Contextual modulations at low- and high-level stages of visual processing hierarchy

Contextual modulations at low- and high-level stages of visual processing hierarchy

M. Umut Canoluk

Thèse présentée en vue de l'obtention du grade de
Docteur en sciences psychologiques et de l'éducation

Promoter
Prof. Dr. Valérie GOFFAUX (UCLouvain, Belgium)

Accompanying committee
Prof. Dr. Michael ANDRES (UCLouvain, Belgium)
Prof. Dr. Gilles VANNUSCORPS (UCLouvain, Belgium)

Collection de thèses de l'Université catholique de Louvain, 2023

President

Prof. Dr. Michael ANDRES (UCLouvain, Belgium)

Accompanying committee and jury

Prof. Dr. Valérie GOFFAUX (UCLouvain, Belgium)

Prof. Dr. Gilles VANNUSCORPS (UCLouvain, Belgium)

Prof. Dr. Chen SONG (Cardiff University, Wales)

Prof. Dr. John GREENWOOD (University College London, United Kingdom)

Contents

Abstract

This thesis examines the contextual modulations at both the low and high stages of the visual processing hierarchy. Holistic processing, a high-level contextual modulation mechanism, is a hallmark of face processing, and is proposed to mark the recruitment of face-specialized regions. It can be easily disrupted by inversion, an image manipulation that largely preserves image properties processed at primary stages of vision. The preservation of low-level properties with inversion directed the traditional focus to high-level mechanisms to account for face holistic processing. Due to this focus, holistic processing has been considered independent from low-level mechanisms and its connection with the early regions was not adequately explored. In the present thesis, we investigated the link of high-level contextual modulations to low-level regions through behavioral and neuroimaging methods. Contextual modulations at both of these distinct stages demonstrate a reliance on local input intensity. This shared fundamental principle provided a refined lens for our combined studies. Inversion reduces the strength of face contextual modulations, decreases the response of face-specific brain regions, and leads to brain activation patterns more similar to non-face object categories. This makes inversion a valuable tool for exploring contextual modulations in advanced, non-specialized processing stages. Therefore, by testing inverted faces alongside upright faces, we were able to investigate the interactions between low level and non-specialized, high-level contextual mechanisms. The interactions between these mechanisms were investigated through an inter-individual differences framework. High-level, face-specialized contextual mechanisms seem to be working independently of the early regions. However, non face-specialized high-level contextual mechanisms seem to be influenced by the constraints presented by lower-levels of the processing hierarchy. These findings pose new questions in the literature such as whether contextual modulations at low-levels would influence the contextual modulations of other complex images that are not handled by dedicated mechanisms, and further adds to the large literature on whether faces are truly processed in a special manner or not.

Acknowledgments

I would like to express my deepest thanks and acknowledge all the people who directly or indirectly helped make this happen.

First of all, I would like to thank Valerie, for giving me the opportunity to work with her. Your guidance and support made this notoriously stressful and difficult period so much more enjoyable, and I felt truly privileged to have you as a supervisor. You created an atmosphere where I felt like I could progress in my own pace, while making sure I can always depend on your unwavering support. I really appreciate you, and all our time spent talking about everything, from Nicolas Cage movies to dogs and UFOs, and of course, occasional science.

Thank you to Pieter, who has been a co-supervisor in my project from the beginning, and unfortunately spent very little time with us. Despite our short time, I learned so much from you over the years, and your rigorous approach to science and robust psychophysics inspired me to always push a little bit further and ensure my methodology is watertight. My accompanying committee, Michael and Gilles. Thank you for taking time out of your schedules to support, discuss and improve our science. Your constructive criticism and unique perspective made this thesis better by helping me see my own work from different points of view.

To Jolien, for not only being just a lab member with whom I share an office, but for being the best friend one can ask for to share all the suffering and pain that came with doing a PhD. Our conversations about life, imposter syndrome, our accomplishments, bananas, and the times spent “not working” made me feel a little less lost, and a little more understood throughout our work and writing journey. To Matthew, who made every single day in the lab a little less boring than it could be, and possibly a little more technical than it should be. I will always remember your captivating stories and the enthusiasm you instilled in me, encouraging me not to fear starting to learn new things from the very beginning. To Alexia, for always making me feel like I am not alone in the lab when I struggle with Bayesian statistics, and for getting lost with me in the technicalities where it sometimes felt impossible to understand. To Hélène, for being a good friend throughout these years who made Belgium and especially Brussels a little more interesting and fun than it could be. To Marius, Mrittika, Yu-Fang, Kirsten and Ece, for making our work and lunch break conversations range from anywhere between lobsters to history of the world, and making me never regret being in the office to join those conversations.

To my friends back at home, Kıvanç, Simge, Ulaş, Müge for providing me with all kinds of support, from the struggles of adult life to being alone and starting in a new country, and of course by being amazing friends that never made me feel lonely even when I am thousands of kilometers away.

And lastly, to my family for their belief in my enthusiasm and giving me the chance to try the unusual and unexplored territory. I will be grateful for your support all my life, and I appreciate you all.

Chapter 1

General Introduction

Vision is a dominant sense in primates that utilizes roughly half of the cerebral cortex; its study is central to cognitive neuroscience. Humans rely heavily on visual perception to interact with their surroundings, adapting their behavior based on the interpretation of visual signals. Among these signals, faces play a crucial role. They serve as a potent and socially significant source of information, communicating an array of attributes including age, gender, ethnicity, health status, emotional states, and trustworthiness. Our evolutionary history and life experiences have equipped us with mechanisms to process the abundant information conveyed by faces. Through the use of these mechanisms, face perception is achieved effortlessly and quickly, with an efficiency that outperforms even the most advanced artificial recognition systems to date (Fleuret et al., 2011; Phillips & O'Toole, 2014).

Are faces special?

Certain cognitive processes have been proposed to be served by domain-specific systems, or "modules", as described by Fodor's modularity hypothesis (1983). It has been used to explain various cognitive phenomena, such as the speed and efficiency of information processing in specific domains, the ease with which humans acquire language, and the potential existence of evolved cognitive adaptations. According to this hypothesis, modules have particular functions, and they are informationally encapsulated (i.e., they do not have access to all information in the system and only process specific types of information). This encapsulation further implies that they remain impervious to variations in elementary properties such as local color, luminance, contrast, spatial frequency and orientation information, which are considered low-level features within the context of vision. The engagement of a module is mandatory and automatic when the relevant kind of information is encountered, and this engagement recruits specific brain regions rapidly without conscious control or effort due to their specialized nature. Furthermore, their structure or architecture is proposed to be hard-wired through evolution, and not malleable by learning or experience. Finally, Fodor's hypothesis suggests that interactions between these modules are limited, and information is typically transferred between them through specific mechanisms or interfaces, rather than through direct communication. This limited interaction is thought to contribute to the modularity of the mind. Several lines of evidence has brought support to the notion that face perception may be processed by such a dedicated module, characterized by specialized cognitive and neural mechanisms that are distinct from those involved in perceiving other visual stimuli (Kanwisher et al., 1997; Kanwisher,

2000). This module is thought to have evolved specifically to handle encoding and processing of facial features, such as the eyes, nose, and mouth, as well as the configuration and arrangement of these features. Evidence supporting the domain specificity of faces comes from studies on development, neuropsychology, neuroimaging and psychophysics.

Developmental evidence for face processing specificity

Newborns show a predisposition for face processing even in the absence of extensive visual experience, suggesting the existence of innate and hard-wired mechanisms for faces. First findings in this regard come from a study of forty newborns (with a median age of 9 minutes) that were found to be significantly more responsive to a face pattern (i.e., two eyes positioned on top of a nose, which is above the mouth) compared to its scrambled version or a blank stimulus (Goren et al., 1975; Johnson et al., 1991, see Figure 1a.). However, this preference declines during the second month postnatal, and instead of a responsiveness to the face pattern, infants develop more sophisticated abilities to discriminate between different faces and show a preference for familiar faces, such as their mother's face (Mondloch et al., 1999; Johnson et al., 2000). Taken together, these two findings formed the two-process theory of newborn face perception, coined CONSPEC-CONLERN. This theory posits that early on, an innate subcortical device (CONSPEC) orients the newborn to stimulate themselves with a face-rich environment, which causes the development of CONLERN, the cortical specialization for face perception. Later on, CONLERN takes over and suppresses CONSPEC, and then becomes responsible from face processing abilities (Johnson & De Haan, 2011).

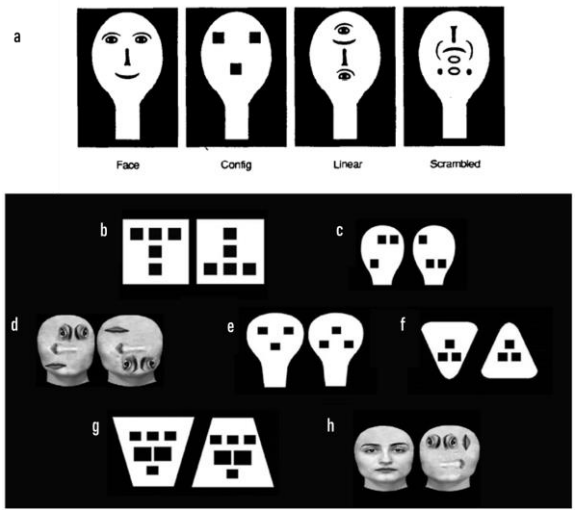

Figure 1. ***a****. The images shown to infants in Johnson et al., (1991). While there is a strong preference for the face pattern over other stimuli in the first month, this preference declines during the second month postnatal.* ***b-h****. Examples of stimuli employed by various experiments in Simion & Giorgio, (2015) to test the inborn predispositions to general structural properties of faces.* ***b-c.*** *Stimuli used to test up-down asymmetry.* ***d.*** *Real faces employed to test up-down asymmetry.* ***e-g.*** *Stimuli used to test congruency.* ***h.*** *Real faces employed to test up-down asymmetry and congruency.*

An alternative perspective suggests that the proclivity of newborns towards faces can be attributed to the inherent structural attributes present in a face, attributes that may also be present in various non-face objects (Valenza et al., 1996; Simion et al., 2001, 2003, 2006; Turati, 2004; Simion & Giorgio, 2015, see Figure 1b.). This view proposes that faces attract the attention of newborns due to a specific constellation of structural properties. These properties include bilateral symmetry, areas of stark contrast (notably around the eyes), and a greater concentration of elements in the upper than the lower portion of the face. Beyond these static features, faces are also three-dimensional, exhibit movement, and display behavior that is contingent upon the baby's activities. The simultaneous presence of these dynamic and static characteristics potentially makes faces one of the most captivating stimuli for newborns. While the two-process theory (CONSPEC-CONLERN) and alternative perspectives shed light on newborns' predisposition for face processing, congenital prosopagnosia (Behrmann & Avidan, 2005) introduces additional nuance to this understanding. Marked by impaired face recognition abilities from birth without any brain damage, this condition presents an exception. It emphasizes the role of biological factors interwoven with our ability to process faces effectively, a delicate balance that, if disrupted, can lead to specific deficits.

Neural evidence

A different line of evidence comes from the neuropsychological literature in the case of prosopagnosia, in which individuals have difficulty perceiving or recognizing faces despite a largely preserved ability to recognize objects of other categories either from birth (congenital prosopagnosia; Behrmann & Avidan, 2005) or after a brain lesion (acquired prosopagnosia; Bodamer, 1947). The opposite syndrome has also been observed with impaired reading and object recognition, but completely intact face recognition abilities (Moscovitch et al., 1997). This double dissociation is a strong argument for the domain-specificity of the mechanisms engaged for face perception and the existence of dedicated brain structures. Prosopagnosia can be categorized into two types: apperceptive and associative. Apperceptive prosopagnosia refers to a deficit in the fundamental ability to visually process and perceive faces and facial features, whereas associative prosopagnosia involves the inability to connect facial information with personal memories or semantic information. Studies of the lesions in the brains of acquired prosopagnosic patients revealed its origins mainly in anterior inferior occipital lobe in both hemispheres (Meadows, 1974; Damasio et al., 1982). Although rare cases with left hemispheric lesions were also noted to contribute to it (Barton et al., 2002; Mattson et al., 2000), in later observations, a higher prevalence of this condition was detected in the right hemisphere (Landis et al., 1986; Renzi, 1986). This aligns with data from functional neuroimaging showing that face processing leads to heightened activation in the right hemisphere in general (e.g., Kanwisher et al., 1997) mainly in the ventral occipito-temporal cortex, fusiform gyrus, and anterior temporal cortex (Grill-Spector et al., 2017; see Figure 2). Individuals with damage in the occipito-temporal cortex or fusiform gyrus typically manifest symptoms indicative of the apperceptive subtype (Barton et al.,

2002), while those with lesions in the anterior temporal areas tend to display characteristics associated with the associative subtype (Barton, 2008; Albonico & Barton, 2019).

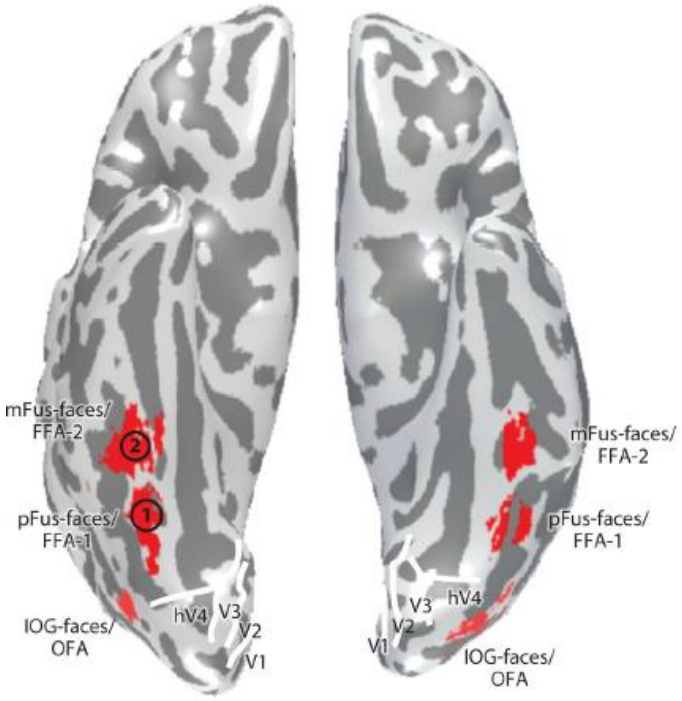

Figure 2. Face-selective regions in human ventral occipito-temporal cortex. Figure from Grill-Spector et al., (2017)

Apart from the mentioned disabilities, starting with the seminal work of Sergent et al., (1992), neuroimaging studies have consistently revealed the existence of brain regions responding preferentially to faces over other visual categories in the normal population. Face-specific fMRI responses were observed in many individuals in the occipital lobe, specifically in the area termed "occipital face area" (OFA). However, the most stable and pronounced face-selective response was found on the lateral aspect of the mid-fusiform gyrus, an area designated as the "fusiform face area" or FFA (Kanwisher et al., 1997; McCarthy et al., 1997; Haxby et al., 2000; Tong et al., 2000; Iidaka, 2014; Grill-Spector et al., 2017). Other regions also demonstrate face-preferential responses, including the pulvinar, which is thought to help with the encoding of novel faces and face-name associations (Morris et al., 1997; Rossion et al., 2012; Vuilleumier et al., 2003); posterior superior temporal sulcus, which reacts to dynamic features related to faces such as gaze direction and facial expressions (Iidaka, 2014); medial prefrontal cortex (Richardson et al., 2021), which is involved particularly when the difficulty of a face related task increases (Schiltz et al., 2010); anterior temporal lobe (Collins & Olson, 2014), which is hypothesized to act as a bridge between face recognition and face memory, connecting visual representations of individual identities with specific personal information. These regions exhibit significantly stronger activation when processing faces compared to various non-face stimuli including (but not only limited to) letter strings (Puce et al., 1996), miscellaneous objects (McCarthy et al., 1997; Fox et al., 2009; Haxby et al., 2000; Tsao et al., 2008;

Rajimehr et al., 2009), headless animals (Kanwisher et al., 1999; although see Chao et al., 1999), scenes (Pitcher et al., 2019), the posterior side of human heads (Tong et al., 2000) and cars (Rossion et al., 2012).

Further evidence supporting face-specific brain mechanisms has been accumulated through electrophysiological measures, including intracranial recordings (Allison, 1999) and scalp event-related brain potentials (ERPs). The studies using ERPs have identified distinct components associated with various stages of face detection/categorization (i.e., interpretation of a visual stimulus as a face), face identity recognition (e.g., the ability to determine that a specific individual has been seen before), and the processing of emotional facial expressions (i.e. recognizing a given expression and distinguishing it from other expressions; e.g., Eimer, 2000; Eimer & Holmes, 2007). Notably, the N170 component, a negative peak occurring around 170 milliseconds after stimulus onset, is consistently observed during face processing tasks, or even while passively viewing faces (Bentin et al., 1996; D. A. Jeffreys, 1996). N170 response is larger and faster for faces compared to other visual objects (Rossion & Jacques, 2008) and is believed to reflect the time course (Jacques & Rossion, 2006) and neural processes associated with the early stages of face perception. It is associated with recognizing the unique features and configurations that make a stimulus distinctly "face-like". Furthermore, simultaneous EEG-fMRI studies identify FFA as the major contributor of the N170 component (Gao et al., 2019).

The "expertise hypothesis" challenges the idea that face processing is domain-specific and suggests that our proficiency and unique neural responses when dealing with faces result from extensive experience and training. It proposes that the apparent specialization of visual mechanisms used for face processing may be acquired through a lifetime of exposure and practice in face individuation, and we would have similar effects to any other object we have such expertise at individuating at a fine identity level (e.g. dogs and dog experts; Diamond & Carey (1986)). We see faces every day, from birth onwards, and learn to pick up subtle cues and differences that allow us to tell individuals apart. This extensive training is what the expertise hypothesis claims underlies our exceptional ability to recognize faces, and that the apparent 'face selectivity' of the brain regions such as FFA are more likely to reflect a more generalized form of processing (Tarr & Gauthier, 2000). Greebles, artificial stimuli that share similarities with human faces (Gauthier et al., 1999; see Figure 3.), were employed to illustrate that developing expertise in distinguishing the members of a new category can also enhance the responsiveness of the FFA. Initially hard to distinguish, proficiency in recognizing Greebles increased with training, showing stronger activation in FFA than other object categories, similar to the activation that is elicited by faces (also seen with other objects of expertise, e.g., birds or cars; Gauthier et al., 2000). These findings have been controverted since they failed to replicate (Brants et al., 2011).

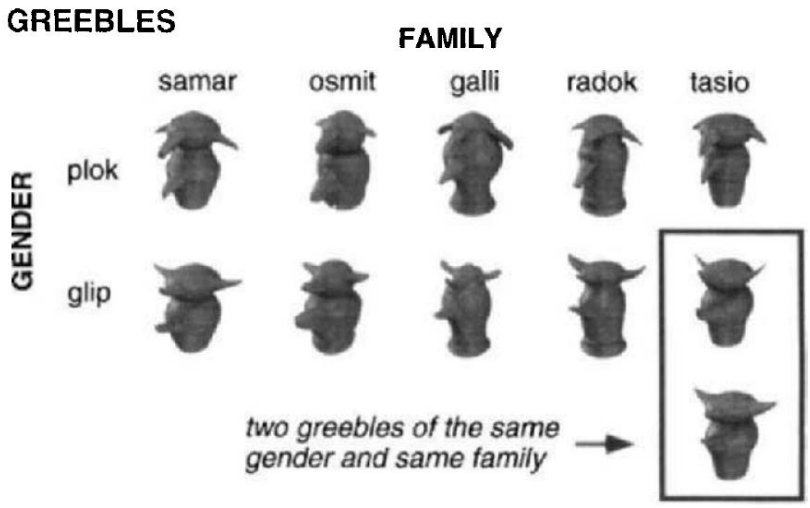

Figure 3. Greebles, the computer-generated stimuli that are designed to look similar to each other, much like human faces, with a similar configuration but varying in details. Figure from Tanaka & Gauthier (1997).

Face inversion effect

Expanding upon Fodor's proposition of domain-specific modularity within face processing, the face inversion effect (FIE) provides compelling evidence to suggest a distinct neural substrate dedicated to face recognition. FIE refers to the phenomenon in which the picture-plane inversion of a face leads to a substantial and disproportionate drop in recognition performance compared to other mono-oriented object categories (i.e., categories usually seen upright; e.g. houses, airplanes; Yin, 1969; cars; Cassia et al., 2009; birds, dogs Bruyer & Velge, 1981; Diamond & Carey, 1986; Robbins & McKone, 2007; scenes; Epstein et al., 2006). It is observed when orientation is manipulated in separate blocks or in-between subjects (Toyama, 1977; Valentine & Bruce, 1986) and randomly interleaved (Carey & Diamond, 1977; Yin, 1969; Goffaux, 2009, 2012). FIE serves as a strong indicator of the specialized mechanisms at work during the processing of upright faces, substantiating the distinct nature of face processing and the presence of a dedicated module for this function.

Valentine (1988) proposed an alternative explanation, namely that the origin of the FIE could be linked to how faces are encoded in memory. They found evidence for FIE in a recognition memory task, whereas no such effect was observed in a face-matching task (Valentine & Bruce, 1986), similar to Bruyer & Velge (1981). According to this theory, the perceptual processing of inverted faces may not necessarily differ significantly from that of upright faces. However, the process of encoding them into memory appears to be less efficient for inverted faces, primarily due to the distinct qualitative nature of forming memory representations. However, the FIE is observed in paradigms that used simultaneous presentation of faces (Farah et al., 1998; Moscovitch et al., 1997) as well as old-new recognition paradigms (Scapinello & Yarmey, 1970; Phillips & Rawles, 1979; Carey et al., 1980), two-alternative-forced-choice paradigms (Diamond & Carey, 1986; Leder & Bruce, 2000; Tanaka &

Farah, 1993; Yin, 1969) and with familiar and unfamiliar faces (Collishaw & Hole, 2000; Rock, 1974). Therefore, an increasing agreement exists that the FIE primarily takes place during perceptual encoding rather than being influenced by the structure of long-term memory. This notion is supported by the observation of FIE on the ERP component N170 (Rossion et al., 2000) and on reduced activity in FFA and OFA (Haxby et al., 1999a; Yovel & Kanwisher, 2005).

Inverting faces leads to small increases in activation within the face-selective areas of the inferior and mid occipital gyri, while causing a small decrease in activation within face-selective regions of the lateral fusiform gyrus and superior temporal sulcus (Haxby et al., 1999b). Interestingly, the same occipital increases and temporal decreases are observed when houses are inverted in house-selective regions. Consequently, these changes do not appear to be connected to the failure of, or disengagement of face perception mechanisms, hinting that inverted faces are still perceived as faces in the visual system. While the processing of inverted faces still engages these face-specific regions, it also involves mid-level and high-level, but non-face-specialized visual areas, showing similar activity patterns to how other high level object categories are processed (Haxby et al., 1999a; Gilaie-Dotan et al., 2010; Taubert et al., 2015; Liu-Shuang et al., 2022). The principal effect of face inversion is an increased response in ventral extrastriate regions that respond preferentially to other classes of objects (such as houses), removes the right hemisphere advantage typically seen in recognizing upright faces (Leehey et al., 1978; Levine et al., 1988; Rossion, 1999; Passarotti et al., 2007), therefore leading to a relatively less-specialized and less sparse activation pattern (Young & Yamane, 1992; Rolls & Tovee, 1995). The involvement of non-specialized regions suggests that they are recruited to aid in the perception of inverted faces. The reduced capacity to discern distinguishing features in inverted faces implies that the representations formed by activity in face-selective regions during inverted face perception are less distinctive than those formed during the perception of upright faces. This becomes evident in studies where a lack of activation in face-specialized regions is only observed when subjects are unable to discern a face within an inverted Mooney face (Kanwisher et al., 1998). Moreover, when faces are inverted, the low-level properties of the stimulus (i.e. contrast, luminance, spatial frequency) remain unchanged (Willenbockel et al., 2010). Hence, the resulting impairment is typically attributed to high-level mechanisms that are finely tuned to the statistical regularities of how upright faces are commonly encountered in everyday life (Andrews et al., 2010; Goffaux et al., 2013; Liu-Shuang et al., 2015; Mazard et al., 2006; Schiltz & Rossion, 2006).

