) = -.34, p = .007^*$	(1) [-0.5636, -0.1262] (2) [-0.5243, -0.1321]	yes	.025

Correlation between rFFA cluster size and face distinctiveness in right lateral VTC	$r(59) = .55, p < .001^*$	[0.3634, 0.6963]	yes	.05
Correlation between rFFA cluster size and CFMT score	$r(32) = .37$, one-tailed $p = .014^*$	[0.0923, 0.5983]	no	.05
Correlation between IVWFA cluster size and (1) first fixations and (2) dwell time on text	(1) $r(59) = .35, p = .006^*$ (2) $r(59) = .26, p = .039$	(1) [0.1453, 0.5411] (2) [0.0642, 0.4542]	yes	.025
Cross-correlation between IVWFA cluster size and (1) first fixations and (2) dwell time on faces	(1) $r(59) = .09, p = .514$ (2) $r(59) = .01, p = .962$	(1) [-0.1425, 0.3056] (2) [-0.2147, 0.2132]	yes	.025
Correlation between IVWFA cluster size and word distinctiveness in left lateral VTC	$r(59) = .75, p < .001^*$	[0.6319, 0.8331]	yes	.05
Correlation between IVWFA cluster size and reading-speed scores for (1) words and (2) pseudowords	(1) $r(23) = .10$, one-tailed $p = .319$ (2) $r(23) = .32$, one-tailed $p = .060$	(1) [-0.3027, 0.5117] (2) [-0.0507, 0.5939]	yes	.025
Result section: Face salience is linked to face recognition performance				
Max-min ratio of CFMT performance scores	<i>subsample:</i> 1.98; $M = 64.08, SD = 11.87$ <i>extended sample:</i> 2.33; $M = 64.37, SD = 11.95$			
correlation between CFMT scores and (1) proportion of first fixations and (2) proportion of dwell time on faces	(1) <i>subsample:</i> $r(32) = .23$, one-tailed $p = .094$ <i>extended sample:</i> $r(67) = .39$, one-tailed $p < .001^*$ (2) <i>subsample:</i> $r(32) = .16$, one-tailed $p = .174$ <i>extended sample:</i> $r(67) = .18$, one-tailed $p = .077$	(1) <i>subsample:</i> [-0.0662, 0.4925] <i>extended sample:</i> [0.1679, 0.5767] (2) <i>subsample:</i> [-0.1726, 0.4917] <i>extended sample:</i> [-0.0438, 0.3849]	No	.025

Max-min ratio of speed-reading test scores for (1) words and (2) pseudowords	(1) <i>subsample</i>: 1.97; $M = 118.40$, $SD = 16.17$ <i>extended sample</i>: 2.11; $M = 111.35$, $SD = 18.07$ (2) <i>subsample</i>: 2.15; $M = 77.92$, $SD = 14.57$ <i>extended sample</i>: 3.52; $M = 74.68$, $SD = 16.36$			
correlation between speed-reading test scores for (a) words and (b) pseudowords and (1) first fixations and (2) proportion of dwell time on text	(1a) <i>subsample</i>: $r(23) = .50$, one-tailed $p = .004^*$ <i>extended sample</i>: $r(63) = -.04$, one-tailed $p = .386$ (1b) <i>subsample</i>: $r(23) = .69$, one-tailed $p < .001^*$ <i>extended sample</i>: $r(63) = .19$, one-tailed $p = .065$ 2a) <i>subsample</i>: $r(23) = .34$, one-tailed $p = .042$ <i>extended sample</i>: $r(63) = -.17$, one-tailed $p = .090$ (2b) <i>subsample</i>: $r(23) = .39$, one-tailed $p = .025$ <i>extended sample</i>: $r(63) = .04$, one-tailed $p = .383$	(1a) <i>subsample</i>: [0.2336, 0.7410] <i>extended sample</i>: [-0.3098, 0.2390] (1b) <i>subsample</i>: [0.4264, 0.8396] <i>extended sample</i>: [-0.0448, 0.4249] (2a) <i>subsample</i>: [-0.0714, 0.6644] <i>extended sample</i>: [-0.4428, 0.1139] (2b) <i>subsample</i>: [-0.1370, 0.6730] <i>extended sample</i>: [-0.2158, 0.2636]	No	.0125

Note. * indicates significant correlations. The significance level was Bonferroni-corrected to account for the use of two salience metrics (proportion of first fixations and dwell time), and two speed reading scores (words and pseudowords). Adjusted significance levels and test directionality are denoted for each comparison.

Reliability of multivariate patterns

We compared the reliability of category-specific multivariate patterns in the individual brain before and after smoothing with a kernel of 5 mm. Averaged across participants, mean pattern reliability in *right* lateral VTC *without smoothing* was .83 ($SD = .08$) for faces, .71 ($SD = .12$) for bodies, .69 ($SD = .11$) for words, .67 ($SD = .15$) for houses, .45 ($SD = .17$) for cars, and .64 ($SD = .14$) for limbs.

With smoothing, mean category reliability in *right* lateral VTC was .94 ($SD = .04$) for faces, .87 ($SD = .08$) for bodies, .86 ($SD = .09$) for words, .86 ($SD = .09$) for houses, .67 ($SD = .19$) for cars, and .84 ($SD = .10$) for limbs.

Without smoothing, mean category reliability in *left* lateral VTC was .81 ($SD = .08$) for faces, .64 ($SD = .14$) for bodies, .78 ($SD = .10$) for words, .66 ($SD = .16$) for houses, .48 ($SD = .15$) for cars, and .67 ($SD = .14$) for limbs.

With smoothing, mean category reliability in *left* lateral VTC was .92 ($SD = .05$) for faces, .82 ($SD = .10$) for bodies, .90 ($SD = .05$) for words, .86 ($SD = .09$) for houses, .68 ($SD = .13$) for cars, and .85 ($SD = .12$) for limbs.

No observed correlation between oddball task accuracy and gaze tendencies or category distinctiveness

Here we tested whether performance in the oddball task during category-specific blocks of the fMRI localiser session correlated with corresponding gaze tendencies or category distinctiveness. This was not the case (see Table S2). Note, two participants had to be excluded from this analysis due to button malfunction during the fMRI localizer experiment.

Table S2

Correlations between Oddball Task Accuracy, Gaze Tendencies and Category Distinctiveness

Correlation	Result
correlation between Oddball accuracy in face blocks and (1) proportion of first	(1) $r(56) = .14$, $p = .287$ (2) $r(56) = .07$, $p = .624$

fixations and (2) proportion of dwell time on faces	
correlation between Oddball accuracy in word blocks and (1) proportion of first fixations and (2) proportion of dwell time on text	(1) $r(56) = .01, p = .945$ (2) $r(56) = -.03, p = .808$
correlation between Oddball accuracy in face blocks and face distinctiveness in the right lateral VTC	$r(56) = .05, p = .728$
correlation between Oddball accuracy in word blocks and word distinctiveness in the left lateral VTC	$r(56) = .15, p = .264$

No robust correlations between lateralization of category-specific responses and gaze tendencies or behavioural performance

We explored hemispheric laterality indices for face- and word- selective responses in VTC (cf. ¹) and tested whether individual differences in laterality were associated with gaze tendencies and behavioural performance. As expected, face-selective activation was on average right lateralized (mean LI_faces = -0.15, SD = 0.28; negative values indicate right hemisphere dominance; $t(60) = -4.14, p < .001$), whereas word-selective activation was left-lateralized (mean LI_words = 0.45, SD = 0.26; $t(60) = 13.34, p < .001$). Importantly, participants varied substantially in the degree of category laterality.

To illustrate this variability, Figure 2 highlights two participants with opposite gaze tendencies. For face selectivity, participant 40 (who showed a stronger tendency to look at faces and a relatively high CFMT score) exhibited a stronger right hemispheric bias (LI_faces = -0.10) relative to participant 4 (who showed a stronger tendency to look at text and relatively high pseudoword reading score; LI_faces = -.02).

However, across participants, we found no robust evidence for relationships between laterality indices and gaze tendencies or behavioural performance (see Table S3). The only (negative) association was between LI_words and the proportion of first fixations on text, $r(58) = -.26$, two-tailed $p = .041$, but this effect did not survive correction for multiple comparisons (Bonferroni-corrected significance threshold $\alpha = .025$).

Furthermore, we found no meaningful association between LI_faces and LI_words ($r(58) = -.10$, $p = .433$), nor between word distinctiveness in left lateral VTC and face distinctiveness in right lateral VTC ($r(58) = -.04$, $p = .768$). Together, these results align with prior work reporting no evidence for local face–word competition in VTC¹.

Table S3

Correlations between Category Lateralization Index and Gaze Tendencies and Behavioural Performance