Configural and Holistic Processing of Faces

Essentially, the FIE demonstrates that the human brain is particularly skilled at processing faces in their normal upright orientation. However, the exact nature of the processes underlying upright face perception remains a subject of continuous discussion. A face's individuality is defined by the specific relationships and spatial distances between its distinct features, forming a configuration particular to

that face (Diamond & Carey, 1986; Tsao & Livingstone, 2008). The processing of such a configuration is optimized for upright faces, while inverting the face disrupts the specialized processing of this configuration (Sergent, 1984; Rhodes et al., 1993; Bartlett & Searcy, 1993; Tanaka & Gauthier, 1997; Leder & Bruce, 2000) The term configuration refers to the first-order and second-order relational properties of the face stimulus (see Figure 4; Rhodes, 1988; Maurer et al., 2002; Civile et al., 2016). First-order relational properties are spatial arrangements of parts within a stimulus, like the placement of eyes over the nose, whereas second-order relational properties pertain to the slight variations in these arrangements when contrasting various faces. Therefore, second-order properties relate not only to the positioning of individual elements but also to how those positions compare to a typical or standard spatial configuration of the parts within that category or group. Diamond & Carey (1986) claimed that faces bear both of these two types of relational properties, and the inversion effect would be exhibited in any stimulus category where individuals have developed expertise in discerning second-order relations.

Figure 4. ***a.*** *A painting by* Giuseppe Arcimboldo (The Vegetable Gardener, 1587–1590, Museo Civico Ala Ponzone, Cremona, Italy) *demonstrates first-order relationships. When viewed in an upright position, the image resembles a face even when the features are made up of vegetables, since the features align to create the primary structure of a face. It does not appear face-like when presented upside-down due to the disruption of such arrangement.* ***b.*** *Manipulations in the second-order relationships (i.e.., the slight variations in the original arrangement of features such as distances between the eyes; images on the left and right) of a face (center, actor Nicolas Cage), changes the appearance and identity drastically.*

Among various interpretations of how a face is processed (e.g., configural, relational, integrative; Carey & Diamond, 1977; Diamond & Carey, 1986; Young et al., 1987; Hole, 1994; Calder et al., 2000; Collishaw & Hole, 2000; Leder & Bruce, 2000; Goffaux, 2012), one distinction has gained prominence: facial features and their configuration are processed “holistically” (Tanaka & Farah, 1993; Farah et al., 1998; Hole et al., 1999), as undivided "gestalts" in contrast to other objects, which are suggested to be processed in a more analytical or part-based manner. While configural processing pertains to the sensitivity towards spatial relationships between components of an object, such as the distance between the eyes (Leder & Bruce, 2000; Maurer et al., 2002; Mondloch et al., 2002), holistic processing is about the tendency to process all parts of an object simultaneously as a unified whole (Andrews et al., 2010; Richler et al., 2012; Richler & Gauthier, 2014; Meinhardt-Injac et al., 2014). While it is often associated with faces or objects of expertise, holistic processing can also be observed in non-face objects of expertise, or with line patterns with salient Gestalt information (Zhao et al., 2016). Holistic processing is therefore not only limited to higher-levels of the visual processing hierarchy, and there is a certain amount of overlap in terms of holistic processing between the processing of these stimuli of distinct levels of complexity at a perceptual level (Curby & Moerel, 2019).

Recognition of individual facial features is better when they are presented as part of a complete face compared to when they are presented in isolation or within a scrambled configuration (see Figure 5a.; Tanaka & Farah, 1993; Davidoff & Donnelly, 1990; Farah et al., 1998; Carbon & Leder, 2005). This contextual advantage, coined “part-whole effect” demonstrates the degree of reliance even on the unattended parts of the face, and how upright faces are processed holistically. It implies that during the recognition of individual facial features, the entire facial configuration plays a crucial role. Indeed, face inversion hinders the recognition of this overall structure, disrupting the benefit caused by the face configuration. Additionally, inversion more profoundly interferes with the perception of the face's configuration than it does with the recognition of individual local facial features. The rationale behind this process is linked to the need to differentiate individual faces (e.g., Mary, John, or Nicolas Cage) while most other objects are typically distinguished at a broader category level (e.g., car, bird, etc.; Rosch et al., 1976). Therefore, holistic processing is thought to enable the fast and efficient processing of faces despite their high visual similarity within a subordinate level. Although holistic processing likely relies on the processing of the face configuration, configural and holistic processing may partly function independently. For example, the depiction of half a face might activate configural processing, as the connections between the visible parts can be understood in terms of their relative positions, while it might not be sufficient to form the holistic template (Rossion & Gauthier, 2002). Still, these terms have been used interchangeably in the literature, mostly under the name ‘holistic processing’, referring to a generic, common mechanism.

Composite Task as a measure of holistic processing

Two completely identical top halves of a face are perceived as being different when they are paired with bottom halves that belong to different faces. This phenomenon, named "composite effect", illustrates that the parts of a face cannot be processed independently from face context, even when the observers explicitly try to ignore the other parts (Young et al., 1987; see Figure 5b.). In other words, the presence of one half alters the appearance of the other, due to the holistic nature of face processing. When the two halves are misaligned, the face is no longer perceived as a whole, and it becomes easier to tell they are identical. The composite effect was replicated in many studies in the literature, using diverse methodologies, and was extended to unfamiliar faces as well (Hole, 1994; Hole et al., 1999; Le Grand et al., 2004; Goffaux & Rossion, 2006). Its neural basis was revealed through fMRI adaptation studies in which the same top half of a face elicits a larger response (a bigger release from adaptation) in FFA when aligned with a bottom half that belongs to a different face (Schiltz et al., 2010; Schiltz & Rossion, 2006) and through EEG where consistent evidence shows that misalignment of a composite face elicits a delayed and enhanced N170 component ((Jacques & Rossion, 2009, 2010; Kuefner, 2010; Wiese et al., 2013).

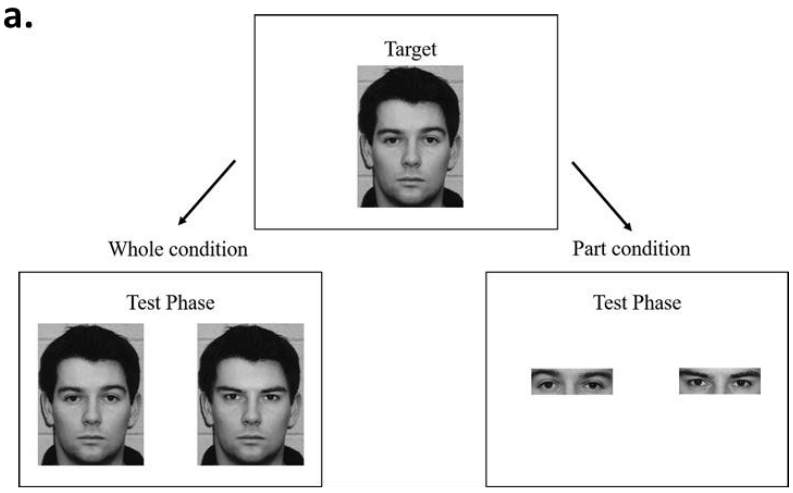

Figure 5. Part-whole effect and Composite effect. ***a. Part-whole effect.*** *When asked to identify the individual features, participants perform better when the feature is presented in the whole condition. Figure from Belanova et al. (2021).* ***b. Composite effect.*** *The vertical alignment of two halves lead to their perceptual merging, making it difficult to tell that the two top halves are identical. Misalignment or inversion drastically reduces this effect. Figure from Murphy et al. (2017)*

Holistic processing is considered a hallmark of specialized face processing mainly because of its drastic disruption by inversion, with the composite task emerging as a standard paradigm to measure it. It has been used to track the development of face recognition (Cassia et al., 2009), to address the functional roles of cortical regions specialized for face processing (e.g.; Schiltz & Rossion, 2006; Schiltz et al., 2010; Goffaux et al., 2013; Zhou et al., 2018; Foster et al., 2021), to study abnormal development of face recognition (Le Grand et al., 2004), and to test generic face recognition abilities in the normal population (Richler et al., 2011; although see Konar et al., 2010). However, the validity of the composite paradigm has been challenged (Richler, Gauthier, et al., 2008; Richler, Cheung, et al., 2009, 2011a).

A trial of the composite discrimination task comprises presentation of two faces, either side by side (Hole, 1994; Hole et al., 1999) or subsequently in a delayed matching task (e.g. Le Grand et al., 2004; Goffaux, 2012; Murphy et al., 2017). The participant is asked to report whether the task-relevant, two top halves are identical or different, while ignoring the task-irrelevant, bottom halves (although holistic processing has also been tested with bottom halves as task-relevant targets, but less frequently, e.g.; Nishimura et al., 2008; Ramon et al., 2010). There are two distinct versions of the composite task in the literature: partial and complete (Gauthier & Bukach, 2007). In the partial design, there are only two categories of trials: the congruent-different trials, where both the target and context vary between two faces, and the incongruent-same trials, where only the context changes while the target feature remains constant (see Figure 6). In same trials, *the alignment effect* is used as an index of holistic processing. In misaligned trials, either accuracy is greater, and/or reaction time (RT) is faster than in aligned trials due to the spatial offsetting, which breaks the whole stimulus configuration, substantiating that it is holistic processing (Hole, 1994; Michel et al., 2006; Goffaux & Rossion, 2006; Robbins & McKone, 2007; Cassia et al., 2009; Boutet et al., 2003; Calder et al., 2000; Schiltz & Rossion, 2006; Anaki et al., 2011; Nishimura et al., 2008; Konar et al., 2010a). Since it infers holistic processing from the interference caused by "different" contexts on "same" target features only, the validity of the partial design has been questioned due to this imbalance and potential confound with response biases (Richler, Gauthier, et al., 2008).

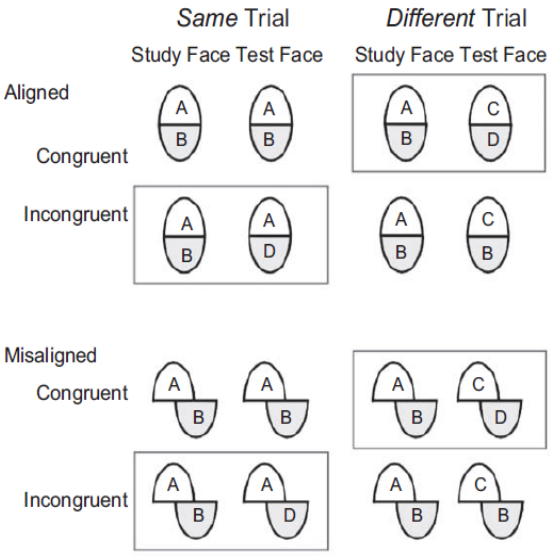

Figure 6. Trial types in the composite task. *Letters represent facial identities. Top halves (white) are the task-relevant targets, whereas bottom halves (grey) are explicitly instructed to be ignored (task-irrelevant). Congruency is defined by the agreement between the answers an individual would give to two halves, e.g., both top and bottom halves would receive a "same" answer in the congruent-same trial, while they would both receive a "different" answer in the congruent-different trial. In the Partial Design, only the trial types marked with gray boxes would be used, while the Complete Design would include all the trial types. Figure from Richler et al. (2011b).*

The complete composite design has been developed to address these issues. It includes all possible congruency by similarity combinations: congruent-same, congruent-different, incongruent-same, and incongruent-different. In this design, holistic processing is indexed by the *congruency effect*. Misalignment is again expected to reduce the overall congruency effect but several studies employ only the aligned version of this paradigm using the congruency effect as the primary index of holistic processing (e.g. Richler, Gauthier, et al., 2008; Richler et al., 2009; Goffaux, 2012; Canoluk et al., 2023). The performance (as measured by accuracy or RTs) is better in congruent trials than in incongruent trials (Farah et al., 1998; Cheung et al., 2008; Goffaux, 2009; Richler, Cheung, et al., 2011b; Anaki et al., 2011; Richler, Tanaka, et al., 2008). In contrast to the partial design, the complete design allows predictions to be made about holistic processing in both "same" and "different" trials. If participants are unable to focus solely on the target part, their performance will suffer on incongruent compared to congruent trials. In same-incongruent trials (those where the correct response is "same," but the irrelevant part is different), a failure to process the upper part in isolation can lead participants to be more inclined to answer "different." Conversely, in different-incongruent trials (where the correct response is "different," but the relevant part is the same), such failure inclines participants to respond "same". When information is contradictory, it disrupts performance, emphasizing the complex interplay between different elements of faces. An alternative explanation for this interference is attributed to attentional strategies that became automated with experience (Richler, Cheung, et al., 2009; Richler et al., 2012; Richler, Wong, et al., 2011). From this perspective, individual facial elements are independently processed and stored, and holistic processing naturally emerges because each facial component is automatically encoded all at once.

The proponents of the partial design compared the complete design to the Stroop task (Stroop, 1935), originally designed to gauge response interferences caused by incongruency. Whether holistic

processing arises from perceptual or decisional (response-based) mechanisms lead to a debate on which of the two designs capture the effect more efficiently. However, it was demonstrated that the interference in holistic processing actually originates at a prior, perceptual stage and is not indicative of response interference (Schiltz & Rossion, 2006; De Heering et al., 2008; Richler, Cheung, et al., 2009; Jacques & Rossion, 2009; Kimchi & Amishav, 2010; Kuefner, 2010; Schiltz et al., 2010; Cheung et al., 2011; Taubert et al., 2011). Although many reviews combine various tasks to draw overarching conclusions about holistic processing (e.g., McKone et al., 2007), there is limited evidence on whether various paradigms measure the same construct (see Richler et al., 2012; Rezlescu et al., 2017; Sunday et al., 2017) due to the lack of correlations among these tasks.

The magnitude of holistic processing shows consistent variation across individuals. These variations were considered to explain the variations in the generic face recognition abilities. Studies on acquired (Ramon, Busigny, et al., 2010; Busigny & Rossion, 2011) and congenital prosopagnosics (Zhu et al., 2009) suggest that prosopagnosic individuals may suffer deficits in holistic processing as well. In normal populations, Konar et al. (2010a) investigated the source of these variations by using the partial version of the composite task in conjunction with a generic face-identification task in the same group of observers, and did not observe any correlation. Richler et al. (2011b) on the other hand, observed a correlation between measures of face identification, face memory and holistic processing when they employed the complete version of the composite task. There has been attempts to find a single factor that would account for general face processing abilities (such as *f;* Verhallen et al., 2017) or for the measures of holistic processing (McCaffery et al., 2018; Wang et al., 2012). However, even when such relationships have been identified (e.g., holistic processing predicting general face recognition ability) the correlations tend to be only modest in strength (DeGutis et al., 2013). Barton et al. (2004) investigated the processing of luminance, saturation, curvature and line-orientation discrimination, contrast sensitivity, and perception of spatial relations in individuals with apperceptive prosopagnosia. Contrast sensitivity at high spatial frequencies and the perception of spatial relations were outside of the normal range with this group, but the deficits in other abilities (processing of low-level properties such as luminance, saturation, curvature) were unlikely to contribute to their apperceptive prosopagnosia. Studies that examine the impact of low-level features on the activation of holistic processing mechanisms deepen our understanding of this process.

Low-level visual influences on holistic processing

Holistic processing, which is thought to make face processing unique, is largely attributed to high-level regions of the visual processing hierarchy. This attribution is primarily due to the impact of FIE, a reversible manipulation where the low-level, basic properties of the inverted image remain the same, and to neuroimaging evidence indicating that high-level face-specialized regions such as the FFA responds to faces in a holistic manner (Goffaux et al., 2013; Zhou et al., 2018; Foster et al., 2021). Despite the emphasis on high-level regions, it's improbable that holistic processing relies exclusively

on these dedicated mechanisms. Visual signals from the retina and lateral geniculate nuclei are initially processed at lower-level cortical areas (e.g., V1, V2, V4), and then projected to higher level regions where the extracted image is processed further (Lamme et al., 1998; Bar, 2003; Kravitz et al., 2013). Various high-level visual mechanisms such as scene categorization or navigation (Groen et al., 2017) exhibit sensitivity to low-level (e.g., spatial frequency) and mid-level (e.g. spatial layout) properties. A dedicated module for face processing would require its encapsulation and independence of the restraints posed by the rest of the visual system. This module would be automatically and fully engaged whenever the visual system is presented with upright faces. However, evidence suggests that holistic processing in upright face processing depends on fundamental image characteristics like orientation and spatial frequency content. Moreover, it is not fully impenetrable as shown by its modulation by experience-dependent grouping (Curby et al., 2016). Thus, it can be influenced or constrained by processes occurring in the earlier stages or lower levels of the visual processing hierarchy, as well as the contextual susceptibility seen at such levels.

Face processing is contingent upon a specific orientation structure, indeed, the recognition of familiar faces is preferentially driven by the presence of horizontal information when compared to other orientation bands (Dakin & Watt, 2009). Specifically, when the information in a face image is restricted to a narrow band of orientations, observers exhibit peak recognition performance when the band spans horizontal angles, while performance steadily declines towards vertical angles (e.g., Goffaux & Greenwood, 2016). The horizontal structure of a face forms a so-called "bar code", i.e., vertically aligned clusters of horizontal stripes (marked mainly by the darker region (eyes), on top of a bright region (from the cheeks), followed by another darker region (mouth), distinguishing them from scenes and objects lacking this structural regularity (Dakin and Watt, 2009). Therefore, the special characteristic of faces is attributed to the presence and vertical alignment of this horizontal structure. Additional evidence supporting this notion comes from Goffaux and Rossion (2007), wherein they demonstrated that face inversion impairs perception only when facial features are vertically displaced, but not when they are horizontally displaced. In a more comprehensive study, Goffaux & Dakin (2010) tested the effect of orientation filtering on various behavioral markers of face processing, such as FIE, facial identity after-effect, viewpoint tolerance of face identity recognition, and holistic processing. When all orientations were removed, but horizontal information was maintained, these four indicators remained mostly intact. In contrast, preserving only vertical information greatly weakened them. The orientation dependence of holistic processing showcases the significant role of low-level properties play in its engagement.

Like all visual stimuli, a face is represented on the retina through luminance gradients of various spatial frequencies (SF). Low SF represents coarse, gradual luminance gradients, while high SF represents fine, rapid luminance changes in retinal input. For instance, face shading is captured in low SF, while the details of facial features and skin texture are encoded in high SF. Moreover, face

perception exhibits higher sensitivity to manipulation of spatial frequencies compared to other object categories, with different kinds of facial information (e.g., identity, expression) being driven by distinct SF ranges (Biederman & Kalocsai, 1997; Collin et al., 2004; Sowden & Schyns, 2006; Williams et al., 2009). While intermediate SFs are required for face identification (7-16 cycles per face; Gold et al., 1999; Näsänen, 1999), the coarse structure provided by low SFs were found to contribute more to processing of distances between features than processing of the individual features (Goffaux et al., 2005). With high SFs, this pattern is reversed and feature processing becomes easier. Further studies indeed corroborated the importance of SF information for the engagement of holistic processing mechanisms through the use of classical whole-part, composite and congruency paradigms. Holistic processing was strongly attenuated when the image only contained high SFs, while it was stronger with low SFs (Goffaux & Rossion, 2006; Goffaux, 2009), another signature that the engagement of these mechanisms are susceptible to the low-level properties of the face stimulus. This was also evident from the enhanced engagement of N170 component with faces presented in low SF. Studies have consistently shown that these early event-related potential (ERP) signatures of face-specific processing primarily rely on low SF information, as inversion effects on the N170 are absent and categorical differences are weak when only high SF are presented in the stimulus (Goffaux et al., 2003; Halit et al., 2006). However, at the behavioral level, inversion does not exclusively disrupt low SF-driven holistic processing; the processing of face low and high SF is equally affected by inversion (Boutet et al., 2003; Gaspar et al., 2008; Goffaux, 2008). This indicates that inversion effect is not simply caused by the loss of low SF information: when a face is upside down, it's not just the overall configuration of the face (holistic processing driven by low SF) that is disrupted, but also the detailed features and fine-grained information that are crucial for face recognition. Both low and high SF components of face processing are impacted when a face is inverted, highlighting the importance of holistic processing that combines information from different spatial scales for effective face recognition.

Contextual modulations from the primary stages of visual processing

Holistic processing is a kind of contextual modulation (CM) where the perception of a local feature is influenced by the face context. In the remaining sections of this thesis, holistic/configural processing will be referred to as face contextual modulations, as they encompass the integration of (and modulation by) the surrounding elements to (on) the processing of the local facial features. The engagement of face contextual modulations seem to depend on features predominantly resolved at the early stages of visual processing (e.g., orientation and spatial frequency), which makes it highly probable that they are connected to the contextual modulations seen in these early stages.

Historically, vision scientists first characterized the responses of single neurons to isolated local

stimuli, which defined the classical receptive field (cRF) of a neuron. The cRF refers to the visual field region where presentation of a stimulus of optimal parameters evokes spiking responses from a single neuron (Hubel & Wiesel, 1965). However, it was later discovered that a neuron's response not only depends on what is presented within its cRF, but is modulated by what is presented outside of its cRF, i.e., its (spatial) context (Blakemore & Tobin, 1972; Maffei & Fiorentini, 1976; Nelson & Frost, 1978). CM is generated through feedforward (from the retina (Solomon, 2006) and lateral geniculate nucleus (LGN; Levick et al., 1972; Solomon et al., 2002)), horizontal (Gilbert & Wiesel, 1990), and feedback connections (Angelucci et al., 2002, 2017) to V1. Feedforward and horizontal connections primarily influence modulation in the near (or immediate)-surround of the cRF. Feedback connections originating from higher hierarchical levels on the other hand contribute to modulation in both near and far-surrounding context. Proximal regions of the far-surround experience modulation from closer hierarchy levels to V1, such as V2, while the most distal regions are influenced by higher stages of hierarchy such as MT (Gonchar & Burkhalter, 2003; Anderson & Martin, 2009; S. Zhang et al., 2014). Axons of the feedback connections transmit signals with a speed 10 times greater (2–6 m/s) than the intra-V1 horizontal axons (Girard et al., 2001). This characteristic allows them to facilitate the rapid initiation of far-SM. Speculatively, feedback connections from various areas likely modulate V1 responses in a stimulus-specific manner, contingent on the functional role of the specific cortical area supplying feedback inputs to V1.

Contextual modulation is a fundamental characteristic of visual neurons observed across various species, such as mice, cats, monkeys, and humans, at different levels of the visual system. It has been documented in the retina (McIlwain, 1964; Solomon, 2006), superior colliculus (Goldberg & Wurtz, 1972; Sterling & Wickelgren, 1969), LGN (Alitto & Usrey, 2008; Bonin et al., 2005; Levick et al., 1972; Marrocco et al., 1982; Nolt et al., 2004; Sceniak et al., 2006; Solomon et al., 2002), pulvinar (Berman & Wurtz, 2011; Chalupa et al., 1983), primary visual cortex (V1; Versavel et al., 1990; Gilbert & Wiesel, 1990; Levitt & Lund, 1997; Sceniak et al., 2001; Shushruth et al., 2009; Angelucci et al., 2002; Cavanaugh et al., 2002; Nauhaus et al., 2009), and extrastriate cortex (Allman et al., 1985; Brincat & Connor, 2004; Desimone, 1991; Krause & Pack, 2014). While extensively studied in the visual cortex, CM is not exclusive to vision and has also been observed in other sensory systems, including audition (Knudsen & Konishi, 1978; Sutter et al., 1999), somatosensation (Sachdev et al., 2012; Vega-Bermudez & Johnson, 1999), and olfaction (Olsen & Wilson, 2008). This widespread presence of CM across sensory domains suggests its fundamental role in sensory processing.

Contextual modulation as an efficient neural information processing mechanism

Numerous studies have contributed to the understanding of CM and its implications for neural information processing. For example, dissimilar stimuli inside and outside of a neuron evokes

stronger responses from a neuron, as opposed to similar stimuli. This observation suggested that they may play a role in saliency, pop-out, and figure-ground segmentation (Knierim & van Essen, 1992; Lamme, 1995). Vinje and Gallant's (2000) findings revealed that CM has a profound impact on the sparseness of neural activity, which refers to the activation of only a few narrowly-tuned neurons at a given moment. This phenomenon is of great importance as it influences various aspects of visual processing, including metabolic efficiency, feature selectivity, and information transmission. By making the neural code sparser, CM optimizes the allocation of neural resources, ensuring efficient energy utilization in the visual system. This metabolic efficiency is crucial for sustaining the high demands of visual information processing, as it minimizes unnecessary neural activity and conserves energy.

Moreover, CM contribute to enhanced feature selectivity by suppressing irrelevant information, enhancing contrast and sensitivity, and adapting the response based on the context, therefore allowing neurons to selectively respond to specific visual features or stimuli. This selectivity is vital for the precise representation and discrimination of different visual attributes, such as shape, orientation, and spatial frequency. Through CM, the visual system can effectively filter out irrelevant information and focus on the salient features of the visual scene, enabling a more accurate and refined representation (T. E. Palmer, 1975). Furthermore, they play a pivotal role in optimizing information transmission within the visual system (Bar & Aminoff, 2003; Bar, 2003, 2004). This optimized coding scheme ensures efficient and reliable transmission of relevant visual signals, leading to improved perceptual performance.

While contextual modulations play various crucial roles in vision, they are not without potential negative effects or limitations. This is especially evident in visual illusions, as they lead to inaccurate judgments and an inaccurate representation of the world, where things no longer appear "as they are". Greater contextual modulations are associated with a tradeoff, typically resulting in reduced accuracy in local processing (Song et al., 2013). Of course, within the context of cortical processing, it may not necessarily mean that leaving out a certain part of information unprocessed, or having a biased perception is negative. In order to address this, it is essential to comprehend visual perception as a biological system that has evolved through natural selection. As with any system in the body, it has been molded by natural selection to fulfill distinct roles in the individual's overall fitness. The visual system is just one of the numerous organs in the body, playing a specialized role in contributing to the overall fitness of the organism (S. E. Palmer, 2002). As we are constantly bombarded by sensory stimulation, over the course of evolution these mechanisms were developed for such purpose. Indeed, this is not specific to human visual system, as various non-human species exhibit visual illusions as well (e.g., dogs (Byosiere et al., 2017); cats (Bååth et al., 2014; Regaiolli et al., 2019), flies (Agrochao et al., 2020), monkeys (Agrillo et al., 2015) and fish (Gori et al., 2014)). Considering the widespread existence of illusions, a decrease in precision caused by contextual modulations may not necessarily

mean a loss of function. Conversely, one could speculate that this might indicate areas where information was not essential for species' capacity and survival, leading to their neglect. This selective functionality more likely represents its overall advantage.

Natural visual signals (as opposed to lab-controlled, artificial stimuli) are always viewed within context; therefore, this mechanism is always active during daily, natural vision. This emphasizes the relevance of studying contextual modulations with natural stimuli in understanding how our visual system processes and interprets visual information in real-world scenarios. The majority of quantitative neuronal models in visual processing rely on neurophysiological data obtained from experiments using basic stimuli like bars, gratings, and spots of light (e.g. Hubel & Wiesel, 1965). These simplified stimuli allow researchers to test specific hypotheses about how neurons function. However, neurons exhibit different behavior under natural viewing conditions (Carandini, 2005), therefore these models based on simple stimuli may not accurately represent neural processes. After all, the main purpose behind the investigation of the neural basis of visual perception is to predict neuronal responses during natural vision. To truly evaluate the accuracy of these models, they should be tested with natural stimuli like faces. Such studies will reveal how the models correspond to real-world visual intricacies, especially given the non-linearity of transitions between visual areas (Olshausen & Field, 2004). Chapters 2 and 3 will touch upon this issue by investigating contextual modulations with (relatively more) naturalistic stimuli and basic stimuli in conjunction.

Local input strength dependence of contextual modulations

Context can influence the perception of information presented in the cRF. Yet, the intensity and nature of this influence are determined not just by the context, but also by the local information and the interplay between the two. For example, neurons in the primary visual cortex are more likely to integrate information across the visual field when the contrast of the local input is weak (Levitt & Lund, 1997; Angelucci et al., 2002; Krause & Pack, 2014; see Figure 7a). Similarly, the extent of spatial summation is larger with weaker local contrast (Sceniak et al., 1999), and local field potential (LFP) waves are stronger in magnitude and travel a longer distance with weak contrast in cat and monkey primary visual cortex (see Figure 7b.; Nauhaus et al., 2009). Therefore, the quality and magnitude of contextual modulations are dependent on the strength of the local input. This phenomenon has been documented in terms of the spiking behavior of inhibitory and excitatory neurons in LGN and V1 as a function of synaptic input strength (see Figure 7c). The excitatory neurons are active even with zero synaptic input, but their spiking output has a gentle slope (i.e., changes gradually or slowly in response to changes in input). On the other hand, inhibitory neurons require a certain level of synaptic input to be activated, but once activated their spiking output raises with a steep slope (i.e., changes rapidly or sharply in response to changes in input). This behavior is demonstrated in Figure 7d where a high contrast input activates the inhibitory neurons, inhibiting the horizontal connections between the feed-forward afferents, whereas with a low contrast input, the

synaptic input never reaches the minimum necessary strength to activate the inhibitory neurons (Figure 7e). With low contrast, the horizontal connections are retained, resulting in stronger influence from the surround (Angelucci et al., 2002; Krause & Pack, 2014).

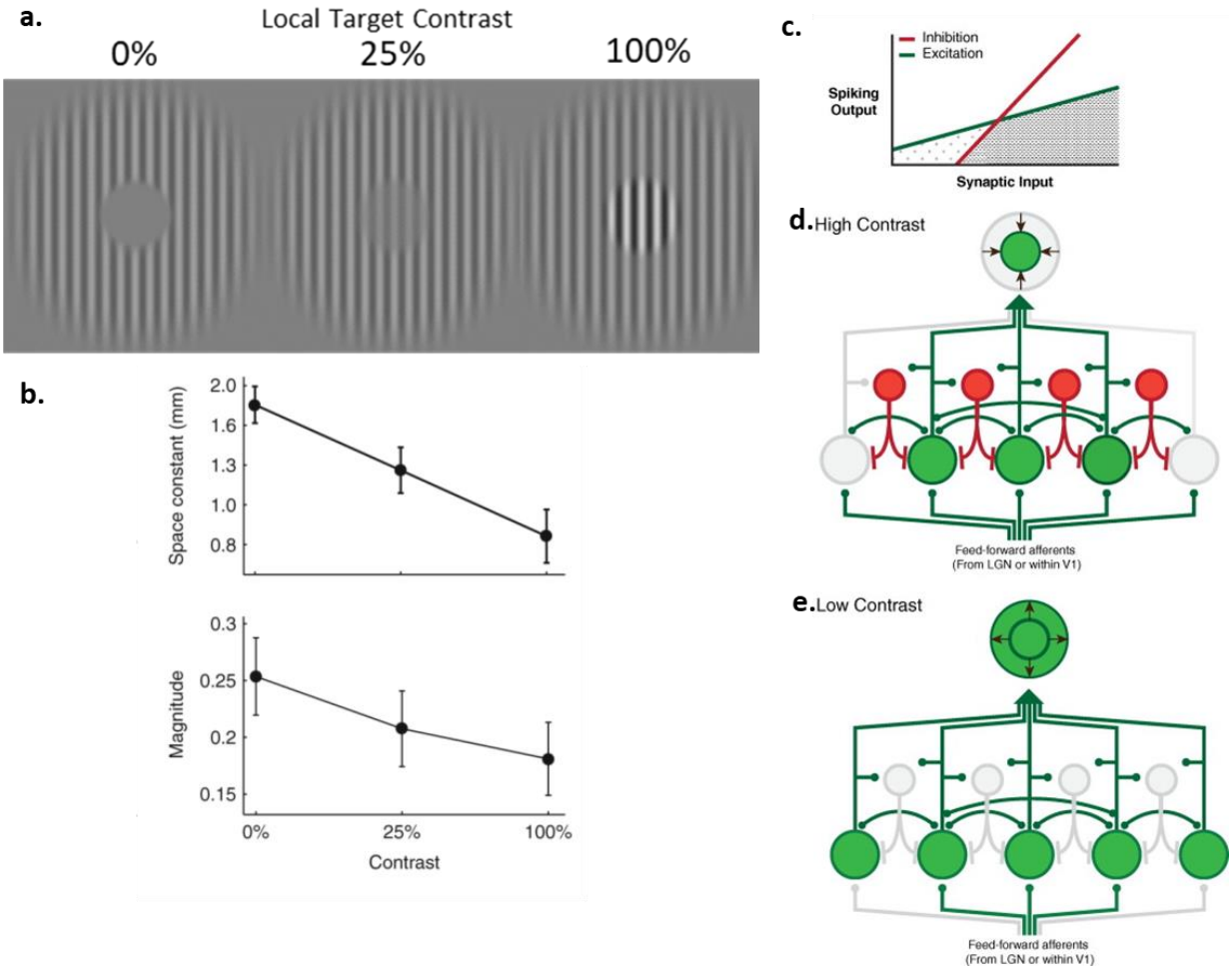

Figure 7. Local input strength dependence of contextual modulations. ***a.*** *The strength of the local input (defined as local target contrast here) can influence the quality of contextual modulations. Low contrast (0%) leads to stronger CM, whereas high contrast (100%) leads to weaker CM.* ***b.*** *Spatial extent and magnitude of local field potential (LFP) waves decrease as a function of local contrast. Figure adapted from Nauhaus (2009).* ***c.*** *The spiking behavior of inhibitory and excitatory neurons in LGN and V1 as a function of synaptic input strength. The excitatory neurons are active even with zero synaptic input, but their spiking output has a gentle slope. On the other hand, inhibitory neurons require a certain level of synaptic input to be activated, but once activated their spiking output raises with a steep slope. This behavior is demonstrated in* ***d.*** *where a high contrast input activates the inhibitory neurons, inhibiting the horizontal connections between the feed-forward afferents, whereas in* ***e.****, with a low contrast input, the synaptic input never reaches the minimum necessary strength to activate the inhibitory neurons. The horizontal connections are retained, resulting in stronger influence from the surround. Figures c, d, e from Krause & Pack (2014).*

Such dependence of CM on local input strength also seems to govern face holistic processing. Inspired by this primary principle, Goffaux (2012) addressed the influence of local input strength on holistic face processing. They used a congruency paradigm and parametrically manipulated the discriminability of target features (specifically, eyes) (e.g. 25%, 50%, 75% difference through

morphing between two features), thus extending the complete design of the composite task where the targets could only be the same (0% different) or fully different (100% different). By using this version of the composite congruency task, they demonstrated that the engagement of holistic mechanisms as indexed by the magnitude of the congruency effect diminishes as a function of local feature discriminability, mimicking the local input strength dependent contextual responses of the neurons and their horizontal connections in the primary visual cortex. This finding demonstrates that indeed holistic processing is not always fully engaged with upright faces, but its engagement depends on the local feature properties akin to CM observed at primary stages of visual processing. This functional similarity indicates that holistic processing may be intrinsically linked to (and possibly constrained by) primary contextual mechanisms.

V1 anatomy and scaling principles

The size of the primary visual cortex (V1) varies dramatically across healthy young adults (up to 3.5-fold; Filimonoff, 1933; Stensaas et al., 1974; Andrews et al., 1997; Duncan & Boynton, 2003; Dougherty et al., 2003; Benson et al., 2022), and its size is genetically heritable (Bakken et al., 2012). Individual variations in V1 cortical size has been shown to account for individual differences in visual perception, and conscious experience. Due to its retinotopic nature, as V1 becomes larger, the neurons need to spread their axons farther away to maintain the same extent of visual field coverage. However, such spread brings disadvantages as with larger spread, electrical signal gets attenuated. Achieving the same signal quality becomes challenging and costly because doubling the dendritic arborization would necessitate a four-fold increase in dendrite thickness, which is biophysically unfeasible. Consequently, the size of neurons cannot scale proportionally to the size of brain areas, resulting in an imbalance where the increase in the extent of dendrite arborization is always comparatively smaller as brain size increases (Ringo, 1991; Kaas, 2000). The perceptual implications of this imbalanced scaling principle manifest as follows: In a larger V1, the lateral spread of dendrites covers a proportionately reduced area in retinotopic space when compared to a smaller V1 (see Figure 8.). This leads to weaker lateral interactions and, accordingly, diminishes the impact of surrounding stimuli on the perception of a local target stimulus.

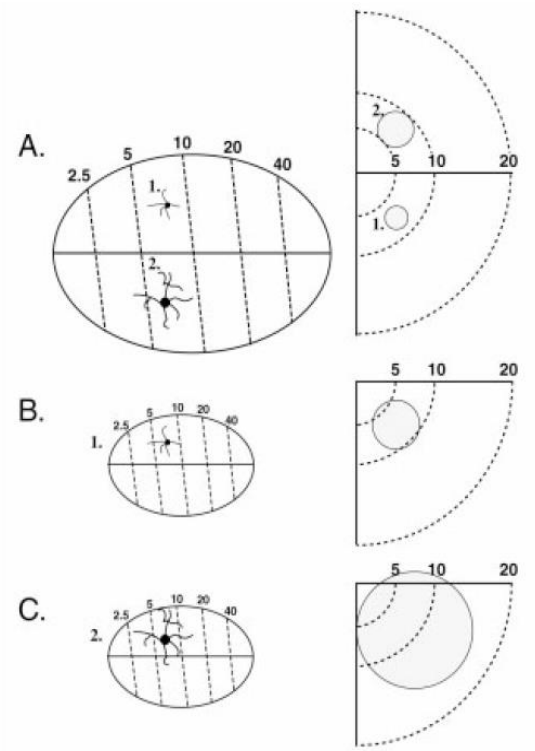

Figure 8. Differing sizes of V1s, neurons, and their visual field coverage. ***A.*** *In a large V1, a neuron of size 1 covers roughly 2-3 degrees of visual angle (deg/va), while a neuron of size 2 covers ~5 deg/va.* ***B.*** *The visual field coverage of a neuron of size 1 is larger in a smaller V1, ~5 deg/va, and* ***C.*** *even bigger with a neuron of size 2, ~12 deg/va. Due to the biophysical limitations of neurons, in a large V1, the same neuron cannot arborize its dendrites to such a big area and this leads to better local processing and weaker contextual modulations. A smaller V1 allows for a larger coverage over the visual field, therefore leading to stronger contextual modulations. Figure from Kaas, (2000).*

V1 size predicts the magnitude of contextual modulations

The magnitude of visual contextual modulations show considerable variability across individuals (Coren & Porac, 1987; DeGutis et al., 2013; Cretenoud, Francis, et al., 2020). The observed variability is not attributed to measurement error, as observers exhibit consistency within their settings and reliably reproduce a similar magnitude of illusion across repeated measurements. The source of these variations were attempted to be explained by various factors, such as age and schooling (Wagner, 1977), spatial-cognitive abilities (Coren & Porac, 1987) or sex (Grzeczkowski et al., 2017; Shaqiri et al., 2018), but perhaps the most tangible and striking factor explaining their magnitude is the functionally defined size of the primary visual cortex, V1.

Schwarzkopf et al. (2011) tested the impact of variations in the cortical size of V1 on two size illusions: Ponzo and Ebbinghaus. In these illusions, the apparent size of a target stimulus is perceived to be different than its physical size due to the context it is presented.. They found an inverse correlation between the magnitude of these two shape illusions and the functionally defined size of V1. A two-fold increase in the size of V1 resulted in a three-fold reduction in the magnitude of Ebbinghaus illusion and a four-fold reduction in Ponzo illusion.

In the Ebbinghaus illusion, two central circles with identical physical dimensions can be perceived as larger or smaller relative to one another, solely based on the size of the surrounding circles. Translating this to the realm of face contextual modulations, we observe a similar principle at play. In this context, specific facial features, like the eyes, can be perceived differently depending on the context of the surrounding facial elements. Just as the Ebbinghaus illusion demonstrates the

significant impact of contextual factors on size perception, it's conceivable that the size of V1, a critical region in the visual processing hierarchy, may exert analogous influences on the contextual modulations associated with faces as documented using various paradigms such as the part-whole effect (Tanaka & Farah, 1993) or composite effect (Young et al., 1987). These contextual modulations pertain to the processing of specific facial features (e.g., eyes) which are affected by the context they are presented in (e.g., the other elements within the face besides the eyes). Considering that V1 plays a pivotal role in early visual processing and the integration of visual information, its size might indeed be a key factor influencing the way the visual system processes facial features within their surrounding context. Such parallels between the Ebbinghaus illusion and face contextual modulations makes further research and exploration into the mechanisms that underlie these high-level aspects in visual perception essential.

Thesis Outline

Several lines of research, by means of developmental, neuroimaging, and psychophysical methods, demonstrated evidence both for and against face-specificity, and lead to a debate with no clear conclusion. Particularly, holistic processing has been an important topic of discussion in this debate: it seems to be a specific mechanism for faces. While the existence of a dedicated face processing module is conceivable, the way holistic processing is influenced by low-level properties raises questions about its "informational encapsulation." An informationally encapsulated module is expected to function independently of constraints imposed by other mechanisms in the system. This presumption instigates additional investigation into its roots within the lower levels of the visual processing hierarchy. Contextual modulations seem to be ubiquitous in sensory processing. Holistic processing is another contextual mechanism that involves the integration of spatial context into the perception of local target. Furthermore, the engagement of holistic processing depends on the strength of the local input, analogous to the primary contextual mechanisms. Considering the similar working principles of low-level and face-specialized contextual mechanisms, especially their shared dependence to local input strength, their conjoint investigation becomes a necessity to understand face holistic processing. This will further enrich the face-specificity debate, and provide further understanding of the underlying mechanisms.

By studying low-level and high-level mechanisms in isolation, most studies in sensation, perception, and cognition have done little to reveal the relationships between the low-level and high-level mechanisms, nor they have provided much insight into which mechanisms are more important for predicting performance in complex real-world tasks. In an influential paper, Geisler & Chou (1995) demonstrated that by simultaneously studying low- and high-level processes, the inter-individual variability shared between tasks can be used to explain the contributions of low-level mechanisms on the high-level process. In order to elucidate the mechanisms underlying face CM, it is important to investigate them in conjunction with CM from the primary, low-level stages of visual processing.

Through the joint behavioral investigation of low- and high-level mechanisms, Chapter 2 investigates the contribution of low-level contextual mechanisms on the perception of a more naturalistic stimuli, faces, and face contextual modulations. Contextual modulations will be examined through the lens of the shared local input strength dependence at low- and high levels of visual processing hierarchy. Inversion was found to weaken both the magnitude of face contextual modulations and its relationship to local input strength. Furthermore, it was shown to cause activation more similar to other (non-face) object categories in the brain. Thus, inversion provides an avenue to study the susceptibility of contextual modulations to local input strength in high-level, yet non-specialized processing stages. Chapter 3 proposes a registered report that seeks the refine the design presented in Chapter 2. The refinements have been implemented with the objective of eliminating any potential sources of noise stemming from methodological differences between the two tasks employed in Chapter 2. Chapter 4 will further extend this investigation by employing fMRI to measure the functionally defined surface area of V1, and its impact on low- to high-level, face-specialized and non-specialized CM. Ultimately, in Chapter 5, the findings will be discussed collectively to gain a comprehensive understanding of the interplay between low- and high-level mechanisms in perception.

Chapter 2

Contributions of low- and high-level contextual mechanisms to human face perception

Mehmet Umut Canoluk[1*], Pieter Moors[2], Valerie Goffaux[1,3,4]

1. Research Institute for Psychological Science (IPSY), UCLouvain, Louvain-la-Neuve, Belgium,

2. Department of Brain and Cognition, Laboratory of Experimental Psychology, KU Leuven, Leuven, Belgium,

3. Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, the Netherlands,

4. Institute of Neuroscience (IoNS), UCLouvain, Louvain-la-Neuve, Belgium

* Corresponding Author

E-mail: umut.canoluk@uclouvain.be

Corresponding publication: *Canoluk, M.U., Moors, P., Goffaux, V. (2023). Contributions of low- and high-level contextual mechanisms to human face perception. PLoS ONE 18(5): e0285255. https://doi.org/10.1371/journal.pone.0285255*

Author Note

The data and study materials are available at https://osf.io/tnfmg/. This study was not preregistered.

Abstract

Contextual modulations at primary stages of visual processing depend on the strength of local input. Contextual modulations at high-level stages of (face) processing show a similar dependence to local input strength. Namely, the discriminability of a facial feature determines the amount of influence of the face context on that feature. How high-level contextual modulations emerge from primary mechanisms is unclear due to the scarcity of empirical research systematically addressing the functional link between the two. We tested (62) young adults' ability to process local input independent of the context using contrast detection and (upright and inverted) morphed facial feature matching tasks. We first investigated contextual modulation magnitudes across tasks to address their shared variance. A second analysis focused on the profile of performance across contextual conditions. In upright eye matching and contrast detection tasks, contextual modulations only correlated at the level of their profile (averaged Fisher-Z transformed r = 1.18, $BF_{10} > 100$), but not magnitude ($r = .15$, $BF_{10} = .61$), suggesting the functional independence but similar working principles of the mechanisms involved. Both the profile (averaged Fisher-Z transformed $r = .32$, $BF_{10} = 9.7$) and magnitude ($r = .28$, $BF_{10} = 4.58$) of the contextual modulations correlated between inverted eye matching and contrast detection tasks. Our results suggest that non-face-specialized high-level contextual mechanisms (inverted faces) work in connection to primary contextual mechanisms, but that the engagement of face-specialized mechanisms for upright faces obscures this connection. Such combined study of low- and high-level contextual modulations sheds new light on the functional relationship between different levels of the visual processing hierarchy, and thus on its functional organization.

KEYWORDS

Contextual modulation, face perception, holistic processing, individual differences

Introduction

Evolution and lifespan experience equip biological systems with the sensory means to interact efficiently with their natural environment (Olshausen & Field, 2004; Vinje & Gallant, 2000). One of the most crucial challenges for humans is to process the dense array of visual information transmitted by the faces of their conspecifics (e.g., familiarity, individual identity, emotional expression, speech movements, physical health; Jack & Schyns, 2017; García & Ibáñez, 2023). A hallmark of human face perception resides in the robust influence of context on the processing of the local facial features (e.g., eyes, nose, or mouth). For example, humans discriminate the local features more accurately when embedded in a face than presented in isolation or in an atypical (e.g., scrambled) face context (Sergent, 1984; Young et al., 1987; Bruce, 1988; Tanaka & Farah, 1993; Davidoff & Donnelly, 1990; Harris & Aguirre, 2008). In the composite face and congruency paradigms, the observer's sensitivity (in terms of accuracy, d', and/or reaction times) to a local feature decreases if it is artificially inserted into a face context of a different similarity status (in case the observer performs a discrimination task), or of a different identity (in the case of a recognition task) from the local target (Young et al., 1987; Hole, 1994; Farah et al., 1998; Goffaux, 2009, 2012; Anaki et al., 2011; Goffaux et al., 2013; Richler et al., 2008; Rossion, 2013; Richler & Gauthier, 2014).