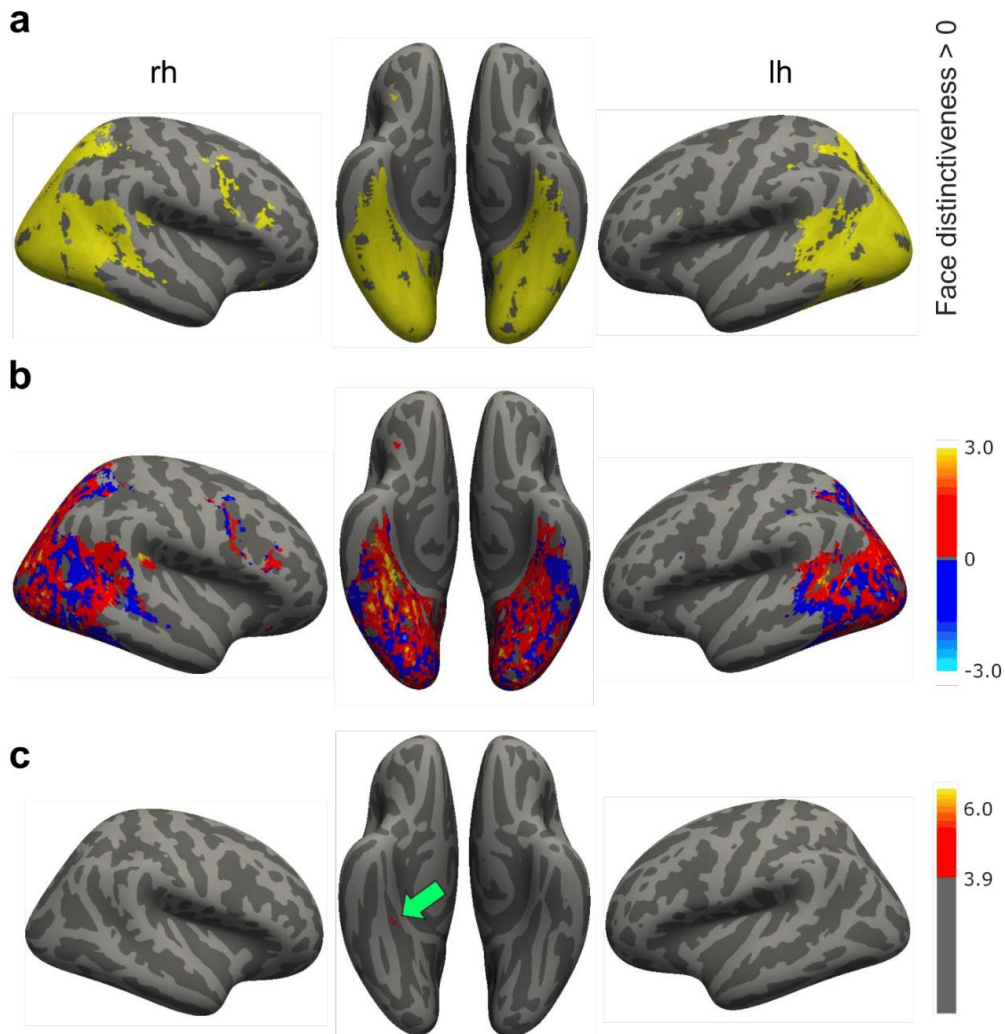
Correlation	Result
correlation between LI_faces and (1) proportion of first fixations and (2) proportion of dwell time on faces	(1) $r(58) = -.15$, two-tailed $p = .233$ (2) $r(58) = -.20$, two-tailed $p = .115$
correlation between LI_words and (1) proportion of first fixations and (2) proportion of dwell time on text	(1) $r(58) = -.26$, two-tailed $p = .041$ (2) $r(58) = -.12$, two-tailed $p = .353$
cross-correlation between LI_faces and (1) proportion of first fixations and (2) proportion of dwell time on text	(1) $r(58) = .25$, two-tailed $p = .055$ (2) $r(58) = .12$, two-tailed $p = .367$
cross-correlation between LI_words and (1) proportion of first fixations and (2) proportion of dwell time on faces	(1) $r(58) = .09$, two-tailed $p = .477$ (2) $r(58) = .20$, two-tailed $p = .132$
correlation between LI_faces and CFMT scores (N = 34)	$r(32) = -.10$, two-tailed $p = .588$
correlation between LI_words and reading test scores for pseudowords (N = 25)	$r(23) = -.20$, two-tailed $p = .329$
correlation between LI_words and reading test scores for words (N = 25)	$r(23) = -.14$, two-tailed $p = .508$

Whole brain searchlight analysis of category distinctiveness and its relation to gaze tendencies

Face distinctiveness was reliably greater than zero across participants in large parts of the inferior temporal cortex and parietal areas (Figures S1a). In the exploratory analysis relating face distinctiveness to the proportion of first fixations on faces, no significant clusters survived family-wise error (FWE) correction. For the correlation with dwell time on faces, a

cluster located in the right temporal occipital fusiform cortex (MNI peak coordinates xyz = 32, -48, -22; $t(59) = 3.98$; $p(\text{corrected}) = .007$; cluster size: $k = 20$; cf. Figure S1c) showed a correlation. The location of this peak is more medial than that we identified as rFFA using the automated Group-constrained Subject-Specific (GSS) approach, but corresponds to rFFA coordinates reported by others (cf. Spiridon et al., 2006⁵).

Figure S1: Searchlight for face distinctiveness and gaze tendencies

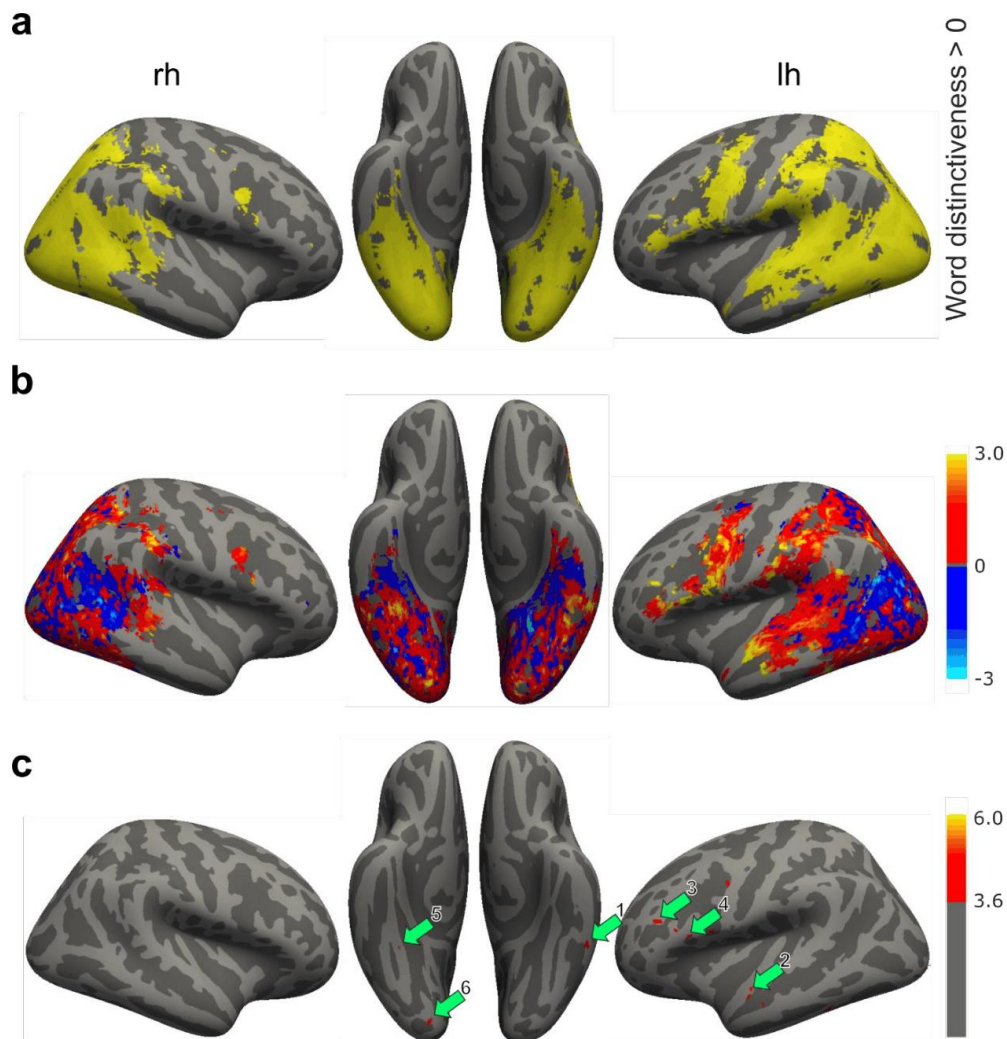


All panels show group-level patterns ($N = 61$). (a) Second-level GLM contrast testing face distinctiveness > 0 in the left (lh) and right (rh) hemispheres. The resulting binary map was used as a mask for the subsequent GLM analyses. (b, c) Overlays of t-contrast corresponding to the correlation between face distinctiveness and the proportion of dwell time on faces. The same contrast is shown both, without (b) and with (c) a minimum threshold applied, to give the complete picture and highlight statistically significant results (cf.

different inset colour bars on the right, showing corresponding t-values). The single significant cluster in rFFA that survived FWE corrections ($p(\text{corrected}) = .007$) is marked by a green arrow in (c); MNI peak coordinates: 32, -48, -22; temporal occipital fusiform cortex (TOF).

Word distinctiveness was reliably greater than zero across participants in large parts of the inferior temporal cortex, parietal and frontal areas (Figures S2a). For the relationship between word distinctiveness and the proportion of first fixations on text, we observed six significant clusters. The first temporooccipital cluster corresponded well to the left visual word from area (IVWFA; cf. ⁶; MNI peak coordinates xyz = -50, -46, -28; $t(59) = 5.18$; $p(\text{corrected}) < .001$; cluster size: $k = 34$; cf. no. 1 in Figure S2c). A second significant cluster emerged in the left superior temporal gyrus (STGa; xyz = -62, -2, -4; $t(59) = 4.11$; $p(\text{corrected}) = .026$; cluster size: $k = 53$; cf. no. 2 in Figure S2c). Additionally, two significant clusters were identified in the left inferior frontal gyrus (IFGpt; xyz = -50, 30, 18; $t(59) = 3.95$; $p(\text{corrected}) = .002$; cluster size: $k = 29$; cf. no. 3 in Figure S2c, and IFGpo; xyz = -50, 8, 12; $t(59) = 3.62$; $p(\text{corrected}) = .011$; cluster size: $k = 21$; cf. no. 4 in Figure S2c). These regions are established components of the reading network^{7,8}. A fifth and sixth significant cluster emerged in locations of the right temporal occipital fusiform cortex (TOF; xyz = 30, -46, -20; $t(59) = 3.74$; $p(\text{corrected}) = .032$; cluster size: $k = 17$; cf. no. 5 in Figure S2c) and lingual gyrus (xyz = 16, -86, -8; $t(59) = 4.50$; $p(\text{corrected}) = .007$; cluster size: $k = 23$; cf. no. 6 in Figure S2c). Previous reports found these regions to be active during letter and word processing as well⁹. No cluster survived FWE correction for the relationship between word distinctiveness and proportion of dwell time on text.