Susceptibility to contextual modulation is a fundamental functional property of human face processing; it is thought to enable the fast and efficient processing of the whole stimulus for the accurate identification of faces despite their high visual similarity (e.g., Diamond & Carey, 1986; Leder & Bruce, 2000; McKone et al., 2007; Richler et al., 2011; DeGutis et al., 2013). Surround influences on local feature processing are thought to be part of what makes face processing special, as they are systematically stronger for faces than for other visual categories (e.g., Cassia et al., 2009; Robbins & McKone, 2007). Consistent with this notion of specialization, the mechanisms underlying face contextual modulations significantly recede when the face stimulus is turned upside down. The selective vulnerability of face contextual modulations to a reversible image manipulation such as inversion (Sergent, 1984; Young et al., 1987; Farah et al., 1995; Rossion, 2008, 2009; Goffaux et al., 2009; Van Belle et al., 2015) suggests a functional locus in the high-level visual cortex tuned to the natural statistics of the upright face (i.e. in the fusiform gyrus: Andrews et al., 2010; Goffaux et al., 2013; Kamps et al., 2019; Schiltz et al., 2010; Schiltz & Rossion, 2006; see also Todorov, 2012; Puce et al., 1995; de Gelder et al., 2003; Pinsk et al., 2009; Iidaka, 2014; Schobert et al., 2018; Zachariou et al., 2016 for the contribution of amygdala, parietal, and superior temporal cortex to face processing). As a result, research on contextual modulations for faces has developed as a relatively independent field with its own terminology ('holistic' or 'configural' processing) and paradigms (cf. Burton, 2013).

Yet, while faces trigger particularly strong contextual modulations, these are observed for most types of stimuli, across the sensory systems of many species, and have been documented using a wide

variety of techniques (from cellular electrophysiological recordings (Levitt & Lund, 1997; Sceniak et al., 1999; Angelucci et al., 2002; Shushruth et al., 2009; Nauhaus et al., 2009) to psychophysics (Polat & Sagi, 1994; Xing & Heeger, 2000, 2001; Petrov & McKee, 2006)). Contextual modulations are observed from the earliest levels of sensory processing, in retinal ganglion cells, the lateral geniculate nuclei (McIlwain, 1964; Levick et al., 1972; Nolt et al., 2004; Solomon, 2006), as well as in the primary visual cortex (V1; Dobbins et al., 1987; Hansen & Neumann, 2008; Nauhaus et al., 2009, 2009; Sceniak et al., 1999; Versavel et al., 1990), extrastriate cortex (V2 (Hubel & Wiesel, 1965), V4 (Cadieu et al., 2007), MT (Krause & Pack, 2014), and inferotemporal cortex (Brincat & Connor, 2004; Desimone, 1991).

Contextual modulations are proposed to be critical for the representation of contours, corners, local curvature as well as for figure-ground segmentation (Fitzpatrick, 2000; Gilbert & Wiesel, 1990; Knierim & van Essen, 1992; Krieger & Zetzsche, 1996; Sillito et al., 1995; Wilson & Richards, 1992; Vinje & Gallant, 2002). Vinje & Gallant (2000; 2002) showed that contextual modulations increase the sparseness of neural responses (i.e., a sparse neural response is when a few narrowly-tuned neurons are active at any moment; see also Allman et al., 1985; Atick & Redlich, 1990; Dong & Atick, 1995). By making the neural code sparser, contextual modulations are thought to improve metabolic efficiency, feature selectivity as well as information transmission in the visual system, for an optimal coding of sensory information (Angelucci et al., 2017).

Given the functional importance and pervasiveness of contextual modulations in vision, it seems unlikely that contextual modulations for upright faces are exclusively mediated by high-level category-specialized mechanisms. A first indication comes from the observation that contextual modulations are not mandatory for upright faces; indeed their engagement depends on so-called basic image properties such as spatial frequency, and orientation content (Biederman & Kalocsai, 1997; Collin et al., 2004; Goffaux et al., 2003, 2005; Goffaux & Rossion, 2006; Goffaux, 2009; Goffaux & Dakin, 2010; but see Cheung et al., 2008). The strength of contextual influence in low-level stimuli is similarly influenced by spatial frequency, orientation content and contrast (Angelucci et al., 2017). Another indication that contextual mechanisms for basic and face stimuli may be tightly linked resides in their common dependence on the strength of the local input. At low-level stages of visual processing, contextual modulations are most extensive when local visual input is weak (due to e.g., low contrast or ambiguity; evidence in retinal ganglion cells, the Lateral Geniculate Nuclei: Nolt et al., 2004; primary visual cortex: Nauhaus et al., 2009; Sceniak et al., 1999; Sengpiel et al., 1997, area MT: Huang et al., 2007; Pack et al., 2005). A similar phenomenon rules the contextual modulations for faces. Goffaux (2012) demonstrated that, when human observers discriminate a local face feature, the strength of contextual modulations decreases as a function of local feature dissimilarity. In other words, the more dissimilar the feature characteristics, the less influence the context has on the processing of a given feature (Busigny et al., 2012; Goffaux et al., 2013; McKone & Yovel, 2009;

Peters et al., 2018; see also Cabeza & Kato, 2000; Carbon & Leder, 2005; Farah et al., 1998; Foster et al., 2021; Yovel & Duchaine, 2006 for indirect evidence).

These results suggest that contextual modulations at the general low-level and specialized high-level stages of processing are governed by the same principle: when the local sensory signals carried by a given stimulus are weak, the visual system integrates information across a larger portion of visual space. The evidence for this shared principle comes from studies that focus on a given processing level (e.g., only low, or high; Levitt & Lund, 1997; Sceniak et al., 1999; Nauhaus et al., 2009; Goffaux, 2012; Angelucci et al., 2017), and the functional link of contextual modulations at low- and high-levels of visual processing level has yet to be determined. We take this functional commonality as potentially reflecting the influence of general low-level mechanisms on holistic face processing, and address it systematically across levels using an interindividual differences approach. It is important to note that our aim here is not to question the specialization of contextual mechanisms for the processing of the upright face; we cast no doubt on the rich and consistent evidence that this category recruits special contextual mechanisms at high-level stages of processing. Rather, our aim is to examine the extent of this specialization, namely the inter-individual variability in face contextual modulations (i.e., holistic/interactive processing abilities; Russell et al., 2009; Wilmer, 2017; Yovel et al., 2014) that is explained by general low-level mechanisms (see Geisler & Chou, 1995). An alternative explanation to their common susceptibility to local input strength is that low- and high-level contextual modulations involve comparable but independent principles (see Krause & Pack, 2014).

The present study addresses these questions by testing the same group of participants with experimental paradigms tapping into the susceptibility of contextual modulations to local input strength in a basic and a complex face task. In the basic, contrast detection task, participants detected the presence of a grating; in the complex face task, they discriminated a local facial feature (i.e., the eye region) in pairs of faces. We parametrically manipulated the strength of the local input by modulating the target grating's contrast in the contrast detection task and the amount of physical difference in the eye region using morphing in the eye matching task (as in Goffaux, 2012). The influence of context was investigated by presenting the local target (grating or eye region) in isolation, in same context (isotropic contextual grating or two identical face contexts) or in different context(s) (orthogonal contextual grating or two different face contexts). The contextual modulations in the eye matching task were investigated using upright and inverted faces. Inverted face stimuli are as complex as upright faces but recruit less specialized high-level mechanisms (e.g. Farah et al., 1995; Goffaux, 2009; Taubert et al., 2015). Inversion was found to weaken both the magnitude of face contextual modulations (e.g., Farah et al., 1995; McKone & Yovel, 2009; Rossion & Boremanse, 2008) and its relationship to local input strength (Goffaux, 2012). Therefore, face inversion allows studying the

susceptibility of contextual modulations to local input strength at high-level but non-specialized processing stages.

Across tasks, we evaluated the influence of context on a similar metric, namely the target signal strength necessary to reach threshold performance and analyzed the relationship of contextual modulation magnitudes across tasks in order to quantify the functional link between low- and high-level contextual mechanisms. By analyzing the relative *profile* of input strength thresholds across context conditions in each task, we further investigated potential qualitative similarities in the way contextual modulations emerge at low- and high-level (non-)specialized stages of visual processing. The (absence of a) correlation in contextual modulation magnitudes would suggest functional (in)dependence of these mechanisms, whereas a(n absence of) correlation in profiles would indicate that low- and high-level contextual mechanisms are governed by similar (independent) principles. A significant correlation in both the magnitudes and profiles would indicate that low-level mechanisms contribute to the high-level contextual processing.

Methods

Participants

Sixty-two participants (psychology students) from UCLouvain, Belgium (50 females, mean age 21.8 ± 3.89 years, 6 left-handed) participated in the experiment in exchange for monetary reimbursement or class credits. Handedness was measured using Edinburgh Handedness Inventory (Oldfield, 1971). Prior to the experiment, the Freiburg Visual Acuity Test (FrACT; Bach, 1996) was conducted for each participant to ensure their visual acuity was within the normal range. All participants had normal or corrected-to-normal vision (mean acuity of 1.6 ± .27 dec, mean logMAR = -.19 ±.09). Additionally, facial recognition abilities were tested using a computerized version of the Benton Facial Recognition Test (BFRT; Benton & Van Allen, 1968; Rossion & Michel, 2018). All participants performed above or at the borderline of a normal score of 39/54 (Duchaine & Weidenfeld, 2003) (mean score: 45/54 ± 3.44). The experimental protocol adhered to the Declaration of Helsinki and was approved by the ethical committee of the UCLouvain (approval number: 2016/13SEP/393). Prior to participating in the study, all participants provided informed consent. Three participants could not reach the required performance level in the eye matching task's practice session (65% accuracy) and were removed from the experiment. Remaining 59 participants (47 females, mean age 21.83 ± 3.85 years, 6 left-handed) did both tasks.

Stimuli

Participants performed both tasks seated in a dark room (whose walls are painted black to ensure the lux/cd ^ m2 are lowest when the doors are closed) with their head position restrained by a chinrest, at a distance of 57 cm from the computer monitor. Prior to starting, all participants were dark adapted for at least 30 seconds. We used a 27" Dell LCD monitor with a spatial resolution of 1920 x 1080

pixels, a temporal resolution of 60 Hz, and a maximum luminance of 100 cd/m^2. Monitor luminance levels were linearized using a ColorCAL MKII colorimeter **(**Cambridge Research Systems, Kent, UK). Luminance levels, gamma correction and the temporal resolution of the monitor were periodically controlled to ensure the same values were kept constant throughout the experiment. Both tasks were run on PsychoPy (2020.2.4post1; Peirce, 2007) and Python 3.8.

In the eye matching task, we presented pairs of stimuli (either full face or isolated eye regions). Models on the photographs were young adult (psychology) students and alumni (aged 18–25 years) of the UCLouvain, who gave written consent for the use of their image (Ramon, Dricot, et al., 2010). In a given pair, the dissimilarity of the local face region relevant to the task (i.e., eyes and brows) was manipulated parametrically using morphing (similar to Goffaux, 2012, see below for a detailed description).

In order to generate the different levels of feature dissimilarity, morphs were created using 32 full-frontal grayscale pictures (on a scale from 0 to 1) of unfamiliar white neutral faces from each gender without glasses, facial hair, or makeup. We morphed 16 pairs of same-gender faces (eight pairs for each gender). Prior to morphing, the luminance and root-mean-square (RMS) contrast of all face images were set to a mean of .52 and .12, respectively. Next, we ensured that the eye regions of morphed faces were comparable in terms of inter-pupillary distance and physical similarity and warranted a smooth morphing within pairs by measuring inter-pupillary distance in each face image, and pairing the faces with comparable values (average pixel difference = 1.32 ± 1.31 pixels). After this initial pairing, pixelwise similarity within the eye region was calculated within each pair using *corr2* function. The pixelwise similarity of the target eye region was .8 on average (SD = .014). For each gender, the eight pairs lying within two standard deviations (Howell, 2013) around the mean pixelwise similarity distribution were used for morphing; the remaining eight pairs were not morphed and used as "face contexts". An additional set of 12 female faces (six pairs of female faces) which had undergone the same procedure were used for practice trials. These image manipulations and calculations to select optimal pairings were performed using MATLAB (Natick, MA).

We created morphing continua between the two faces of all pre-selected pairs using Morpheus Photo Morpher 3.17 (Morpheus Development, LLC, 2020). For each face, ~110 control points were used for morphing (contour: ~30 points, mouth: ~15 points, eyes and eyebrows: ~25 points each, and nose: ~15 points; Goffaux, 2012; Jacques & Rossion, 2006). In each continuum, we selected faces at eight morphing levels: 90% of Face A (10% of Face B), 77%A (23%B), 68%A (32%B), 62%A (38%B), 38%A (62%B), 32%A (68%B), 24%A (76%B), 10%A (90%B). These levels were selected specifically so that for each non-zero dissimilarity level in the task (24%, 36%, 53% and 80%), all pairs crossed the midpoint (50%) of the morphing continuum. We extracted the target region (eyes and brows) of each selected morphed pair, and blended it onto two unmorphed "face contexts" whose

eye areas were *a priori* replaced with a smooth skin texture (using Adobe Photoshop CS., Berkeley, CA; see **Fig 1a**). Since all stimuli were artificially generated, all were subject to same potential amount of artifacts across conditions.

Each face subtended a visual angle of ~7° in height, and ~5.2° in width. This size corresponds to the size of a human face seen at an average social distance (a human face is approximately of 18 cm in height and social distance is of 1.5 m approximately, e.g. McKone (2009)). The target eye region subtended a visual angle of ~1.4° in height, and ~4.1° in width. Center of the face image was 1.18° below the screen center in Upright trials, and 1.18° above in Inverted trials to ensure that the target eye region aligned with fixation. Stimuli were shown in their canonical Upright (0° rotation) or Inverted (180° rotation) position.

We additionally created noise masks by iteratively phase-scrambling each face image in Fourier space (in 500 iterations; see Jacobs et al., 2020; Petras et al., 2019, 2021). Iterative scrambling is used to prevent noise masks from being contaminated by uniform background pixels, and thus ensures that the spectral properties of the stimulus and mask are similar.

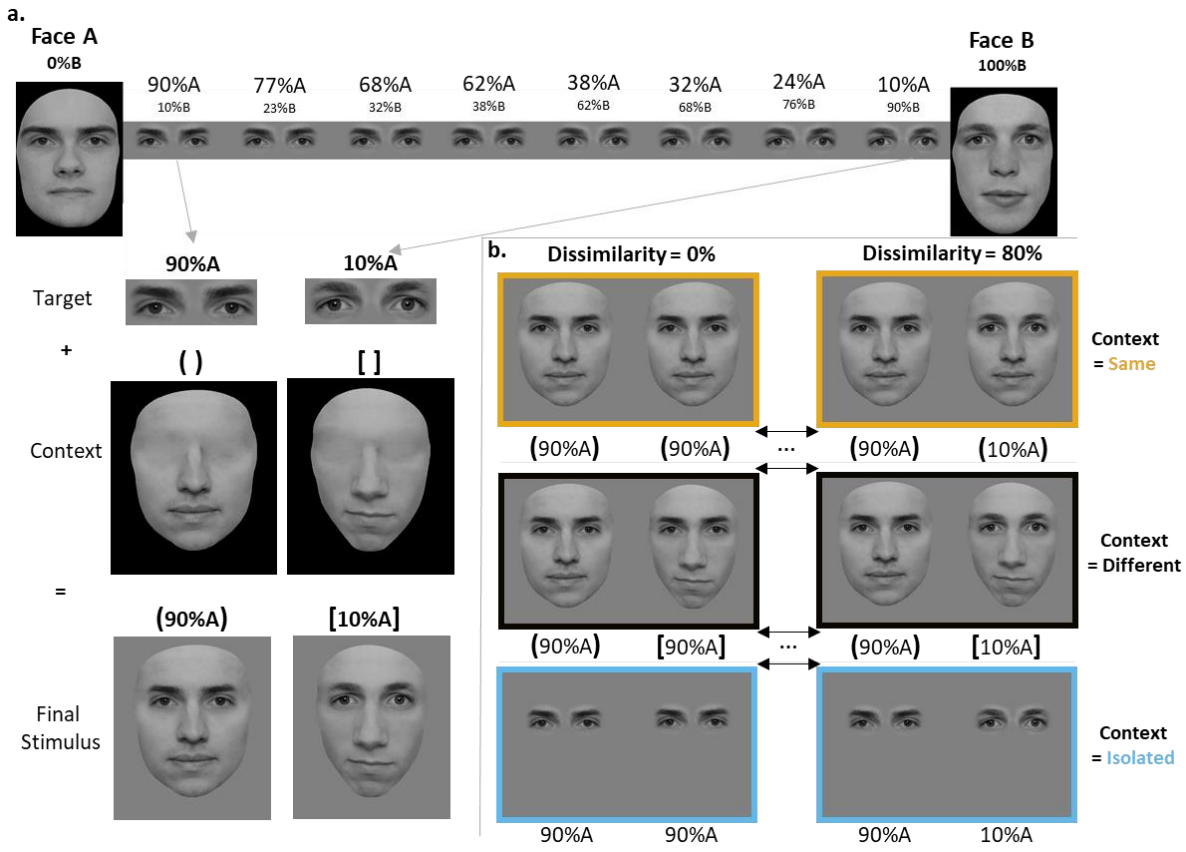

Figure 1. **Eye matching task stimuli and conditions. a.** Morphing procedure and stimulus creation. Two morphed eye regions are selected from the morphing continuum and pasted on two distinct face contexts. **b.** Same, Different and Isolated context conditions. Bracket shape (round, square) defines

distinct face contexts, percentage values define morphing percentage in the eye region. Target regions within a pair had 0%, 24%, 36%, 53% or 80% dissimilarity. It can be seen here that when the target regions are the same (0% dissimilarity) and context differs, the contextual modulations are stronger compared to when target regions differ largely (80% dissimilarity).

The stimuli and procedure of the contrast detection task were adapted from Mannion et al. (2017). Stimuli consisted of four circular target regions (3° of visual angle each) in a circular context (diameter of 20°) (see **Fig 2**). Both the context and the target had a spatial frequency of 1 cycle per degree. The context was made up of a horizontally or vertically oriented grating with a contrast of 25%. Target regions were positioned at 5° in eccentricity, one in each quadrant of the visual field (at 45°, 135°, 225°, 315° from the horizontal axis).

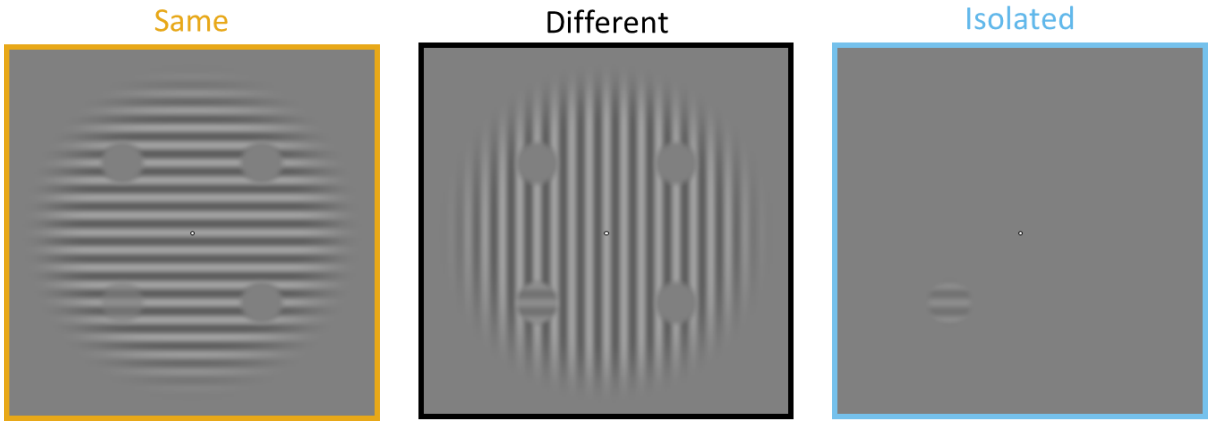

Figure 2. **Same, Different, and Isolated context conditions of the contrast detection task.** Participants were asked to locate the horizontal grating appearing in one of the four target regions. It can be seen here that although target contrast is identical in all three examples, target appears to be less visible in Same condition compared to other two conditions.

On each trial, a horizontal grating with the same spatial frequency and phase as the context and with a variable, non-zero contrast was presented in one of the four target regions. The contrast of both the context and target regions ramped smoothly to zero at the edges with the use of a raised cosine filter. Additionally, a small fixation dot (diameter of 0.25°) was always present at the center of the screen during the task. The contrast of the target grating was determined using a Psi adaptive staircase procedure (Kontsevich & Tyler, 1999; for details, see Mannion et al., 2017), which used a pool of 350 logarithmically spaced contrast values ranging from 0.001 to 1. There were two separate, interleaved staircases for each context condition. This interleaving procedure ensured that the algorithm did not settle into a narrow range of contrast levels within a condition.

Procedure

The experiment consisted of three different sessions of ~45 minutes each. The eye matching task was relatively long, and was divided into two equal sessions. The participants performed the eye matching task on the first and third sessions, and the contrast detection task on the second. This task order aimed at keeping participants motivated by alternating tasks between sessions. Participants took a break of at least 10 minutes between sessions.

In the eye matching task, a pair of faces (or eye regions) was presented successively in the center of the screen. Participants had to focus only on the eye region, ignore the rest of the face (when there was one), and determine whether the two eye regions shown in succession were strictly identical or differed in some way (i.e., two alternative forced choice). On each trial, participants first saw a fixation cross on a gray background for 500 ms, followed by a 500 ms blank screen and the first face for 500ms. The phase-scrambled version of the first face was then presented for 200 ms as a mask, directly followed by the second face which stayed on the screen until the participant gave a "Same" or "Different" response by button press (see **Figure 3.**). After response, a blank screen was shown for an interval of random duration (ranging from 700 to 1200 ms) before the next trial started.

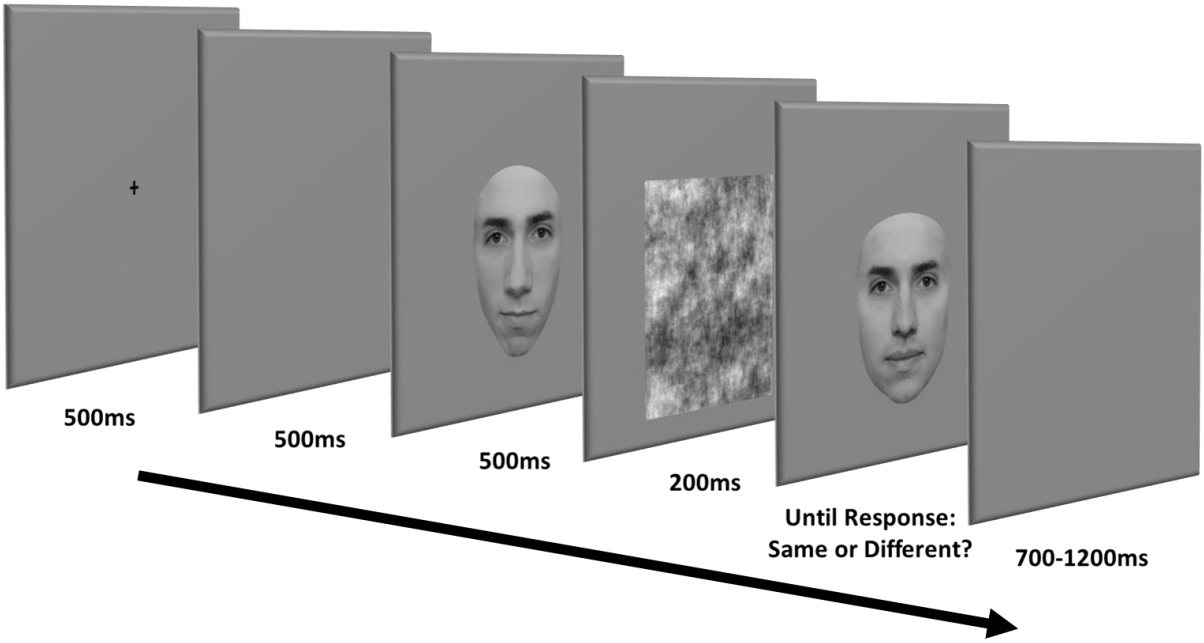

Figure 3. **Eye matching task trial sequence.** An example of 0% eye region dissimilarity and Different context condition trial. The participant sees a fixation cross for 500 ms at the beginning of each trial, followed by a blank screen. Afterwards, participant sees the first stimulus for 500 ms, then the noise mask for 200 ms, and finally, the second stimulus. The second stimulus stays on the screen until the participant indicates whether the eye regions were "Same" or "Different" between the two stimuli by button press, followed by a blank screen before the next trial starts.