Figure S2: Searchlight for word distinctiveness and gaze tendencies



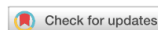
All panels show group-level patterns ($N = 61$). (a) Second-level GLM contrast testing word distinctiveness > 0 in the left (lh) and right (rh) hemispheres. The resulting binary map was used as a mask for the subsequent GLM analyses. (b, c) Overlays of t-contrasts showing the correlation between word distinctiveness and the proportion of first fixations on text. The same contrast is shown both, without (b) and with (c) a minimum threshold applied, to give the complete picture and highlight statistically significant results (cf. different inset colour bars on the right, showing corresponding t-values). The clusters in (c) that survived FWE corrections are marked by green arrows and corresponding numbers; MNI peak coordinates: 1) -50, -46, -28; inferior temporal gyrus, temporooccipital part (ITGtp), $p(\text{corrected}) < .001$; 2) -62, -2, -4; superior temporal gyrus, anterior division (STGa), $p(\text{corrected}) = .026$; 3) -50, 30, 18; inferior frontal gyrus, pars triangularis (IFGpt), $p(\text{corrected}) = .002$; 4) -50, 8, 12; inferior frontal gyrus, pars opercularis (IFGpo), $p(\text{corrected}) = .011$; 5) 30, -46, -20, temporal occipital fusiform cortex (TOF), $p(\text{corrected}) = .032$; 6) 16, -86, -8, occipital fusiform gyrus (OFG), $p(\text{corrected}) = .007$.

References

1. Blauch, N. M., Plaut, D. C., Vin, R. & Behrmann, M. Individual variation in the functional lateralization of human ventral temporal cortex: Local competition and long-range coupling. *Imaging Neurosci.* **3**, imag_a_00488 (2025).
2. Frontiers | CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave.
<https://www.frontiersin.org/journals/neuroinformatics/articles/10.3389/fninf.2016.00027/full>.
3. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006).
4. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *NeuroImage* **62**, 782–790 (2012).
5. Spiridon, M., Fischl, B. & Kanwisher, N. Location and spatial profile of category-specific regions in human extrastriate cortex. *Hum. Brain Mapp.* **27**, 77–89 (2006).
6. Dehaene, S. *et al.* How Learning to Read Changes the Cortical Networks for Vision and Language. *Science* **330**, 1359–1364 (2010).
7. Turker, S., Fumagalli, B., Kuhnke, P. & Hartwigsen, G. The ‘reading’ brain: Meta-analytic insight into functional activation during reading in adults. *Neurosci. Biobehav. Rev.* **173**, 106166 (2025).
8. Jobard, G., Crivello, F. & Tzourio-Mazoyer, N. Evaluation of the dual route theory of reading: a metanalysis of 35 neuroimaging studies. *NeuroImage* **20**, 693–712 (2003).
9. Protopapas, A. *et al.* Evaluating cognitive models of visual word recognition using fMRI: Effects of lexical and sublexical variables. *NeuroImage* **128**, 328–341 (2016).

Study 2: Individual gaze predicts individual scene descriptions

Kollenda, D., Reher, A. S., & de Haas, B. (2025). Individual gaze predicts individual scene descriptions. *Scientific Reports*, 15(1), 9443.



OPEN Individual gaze predicts individual scene descriptions

Diana Kollenda^{1,2}, Anna-Sophia Reher¹ & Benjamin de Haas^{1,2}

Do different people looking at the same scene perceive individual versions of what's in front of them? If perception is individual, which mechanisms mediate our particular view of the world? Recent findings have shown systematic observer differences in gaze, but it is unclear whether individual fixation biases translate to divergent impressions of the same scene. Here, we find systematic differences in the scene descriptions individual observers provide for identical complex scenes. Crucially, observer differences in fixation patterns predicted pairwise differences in scene descriptions, particularly the use of nouns, even for out-of-sample images. Part of this could be explained by the individual tendency to fixate text and people predicting corresponding description references. Our results strongly suggest that subjective scene perception is shaped by individual gaze.

Keywords Individual differences, Perceptual consequences, Gaze behaviour, Scene descriptions

What does it mean to see with your own eyes? When presented with the same scene, how do our individual impressions diverge from each other? While the subjective quality of perceptual experience is private¹, neuroimaging work suggests a high degree of observer-similarity in the geometry of ventral stream representations². At the same time, neural responses to identical stimuli can diverge in ways scaling with social distance³ and the individual functional anatomy of the visual system predicts low-level psychophysical biases^{4,5}. How individual perception varies for complex natural scenes and which mechanisms mediate such differences is less well understood.

When free-viewing complex scenes, human fixations are systematically drawn towards objects, faces, people and text^{6–9}. Recent findings show that these general tendencies are modulated by substantial individual variation: Observers show large and robust individual differences in the tendency to fixate faces, people and text as well as other semantic salience dimensions^{10–14}. Moreover, twin studies indicate a surprising degree of heritability for individual gaze, suggesting it is shaped by individual biology^{13,15}. Do such individual differences in gaze bear consequences for scene understanding, perception and memory?¹⁶ Previous studies have shown that individual face and head salience are related to skills in face recognition^{10,17} and to prioritising social elements for scene understanding¹⁴. Nevertheless, it is largely unclear whether individual gaze is systematically related to the individual way we perceive and communicate natural scenes and their content.

At the group level, previous research shows that objects which are critical to scene understanding are fixated (preprint by¹⁸) and objects which are fixated (longer), in turn are remembered and reported more frequently and in more detail^{19–22}. For example, one study demonstrated that the more fixations are made on an object, the better the memory for its position in space¹⁹. Another study found a positive correlation between the naming of objects in image descriptions and previous fixations on these²².

When describing pictures, artwork or our surroundings, verbal communication is a tool to convey our observations to others. Such descriptions can be rich in detail, even for short presentation times²³. When asking observers to focus on the most relevant aspects and limit their descriptions to one or two sentences, individual descriptions can vary substantially²⁴, especially if there are many objects in a scene^{21,22}, as is typical for complex natural scenes. A study by Coco & Keller (2012)²⁵ combined a target search and sentence production task to investigate the relationship between fixation locations and scene descriptions. Visual attention towards a cued visual target influenced scene descriptions, and descriptions were more similar between participants when the order and location of fixations were also more similar. Importantly, the observed cross-modal similarity for a given scene was greater within an individual than between different individuals. This finding enabled the researchers to predict a participant's scene description based on their respective gaze pattern. However, it is unclear whether such individual differences in scene descriptions are systematic and reproducible, not only for a particular scene but also across different sets of scenes, as suggested by the findings of trait-like differences in gaze^{10,11,26}. We aimed to investigate this gap and hypothesized that stable individual differences in fixation tendencies predict corresponding differences in descriptions¹⁶.

¹Experimental Psychology, Justus Liebig University, Giessen, Germany. ²Center for Mind, Brain and Behavior (CMBB), Marburg and Giessen, Germany. email: diana.kollenda@googlemail.com

Here, we find that individual differences in briefly describing 100 natural scenes reflect stable observer differences, which are explained and predicted by individual differences in gaze behaviour towards those scenes. Specifically, we computed pairwise observer similarities, based on the semantic similarity of complete scene descriptions as well as in the use of nouns, verbs or adjectives specifically. Similarly, we quantified pairwise similarities of dwell time distributions across objects in scenes. This abstraction to pairwise observer similarities enabled us to test a relationship between individual gaze and descriptions independent of a priori assumptions regarding the precise mapping of the two modalities onto each other (cf. ref.²⁷). Observers with more similar patterns of gaze indeed provided more similar scene descriptions. Crucially, these individual tendencies proved robust enough to decode which observer provided which descriptions, based on individual gaze for a *separate* set of scenes, suggesting trait-like differences in gaze result in trait-like differences in perception. The relationship between gaze and descriptions was particularly evident in the idiosyncratic use of nouns, specifically the tendency to fixate and reference people and text.

Results

Observer (dis)similarities in gaze and descriptions are highly consistent across scenes

Participants viewed 100 natural scenes for three seconds each and briefly described what they had just seen after each scene. We recorded each observer's gaze and represented it as a vector of dwell times across all objects and people in the images. We then quantified pairwise observer dissimilarities as the difference between these vectors and entered all pairwise dissimilarities into a Gaze Dissimilarity Matrix (GDM_{dwell}). Repeating this separately for odd and even scenes confirmed highly consistent inter-individual gaze dissimilarities across scenes ($r = 0.72$, $p < 0.001$; Fig. 1c; also see Fig. 1a,b for example overlays of pairs with similar and dissimilar gaze tendencies; cf. ref.¹⁰). We additionally tested individual differences in the tendency to fixate a given object at all. When calculating pairwise dissimilarities based on binary vectors indicating whether an object was fixated or not (coded as 1 or 0; GDM_{fix}), the consistency of inter-individual gaze dissimilarities was reduced, but still significant, $r = 0.58$, $p < 0.001$.

Similarly, we quantified pairwise *description* differences as one minus the semantic similarity of scene descriptions. We did this separately for complete scene descriptions as well as for the similarity of used nouns, verbs or adjectives. To derive a quantitative estimate of semantic similarity, we determined the cosine similarity of description embeddings using SentenceBERT²⁸ for complete descriptions and the genSim library²⁹ and pretrained fastText model^{30,31} for pairs of words (see methods for details). We then averaged the resulting difference values across scenes for each pair of observers to compute Description Dissimilarity Matrices (DDMs). Split-half correlations confirmed high internal consistencies for the DDMs for complete descriptions, $r = 0.79$, $p < 0.001$, nouns, $r = 0.71$, $p < 0.001$; verbs, $r = 0.95$, $p < 0.001$; and adjectives, $r = 0.71$, $p < 0.001$ (see Fig. 2a–b for an example). Thus, pairwise dissimilarities in descriptions reflect systematic observer differences, which generalize across scenes. Notably, we found highly consistent differences in the use of all word types, even though their

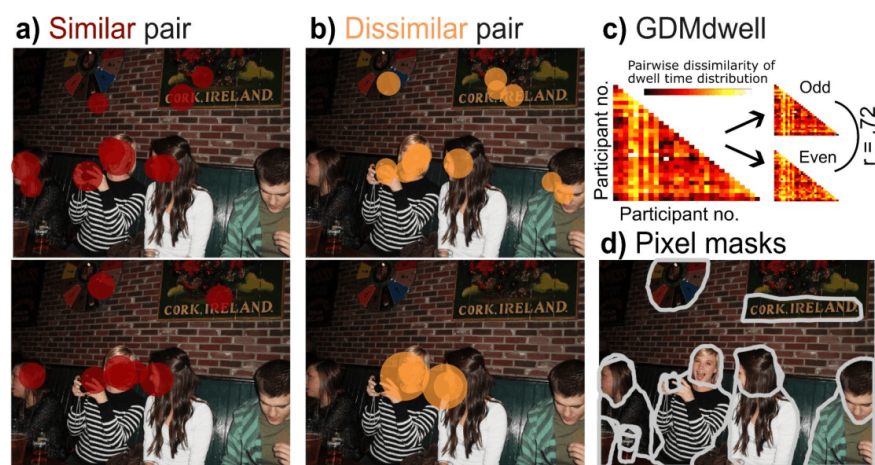


Fig. 1. Pairwise dissimilarities in gaze. (a) Example fixations from a pair of observers with more similar gaze (participants 1 and 11; dissimilarity score of 0.14) and (b) with more dissimilar gaze (participants 9 and 19, dissimilarity score of 0.87). Dissimilarity scores range from 0 to 1 with higher values indicating less similarity. The size of fixation disks scales with fixation duration. (c) The gaze dissimilarity matrix (GDM) shows the pairwise observer dissimilarities based on object dwell times. Hotter colours indicate higher dissimilarity scores. The correlation value corresponds to the split-half reliability of the GDM across odd and even trials ($p < 0.001$). (d) The example image was modified from Xu et al.'s (2014) OSIE dataset, which was published under the Massachusetts Institute of Technology license⁶, to display object pixel masks used for calculating object dwell times.

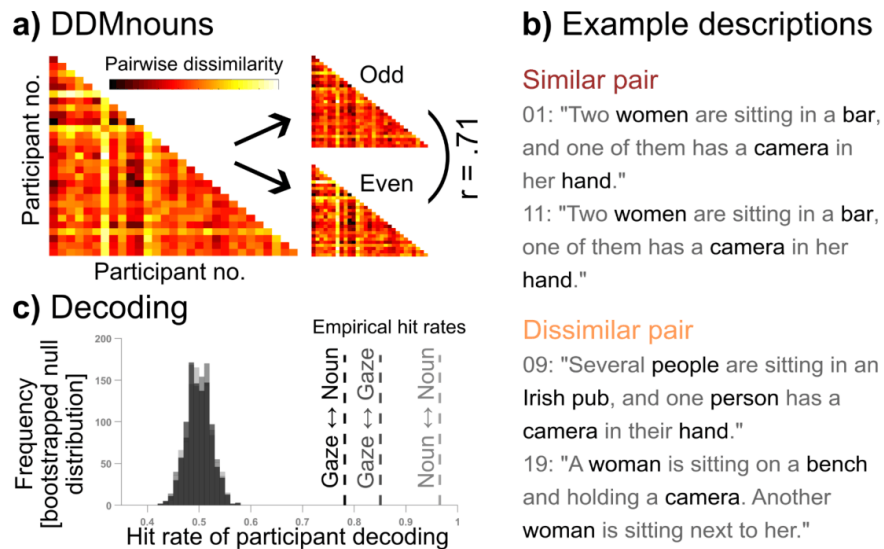


Fig. 2. Pairwise dissimilarities in scene descriptions. **(a)** The example description dissimilarity matrix (DDM) shows pairwise observer dissimilarities based on the average semantic similarities of used nouns in descriptions. Higher values are displayed as hotter colours and indicate less semantic similarity. The correlation shows the split-half reliabilities across odd and even scenes, $p < 0.001$. **(b)** Example descriptions for the scene shown in Fig. 1 from a similar (participants 1 and 11; dissimilarity scores for nouns: 0.39; verbs: 0.34; adjectives: 0.2) and dissimilar pair of observers (participants 9 and 19; dissimilarity scores for nouns: 0.72; verbs: 0.48; adjectives: 0.35). The values range from 0 to 1 with higher values indicating higher dissimilarity. All sentences were translated to English and the original German versions can be found in Supplementary Information B. **(c)** The vertical lines represent the empirical hit rates for decoding individual observers based on out-of-sample training data from different sets of scenes and are categorized by decoding of gaze $\leftrightarrow$ gaze (dark grey), noun $\leftrightarrow$ noun (light grey), and gaze $\leftrightarrow$ noun (black). Each hit rate is compared against its respective bootstrapped null distribution.

average frequency per description varied considerably (nouns: $M = 3.88$, $SD = 1.29$; verbs: $M = 1.91$, $SD = 0.77$; adjectives: $M = 0.53$, $SD = 0.65$).

Taken together, our findings confirm that pairwise similarities in both, gaze and scene descriptions are stable across separate sets of scenes.

Gaze similarity explains description similarity

In order to test the main hypothesis that individual gaze predicts individual descriptions, we correlated the gaze dissimilarity matrix with the description dissimilarity matrices for complete descriptions and each word type. In line with our hypothesis, we observed positive correlations between the dissimilarities for dwell time distributions (GDM_{dwell}) and complete scene description (DDM_{complete}), $r = 0.33$, $p < 0.001$, as well as nouns (DDM_{noun}), $r = 0.42$, $p < 0.001$, but not verbs, $r = -0.07$, $p = 0.128$ and adjectives, $r = -0.06$, $p = 0.210$. Although individual differences in which objects were fixated at all (GDM_{fix}) are less consistent (see above), GDM_{fix} correlated positively with dissimilarities in noun use DDM_{noun}, $r = 0.38$, $p < 0.001$ and to a lesser extent with dissimilarities in complete descriptions DDM_{complete}, $r = 0.21$, $p < 0.001$. All remaining analyses focus on dissimilarities in noun use, but we will return to the use of verbs and adjectives in the discussion section.

To further test the relationship between individual gaze and scene descriptions, we tested whether observer identity can be decoded via dissimilarity patterns across independent sets of scenes. Specifically, we aimed to decode which of two observers provided a given set of fixations or descriptions, based on their dissimilarity to other observers. This tests two assumptions: 1) Every observer has a fairly unique pattern of gaze and description dissimilarity compared to other observers; 2) This unique pattern of dissimilarity generalizes across independent sets of scenes. When splitting the gaze data into those from odd and even scenes, we were able to decode which of two observers provided a given dwell time distribution based on the nearest neighbor correlation of dissimilarity patterns with those observed for the other half of the data (GDM_{dwell}; gaze $\leftrightarrow$ gaze decoding hit rate 85%, $p < 0.001$). The same applied to decoding observer identity based on the dissimilarity pattern of descriptions (DDM_{noun}; description $\leftrightarrow$ description decoding hit rate 97%, $p < 0.001$). Crucially, inter-observer differences did not just generalize across scenes, but also from gaze to descriptions. Decoding which of two observers provided a given set of scene descriptions (or gaze patterns) also worked based on a nearest neighbor correlation between their scene description dissimilarities with their dwell-time dissimilarities for another set of scenes

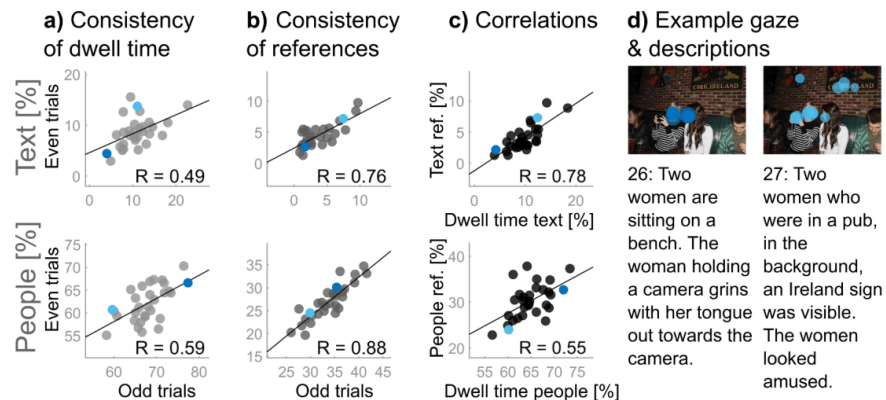


Fig. 3. Fixations predict references to text and people. Correlations show (a) consistency of individual dwell time allocations across odd and even trials for text elements, $p = 0.006$, and people, $p < 0.001$, as well as (b) consistency of individual references in descriptions across odd even trials for text, $p < 0.001$ and people, $p < 0.001$. Correlations in (c) show consistency of individual differences across dwell time allocations and references to text, $p < 0.001$ and people, $p = 0.002$. Coloured dots in each scatter highlight data from the two example participants (26 and 27) whose fixation pattern and descriptions are shown in (d). The original German sentences can be found in Supplementary Information B.

(GDM_{dwell} and DDM_{noun}; gaze $\leftrightarrow$ description decoding hit rate of 78%, $p < 0.001$; see Fig. 2c). All p-values were derived from a statistical comparison of the respective empirical hit rate with their bootstrapped null distribution in a permutation analysis (see Methods for details).

Taken together, individual differences in gaze predict individual differences in descriptions, which generalize to hold-out scenes.

The tendency to fixate text and people explains the tendency to reference them in descriptions

Previous studies found consistent individual differences in the tendency to fixate text and people^{10–14}, which replicated in our experiment (split-half reliabilities of $r = 0.49$, $p = 0.006$, and $r = 0.59$, $p < 0.001$, for text and people, respectively). Some observers spent more than four times (factor of 4.66) as much of their dwell time on text than others and the proportion of dwell time spent on people varied up to a factor of 1.3 (cf. Figure 3a). Furthermore, we found a negative correlation between the individual proportion of dwell time on text and that on people, $r = -0.74$, $p < 0.001$, again in line with earlier results.¹⁰

Interestingly, individual differences in the proportion of text and people references in descriptions were also highly consistent, ($r = 0.76$, $p < 0.001$ and $r = 0.88$, $p < 0.001$, respectively), with some observers referencing text 7.36 as often as others and some observers referencing people 1.65 as often as others (cf. Figure 3b). Again, individual differences in the proportions of text and people references were negatively correlated with each other, $r = -0.58$, $p < 0.001$. Crucially, the individual proportion of dwell time on text was highly correlated with the individual proportion of text references in descriptions, $r = 0.78$, $p < 0.001$, and the same was true for the individual proportion of dwell time on people and the proportion of people references in descriptions, $r = 0.55$, $p = 0.002$ (cf. Figure 3c). That is, a large fraction of the variance in referencing text and people in scene descriptions could be explained by corresponding fixation biases (see Fig. 3d for a gaze and description example).