The eye matching task employed the method of constant stimuli with constant morphing dissimilarity values and was designed to obtain thresholds expressed in morphing units. This task consisted of 30 conditions: two orientations (Upright, Inverted) by three context conditions (Same, Different, Isolated) by five dissimilarity levels (0%, 24%, 36%, 53%, 80%). The 24%, 36%, 53% or 80% dissimilarity trials called for ”Different” response; they were presented 16 times in each orientation and context condition (384 trials in total). In order to balance the responses, we prepared 384 “Same” response trials made out of 64 0% dissimilarity trials in each orientation and context condition. In “Same” response trials, 10%, 32%, 62% or 77% target morphs were repeated twice. The 24% dissimilarity trials comprised 38% and 62% morphs, the 36% dissimilarity trials comprised 32% and 68% morphs, the 53% dissimilarity trials comprised 24% and 77% morphs, and the 80% dissimilarity trials comprised 10% and 90% morphs. Doing so, we ascertained that half of the morphs used in the non-zero dissimilarity conditions appeared in the 0% dissimilarity condition. In total, there were 768 trials in the eye matching task.

Independent variables in this task were five dissimilarity levels (0%, 24%, 36%, 53%, 80%), three context conditions (same, different, isolated) and two orientations (upright, inverted). A trial was made up of a combination of a level of each of these independent variables. Our dependent variable was the (proportion of) “different” responses given by a participant. In 0% dissimilarity trials, each unique target (i.e., a specific eye region at a specific morph level) appeared six times. In other dissimilarity levels, each unique target appeared three times. We replicated trials across Upright and Inverted conditions.

Each experimental session of the eye matching task was divided into eight blocks of 48 trials. Within each block, there were equal number of “Same” (0% dissimilarity level, 24 trials) and “Different” (24%, 36%, 53% or 80% dissimilarity levels, six trials each) response trials equally distributed across contexts and orientation conditions, in randomized order. If a participant performed below 65% on a given block, a text appeared on a red background, suggesting them to take a longer rest, and stay attentive. This was merely a notification and did not impact the experimental design in any way. It was used to inform the participant about their performance, and to keep them attentive and active during the experiment. Instructions emphasized accuracy over speed. Participants were required to do a short practice round (four short blocks of 24 trials) before each session to familiarize with the task. If necessary, participants re-took the practice round, until they reached a minimum of 65% correct response accuracy. The practice data was not included in any further analyses, and this 65% correct response criterion was only used as a prerequisite to start the experiment.

In the contrast detection task, the target horizontal grating appeared in three distinct context conditions: in Isolation, in an iso-oriented (“Same”) context, and in an orthogonal (“Different”) context. As in Mannion et al., (2017) each context condition was presented in separate blocks. Each

block consisted of 80 trials, made up of two staircases of 40 trials. There were three block repetitions per condition, i.e. a total of 9 blocks (720 trials in total). Block order was randomized across participants, with successive blocks belonging to distinct conditions. Each trial started with a preparatory, black fixation dot of 500ms. The fixation dot then turned to white for another 50ms. Next, the white fixation dot and the target stimulus (embedded in a context or presented in isolation, depending on the condition) was presented in one of the four possible target regions for 100ms. After the stimulus presentation, the fixation dot turned to grey to signal the participant that they could give their response. The target could potentially appear in any of the four target regions as in Mannion et al. (2017), but we asked the participant to report only the side on which it appeared (left or right) so that both tasks involved a 2-AFC. As in the eye matching task, accuracy was emphasized over speed. The contrast detection task was relatively simpler, so the participants were asked only to perform a brief practice round of 10 trials before the experiment to ensure they understood the task, and the data from the practice were not included in any of the analyses.

Analyses

The purpose of the study was to investigate the functional dependence and similarity in working principles between low- and high-level contextual modulations. We first quantified the shared inter-individual variance in contextual modulation magnitude across low- and high-level mechanisms. A second analysis focused on profiles, examining whether these two mechanisms showed similar characteristics. All data were visualized using the *ggplot2* package in R (Wickham, 2016).

Across tasks, we first investigated the function relating performance to local input strength (eye region dissimilarity or grating contrast). Since the definition of local input strength varied between tasks (i.e., 5 levels of target dissimilarity (0%, 24%, 36%, 53%, 80%) in the eye matching task; 350 levels of contrast (between 0.001 and 1) in the contrast detection task), we z-transformed the local input strength values for the eye contrast task, and log10 and z-transformed them for the contrast detection task to bring the two into a comparable scale. After these transformations, local visual input strength was within the range of -3 to 3 for both tasks.

The z-scaled data from both tasks were submitted to a Bayesian generalized linear mixed model (Moscatelli et al., 2012; Yssaad-Fesselier & Knoblauch, 2006) by means of *brms* package available in R (Bürkner, 2017, 2018). We chose a Bayesian approach to determine not only the existence of a correlation and the evidence for it, but also the strength of the evidence for a null correlation when/if it occurs. Using a mixed model makes it possible to include inter-individual variability in the statistical model where the estimates for the individuals are influenced by other individuals (due to the hierarchical nature of the model), yielding better estimates and conferring more statistical power to the analysis compared to other approaches (see Moors et al., 2020 for more details).

The dependent variable in the contrast detection task was accuracy. Participants could give two responses: left or right. In such a 2-AFC detection task, the guess rate (i.e., lower asymptote of the psychometric function) is set at 0.5. The contrast value at .75 was taken as contrast detection "threshold". In the 2-AFC eye matching task, participants could give a "Same" response or "Different" response. Here, the dependent variable was the proportion of "Different" responses instead of accuracy. The reason for that was that in this task, the outcome of interest was not how accurately a participant was telling whether the eyes were "Same" or "Different", but how the 50% threshold (the tipping point between predominantly "different" and "same" responses) was affected in response to varying contexts. The threshold in this task would indicate the amount of target feature dissimilarity (i.e., morphing difference) necessary to reach a rate of 50% "Different" responses. In this case, the guess rate was set at 0 and the threshold was measured at .5 (see **Fig 4**).

The eye matching task consisted of two different orientation conditions: Upright and Inverted. We treated these two orientation conditions separately. With this adjustment, the factor, "task" has three different levels: "Upright Face", "Inverted Face", "Grating".

We fitted a logit link function to model the sigmoidal pattern of the psychometric functions in each task (**Equation 1**). We put a normal distribution prior on all regression weights, and a lognormal distribution prior on the parameters capturing the random effects. All random effects were included, yet are assumed to be uncorrelated to each other (by means of || term in the formula) to increase sampling efficiency and ensure model convergence (Bürkner, 2018). Doing so, we allow a participant to deviate from the average, but we do not assume the size of these deviations to correlate across conditions (Kurz, 2019). Among several models tested, the model formula with the best fit was the one with complete interaction between all variables, and with random effect structure integrated (see Supplementary file 5 for the details of model selection procedure):

$$dependent\ variable \sim .5 * guess + (1 - .5 * guess) * inv\_logit(eta)$$

$$eta \sim local\ input\ strength * task * condition + (local\ input\ strength * task * condition\ ||\ participant)$$

Equation 1. ***brms*** **formula used to fit the sigmoidal functions.** The variable "guess" was set as "1" for the contrast detection task (0.5 guess rate) and "0" for eye matching task (0 guess rate).

All predictor variables were dummy-coded (as in Moors et al., 2020). Using these parameters and the formula, we ran four Markov chain Monte Carlo (MCMC) chains with 6,000 iterations each (3,000 of each were considered warm-up), resulting in 12,000 iterations after warmup. Default values were used for the *adapt_delta* and *max_treedepth* parameters. Running this model yielded MCMC chains that were converging according to Gelman-Rubin diagnostics such as effective sample size, the R-hat statistic, and the absence of divergent transitions.

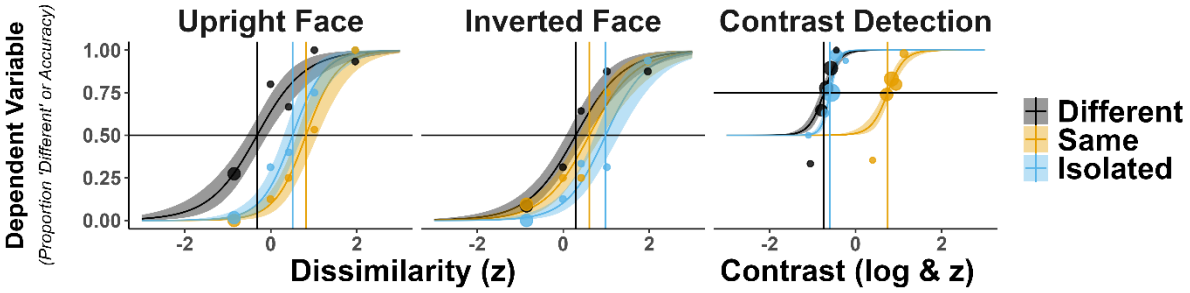

Figure 4. Data and psychometric functions of a representative participant. The dependent variable in the eye matching task is the proportion of 'Different' responses, and accuracy in the contrast detection task. Dots represent actual data (varying in size depending on frequency of presentation) and curves represent the model fits, with 95% HDI shaded ribbons. Vertical lines mark the "***threshold***" for a given condition (at .5 in the eye matching task, at .75 for the contrast detection task).

After fitting the data from all tasks, we obtained the threshold value of the logit link function for each task and context condition, and for each participant using the formula, *threshold = − (intercept / slope)* as outlined in Knoblauch & Maloney (2012). The local input strength threshold values were in terms of morphing units for the eye matching tasks, and in terms of contrast values in the contrast detection task. We compared these threshold values across tasks and contexts using the model's estimates, and each threshold's respective 95% Highest Density Interval (HDI; Kruschke & Liddell, 2018a, 2018b). We interpreted the differences across conditions by testing whether the HDI of the differences between conditions had an overlap with zero. A HDI that excludes zero indicates a "significant", or a non-negligible difference across conditions.

In order to obtain a measure of the contextual modulation magnitude in each participant, we submitted the individual threshold values in the various context conditions of each task to a two-step regression of regressions analysis (DeGutis et al., 2013). Each task comprised a control "Isolated" condition in which the target region (eyes and brows, or grating) was presented on its own. In a first regression, we regressed participants' thresholds in the "Isolated" context condition from "Same" context and "Different" context thresholds, therefore removing all variance related to the control condition from the conditions where a face context was present. In the second regression, we regressed the residuals of the "Different" context condition from the residuals of the "Same" context condition. We chose to use this approach because the residuals of this two-step regression delivers an estimate of each individual's contextual modulation magnitude that was not contaminated by the variance of the compared conditions, as opposed to simple subtraction method where the final outcome contains variance from both of the compared conditions (DeGutis et al., 2013).

We addressed whether contextual modulations for eye matching and grating contrast detection are functionally related by submitting the final product from the two-step regression to a Bayesian correlation analysis. We chose a Bayesian approach because of its ability to determine the strength of

evidence presented by the data in supporting a particular hypothesis (H0 or H1), hence, its potential to interpret a finding, even when it is null. We computed Bayes Factors (BFs) of the correlation using BayesFactor package (Morey & Rouder, 2018) in R. For Bayesian correlation analyses, a uniform default prior (i.e., a stretched beta prior distribution with a width of 1, where any parameter value over the range of correlations between -1 and 1 are equally likely) can produce Bayes factors that can be biased towards null models (Vanpaemel, 2010; Quintana & Williams, 2018), so they are generally not recommended. Therefore, we opted for a stretched beta prior distribution with a width of .33, centered at 0. Furthermore, in order to ensure the robustness of the results to different prior distributions, we ran Bayes Factor robustness analyses (see Supplementary). Combining this prior distribution with the observed data forms the posterior distribution, and the Bayes factor is the ratio between the marginal likelihoods of the null model and the alternative model, given the data. A BF_{10} value indicates the support for H1 (e.g., correlation) over H0 (e.g., no correlation) given the data (conversely, BF_{01} would indicate support for H0 over H1, and can be calculated with $1/BF_{10}$). We followed Jeffreys' guidelines (H. Jeffreys, 1998; Jarosz & Wiley, 2014) to interpret the BF values: a BF_{10} value of 1 would mean equal amount of evidence for both H0 and H1, and therefore inconclusive. A BF_{10} value between 1 and 3 is considered anecdotal evidence in favor of H1 over H0, whereas a BF_{10} between 3 and 10 indicates substantial evidence for H1. $BF_{10} > 10$, $BF_{10} > 30$ and $BF_{10} > 100$ are regarded as strong, very strong and decisive evidence for H1, respectively.

In an additional analysis, instead of regressing out threshold values stemming from the distinct context conditions to obtain a quantitative measure of contextual susceptibility, we investigated whether the modulation of thresholds across the three context conditions was qualitatively similar across tasks on an individual level. To this extent, 3-point vectors of thresholds from each individual (Isolated, Different, Same) were submitted to a Pearson correlation analysis across tasks. This two-dimensional (2D) correlation approach results in a single correlation value per comparison (Upright/Inverted, Upright/Contrast Detection, Inverted/ Contrast Detection) in each individual. These correlation values are then Fisher-Z transformed. Because transforming 1 (and -1) correlation coefficients yields extreme values, we set any correlations with a value of 1 (or -1) to .99 (or -.99) before the transformation. We obtained empirical chance levels by permuting 1000 times (Marozzi, 2007) the 3-point vectors in each task (so that the new ranks differed from initial order) and averaging the correlations of each iteration. Finally, we quantified the profile similarities by testing whether the sample of (Fisher-Z transformed) correlation coefficients differs from empirical chance level by using Bayesian one-sample t-tests (see Jacobs et al., 2020 for a similar analysis). We report the HDI (to interpret whether they overlap with the empirical chance or not) and Bayes Factors of these one-sample t-tests.

Results

In both tasks, contextual influences depended on local signal intensity, and were strongest when the local signal strength was low suggesting that similar functional principles may be at stake for both faces and gratings (see Supplementary.1; Goffaux, 2012; Levitt & Lund, 1997; Nauhaus et al., 2009; Sceniak et al., 1999). Figure 5A illustrates the z-scaled threshold values for eye matching (Upright Face, Inverted Face) and contrast detection tasks. As expected, contrast detection thresholds increased for isotropic (same) oriented grating context, compared to an orthogonal (different) oriented grating context (Mannion et al., 2017; Self et al., 2014; Sengpiel et al., 1997), and a similar mechanism was at play in the eye matching task: eye matching thresholds were higher when presented within a stable (same) face context, compared to different face contexts. As for the Isolated condition, its rank with respect to the Same and Different context conditions was similar in the contrast detection and Upright eye matching tasks, but differed in Inverted eye matching task. The threshold difference between Same and Different contexts was larger in the contrast detection task than in eye matching tasks (HDI [2.26, 2.57]). In the eye matching task, the threshold difference between Same and Different context was more pronounced for Upright Faces (HDI [1.18, 1.84]) compared to Inverted Faces (HDI [.10, .48]), corroborating past evidence that inversion attenuates contextual influences on the encoding of features.

While Same and Different context conditions rank similarly across tasks, the rank of the Isolated varied much more. In the contrast detection task, Isolated and Different thresholds were similarly low with only a mild advantage for Different context condition (HDI [-.13, -.40]), suggesting that the detection of local grating contrast works almost as effectively when the grating is inserted in an orthogonal context as when it is viewed in isolation. In the upright eye matching task, however, thresholds in the Isolated condition were closer to the Same than the Different context condition (HDI [.21, .77] and [-.67, -1.34], respectively). In the Inverted eye matching, the thresholds varied much less across context conditions than at upright, and culminated in the Isolated condition (HDI [-.12, -.30], [-.31, -.70] for Same-Isolated and Different-Isolated comparisons, respectively). This agrees with the proposal that inverted faces are processed more independently from context than upright faces. These similarities in the ranking of various context thresholds are systematically investigated later with the profile correlations below.

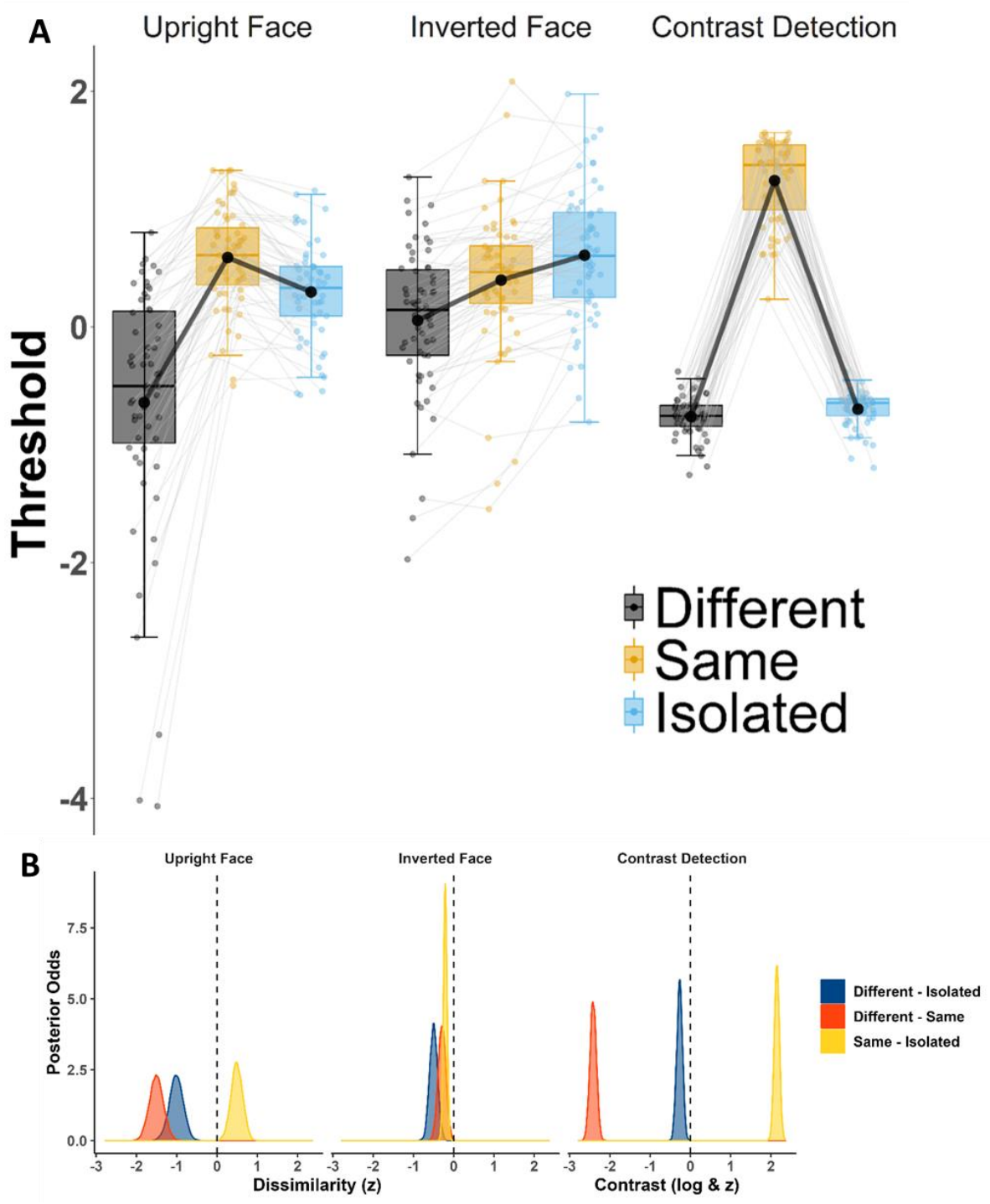

Figure 5. **A. Z-scaled threshold estimates in eye matching (left) and contrast detection (right) tasks**. Each dot represents the z-scaled threshold estimate for a given task and condition for a participant. The top, middle and bottom lines of boxes correspond to the first, second (median) and third quartiles (25th, 50th and 75th percentiles) respectively. The upper and lower whiskers extend until 1.5* Inter-Quartile Range. Faded grey lines connect the thresholds in respective conditions for a given participant. Bold black dots and lines display group means. **B. Posterior odds of differences across**

the three tasks. The density plots display the posterior odds of the threshold differences. 95% interval of the distributions did not overlap with zero, therefore all differences were non-negligible.

In order to examine a possible functional link across low- and high-level (non)specialized contextual mechanisms, we first estimated the magnitude of contextual modulation by submitting the z-scaled individual thresholds to a regression of regression analysis. The residuals resulting from this analysis were then correlated across tasks (see Methods). This magnitude correlation served as a measurement of the quantitative relationship across tasks (see **Fig 6.**).

We found anecdotal evidence for the absence of a correlation in contextual modulation magnitude between contrast detection and Upright eye matching (r = .15, BF_{10}= .61) suggesting the independence of the contextual mechanisms between low-level and high-level face-specialized processing. However, there was substantial evidence in favor of a weak correlation in contextual modulation magnitude between contrast detection and Inverted eye matching tasks (r = .28, BF_{10} = 4.58). This suggests that the contextual mechanisms involved in low-level stages can, to some extent, be tracked at high-level stages. Lastly, contextual modulation magnitude correlated strongly between Upright and Inverted eye matching, with decisive evidence in favor of a correlation (r = .61, BF_{10} > *100*). Bayes Factor Robustness Checks for these correlations are provided in the supplementary (Figure S.4).

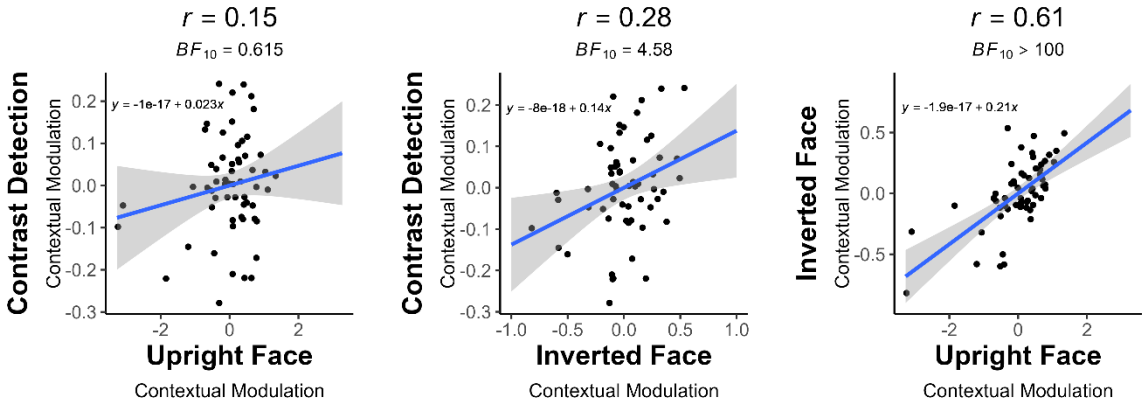

Figure 6. **Correlation of contextual modulation magnitudes across tasks.** Contextual modulation magnitude values are the product of *regression of regressions* method for each task. Blue line indicates the regression line and shaded area covers the standard error bounds.

So far, we have focused on an aggregate metric of contextual modulation magnitude based on a regression of regression of the distinct context conditions. Considering the pervasiveness of contextual modulations in the visual system hierarchy, and their functional similarity across different processing levels, we now examine the profile of threshold variations across context conditions and

investigate their qualitative similarity across tasks. As noted above, across tasks, thresholds increased when the target was presented within the Same context in comparison to Different context condition.