These results suggest that individual tendencies to fixate and reference text and people are an important part of the more general relationship between individual differences in dwell time distributions (GDM_{dwell}) and descriptions (DDM_{noun}). In an additional analysis we tested whether this relationship can be reduced to individual fixation biases towards text and people or captures additional covariance. To juxtapose these possibilities, we computed additional gaze dissimilarity matrices for the pairwise differences in the proportion of dwell time on text (GDM_{text}) and people (GDM_{people}) for which we obtained internal consistencies of respectively $r = 0.31$, $p < 0.001$ and $r = 0.64$, $p < 0.001$. We then fitted a multiple linear regression model with the general GDM_{dwell}, GDM_{text} and GDM_{people} as predictors of the DDM_{noun}, yielding an adjusted R^2 of 0.18, $F(3, 431) = 32.2$, $p < 0.001$. Dropping the general GDM from the model resulted in a significant decrease of model performance (adjusted R^2 of 0.15, $F(2, 432) = 39.1$, $p < 0.001$, $\Delta AIC = 13.55$), revealing a unique contribution of general dwell-time similarities to description similarities beyond individual tendencies to fixate people and text. This suggests that our general and targeted analyses are not redundant and the relationship between individual gaze and descriptions contains further dimensions yet to be identified.

Discussion

We investigated systematic individual differences in gaze and found that they translated to individually divergent descriptions of identical scenes.

We observed positive correlations between pairwise observer dissimilarities in scene descriptions and fixation patterns. This relationship was strongest for the semantic dissimilarity of nouns and not significant for verbs and adjectives. Nouns most directly match the objects across which we quantified individual fixation distributions. To test whether the use of object masks for GDM_{dwel} biased results towards DDM_{nouns}, we conducted an additional control analysis using pixel-based fixation density maps instead (see Supplementary Information C). Reassuringly, this analysis yielded the same pattern of results, reinforcing our conclusion that inter-individual gaze similarity aligns with the usage of nouns, but not verbs or adjectives. Moreover, recent results suggest that scene representations in higher visual cortices are well captured by the nouns used in corresponding descriptions³². Semantic similarity of complete scene descriptions (containing nouns, verbs, adjectives and function words) also correlated positively with fixation patterns, but less strongly than that of nouns. This is further support of the prominent role of nouns in explaining fixation patterns.

Although verbs and adjectives are important for scene understanding, e.g., by disambiguating correspondences³³ and describing relationships, details and actions, it is less clear which pixels in an image or fixations in a gaze pattern would correspond to them. This may be the reason why previous research comparing gaze and descriptions focused on nouns^{21,22}. Interestingly, however, we also observed consistent individual differences in the verbs used in scene descriptions. Future work may quantify action-related differences in gaze more directly. For instance, recent work showed that deep saliency models appear to use semantic action features (e.g. having a meeting, snowboarding, and jumping) to predict visual attention³⁴. It may be possible to estimate the weight of such features on an individual level and relate them to scene descriptions. Additionally, verbs could be grouped into functional categories³⁵ which semantically link objects, people, or both. Individual differences in the use of such verb clusters may be related to corresponding differences in scan path sequences¹⁶.

Probing more specific biases, reliable individual differences in the tendency to look at people and text^{10–12,14} explained corresponding references in scene descriptions. This extends the previous finding that objects which are more frequently fixated are also more likely to be referenced^{19–22} to the level of trait-like individual differences. The *habitual individual* way we look at scenes is predictive of the individual way we communicate their content. Importantly, this variation generalised across independent sets of scenes. How an individual looks at one set of scenes is predictive of how it will describe another. This strongly suggests that individual gaze is linked to trait-like differences in subjective scene perception. The interindividual variance in the proportion of text references and dwell time was very large (ratio of extremes close to factor seven and five, respectively) and that for the proportion of people references and dwell time was smaller (ratio of extremes 1.65 and 1.3, respectively). This resonates with previous findings showing that people and faces have a high probability to be fixated and included in scene descriptions overall²² (see also ref.³⁶), which limits the scale of variation due to ceiling effects. In contrast, reading is an acquired skill (e.g.,³⁷) and text salience may be modulated by education, profession and experience throughout life. Because the group of participants we studied has similar educational backgrounds, the variation in text salience and references may be even larger in more diverse samples. Generally, our study sample was relatively homogeneous, consisting mainly of female German university students. Future studies could use the methods introduced here to study variance across more diverse groups including different educational and cultural backgrounds, children, experts and individuals with clinical conditions, who may perceive and describe scenes differently. Previous studies indicate altered viewing patterns in autism spectrum and affective disorders, such as reduced visual exploration and social salience^{38,39}. However, the connection between these gaze characteristics and scene perception remains to be tested.

A potential limitation of our findings is that it is unclear how far they generalise to other tasks than describing scenes. Interestingly, our paradigm replicated the most important and reliable individual fixation tendencies previously found in free-viewing tasks (i.e. anticorrelated tendencies to fixate text and faces). This speaks against a narrow task-dependence of these individual tendencies. Additionally, the reliability of these gaze tendencies across time and independent sets of scenes^{10,11} suggests they reflect ecological meaningful observer traits. Furthermore, a recent preprint by Murlidaran and Eckstein (2024) found similar patterns of fixations when instructing participants to freely view or describe a scene (but see ref.⁴⁰) and suggests the extraction of global meaning as the default objective of scene viewing. This supports the notion that our description task is suitable to study ecologically valid individual differences in perception. Nevertheless, further research is needed to investigate the robustness of the observed individual differences across tasks and dynamic or real-world stimuli.

Furthermore, there may be aspects of gaze that correlate with nouns, verbs, and adjectives when examined more closely, though we have yet to identify them. A post-hoc exploratory analysis (see Supplementary Information D) suggested that observer dissimilarities in the tendency to fixate “touched” objects was positively correlated not only with noun dissimilarity patterns, but also with description differences in verbs and adjectives. These findings provide a promising starting point for future research, which could further explore these relationships using larger image sets. Finally, while our results provide evidence for systematic individual differences in scene descriptions and for individual gaze as a mediating mechanism, they cannot reveal why people attend to different things in the first place. Twin studies found high heritability for individual differences in gaze^{13,15}, and experiences like learning to read appear to shape scene viewing as well⁴¹. Genetic differences may imply structural and functional differences in the visual system⁴², which in turn may bias individual visual diets and ultimately lead to the individual differences in scene perception suggested by our results. Future studies can probe individual gaze and scene descriptions (or drawings⁴³) in large and age-diverse cohorts to decipher their developmental relationship.

Our findings open a window into individual perception which is potentially relevant beyond biology and vision science. For instance, individual gaze may shape the appreciation of visual stimuli and memorability of artworks (cf. ref.⁴⁴) or contribute to the strikingly individual appraisal of motion pictures⁴⁵. Systematic covariance, such as the anticorrelation between text and people salience point to the possibility of a low-dimensional space

of individual perception (cf. ref.¹⁰), which could inform the design of visual communication from advertisement to education and road safety.

In sum, we found consistent differences in how individuals describe identical natural scenes, which can be explained by corresponding differences in gaze. These results strongly suggest that the individual way we see the world is shaped by the individual way we look at it.

Methods

Participants

We included 30 participants (average age: $M = 22.77$; $SD = 2.87$; 4 male, 26 female), recruited from the participants mailing list at Justus-Liebig University Giessen, Germany, with normal or corrected-to-normal vision, German language proficiency at native level and no acute mental health problems. Participants were compensated with course credit or 20 euros.

The study was approved by the local ethics committee of the Justus-Liebig-University Giessen (LEK-FB06) and was conducted in accordance with the relevant guidelines and regulations. All participants provided written informed consent.

Apparatus

Data were collected using Psychtoolbox⁴⁶ and MATLAB (R2020a, MathWorks, Natick, MA) on a BenQ XL2430T computer monitor (BenQ Corporation, Taipei, Taiwan). The dimensions of the display, which was placed at a distance of 55 cm, were 526×296 mm with a 3840×2160 pixel resolution. Stimuli were presented at a size of 2400×1800 pixels and 34.3×25.7 degrees viewing angle. Movements of the right eye were recorded at a frequency of 1 kHz using a towermount Eyelink 1000 eye tracker (SR Research, Ottawa, Canada).

Stimuli and procedure

Participants were presented with a total of 100 images from the OSIE image set⁶ in fixed order. These images depicted everyday natural indoor and outdoor scenes and came with a total of 845 pixel masks (published by Xu et al., 2014) for various objects ranging from inanimate to animate objects (see Fig. 1d for an example). For our analysis focusing on specific object categories, we considered 84 pixel masks for “text” elements and 237 pixel masks for “people” (the latter were published by Broda & de Haas (2022)[26]). The selection of the 100 images for our experiment was composed of 40 everyday scenes that were found to be particularly suitable for analysing individual gaze behaviour in the study by Linka & de Haas (2020)¹¹ (cf. OSIE40) as well as 60 additional scenes taken from the OSIE200¹¹, which were selected with a preference for dynamic content (all stimuli can be found at <https://osf.io/83mjc>).

After providing written informed consent, participants were given written instructions, familiarized with the procedure and provided the opportunity to ask questions. Instructions (Supplementary Information A) loosely followed Chen et al. (2015)²⁴ and asked participants to explore and describe each image in turn (cf. ref.¹⁰). Participants initiated a trial by simultaneously fixating a fixation point at the centre of the screen and pressing the space bar. This triggered the presentation of a scene image which lasted three seconds, followed by a written instruction to describe the image verbally. Participants’ descriptions were documented by the experimenter in writing into an Excel spreadsheet using a laptop. The instructions were presented until the participant pressed the space bar to indicate that they completed their description and to move on to the next trial.

Participants could take short breaks after every 10 trials, followed by a nine-point calibration of the right eye, including validation with a maximum error of 1.2 and a maximum average error of 0.6 (observed validation error: $M = 0.35$ dva, $SD = 0.06$ dva). On average, the experiment lasted 90 to 120 min in total.

Measured variables and preprocessing

Similarity of gaze behaviour

Following de Haas et al. (2019) and manufacturer recommendations, we excluded fixations with an onset latency or duration < 100 ms. We then quantified the cumulative fixation duration (i.e. dwell time) for each observer and object for which we have pixel masks in a given scene (see Fig. 1d for an example scene in which we visually outlined respective pixel masks obtained from ref.⁶). Moreover, for each observer, we quantified the percentage of overall dwell time across all images which fell on text elements and people, respectively. To account for foveal extent and eye tracking accuracy, a fixation was assigned to each object for which the distance between the fixation coordinate and pixel mask was within a tolerance margin of ~ 0.5 degrees visual angle. Note, also multiple assignments were possible for neighbouring or overlapping pixel masks such that a given fixation duration could be considered for more than one object. Dwell times for single objects were then expressed as the proportion of cumulative dwell time across all fixations in the respective scene; dwell times for object classes (people and text elements) were expressed as the proportion of cumulative dwell time across all fixations in all scenes (cf. ref.^{10,11}).

Furthermore, the proportion of cumulative dwell times for all 845 objects (pixel masks by ref.⁶) across all 100 images were entered into a vector and compared between each possible pair of observers, using the Euclidian distance as a summary measure of pairwise differences. That is, we computed the squared difference between dwell times of observer A and B for each object and then summed these differences across objects. This measure of difference was computed for all possible pairs of observers and entered into a Gaze Dissimilarity Matrix (GDM). This procedure was repeated, considering object pixel masks with a text (GDM_{text}) or people label (GDM_{people}) only. Additionally, we calculated pairwise dissimilarities based on the Euclidian distance between binary vectors indicating whether an object was fixated (coded as 1) or not (coded as 0; GDM_{fix}). Finally, dissimilarities were scaled to a range from 0 to 1. Note that the GDMs reported here were calculated based on gaze behaviour across all 100 images. However, to assess the internal consistency of the resulting matrices, we later conducted a split-half analysis for which we computed independent GDMs for the 50 odd and 50 even

trials. The split-half data (odd and even trials) were also used in the decoding analysis (see the Analysis section for details).

Similarity of scene descriptions

To analyse the similarity of scene descriptions, we first quantified the semantic similarity between participants' complete scene descriptions using the pre-trained paraphrase-MiniLM-L12-v2 model from the Sentence Transformers library (SentenceBERT; ref.²⁸) in Python (v. 3.8.17). This model is well-suited for generating embeddings of sentences and computing their semantic similarity based on the cosine similarity between them (ranging from -1 to 1, with higher values indicating higher semantic similarity). For each image, a 30×30 similarity matrix was initialized to store the pairwise similarity scores between participants' descriptions.

Second, we pooled all descriptions for a given image across participants and automatically identified all unique nouns, verbs and adjectives using the part-of-speech tagging by the spaCy library (v. 3.6.0) and the associated German language model in Python (v. 3.8.17). The assignment was subsequently checked manually, and isolated assignment errors were corrected. Moreover, we estimated the semantic similarity of all possible word pairs of the respective scene descriptions by using the genSim library²⁹ and the pretrained fastText model³⁰, which offers also German language applications³¹ and calculates the similarity values using cosine similarity (ranging from -1 to 1; the higher the value, the higher is the contextual similarity of the words). Finally, we identified the individual words that a subject used to describe a scene and compared the semantic similarity between the words used by this subject and their respective comparison partner. If an identical word was used in a comparison pair, the word was assigned a semantic similarity of 1. If this was not the case, we searched for the semantic 'nearest neighbour' among the words used by the comparison partner according to the fastText model, and then assigned the corresponding similarity value to the respective word. We then calculated the mean semantic similarity value across all 'closest matches' for a given pair of descriptions. Subsequently, we computed the average similarity of scene descriptions across all scenes for a given pair of observers. Finally, these values were scaled to a range from 0 to 1 across all observers and subtracted from 1 so that higher values were associated with more dissimilar pairs. The resulting Description Dissimilarity Matrices (DDMs) for nouns (containing all data or data from the odd and even splits) are displayed in Fig. 2a. Additionally, we grouped all the nouns referring to text elements and people in a separate text file and quantified the proportion of nouns referring to the respective category for each observer.

Analysis

We performed all statistical analyses and created all figures using MATLAB (R2020a, MathWorks, Natick, MA).

Computationally, both, gaze patterns and scene descriptions can be thought of as vectors in high-dimensional spaces, defined by the distribution of dwell time across objects in an image, or the semantic similarity of complete descriptions and words from a library of scene descriptions. Our rationale was to map gaze and description data onto each other via the transformation of both variables into the same abstract space of pairwise observer similarities (see above). This is akin to representational similarity analysis in neuroscience, which has been used successfully to compare representational geometries across modalities²⁷. It has the additional advantage of requiring no a priori knowledge of the most relevant dimensions of individual variance. While previous research has revealed some specific individual gaze biases¹⁰, these may not be exhaustive and individual biases for scene descriptions were unknown a priori.

Once the data were transformed to Gaze or Description Dissimilarity Matrices (GDM or DDM), we assessed the internal consistencies of all corresponding matrices (GDM_{dwell}, GDM_{text}, GDM_{people}, GDM_{fix}, DDM_{complete}, DDM_{noun}, DDM_{verb}, DDM_{adjective}). To obtain a measure of internal consistency, we first split the data into 50 odd and even trials and computed all variables of interest separately for these two halves. The resulting estimates were then compared using split-half correlations. Furthermore, we calculated the inter-subject range (maximum/minimum ratios) of proportions of dwell times and references to text and people (cf. ref.¹¹ and ref.¹⁰).

To test the relationship between individual gaze behaviour and scene descriptions and differentiate between contributions from the general GDM and the a priori dimensions of text and people, we performed a multiple regression with GDM_{dwell}, GDM_{text} and GDM_{people} as predictors and DDM_{noun} as target variable. We then dropped each predictor in turn and tested the effect on the adjusted proportion of variance explained by the model (adjusted R²) and Akaike Information Criterion (AIC) to assess its unique contribution. A $\Delta AIC > 4$ between the full and reduced model was considered indicative of a significant difference.

To decode an observer's identity based on their dissimilarity pattern for a set of images, we derived pairwise difference in gaze (GDM_{dwell}) and descriptions (DDM_{noun}) independently for two halves of the data (odd and even trials). We then extracted the dissimilarity patterns of two observers to all other observers for both halves of the data. Crucially, we provided a decoding algorithm with the corresponding observer labels only for the dissimilarity vectors for one half of the data and tried to determine the observer labels for the corresponding dissimilarity vectors from the other half of the data using nearest neighbour correlations. This was done both, within and between modalities (gaze $\leftrightarrow$ gaze, noun $\leftrightarrow$ noun, and gaze $\leftrightarrow$ noun dissimilarity vector correlations). Empirical hit rates corresponded to the proportion of instances in which the algorithm correctly identified observer identities applying this method across all possible pairs of observers. Finally, we performed 1000 bootstrap iterations to generate a null distribution for the respective hit rates by shuffling observer identities within each split. This allowed us to assess the statistical significance of our empirical hit rates against a baseline distribution obtained under the null hypothesis (see Fig. 2c).

When multiple comparisons were performed, the significance level was adjusted according to Bonferroni by dividing it by the number of comparisons. P-values remain unaffected by this procedure and a correction is only

applied in the interpretation of these values with regard to the adjusted significance level. All data and code is publicly available at [<https://osf.io/83mjc>].

Data availability

The datasets generated and analysed in the current study are available for download from the first author's (Diana Kollenda) OSF profile [<https://osf.io/83mjc>] along with the analysis scripts and all stimuli.