The profile of threshold values indicates a similarity between contrast detection and Upright face tasks, while Inverted face profile seems to be standing out. In order to investigate the profile similarity more systematically, we analyzed the 2D correlation of threshold profiles across tasks at the individual level. We compared the sample of Fisher-Z transformed correlation values obtained from each task comparison against empirical chance level (i.e., correlation coefficient obtained by shuffled permutations) using Bayesian one-sample t-tests (see Analyses section for details). These analyses revealed conclusive evidence that the threshold profile in the Upright eye matching task correlates with the threshold profiles in the contrast detection task (averaged Fisher-Z transformed r = 1.18, HDI [1.02, 1.33], BF_{10} > 100 tested against empirical chance r = .005). Given the lack of a magnitude correlation between the two tasks, these findings suggest that the functional similarity in contextual processing between upright face eye matching and contrast detection is more qualitative than quantitative. We also observed conclusive evidence that the threshold profiles in Upright eye matching task correlates with Inverted eye matching task (averaged Fisher-Z transformed r = 1.20, HDI [.92, 1.40], BF_{10} > 100 tested against empirical chance r = .003; see **Fig 7**).

Inverted eye matching and contrast detection threshold profiles showed a relatively weaker profile similarity, but still with substantial evidence in favor of a correlation (averaged Fisher-Z transformed r = .32, HDI [.12, .48], BF_{10} = 9.7 tested against empirical chance r = .002). We would like to point out the correlations between Inverted Face and Contrast Detection tasks which are relatively smaller, both in terms of averaged Fisher-Z transformed r values and Bayes' Factors, which suggests that this analysis (despite being only on the three points) is not blindly giving extremely large Bayes' factors.

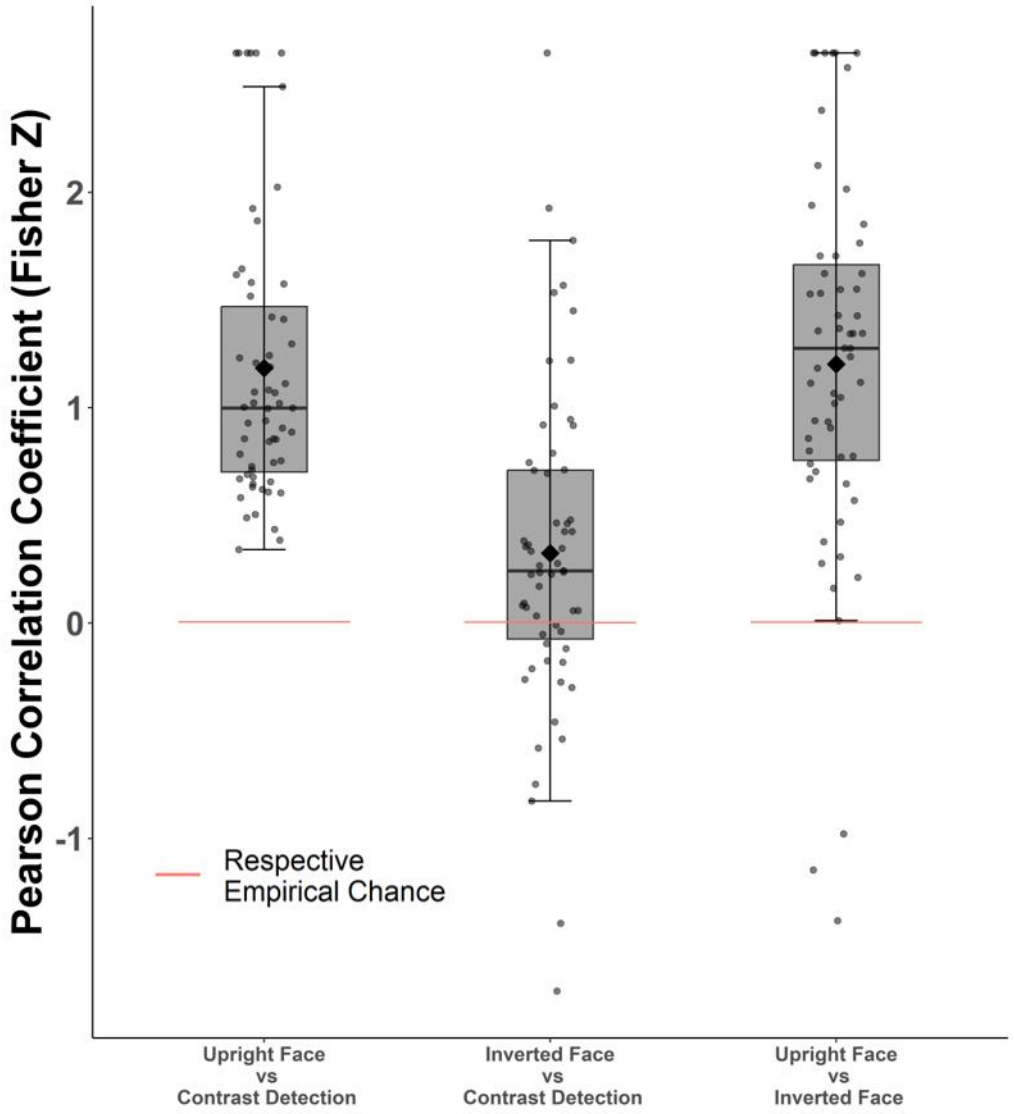

Figure 7. **Correlation of the individual threshold profiles across tasks.** Each dot represents a Fisher-Z transformed Pearson Correlation Coefficient value of an individual's threshold values from the corresponding comparison. The top, middle and bottom lines of boxes correspond to the first, second (median) and third quartiles (25^{th}, 50^{th} and 75^{th} percentiles) respectively. The upper and lower whiskers extend until 1.5* Inter-Quartile Range. Black diamonds display means.

Discussion

Contextual modulations are an integral part of visual information processing (Angelucci et al., 2017). They are ubiquitous in the visual processing hierarchy, and manifest a similar phenomenon at various levels: their impact on perception depends on the local input strength (Sceniak et al., 1999; Nauhaus et al., 2009; Goffaux, 2012). Despite extensive research endeavor to understand contextual modulations at distinct processing levels, little effort has been made to study potential relationships

across levels (see Barton et al., 2004 for a rare example). In this study, we investigated a potential link between low- and high-level contextual modulations through the lens of their dependence to local input strength. We reasoned that relying on this shared phenomenon across processing levels provides a rich metric that reflects a similar functional principle across tasks, and therefore increases the chance of observing a correlation, should there be one.

By testing the same group of participants in a low-level, contrast detection task and a high-level, eye matching task, we investigated contextual modulations separately for gratings, upright faces and inverted faces in the same individual observers. In the contrast detection task, as expected, we observed orientation-dependent contextual modulations as seen in Mannion et al. (2017). Presenting a parallel-oriented context along with the target led to an increase in detection thresholds compared to presenting an orthogonal context or no context. We made our two tasks more similar by halving the number of response choices used in Mannion et al. (2017) while keeping the four possible target locations, so the chance level differed across our and their studies (50% versus 25%). Despite this difference, we were able to replicate the results of Mannion et al. (2017; see Supplementary 1) and more particularly the strength and profile of contextual modulations were similar. In line with past evidence, inversion decreased the magnitude of face contextual modulations (e.g. Farah et al., 1995; Rhodes et al., 1993). Additionally, our findings reaffirmed that inversion weakens the dependence of contextual modulations on the discriminability of local targets (see Supplementary.1; Goffaux, 2012). Given that the inverted face is as complex as the upright face and activates large portions of high-level visual cortices (Haxby et al., 1999b, 2000; Yovel & Kanwisher, 2005; Goffaux et al., 2009, 2016; Rossion et al., 2012), this stimulus condition adds an in-between layer that gives access to high-level but non-specialized processing.

We used a single Bayesian Hierarchical Model to obtain local input strength thresholds for each task. Utilizing a hierarchical model provided a stronger statistical power than running individual psychometric functions, mainly owing to its ability to employ variability of all subjects when making estimations about a parameter (in this case, the thresholds; McElreath, 2016). Applying the *regression of regressions* method (DeGutis et al., 2013) to model the estimated thresholds across the three contexts provided us with a final metric of contextual modulation magnitude, which we compared across tasks. These quantitative analyses showed no relationship between low- and high-level face-specialized contextual modulation magnitude. Evidence for this absence of correlation was however anecdotal, suggesting that it may be due to the strong involvement of face-specialized mechanisms overriding low-level contextual mechanisms, and rendering them inconsequential, or unobservable in our behavioral measures. There was, however, substantial evidence for a weak relationship in the magnitude of low- and high-level non-face-specialized contextual modulations (i.e., between the contrast detection and inverted eye matching tasks). The fact that low-level contextual mechanisms explain a limited portion of the high-level non-specialized contextual modulations supports evidence

for a functional link between simple and complex image contextual modulations. To our knowledge, our paper is the first one to demonstrate such dependence across the two. When the face-specialized mechanisms are engaged, this seems to cause a detachment in this dependence, leading to the absence of correlation between low-level and face-specialized contextual modulations.

Despite the disparity in their relationship to low-level mechanisms, we found that 37% of the variability in upright face contextual modulation is explained by contextual effects in inverted faces. Contextual modulations are expected to correlate more strongly within than across visual processing levels. Indeed, inverted faces still weakly activate face-specialized regions such as the FFA (Mazard et al., 2006; Gilaie-Dotan et al., 2010; Goffaux et al., 2013). Yet, the large variability in upright face contextual modulation that remains to be accounted for after taking into account the variability in inverted face contextual modulations supports the functional specificity of the contextual mechanisms elicited by upright faces. This finding lends support to the existence of high-level mechanisms specialized for the natural statistics of the upright face (e.g.; Yin, 1969; Bentin et al., 1996; Robbins & McKone, 2007; Rossion, 2013).

We found no relationship between upright face eye matching and contrast detection tasks in our quantitative analyses, however, their similar susceptibility to local input strength indicated that they may still share common functional principles. In an additional analysis, we examined the correlation of the profile of thresholds in Same, Different and Isolated conditions (i.e., qualitative relationship) across tasks. Whereas the magnitude analysis examined the functional dependence of contextual mechanisms across tasks, this profile analysis examines their similarity qualitatively, in terms of working principles. The profile of contextual modulation correlated positively across all three tasks suggesting they share some rules. As discussed in Angelucci et al. (2017), and seen in various studies (Li & Li, 1994; Cavanaugh et al., 2002; Goffaux, 2012; Self et al., 2014; Mannion et al., 2017), contextual modulations are strongest when stimulus and context share similar properties. With dissimilar stimulus and context, contextual modulations are either weak, or facilitatory. This profile was replicated in our study, in all tasks. Despite the functional independence of upright face and contrast detection contextual mechanisms, this finding supports our hypothesis that comparable but independent principles govern low-level and face-specialized high-level contextual modulations. Inverted faces showed a relatively weaker profile correlation to contrast detection, possibly due to the high variability of the contextual modulation profile for inverted faces (see Figure 5A). Lastly, the profile of upright and inverted face contextual modulations was tightly linked (48% of shared variance). The positive correlation between upright and inverted faces both in terms of the profile and the magnitude of contextual modulations yields an unprecedented characterization of their functional overlap at the level of perceptual behavior. It is however important to keep in mind that our profile analyses were conducted on three data points and should therefore be interpreted with caution.

The highly robust contextual modulations for the processing of upright faces is attributed to high level mechanisms implemented in face-specialized visual regions such as the FFA (Goffaux et al., 2013; Schiltz et al., 2010; Schiltz & Rossion, 2006). While some studies have examined the connection between face perception and basic visual abilities like contrast sensitivity, luminance discrimination, and spatial resolution (e.g., Barton et al., 2004), the functional link between face-specific contextual mechanisms and lower-level mechanisms has not been extensively explored. Recent work has however shown that the activity in face-specialized mechanisms is sensitive to face image properties known to be processed in V1 (e.g., orientation, spatial frequency, position). Goffaux & Dakin (2010) demonstrated that the specialized processing of the upright face preferentially relies on horizontally-structured content (see also: Duncan et al., 2017; Hashemi et al., 2019; Pachai et al., 2018; Goffaux & Greenwood, 2016). Face processing is also known to depend on the spatial frequency content of the stimulus (Fiorentini et al., 1983; Peli et al., 1994; Costen et al., 1996; Näsänen, 1999; Gold et al., 1999; Goffaux et al., 2003, 2005; Goffaux & Rossion, 2006; Goffaux, 2009) and even more so than other object categories (Biederman & Kalocsai, 1997; Collin et al., 2004, though see Oruc & Barton, 2010).

Past studies reported no correlation, or at best, anecdotal, weak correlation in the magnitude of contextual modulations across tasks irrespective of whether they are measured exclusively within the domain of basic (Song et al., 2011; Samaey et al., 2020; Cretenoud, Francis, et al., 2020), or complex stimuli (Konar et al., 2010; Chamberlain et al., 2017; Boutet et al., 2021; but see DeGutis et al., 2013). Considering these past findings, our finding of a correlation in the magnitude of the contextual modulations across visual domains is remarkable. There were several divergences across tasks, among which was the fact that one task required discriminating two subsequently shown eye regions and therefore involved memory while the other did not. Past evidence however shows that the contextual modulations for faces depend on local input strength irrespective of whether a memory component is involved in the task (cf. Goffaux, 2012). The tasks also differed in the nature of the target signal (related to contrast in one task and to shape in the other) and in terms of the nature of the task, one being a yes/no discrimination task and the other being a 2AFC detection task. In the facial feature matching task, we are taking the tipping point between predominantly ‘no/same’ to predominantly ‘yes/different’ as thresholds. We acknowledge that this approach is possibly contaminated by individual response biases, however, adapting the nature of this task to fit an exact 2AFC description would lead us to diverge too far from the paradigms used in the vast literature on human face perception, which typically uses yes/no tasks to address the influence of context on local feature perception (Young et al., 1987; Rossion, 2013). Despite these divergences, our design equated the implications of context-target relationships as a function of local input strength across domains, which allowed us to measure contextual modulations on a single, and most importantly, similar metric across tasks. This single metric that is calculated identically across tasks, along with the higher power

provided by the Hierarchical Modeling approach may have provided the ideal conditions to observe cross-domain correlation (Rouder & Haaf, 2019; Haines et al., 2020; Van Geert et al., 2022).

Several studies to date have traced back low and mid-level contextual effects to V1 anatomy. For example, the functionally defined size of V1 has been linked to contextual modulations for basic visual properties such as orientation (tilt illusion: Song, Schwarzkopf, Lutti, et al., 2013; Song, Schwarzkopf, & Rees, 2013), and even for shape perception (e.g., Ebbinghaus and Ponzo illusions: Schwarzkopf et al., 2011; Schwarzkopf & Rees, 2013). Whether similar connections can be established between V1 anatomy and high-level, face contextual modulations is an unanswered question. The absence of a correlation between contrast detection and upright face contextual effects may suggest that there may be no such correlation with V1. Nevertheless, we cannot draw any firm conclusions about the neural workings underlying our findings since the data presented in this study are exclusively behavioral.

In conclusion, by comparing the contextual modulations in a contrast detection task and a facial feature matching task, our study revealed shared variability and functional similarity between low- and high-level contextual mechanisms. Our findings show that low- and high-level, face-specialized contextual modulations have similar working principles while maintaining functional independence. When face-specialized mechanisms are disrupted by inversion, high-level contextual modulations are more functionally connected to low-level mechanisms. The combined study of low- and high-level contextual modulations sheds new light on the functional relationship between the different levels of the visual processing hierarchy, and thus on the functional organization of the visual system.

Acknowledgements

We would like to thank Vincent Bremhorst for his expert statistical guidance on our analyses.

Supplementary

Eye matching task: Magnitude of contextual modulations.

The eye matching task was a replication of Goffaux (2012). Here, for sake of empirical replication, we used their approach to analyze the data and confirm the quasi-linear decrease of contextual modulation strength as a function of local input dissimilarity. The magnitude of contextual modulations was estimated using the congruency effect; subtracting the proportion of "different" responses in same context/different target trials from different context/same target trials. by subtracting the proportion of "Different" responses in Same context trials from the proportion of "Different" responses in Different context trials. We then calculated the slope of the contextual modulations in terms of increasing dissimilarity for Upright and Inverted faces.

As seen in Figure S.1, in Upright faces, the magnitude of contextual modulations decreased as a function of eye region dissimilarity (The slope, β = -.0043, 95% CrI [-.005, -.0036], BF10 > 100): the more dissimilar the eyes are, the weaker the contextual modulations. This relative dependence of face contextual modulations on local input strength vanished in Inverted faces (β = -.0013, 95% CrI [-.003, .0003], BF10 > 100). These results replicate the findings of Goffaux (2012), along with the congruency – accuracy scores displayed in Table S.1.

	Upright				
	0%	24%	36%	53%	80%
Congruent	.94 (.03)	.6 (.06)	.75 (.05)	.89 (.04)	.97 (.02)
Incongruent	.6 (.06)	.19 (.05)	.43 (.06)	.71 (.06)	.93 (.03)
Isolated	.9 (.04)	.29 (.05)	.57 (.06)	.83 (.05)	.95 (.03)
	Inverted				
	0%	24%	36%	53%	80%
Congruent	.87 (.04)	.44 (.06)	.59 (.06)	.80 (.05)	.93 (.03)
Incongruent	.77 (.05)	.31 (.06)	.48 (.06)	.74 (.06)	.92 (.03)
Isolated	.9 (.04)	.26 (.06)	.44 (.06)	.66 (.06)	.87 (.04)

Table S.1. *Congruency scores for the eye matching task. Values display accuracy levels, with standard errors for the respective condition in parentheses.*

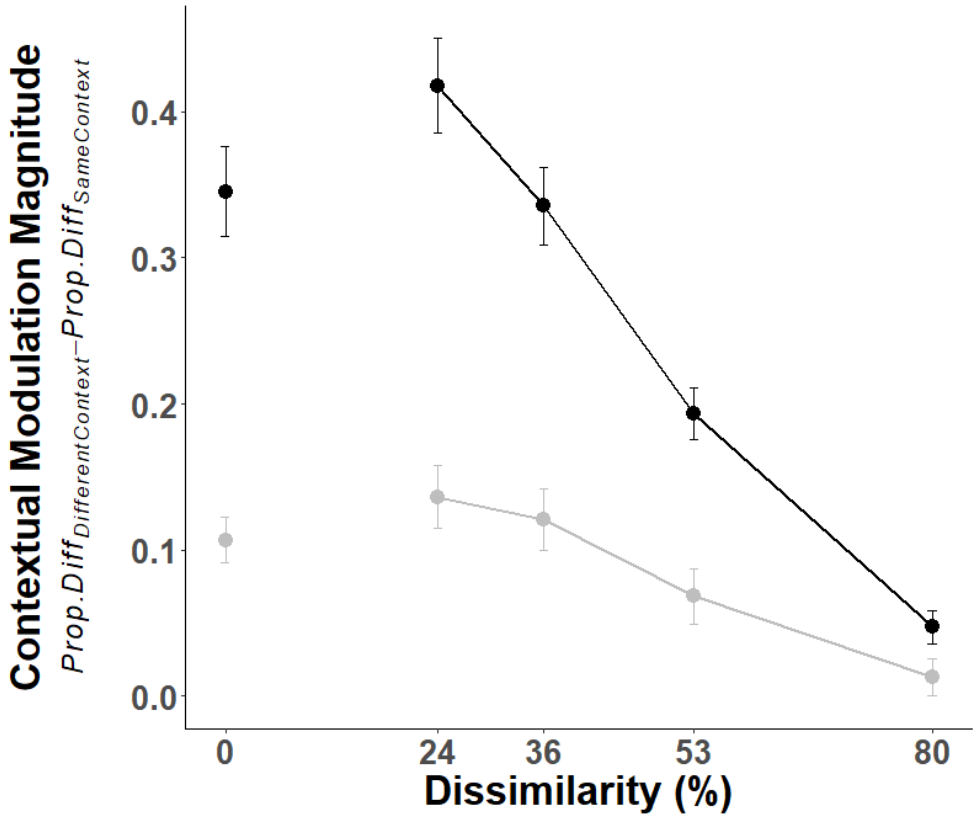

Figure S.1. *Eye matching task. Contextual modulation magnitude decreases as a function of target feature dissimilarity. The decrease is steeper in Upright compared to Inverted condition.*

Contextual Modulation Magnitude Across Sexes

Despite our sample was not properly balanced to address the effect of gender reliably (47 females, 12 males), we ran Bayesian Independent Sample T-tests on the contextual modulation magnitude values we obtained from the regression of regressions method. We found no difference in any of the three tasks across sexes (see Figure S.2). The mean contextual modulation magnitude in Upright Faces were .035 (SE = ±.09) for males and -.009 (±.11) for females (BF10 = 0.32). In Inverted Faces, it was -.014 (±.025) for males and .003 (±.037) for females (BF10 = .32). Lastly, in the Contrast Detection task, mean contextual modulation magnitude was .012 (±.014) for males and -.003 (±.016) for females (BF10 = .34). While these null findings from an unbalanced sample must be interpreted with caution, they agree with past evidence that sex has little or no influence on contextual modulations (Cretenoud et al., 2021; Mannion et al., 2017; Shaqiri et al., 2018).

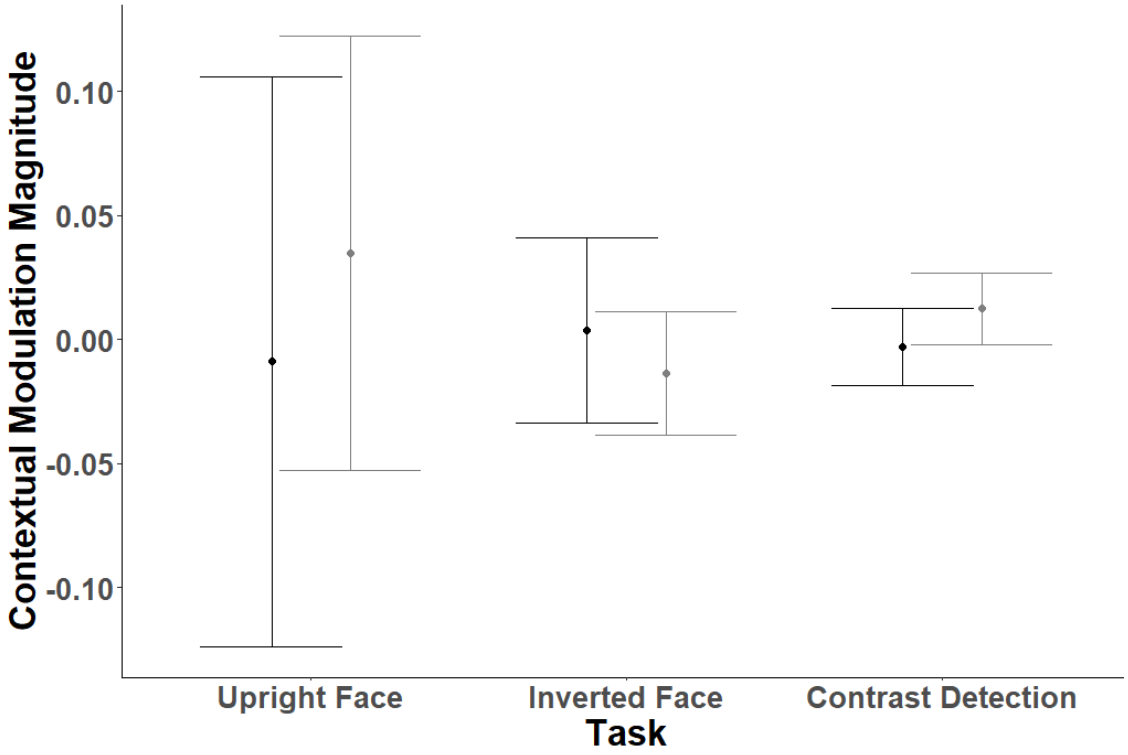

Figure S.2. Contextual Modulation Magnitude across Sexes. *Error bars display standard errors. Contextual modulation magnitudes showed no difference in any of the tasks between males and females.*

Contrast Detection Task – Comparison with Mannion et al. 2017

In our contrast detection task, we utilized a modified version of the paradigm used in Mannion et al. (2017). We halved the number of response choices compared to Mannion et al. (2017) while keeping the 4 possible target locations. Therefore, chance level differed across studies (50% versus 25%). Despite this difference, we were able to replicate their results. Mean contrast threshold in the Different context condition was 1.37% (95% CI [1.31, 1.43]) in their study whereas it was 1.23% (95% CI [1.12,1.27]) in ours. In Same context condition, mean threshold was 25.49% (95% CI [23.23, 27.71]) in Mannion et al. (2017) whereas it was 18.7% (95% CI [17.6, 19.7]) in ours. The overall pattern of the two contextual conditions were identical.