Received: 7 August 2024; Accepted: 11 March 2025

Published online: 19 March 2025

References

- Nagel, T. What is it like to be a bat? In *the language and thought series* 159–168 (Harvard University Press, 1980) <https://doi.org/10.4159/harvard.9780674594623.c15>
- Haxby, J. V. et al. A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* **72**, 404–416. <https://doi.org/10.1016/j.neuron.2011.08.026> (2011).
- Parkinson, C., Kleinbaum, A. M. & Wheatley, T. Similar neural responses predict friendship. *Nat. Commun.* **9**, 332. <https://doi.org/10.1038/s41467-017-02722-7> (2018).
- Moutsiana, C. et al. Cortical idiosyncrasies predict the perception of object size. *Nat. Commun.* **7**, 12110. <https://doi.org/10.1038/ncomms12110> (2016).
- de Haas, B., Kanai, R., Jalkanen, L. & Rees, G. Grey matter volume in early human visual cortex predicts proneness to the sound-induced flash illusion. *Proceedings, biological sciences* **279**, 4955–4961 <https://doi.org/10.1098/rspb.2012.2132> (2012)
- Xu, J., Jiang, M., Wang, S., Kankanhalli, M. S. & Zhao, Q. Predicting human gaze beyond pixels. *J. Vis.* <https://doi.org/10.1167/14.1.28> (2014).
- Cerf, M., Frady, E. P. & Koch, C. Faces and text attract gaze independent of the task: Experimental data and computer model. *J. Vis.* **9**(10), 1–15. <https://doi.org/10.1167/9.12.10> (2009).
- End, A. & Gamer, M. Preferential processing of social features and their interplay with physical saliency in complex naturalistic scenes. *Front. Psychol.* **8**, 418. <https://doi.org/10.3389/fpsyg.2017.00418> (2017).
- Fletcher-Watson, S., Findlay, J. M., Leekam, S. R. & Benson, V. Rapid detection of person information in a naturalistic scene. *Perception* **37**, 571–583. <https://doi.org/10.1068/p5705> (2008).
- de Haas, B., Iakovidis, A. L., Schwarzkopf, D. S. & Gegenfurtner, K. R. Individual differences in visual salience vary along semantic dimensions. *Proceedings of the national academy of sciences of the United States of America* **111**, 11687–11692 <https://doi.org/10.1073/pnas.1820553116> (2019)
- Linka, M. & de Haas, B. OSIShort: A small stimulus set can reliably estimate individual differences in semantic salience. *J. Vis.* **20**(9), 13–13. <https://doi.org/10.1167/jov.20.9.13> (2020).
- Guy, N. et al. A novel perceptual trait: Gaze predilection for faces during visual exploration. *Sci. Rep.* **9**, 10714. <https://doi.org/10.1038/s41598-019-47110-x> (2019).
- Constantino, J. N. et al. Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature* **547**(340), 344. <https://doi.org/10.1038/nature22999> (2017).
- Berlijn, A. M., Hildebrandt, L. K. & Gamer, M. Idiosyncratic viewing patterns of social scenes reflect individual preferences. *J. Vis.* **22**, 10. <https://doi.org/10.1167/jov.22.13.10> (2022).
- Kennedy, D. P. et al. Genetic influence on eye movements to complex scenes at short timescales. *Curr. Biol.* **27**(3554), 3560.e3. <https://doi.org/10.1016/j.cub.2017.10.007> (2017).
- Zelinsky, G. J. Understanding scene understanding. *Front. Psychol.* **4**, 954. <https://doi.org/10.3389/fpsyg.2013.00954> (2013).
- Linka, M., Broda, M. D., Alsheimer, T., de Haas, B. & Ramon, M. Characteristic fixation biases in super-recognizers. *J. Vis.* **22**(8), 17. <https://doi.org/10.1167/jov.22.8.17> (2022).
- Muridaran, S. & Eckstein, M. P. Eye movements during free viewing maximize scene understanding. *J. Vis.* **24**(10), 1189 (2024).
- Tatler, B. W., Gilchrist, I. D. & Land, M. F. Visual memory for objects in natural scenes: From fixations to object files. *Q. J. Exp. Psychol. Hum. Exp. Psychol.* **58**(5), 931–960. <https://doi.org/10.1080/02724980443000430> (2005).
- Pertsov, Y., Avidan, G. & Zohary, E. Accumulation of visual information across multiple fixations. *J. Vis.* **9**(10), 2. <https://doi.org/10.1167/9.10.2> (2009).
- Liang, H., Jiang, M., Liang, R. & Zhao, Q. CapVis: Toward better understanding of visual- verbal saliency consistency. *ACM Trans. Intell. Syst. Technol.* **10**, 1–23. <https://doi.org/10.1145/3200767> (2019).
- Yun, K., Peng, Y., Samaras, D., Zelinsky, G. J. & Berg, T. L. Exploring the role of gaze behavior and object detection in scene understanding. *Front. Psychol.* **4**, 917. <https://doi.org/10.3389/fpsyg.2013.00917> (2013).
- Fei-Fei, L., Iyer, A., Koch, C. & Perona, P. What do we perceive in a glance of a real-world scene?. *J. Vis.* **7**(1), 10. <https://doi.org/10.1167/7.1.10> (2007).
- Chen, X. et al. Microsoft COCO Captions: Data collection and evaluation server. *arXiv preprint* <https://doi.org/10.48550/arXiv.1504.00325> (2015)
- Coco, M. I. & Keller, F. Scan patterns predict sentence production in the cross-modal processing of visual scenes. *Cogn. Sci.* <https://doi.org/10.1111/j.1551-6709.2012.01246.x> (2012).
- Broda, M. D. & de Haas, B. Individual differences in looking at persons in scenes. *J. Vis.* **22**(12), 9. <https://doi.org/10.1167/jov.22.12.9> (2022).
- Kriegeskorte, N., Mur, M. & Bandettini, P. A. Representational similarity analysis-Connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4. <https://doi.org/10.3389/fpsyg.2008.004.2008> (2008).
- Reimers, N. & Gurevych, I. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint* <https://doi.org/10.48550/arXiv.1908.10084> (2019)
- Rehúfek, R. & Sojka, P. Software framework for topic modelling with large corpora (2010).
- Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C. & Joulin, A. Advances in pre-training distributed word representations. *arXiv preprint* <https://doi.org/10.48550/arXiv.1712.09405> (2018)
- Grave, E., Bojanowski, P., Gupta, P., Joulin, A. & Mikolov, T. Learning word vectors for 157 languages. *arXiv preprint* <https://doi.org/10.48550/arXiv.1802.06893> (2018).
- Conwell, C., Prince, J., Alvarez, G. & Konkle, T. The Unreasonable effectiveness of word models in predicting high-level visual cortex responses to natural images. *Conference on Computational Cognitive Neuroscience* (2023)
- Gupta, A. & Davis, L. S. Beyond nouns: Exploiting prepositions and comparative adjectives for learning visual classifiers. *Computer vision—ECCV 2008 10th European Conference on computer vision, Marseille, France, Proceedings 16–29* https://doi.org/10.1007/978-3-540-88682-2_3 (2008)
- Chen, S., Jiang, M. & Zhao, Q. What Do Deep Saliency Models Learn about Visual Attention? *arXiv preprint* <https://doi.org/10.48550/arXiv.2310.09679> (2023)

35. Orban, G. A., Lanzilotto, M. & Bonini, L. From observed action identity to social affordances. *Trends cognitive sciences* **25**, 493–505. <https://doi.org/10.1016/j.tics.2021.02.012> (2021).
36. Berg, A. C. *et al.* Understanding and predicting importance in images. *IEEE Conference on computer vision and pattern recognition* 3562–3569 <https://doi.org/10.1109/CVPR.2012.6248100> (2012).
37. Dehaene, S. *et al.* How learning to read changes the cortical networks for vision and language. *Science* **330**(1359), 1364. <https://doi.org/10.1126/science.1194140> (2010).
38. Wang, S. *et al.* Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. *Neuron* **88**(604), 616. <https://doi.org/10.1016/j.neuron.2015.09.042> (2015).
39. Armstrong, T. & Olatunji, B. O. Eye tracking of attention in the affective disorders: a meta-analytic review and synthesis. *Clin. Psychol. Rev.* **32**, 704–723. <https://doi.org/10.1016/j.cpr.2012.09.004> (2012).
40. Griffin, Z. M. & Bock, K. What the eyes say about speaking. *Psychol. Sci.* **11**(4), 274–279. <https://doi.org/10.1111/1467-9280.00255> (2000).
41. Linka, M., Sensoy, Ö., Karimpur, H., Schwarzer, G. & de Haas, B. Free viewing biases for complex scenes in preschoolers and adults. *Sci. Rep.* **13**(1), 11803. <https://doi.org/10.1038/s41598-023-38854-8> (2023).
42. Broda, M. D. & de Haas, B. Individual differences in human gaze behavior generalize from faces to objects. *Proceedings of the national academy of sciences* **121** 12 e2322149121 <https://doi.org/10.1073/pnas.2322149121> (2024).
43. Long, B., Simson, J., Buxó-Lugo, A., Watson, D. G. & Mehr, S. A. How games can make behavioural science better. *Nature* <https://doi.org/10.1038/d41586-023-00065-6> (2023).
44. Davis, T. M. & Bainbridge, W. A. Memory for artwork is predictable. *Proceedings of the national academy of sciences of the United States of America* **120** e2302389120 <https://doi.org/10.1073/pnas.2302389120> (2023).
45. Wallisch, P. & Whritner, J. A. Strikingly low agreement in the appraisal of motion pictures. *Projections* <https://doi.org/10.3167/proj.2017.110107> (2017).
46. Kleiner, M. *et al.* What's new in Psychtoolbox-3?. *Perception* <https://doi.org/10.1068/v070821> (2007).

Acknowledgements

We thank Diana Weißleder and Sophie Halstenberg for their help with data collection and stimulus editing.

Author contributions

D.K. and B.d.H. conceptualized the experiment. A.S.R. collected the data. D.K., A.S.R. and B.d.H. analysed the data. D.K. drafted the manuscript. D.K., A.S.R. and B.d.H. revised the manuscript and approved its final version.

Funding

Open Access funding enabled and organized by Projekt DEAL.

This work was supported by European Research Council Starting Grant 852885 INDIVISUAL; BdH was further supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project Nos. 222641018-SFB/TRR 135 TP C9 as well as "The Adaptive Mind", funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art.

Declarations

Competing interests

The authors declare that they have no competing interests.

Ethics approval

All participants gave informed consent and the study was performed in line with relevant guidelines and regulations as well as approved by the local ethics review board (LEK-FB06) at Justus-Liebig-University Giessen.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-94056-4>.

Correspondence and requests for materials should be addressed to D.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

Supplementary Information for the manuscript entitled "Individual gaze predicts individual scene descriptions"

Diana Kollenda*, Anna-Sophia Reher & Benjamin de Haas

Supplementary Information A

German original instructions

Wichtig ist, dass Du deinen Kopf in der Kinnstuetze moeglichst nicht bewegst.

In jedem Durchgang siehst Du zuerst ein Fixationskreuz. Sobald Du darauf schaust und die Leertaste drueckst, wirst Du fuer 3 Sekunden ein Bild sehen, das Du frei betrachten kannst.

Im Anschluss bitten wir Dich, dieses Bild muendlich zu beschreiben und ich werde Deine Beschreibung aufschreiben. Es gibt keine richtigen oder falschen Antworten.

Bitte:

- Beschreibe die wichtigsten Dinge in der Szene → Was passiert?
- 1-2 Saeetze sind ausreichend
- Beginne den Satz NICHT mit ‚Es gibt..‘
- Beschreibe KEINE unwichtigen Details
- Beschreibe KEINE Dinge, die moeglicherweise in der Vergangenheit oder Zukunft passieren
- Beschreibe NICHT, was eine Person moeglicherweise sagt
- Gib Personen KEINE Namen
- Jeder Satz sollte mindestens 8 Woerter lang sein
- Achte darauf, Verben zu benutzen

Keine Sorge, ich werde mit darauf achten, dass Du Dich daran erinnerst. Am allerwichtigsten ist, dass Du versuchst, den Kopf in der Kinnstuetze so still wie moeglich zu halten.

Zwei Beispielbilder und die jeweilige Beispielbeschreibung wären:



1. „Ein Pferd trägt eine große Ladung Heu und zwei Personen, die darauf sitzen.“



2. „Der Spieler macht sich bereit, den Ball zu schlagen, während der Schiedsrichter zuschaut.“

Translated instructions

It is important that you do not move your head in the chinrest.

In each trial, you will first see a fixation cross. Once you look at it and press the spacebar, you will see an image for 3 seconds, which you can freely view.

Afterwards, we will ask you to describe this image orally, and I will transcribe your description. There are no right or wrong answers.

Please:

- Describe the most important things in the scene → What is happening?
- 1-2 sentences are sufficient.
- Do not start the sentence with 'There is...'
- Do not describe unimportant details.
- Do not describe things that may happen in the past or future.
- Do not describe what a person might say.
- Do not give names to people.
- Each sentence should be at least 8 words long.
- Make sure to use verbs.

Don't worry; I will remind you to follow these guidelines. Most importantly, try to keep your head as still as possible in the chinrest.

Two example images and their respective example descriptions would be:



1. "A horse carries a large load of hay with two people sitting on it."



2. "The player prepares to hit the ball while the referee watches."

Supplementary Information B

Original sentences that were translated to English in Figure 2.

01: Zwei Frauen sitzen in einer Bar und eine hat eine Kamera in der Hand.

11: Zwei Frauen sitzen in einer Bar, eine von ihnen hat eine Kamera in der Hand.

09: Mehrere Personen sitzen in einem Irish-Pub, eine Person hat eine Kamera in der Hand.

19: Eine Frau sitzt auf einer Bank und hält eine Kamera. Neben ihr sitzt eine weitere Frau.

Original sentences that were translated to English in Figure 3.

26: Zwei Frauen sitzen auf einer Bank. Die Frau, die eine Kamera in der Hand hält, grinst mit ausgestreckter Zunge in die Kamera.

27: Zwei Frauen, die in einer Kneipe waren, im Hintergrund war ein Schild mit Irland zu erkennen. Die Frauen sahen amüsiert aus.

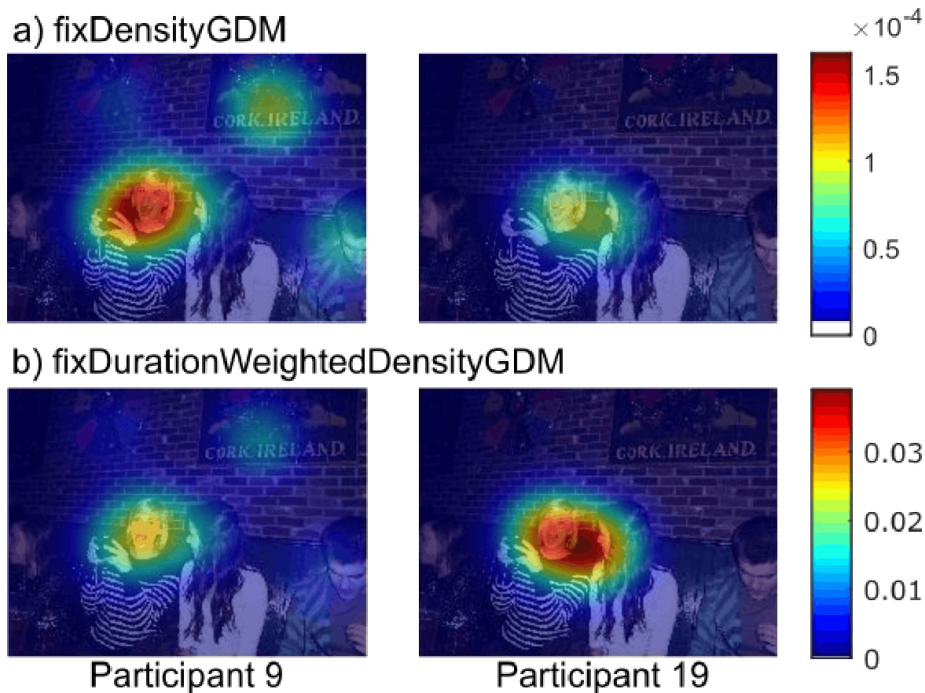
Supplementary Information C

In the main text, the Gaze Dissimilarity Matrix was calculated based on predefined object pixel masks that can be most directly associated with nouns and thus could have biased our results. For this reason, we conducted an additional control analysis using unrestricted Gaussian-smoothed fixation density maps (fixDensityGDM) for each individual. These maps were compared across participants and the correlations averaged across images for each pair of observers. We found high reliability across odd- and even-image splits for this fixDensityGDM ($r = .81$, $p < .001$). Furthermore, fixDensityGDM correlated positively with both the DDM for nouns ($r = .49$, $p < .001$) and complete descriptions ($r = .29$, $p < .001$), but not with the DDMs for verbs ($r = -.12$, $p = .01$) or adjectives ($r = -.07$, $p = .152$), replicating our original results. The same pattern was observed when weighting fixations in the density maps by duration (fixDurationWeightedDensityGDM; split-half reliability $r = .84$, $p < .001$; DDM nouns: $r = .51$, $p < .001$, complete descriptions: $r = .33$, $p < .001$, verbs: $r = -.13$, $p = .007$, adjectives: $r = -.04$, $p = .393$).

Notably, when using fixation density maps (both weighted and unweighted by fixation duration, see Figure S1 for a comparison), the correlation with DDM nouns even increased. However, their correlation with DDM verbs remained insignificant or even negative, which is in line with our main results. Here, the negative correlation was even significant, which is challenging to interpret, suggesting that greater gaze dissimilarity between participants corresponds to greater similarity in verb usage. As with the GDM based on object dwell times, we again found no significant relationship between fixation density-based GDMs and DDMs for adjectives.

We conclude that our main findings are robust to using object or pixel-based gaze similarity metrics: Inter-individual gaze similarity is aligned with the usage of nouns, but not verbs or adjectives.

Figure C1



Example fixation density maps for Participants 9 and 19 (a dissimilar pair), displayed in separate columns. a) shows unweighted fixation density maps, while b) shows maps weighted by fixation duration.

Supplementary Information D

In the main text, we focused our analysis on gaze dissimilarity patterns related to the semantic dimensions of text and people and their correlation with description patterns. Previous research suggests that individual gaze tendencies toward text and faces are particularly robust in smaller stimulus sets, such as the 100 images used in this study (cf. Linka & de Haas, 2020). However, gaze tendencies toward other semantic categories, such as emotion, taste, motion, touch, and gaze, may also correlate with description patterns involving nouns, verbs, or adjectives.

Here, we present an exploratory analysis of the internal consistencies of these semantic dimensions. For dimensions with $r > .25$, we examined their correlations with the respective DDMs and reported these results in Table D1. The GDMs for taste and motion correlated

positively with DDMnoun only, whereas GDMtouched showed positive correlations not only with DDMnoun but also with DDMs for verbs and adjectives.

Table D1

Internal Consistencies of Gaze Dissimilarity Across Five Semantic Dimensions (Emotion, Taste, Motion, Touched, Gazed) and Their Correlations with Description Dissimilarity

Semantic dimension	Number of objects labeled	Internal consistency	Correlation with DDMnoun	Correlation with DDMverb	Correlation with DDMadj
GDMemotion	41	$R = .09, p = .060$			
GDMtaste	50	$R = .32, p < .001$	$R = .41, p < .001$	$R = .06, p = .209$	$R = -.12, p = .010$
GDMmotion	99	$R = .46, p < .001$	$R = .26, p < .001$	$R = -.09, p = .070$	$R = -.02, p = .675$
GDMtouched	105	$R = .25, p < .001$	$R = .13, p = .006$	$R = .13, p = .006$	$R = .10, p = .033$
GDMgazed	62	$R = .14, p = .003$			

Note. GDM = Gaze dissimilarity matrix, DDM = description dissimilarity matrix. P-values are uncorrected for multiple comparisons.

Supplementary Information E

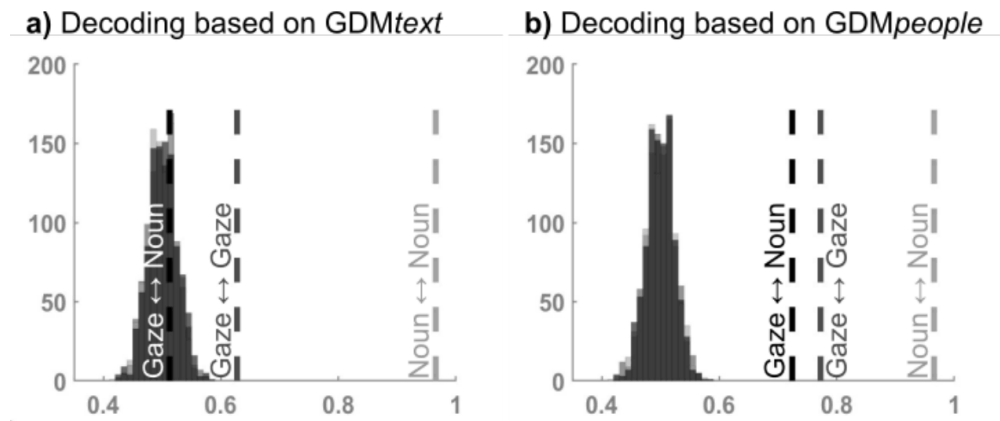
We were able to decode which of two observers provided a given dwell-time distribution based on their descriptions for other images (and vice versa). Here, we explored whether this was also possible when limiting the gaze data to either individual differences in text or people fixations.

We were able to decode which of two observers provided a given dwell-time distribution related to text from one set of scenes to another (GDMtext; gaze <-> gaze, decoding hit rate: 63%, $p < 0.001$). However, we could not decode gaze patterns related to text in one half of the dataset based on scene descriptions in the other (and vice versa; GDMtext and DDMnoun; gaze <-> descriptions, decoding hit rate: 51%).

For dwell-time distributions related to people, we successfully decoded participants both within and across modalities. Specifically, we achieved a gaze <-> gaze decoding hit rate of

77% (GDM_{people}; $p < 0.001$) and a gaze $\leftrightarrow$ descriptions decoding hit rate of 72% (GDM_{people} and DDM_{noun}; $p < 0.001$).

Figure E1



The figure shows bootstrapped null distributions and decoding hit rates based on nearest neighbour correlation between a) GDM_{text} and DDM_{noun} and b) GDM_{text} and DDM_{noun} from independent trials (odd and even splits), respectively. The vertical line of the Noun $\leftrightarrow$ noun hit rate (97%, light grey) is the same in both panels and as reported in the main text.

3. List of all publications

Kollenda, D., Akbari, E., Broda, M. D., & de Haas, B. (2026, April 16). Active vision is linked to category selectivity in the individual brain. *Psyarxiv* (Accepted for publication in *Nature Human Behaviour*)

Kollenda, D., & de Haas, B. (2025, August 31). Memories that remain: Image memorability transfers to scene descriptions. *Psyarxiv*

Kollenda, D., Reher, A. S., & de Haas, B. (2025). Individual gaze predicts individual scene descriptions. *Scientific Reports*, 15(1), 9443.

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.

Kollenda, D., & de Haas, B. (2022). The influence of familiarity on memory for faces and mask wearing. *Cognitive Research: Principles and Implications*, 7(1), 45.

4. Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne unzulässige Hilfe oder Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Im Rahmen der Erstellung dieser Synopsis wurden KI-gestützte Sprachmodelle für Übersetzungen, sprachliche Überarbeitung und zur kritischen Reflexion der Argumentationsstruktur einzelner Passagen verwendet. Alle inhaltlichen Entscheidungen, Interpretationen und wissenschaftlichen Schlussfolgerungen stammen von mir. 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