Goodness-of-fit measurements for the psychometric functions

We ran deviance tests for the goodness-of-fit of our model as described in Wichmann & Hill (2001). We compared the deviance of the fitted values from the actual data to obtain an empirical deviance value for every single fit (59 subjects, 3 tasks, 3 conditions, 531 fits in total). Fig S.3A shows data from the Contrast Detection task, Same condition data of a single subject, together with the best fitting psychometric function. Fig S.3B shows a histogram of deviance values from 1,000 simulations

using the same model's predictions. The deviance of the empirical data set (D) is 62.675, and the simulated cumulative probability estimate (cpe) is 0.355. The summary statistic deviance, hence, does not indicate a lack of fit. We ran these 1000 model-predicted simulations for each subject/task/condition fit and found that in 88% of the fits, the empirical deviance value were within 95% confidence interval of the simulated deviance values, which points to an overall good fit of our model.

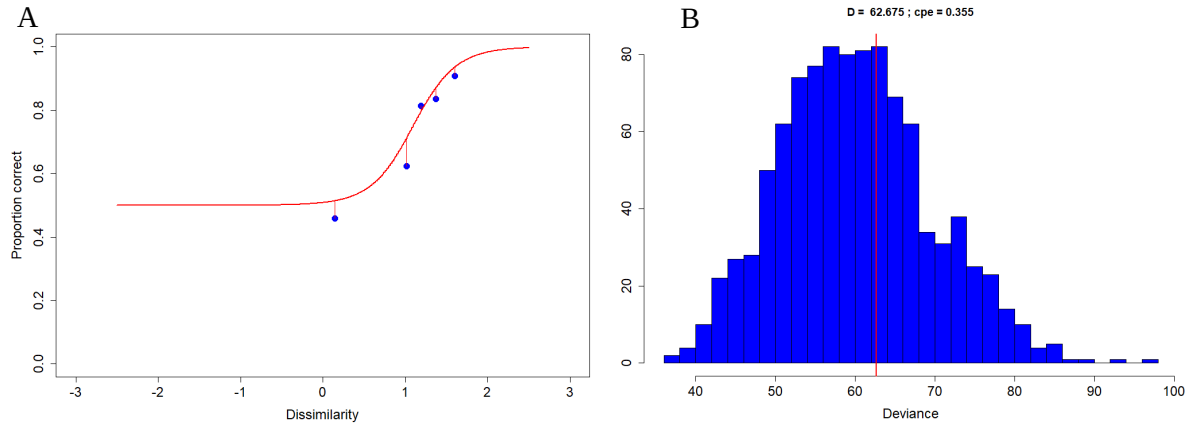

Figure S.3. Goodness-of-fit tests. *A. The example data from a single subject with the best-fitting psychometric function. The segments connecting dots to the psychometric function displays the deviation of the model fit from the data. B. Histogram of the deviance values of 1,000 simulations using the model predictions. Red line indicates the empirical deviance value obtained from the model fit. Empirical deviance within 95% confidence interval points to a good fit.*

Bayes Factor Robustness Analyses

In Bayesian statistics, the interpretability of outcomes and the definitions of H0 and H1 are directly linked to, and relies significantly on the selection of prior distributions. Greater information availability allows for more precise specification of "informed" priors. Nevertheless, when informed priors significantly differ from default priors, it becomes increasingly essential to provide justifications. In such situations, Bayes Factor Robustness checks can be utilized to evaluate the degree to which the results are influenced by the particular choice of prior distribution. In the current study, we opted for a stretched beta distribution with a width of 0.33. Traditionally, a stretched beta prior distribution with a width of 1 is used as a non-informative prior. In this distribution, any parameter value over the range of correlations between -1 and 1 are equally likely. However, it has been shown that this choice of priors can produce Bayes factors that can be biased towards null models (Vanpaemel, 2010; Quintana & Williams, 2018). In order to justify our specification of priors and the robustness of our results to different prior specifications, below we provide the prior/posterior

distributions and Bayes Factor Robustness checks (Figure S.4) for the three main correlation analyses we ran using contextual modulation magnitudes from different tasks.

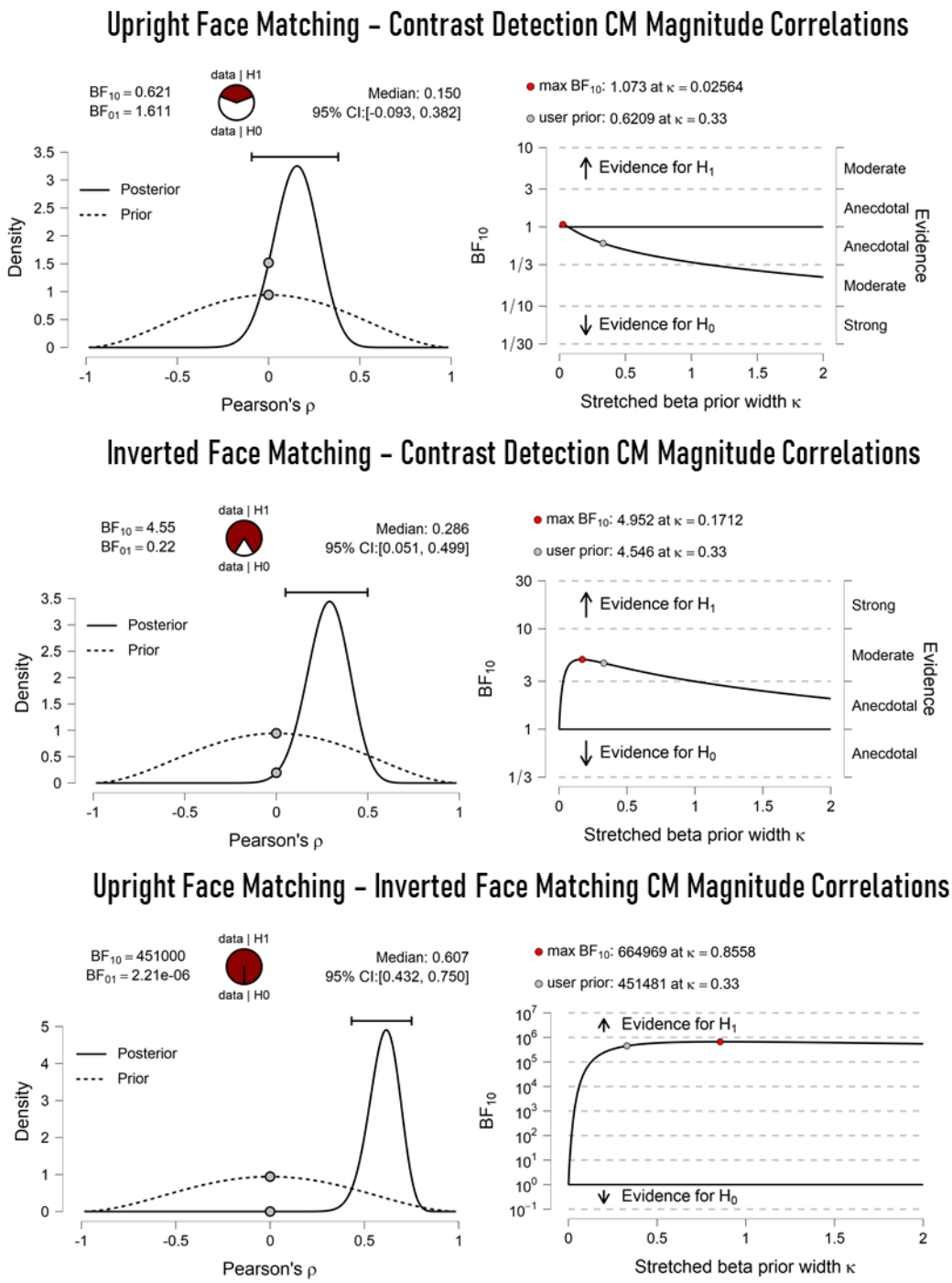

Figure S.4. Prior/posterior distributions and Bayes Factor Robustness checks for the three CM magnitude correlation analyses. *Left column displays the selected prior distribution with dashed lines. Combining this prior distribution with the observed data forms the posterior distribution (solid lines), and the Bayes factor is the ratio between the marginal likelihoods of the null model (no correlation) and the alternative model (correlation), given the data. The pie charts illustrate the ratios for H1 and H0. No specific minimum value for "correlation" is employed. Right column*

displays Bayes Factor robustness checks. The figures display the changes in BF_{01} value as a function of specified beta prior width. As can be seen in all three figures, regardless of the informativeness of the specified prior, the results are in the same direction in most cases.

Chapter 3

Contextual modulations at low and high levels of visual perception: functional (in)dependence and influence of local contrast

Mehmet Umut Canoluk[1], Marius Grandjean[1], Valerie Goffaux[1,2]

1. Research Institute for Psychological Science (IPSY), UCLouvain, Louvain-la-Neuve, Belgium,

2. Institute of Neuroscience (IoNS), UCLouvain, Louvain-la-Neuve, Belgium

KEYWORDS

Contextual modulation, local contrast dependence, contrast detection, face perception, holistic processing, individual differences

This study is a registered report currently under review at PCI:RR

Abstract

Spatial context modulates perception at low- and high levels of visual processing, with a shared dependence on the strength of the local input. In the case of human face perception, contextual modulations are presumed to rely on high-level, specialized mechanisms, with little contribution of low-level contextual mechanisms. Low- and high-level contextual modulations are typically studied independently, further obscuring their possible functional link. The present study will investigate whether consistent individual differences exist in contextual modulations at these two distinct levels of visual processing using contrast detection and facial feature discrimination tasks while manipulating the contrast (i.e., strength) of a local input target (local grating and facial feature, respectively). The context in the contrast detection task refers to a larger surrounding grating while in the facial feature discrimination task it refers to the facial features surrounding the target. The detection/discrimination contrast thresholds will provide a metric to evaluate contextual modulation magnitudes. By investigating the shared variance between the contextual modulation magnitudes across tasks, this study will provide insight into the functional relationship between low- and high-level contextual modulations and re-address the specificity of the contextual mechanisms at stake when processing faces in the light of low-level contextual modulations.

Introduction

Contextual modulations are found at various stages of the visual processing hierarchy, from retinal ganglion cells to IT cortex (e.g., Brincat & Connor, 2004; Cadieu et al., 2007; Desimone, 1991; Dobbins et al., 1987; Hansen & Neumann, 2008; Hubel & Wiesel, 1965; Levick et al., 1972; McIlwain, 1964; Nauhaus et al., 2009; Nolt et al., 2004; Sceniak et al., 1999; Solomon, 2006; Versavel et al., 1990). They modulate the processing of diverse stimulus aspects, ranging from simple properties such as contrast (Angelucci et al., 2002; Levitt & Lund, 1997; Nolt et al., 2004) or orientation (O'Toole & Wenderoth, 1977; Schwartz et al., 2007) to more complex characteristics such as perceived shape (Young et al., 1987; Thompson & Wilson, 2012; Schloss et al., 2014; Fang et al., 2008; Schwarzkopf et al., 2011), facial expression (Calder et al., 2000; Palermo et al., 2011; Tanaka et al., 2012) and face identity (Sergent, 1984; Young et al., 1987; Tanaka & Farah, 1993).

The processing of the human face, among other visual objects, is particularly susceptible to contextual modulations. Facial features are more accurately discriminated or recognized when they are part of a complete face (context) compared to when they are presented in isolation or in an atypical context, such as a scrambled face (Bruce, 1988; Davidoff & Donnelly, 1990; Sergent, 1984; Tanaka & Farah, 1993), resulting in the so-called whole-part advantage. Another manifestation of face contextual modulations is observed with the composite face and congruency paradigms. In these paradigms, the observer's sensitivity to a local feature decreases if it is incongruent with the surrounding context, in terms of either similarity status (when performing a discrimination task) or identity (when performing a recognition task) (Hole, 1994; Farah et al., 1998; Anaki et al., 2011; Goffaux, 2009, 2012; Richler & Gauthier, 2014). Face contextual modulations are hypothesized to enable the fast and efficient processing of the whole stimulus for an accurate identification despite faces' high visual similarity (e.g., DeGutis et al., 2013; Diamond & Carey, 1986; Leder & Bruce, 2000; McKone et al., 2007; Richler et al., 2011).

Contextual modulations with human faces are disrupted by simple and reversible image manipulations such as inversion (Sergent, 1984; Young et al., 1987; Farah et al., 1995; Yovel & Kanwisher, 2005). Inversion causes weaker activation in face-specialized regions while still recruiting a significant portion of non face-specialized high-level visual cortex (Haxby et al., 1999b; Yovel & Kanwisher, 2005; Taubert et al., 2015). This has been taken to suggest that the face contextual, so-called holistic, mechanisms are implemented at high-level stages of visual processing specifically tuned to the natural statistics of the human face (i.e., the regularities in the way faces appear to us in everyday life; e.g., Schiltz & Rossion, 2006; Mazard et al., 2006; Andrews et al., 2010; Schiltz et al., 2010; Goffaux et al., 2013; Liu-Shuang et al., 2015).

However, empirical evidence challenges a purely high-level account for face contextual modulations. The strength of contextual modulations for faces depends on image properties processed at primary

stages of visual processing, such as spatial frequency (Boutet et al., 2003; Goffaux & Rossion, 2006; Goffaux, 2009) and orientation (Dakin & Watt, 2009; Goffaux & Dakin, 2010). Moreover, face contextual modulations share similarities with low level mechanisms in their dependence on local input strength. Specifically, recent evidence indicates that the impact of the (face) context in a local feature discrimination task is inversely related to the physical difference of the features at stake (Goffaux, 2012; Goffaux et al., 2013; Canoluk et al., 2023).

This inverse relationship between contextual influence and local input strength is shared with low-level contextual mechanisms. For example, neurons in the primate visual cortex are more likely to integrate information across the visual field when the contrast of the local input is weak (Levitt & Lund, 1997; Angelucci et al., 2002, 2017). Similarly, the extent of spatial summation is larger with weaker local contrast in retina, lateral geniculate nucleus (LGN; Nolt et al., 2004) and primary visual cortex (V1; Sceniak et al., 1999), and local field potential (LFP) waves, which reflect post-synaptic lateral connectivity, are stronger in magnitude and travel a longer distance with weak contrast in cat and monkey primary visual cortex (Nauhaus et al., 2009). Face contextual modulations' reliance on low-level properties, as well as this shared dependence on local input strength raises the rarely asked question of whether the inter-individual variability in face contextual mechanisms can be explained by low-level contextual mechanisms to some extent.

In a recent study, we aimed to answer this question by measuring the amount of shared variance in the magnitude of contextual modulations for low-level contrast detection and morphed facial feature discrimination (in upright and inverted faces) by exploiting their shared dependence on local input strength (Canoluk et al., 2023). We obtained contrast detection thresholds for the low-level task and discrimination thresholds of two subsequently displayed, morphed facial features (eyes) in matched contextual conditions using a single Bayesian Hierarchical Model. We found no evidence for the covariance of the contextual modulation magnitude in upright facial feature discrimination and contrast detection, which indicates they are governed by separate mechanisms. On the other hand, there was substantial evidence that contextual modulation magnitude weakly covaried between contrast detection and inverted facial feature discrimination, suggesting a relationship between low-level and high-level, non face-specialized contextual modulations. Lastly, there was decisive evidence for a strong correlation between upright and inverted face contextual modulations which can be explained by the overlapping mechanisms that these two kinds of stimuli recruit, especially when the depicted identities are unfamiliar to the participant (Megreya & Burton, 2006; Burton, 2013).

It is important to consider the fact that in Canoluk et al. (2023) local input strength was defined using a different property of the stimulus across the two tasks (contrast versus feature shape difference defined through morphing). This difference in the local input properties of the two tasks may have reduced the dependence of the contextual mechanisms recruited by facial feature discrimination and

contrast detection. Past findings indeed showed that contextual illusions employing a similar local input property (e.g., two different size illusions) are more likely to correlate (Grzeczkowski et al., 2017; see also Haberman et al., 2015 for a similar comparison in ensemble perception). Furthermore, in the facial feature discrimination task, the target (eye region) was displayed foveally, at ~1.3° eccentricity whereas the targets appeared at 5° of eccentricity in the contrast detection task. Past evidence indicates that contextual influences across fovea and periphery rely on non-overlapping mechanisms (Xing & Heeger, 2000). Therefore, we reasoned that the differences in the employed local input property and foveal position of the targets could be the reason that we could not observe a dependence between contrast detection and facial feature discrimination tasks.

In this study, we aim to re-address the existence of a functional link between low- and high-level contextual mechanisms by circumventing possible confounds caused by these differences across tasks in the previous study. First, we will manipulate contrast at the eye region for the facial feature discrimination task. By doing so, the two tasks will have the same local input property. By measuring the magnitude of face contextual modulations at varying degrees of contrast, we will investigate whether contextual modulations exhibit a dependence on local contrast during face processing like they do in basic visual tasks. In the contrast detection task, the targets will be displayed closer to the fovea, aiming to tap into the foveal contextual mechanisms as in the facial feature discrimination task. The comparison of the discrimination/detection contrast thresholds across different context conditions will provide a metric of contextual modulation magnitude. A correlation of contextual modulation magnitudes between upright facial feature discrimination and contrast detection would demonstrate how much of the inter-subject variability in face contextual modulations can be explained by low-level contextual mechanisms (H1). An absence of correlation would bring further support to the specialization and independence of face contextual mechanisms. Second, we will investigate the correlation between inverted facial feature discrimination and contrast detection to see how the partial disruption of the face-specialized mechanisms caused by inversion influences the amount of shared variance with low-level mechanisms (H2). Finally, we will investigate the correlation between upright facial feature discrimination and inverted facial feature discrimination contextual modulations, in order to quantify the functional overlap of the contextual mechanisms involved for these two stimulus types (H3).

Methods

Participants

In this study, we will recruit participants from the young adult population, specifically targeting individuals between the ages of 18 and 35. Prior to the experiment, participants' handedness will be measured using Edinburgh Handedness Inventory (EHI; Oldfield, 1971), and their visual acuity will be tested with the Freiburg Visual Acuity Test (FrACT; Bach, 1996) to ensure they have normal or

corrected-to-normal vision (LogMAR ≤ 0.0). Participants with a LogMAR value larger than 0 will be excluded and will be replaced with new participants. Additionally, facial recognition abilities will be tested using a computerized version of the Benton Facial Recognition Test (BFRT; Benton & Van Allen, 1968; Rossion & Michel, 2018), to ensure their face recognition performance is in the normal range (i.e., above the borderline score of 39/54; (Duchaine & Weidenfeld, 2003). The experimental protocol was approved by the ethical committee of the UCLouvain (approval number: 2016/13SEP/393). The participants who fail to reach the required performance level in the practice (80% accuracy) will be removed from the experiment and new participants will be added until the desirable sample size is reached. Additional exclusion criteria are described in the Analyses section.

To determine the necessary sample size, we performed an a priori Bayes Factor (BF) design analysis (Schönbrodt & Wagenmakers, 2018). For data collection, we decided to opt for a sequential design with a maximal sample size, which entails that data be gathered until enough evidence has been accumulated or the predefined maximum number of participants has been tested. Thus, we will continuously calculate the BF of the correlations between the contextual effects in the two tasks: facial feature discrimination (upright/inverted), and the contrast detection (for more details, see the Data Analysis section). Such a sequential design requires one to specify the BF_{10} boundaries, the maximal sample size (N_{max}), the sample size at which the BF_{10} is computed for the first time (N_{min}), and the number of participants to be added each time until N_{max} or one of the BF_{10} boundaries is reached ($N_{increment}$). We opted for the BF_{10} boundaries of 6 and 1/6, and N_{min} was set to 40 as in Schönbrodt & Wagenmakers (2018). $N_{increment}$ was set to 5 for practical purposes (i.e., scheduling participants). N_{max} was determined through simulations with the BFDA package (Schönbrodt & Wagenmakers, 2018) in R (R Core Team, 2020). We chose an informative kappa prior of 0.7 for BFDA simulations. A kappa prior of 0.7 creates a prior distribution for the correlation simulations in which the distribution resembles the one of our previous study, with a mean of roughly r = 0.3, 95% confidence intervals ranging from .05 to .49. This reflects the weak significant correlation we found in our previous study between low-and high-level non face-specialized contextual modulations (Canoluk et al., 2023). We conducted simulations using 10,000 hypothetical datasets to examine sequential testing with the specified kappa prior. Two scenarios were simulated: one assuming a correlation of 0.3 and the other assuming no effect (i.e. the null hypothesis, with an effect size of r = 0). Our goal was to estimate the maximum sample size necessary to reach one of the Bayes Factor (BF) boundaries in 80% of the cases. This analysis was performed for both scenarios to compare the required sample sizes. We selected the maximum value between these two values as N_{max}, resulting in 130. Next, we evaluated the rates of false-negative and false-positive evidence using this N_{min}/N_{max} combination. Both rates were < 0.05, therefore we opted for an N_{min} of 40 and an N_{max} of 130.

In summary, with this design, we will start calculating our BF_{10} values at every 5 subjects starting from 40 subjects (N_{min}). We will terminate the data collection as soon as a certain BF_{10} boundary (1/6

or 6) for all three correlations (Upright Face vs. Grating, Inverted Face vs. Grating, Upright Face vs. Inverted Face), or N_{max} (130 participants) is reached.

Stimuli

In the facial feature discrimination task, we will present pairs of whole faces or isolated eye regions. The original stimulus set consists of 32 full-frontal grayscale pictures (on a scale from 0 to 1; 600 x 700 pixels) of unfamiliar white neutral faces from each gender without glasses, facial hair, or makeup. Face models are young adult students and alumni (aged 18–25 years) of the UCLouvain (Belgium), who gave written consent for the use of their image (Laguesse et al., 2012).

The luminance and root-mean-square (RMS) contrast of all face images are set to a mean of .52 and .12, respectively (using MatLab, Natick, MA). Next, we ensure that the eye regions of faces are comparable in terms of inter-pupillary distance and physical similarity. We measured inter-pupillary distance in each face image, and paired the faces with comparable values (average pixel difference of 1.32 ± 1.31 pixels). After this initial pairing, pixelwise similarity within the eye region was calculated within each pair using *corr2* function in MatLab. The pixelwise similarity of the target eye region was .8 on average (SD = .014). For each gender, the eight pairs lying within two standard deviations around the mean pixelwise similarity distribution are paired; the remaining eight pairs are only used as "face contexts". An additional set of 12 female faces (six pairs of female faces) which had undergone the same procedure are selected for practice trials. We extract the target region (eyes and brows) of each selected pair using a single aperture within the same pair (so the subjects could not rely on the shape of the aperture between two images within a pair) and blend it onto two "face contexts" whose eye areas were *a priori* replaced with a smooth skin texture (using Adobe Photoshop CS., Berkeley, CA), therefore all presented stimuli are edited, to ensure similarity across experimental trials.

Each face subtends a visual angle of ~7° in height, and ~5.2° in width. This size corresponds to the size of a human face seen at an average social distance (a human face is approximately of 18cm in height and social distance is of 1.5m approximately, e.g. McKone, 2009). The target eye region subtends a visual angle of ~1.4° in height, and ~4.1° in width. In this arrangement, the pupils are roughly at ~1.3° lateral eccentricity from the image center. Stimuli will be shown in their canonical Upright (0° rotation) and Inverted (180° rotation) position, and the face image will be offset 1.18° vertically to ensure that the target eye region is always at the center of the screen regardless of the orientation.

We additionally created noise masks by iteratively phase-scrambling each face image in Fourier space (in 500 iterations; e.g. Jacobs et al., 2020; Petras et al., 2019, 2021; Canoluk et al., 2023; Roux-Sibilon et al., 2023). Iterative scrambling is used to prevent noise masks from being contaminated by

uniform background pixels, and thus ensures that the spectral properties of the face and mask are similar.

The preparation steps and all the settings of the face stimulus set described up until this point are identical to what we used in Canoluk et al. (2023), hence were already tested and validated. The main novelty in the current design is that the manipulation of local input strength is better matched across tasks. To that end, the contrast levels of the target regions (i.e., eyes and brows) will be manipulated with the use of a rectangular, grey occluder, covering 1.8° in height and 4.15° in width (see Figure **2)** and an adaptive staircase, as detailed in the following section. We opted for an occluder instead of directly manipulating the eye region contrast because it provides a more naturalistic view of the face (similar to seeing faces through sunglasses) as opposed to faces with low-contrasted eyes. Occluder's edges will have a thickness of .1° and a pre-defined luminance of .35.

The stimuli and procedure of the contrast detection task were adapted from Mannion et al. (2017) to make the stimulus appearance more similar to the stimuli of facial discrimination task. Stimuli will consist of a circular target grating (diameter of 0.75° instead of the original 3°) that will appear in one of two target regions (positioned at 1.5° lateral eccentricity, similar to eye targets instead of the original four target regions at 5° eccentricity) and a circular context grating (diameter of 20°). The target will always be a horizontal grating whose contrast will be defined by a Psi Marginal staircase (as detailed further in the following section), while the context will be made up of a horizontally (parallel-) or vertically (orthogonal+ to the target orientation) oriented grating with a contrast of 25%. The contrast of both the context and target regions will be ramped smoothly to zero at the edges with the use of a raised cosine filter. Additionally, a small fixation dot (diameter of 0.25°) will be present at the center of the screen before and after the stimulus presentation.

Procedure

The experiment consists of two different sessions that last ~45 minutes for each task. The order of the sessions will be counterbalanced across participants. Participants will perform both tasks seated in a dark room (whose walls are painted black to ensure the lux/cd ^ m2 are lowest when the doors are closed) with their head position restrained by a forehead/chin rest, at a distance of 57 cm from the computer monitor. Prior to starting, all participants will be dark adapted for at least 30 seconds. Stimuli will be displayed on a ViewPixx monitor with a spatial resolution of 1920 x 1200 pixels, a temporal resolution of 120 Hz, a bit depth of up to 12 bits (achieved through M16 mode of ViewPixx), and a maximum luminance of 27 cd/m^2. Monitor luminance levels are linearized using a ColorCAL MKII colorimeter (Cambridge Research Systems, Kent, UK). Luminance levels, gamma correction and the temporal resolution of the monitor will be periodically controlled to ensure the same values are kept constant throughout the experiment. Both tasks will be run on PsychoPy (2020.2.4post1; Peirce, 2007) and Python 3.8.

The facial feature discrimination task consists of six conditions: 2 Orientation (Upright, Inverted) and 3 Context (Congruent+, Incongruent-, Isolated). The term "context" in this task includes all the elements of the face apart from the targeted eyes and brows, such as the nose, mouth, cheeks and face outline. The target regions will be presented either out of face context (i.e., Isolated), in a Congruent+ context, where the context and the target in a pair call for a similar response (e.g., “same” targets displayed with “same” face contexts, or “different” targets displayed with “different” face contexts), or in an Incongruent- context, where the context and target call for opposing responses (e.g., “same” targets displayed with “different” face contexts, or vice versa; **see Figure 1**).

For both tasks, the contextual conditions were assigned designations of + (facilitatory) and – (suppressive) indicators as they align with the anticipated effects (if any) of the context in relation to the Isolated condition in accordance with existing literature.

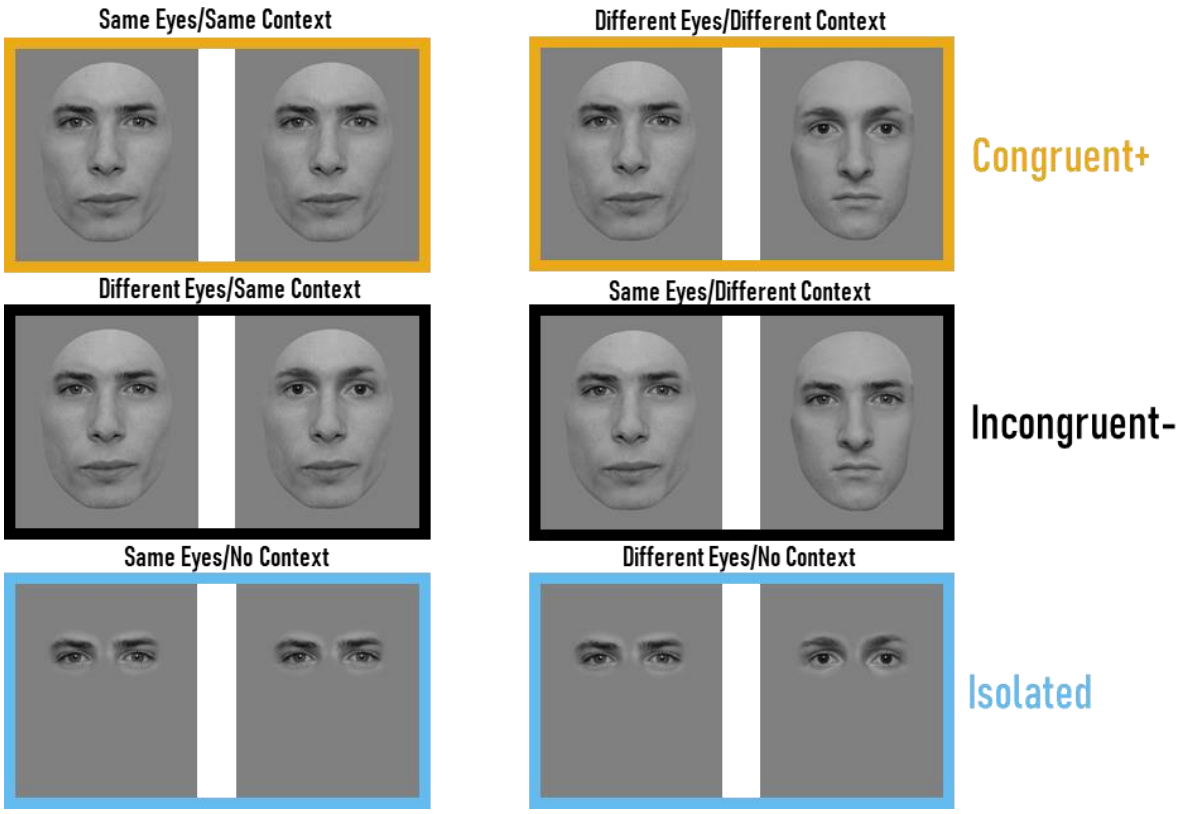

Figure 1. **Facial feature discrimination task contextual conditions.** The interaction of the context (face) and target (eyes) defines the conditions. In the congruent+ condition, the targets and the contexts of a pair are in accordance in terms of similarity (Same/Same or Different/Different) whereas there is a disagreement between them in the incongruent- condition (Same/Different or Different/Same). Target will be presented out of face context in the isolated condition. It can be seen here in Same Eyes/Different Context condition that context differences modulate the perception of the eyes, making them appear different when they are strictly identical (Young et al., 1987; Canoluk et al., 2023).

In each trial, the luminance of the occluder will be equalized to the average luminance of the region of the stimulus to be occluded. The occluder will be alpha-blended into the image. The amount of blending (i.e., transparency; e.g., 20% image, 80% occluder) will define the contrast (ranging from 2-40% contrast, or 98 - 60% occluder respectively) of the target region. Target contrast will be determined in each trial by two interleaved Psi Marginal adaptive staircases (Kontsevich & Tyler, 1999) for each condition. This interleaving procedure ensures that the algorithm does not settle into a narrow range of contrast levels within a condition.

The experimental session of the facial feature discrimination task will be divided into 16 blocks of 48 trials (768 trials in total), where the context conditions will be pseudo-randomized. Each block will comprise equal number of trials from each condition, therefore also equal number of trials that call for "Same" and "Different" responses in random order. Each trial will start with a fixation cross of 500 ms, followed by a blank screen of 500 ms. The first stimulus is then displayed for 500 ms, immediately followed by the phase-scrambled mask for 200 ms and the second stimulus will stay on the screen until the participant responds through keypress whether the eye regions of the first and second stimulus were same or different. The occluder will be overlaid on both the stimuli and the noise mask for the duration of a trial.

Prior to the experiment, participants will be required to do a practice run of 4 blocks (of 48 trials each) with 20% contrast in the eyes for the first two blocks and 10% contrast in the eyes for the second two blocks to familiarize with the task. They will be asked to re-do it until they reach 80% accuracy.

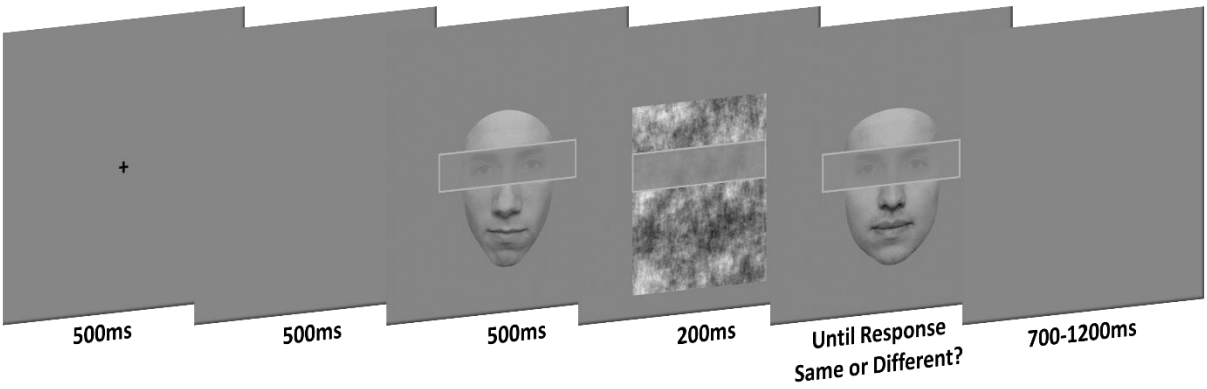

Figure 2. **Facial feature discrimination task trial sequence.** An example of an Incongruent- condition trial (different context, same eyes). Each trial starts with a fixation cross for 500 ms, followed by a blank screen. The first stimulus then appears for 500 ms, followed by the noise mask for 200 ms, and finally, by the second stimulus. The second stimulus stays on the screen until the participant indicates whether the target eye regions

were “Same” or “Different” between the two stimuli by button press. This will be followed by a blank screen (randomly ranging between 700-1200 ms) before the next trial starts.

Similar to Mannion et al. (2017) and Canoluk et al. (2023), in the contrast detection task, the local target is a horizontal grating with the same spatial frequency (2 cycles per degree, and same phase as the context in the Parallel- condition) as the context, and with a variable, non-zero contrast will be presented in one of the two lateralized regions, within one of three context conditions: Isolated, Parallel- context, Orthogonal+ context (see **Figure 3**). The term "context" in this task refers to a surrounding, larger grating. Each context condition will be presented in separate blocks. Each block consists of 80 trials, with two staircases (as in the facial feature discrimination task) interleaved for 40 trials each. The contrast of the target grating will be determined by these staircases, which use a pool of 350 logarithmically spaced contrast values ranging from 0.001 to 1. There will be three block repetitions per condition, i.e. a total of 9 blocks (720 trials in total). Block order will be pseudo-randomized across participants, ensuring successive blocks belong to distinct context conditions. Each trial starts with a preparatory, black fixation dot of 500ms. Next, the local target (embedded in a context or presented in isolation, depending on the condition) is presented in one of the two possible locations for 100ms. After the stimulus presentation, the fixation dot turns to dark grey to signal the participant that they can give their response. We will ask the participant to report the side on which the target appeared (left or right). Before the experiment, participants will perform a short practice of 20 trials to familiarize with the task. No upper/lower limit is defined to pass the practice due to the simplicity of the task and instructions. Both tasks will emphasize accuracy over speed.

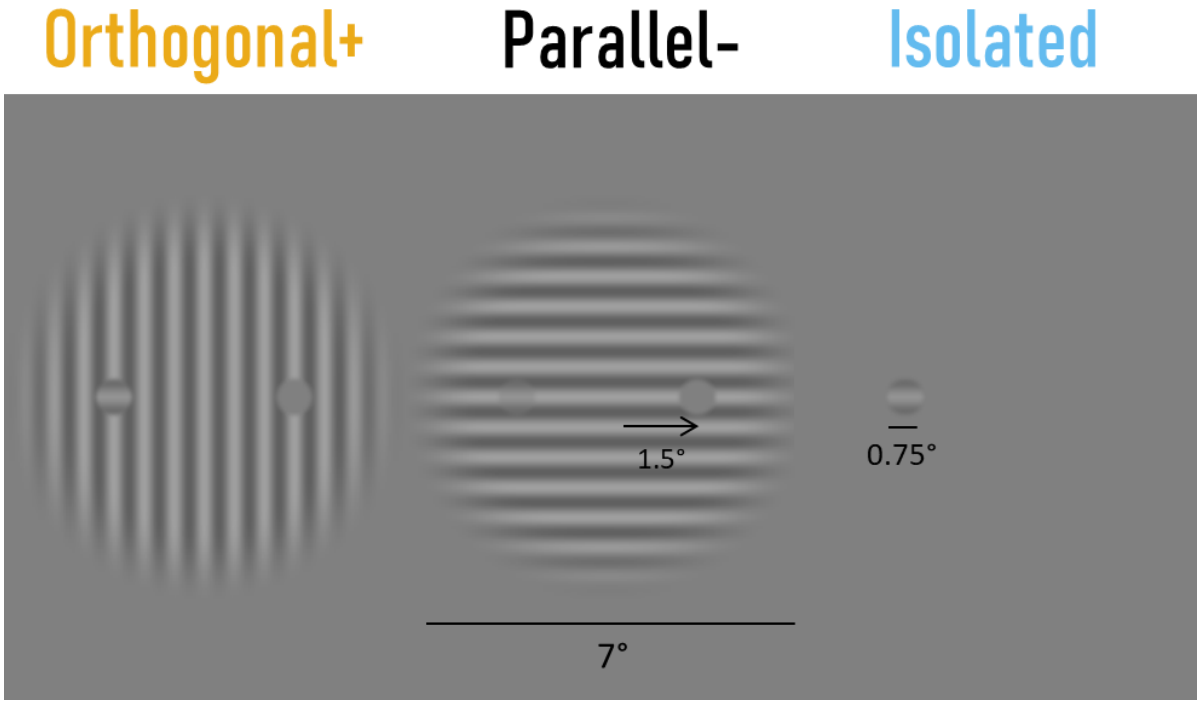

Figure 3. **Orthogonal+, Parallel- and Isolated context conditions of the contrast detection task, example of a left-lateralized local target.** It can be seen here that despite having the identical contrast value, the target in the parallel- condition appears to have lower contrast because of the parallel context it is presented in.

Analyses

The present study aims at quantifying the variance of contextual modulation magnitude that is shared across tasks (i.e., processing levels). We will do so by comparing the detection/discrimination of each task thresholds in various context conditions. This approach largely replicates the one we successfully applied in Canoluk et al. (2023).

We will first log10 transform the contrast values from both tasks. Second, these values will be z-transformed for each task separately. After these transformations, contrast values will be within the range of -3 to 3 for both tasks. This normalization procedure on independent variables is a common practice when using complex hierarchical models, as it facilitates model convergence (Stan Development Team, 2023). Across tasks, we first investigate the function relating performance to local input strength (target eye region contrast or target grating contrast). Therefore, the z-scaled data from both tasks will be submitted to a Bayesian generalized linear mixed model (Yssaad-Fesselier & Knoblauch, 2006; Moscatelli et al., 2012) by means of *brms* package available in R (Bürkner, 2017, 2018). Using a mixed model makes it possible to include inter-individual variability in the statistical model where the estimates for the individuals are shrunk towards the general mean (due to the hierarchical nature of the model) based on trial-by-trial data, yielding richer (better informed) and more sensitive independent estimates and conferring more statistical power to the analysis compared to other analyses that rely on aggregated/averaged data (see Rouder & Haaf, 2019; Moors et al., 2020 for more details).

The dependent variable will be accuracy. The guess rate (i.e., lower asymptote of the psychometric function) is set at .5 for the contrast detection task (2-AFC, Knoblauch & Maloney, 2012) and the contrast value at .75 will be taken as "threshold". Facial feature discrimination task being close to a yes-no task, the guess rate will be set at 0 and the performance at .5 will be taken as "threshold" in this task (similar to Canoluk et al., 2023). Using .5 as the threshold, and choosing such a yes/no task is likely to be contaminated by response biases. However, straying from this approach would require bigger modifications in the design of the experiment, which would result in divergence from the past literature on face perception. Another approach to circumvent this would be using signal detection theory and obtaining d' values for each condition, but in that case, we would not be assessing exact threshold locations for the facial feature discrimination task. Using d' for one task and contrast detection thresholds for another makes the measures we take from the two tasks even less comparable, therefore going against our initial goal to make the measurements from the two tasks as similar as

possible. Therefore, we have decided to stick with this methodology despite possible confounds. Facial feature discrimination performance with Upright and Inverted faces will be treated separately. With this adjustment, the factor, "stimulus" has three different levels: "Upright Face", "Inverted Face", "Grating". We will fit an inverse logit link function to model the sigmoidal pattern of the psychometric functions in each stimulus type (**Equation 1**). We put a normal distribution prior ($\mu = 0$, $\sigma = 5$) on all regression weights, and a lognormal distribution prior ($\mu = 0$, $\sigma = .1$) on the parameters capturing the random effects. All random effects will be included, yet will be assumed to be uncorrelated to each other (by means of || term in the formula) to increase sampling efficiency and ensure model convergence (Bürkner, 2018). Doing so, we allow a participant to deviate from the average, but we do not assume the size of these deviations to correlate across conditions (Kurz, 2019). The model will be fit with the following formula:

$$accuracy \sim .5 * \text{guess} + (1 - .5 * \text{guess}) * inv\_logit(\eta)$$

$$\eta \sim contrast * stimulus * context + (contrast * stimulus * context \,||\, participant)$$

Equation 1. ***brms*** **formula to fit the sigmoidal functions.** The first formula implements the guess rate for both tasks (parameter "guess" is set to 1 for the contrast detection task, and 0 for the facial feature discrimination task, resulting in .5 and 0 guess rates respectively), to the inverse logit link function with the *η(eta)* hyperparameter, which is estimated by using the fixed (contrast, stimulus, context) and random (participant) effects described in the second formula.

Using these parameters and the formulae, we will run four Markov chain Monte Carlo (MCMC) with 6,000 iterations each (3,000 of each are considered warm-up), resulting in 12,000 iterations after warmup. Default values will be used for the *adapt_delta* and *max_treedepth* parameters, and they will be increased whenever it is deemed necessary for a better model fit. After fitting the data, we will run deviance tests as described in Wichmann and Hill (2001), to measure the goodness-of-fit of each individual psychometric function, in addition to the generic model diagnostics such as the R-hat value, effective sample size, and number of divergent transitions. The participants whose empirical deviance values of the psychometric functions are outside of the 95% confidence interval of the simulated deviance values will not be included in the analyses. Instead, they will be replaced with new participants until the maximum sample size is reached.

After fitting the data from both tasks, we will obtain the threshold value of the inverse logit link function for each stimulus and context condition, and for each participant using the formula, *threshold = − (intercept / slope)* as outlined in Knoblauch & Maloney (2012). This formula is used to transform intercept and slope values obtained from classical regression models to a threshold value. We will compare these threshold values across tasks and contexts using the model's threshold estimates, and the Highest Density Interval (HDI) of the differences in threshold comparisons. An HDI excluding

zero indicates that the difference is non-negligible, or states that the two sets of compared values are practically unlikely to be equivalent. An HDI including zero on the other hand suggests the values are likely equivalent and the difference between the two sets of values is practically zero (Kruschke, 2018). In the facial feature discrimination task, we expect highest thresholds in the incongruent-condition than in the congruent+ condition (Outcome-Neutral Tests 1 (Upright Face) & 2 (Inverted Face); Richler et al., 2008; Goffaux, 2009; Meinhardt-Injac et al., 2014) and a smaller difference in thresholds between these conditions for inverted faces compared to upright faces (Outcome-Neutral Test 3), while in the contrast detection task, we expect the thresholds to be highest in the parallel-condition (Outcome-Neutral Test 4) as documented by the orientation-dependent contextual modulation effects . These differences will be used as confirmatory tools for replication of past findings and tests of soundness and feasibility of our methodology.

In order to quantify the contextual modulation magnitude of each participant, we will submit the individual threshold values from the three context conditions of each stimulus type to a two-step regression of regressions method (DeGutis et al., 2013; Canoluk et al., 2023). We chose to use this approach because the residuals of this two-step regression deliver an estimate of each individual contextual modulation that is not contaminated by the variance of the compared conditions, as opposed to simple subtraction method where the final outcome contains variance from both of the compared conditions. Each task comprises a control "Isolated" condition in which the target region (eyes and brows, or local grating) is presented on its own. In a first regression, we will regress participants' thresholds in the "Isolated" context condition from "+" context and "-" context thresholds, therefore removing all variance related to the control condition from the conditions where a context was present. In the second regression, we will regress the residuals of the "-" context condition from the residuals of the "+" context condition residuals. We confirmed in Canoluk et al. (2023) that this method returns sufficiently large intersubject variability, as well as good reliability.

Finally, we will conduct Bayesian correlation analyses on the final products of the two-step regressions from each stimulus type (Upright Face vs. Grating (H1), Inverted Face vs. Grating (H2), Upright Face vs. Inverted Face (H3)). We choose a Bayesian approach because of its ability to determine the strength of evidence presented by the data in supporting a particular hypothesis, hence, its potential to interpret a finding, even when it is null. We will compute Bayes Factors (BFs) of the correlation using BayesFactor package (Morey & Rouder, 2018) in R. For Bayesian correlation analyses, a uniform default prior (i.e., a stretched beta prior distribution with a width of 1, where any parameter value over the range of correlations between -1 and 1 are equally likely) can produce Bayes factors that can be biased towards null models (Vanpaemel, 2010; Quintana & Williams, 2018), so they are generally not recommended. Therefore, we will opt for a stretched beta prior distribution with a width of .33, centered at 0. Combining this prior distribution with the observed data forms the posterior distribution, and the Bayes factor is the ratio between the marginal likelihoods of the null

model and the alternative model, given the data. A BF_{10} value indicates support for H_1 (e.g., correlation) over H_0 (e.g., no correlation) given the data (conversely, BF_{01} would indicate support for H_0 over H_1, and can be calculated with $1/BF_{10}$). We follow Jeffreys' guidelines (H. Jeffreys, 1998; Jarosz & Wiley, 2014) to interpret the BF values: a BF_{10} value of 1 would mean equal amount of evidence for both H_0 and H_1, and therefore inconclusive. A BF_{10} value between 1 and 3 is considered anecdotal evidence in favor of H_1 over H_0, whereas a BF_{10} between 3 and 10 indicates substantial evidence for H_1. $BF_{10} > 10$, $BF_{10} > 30$ and $BF_{10} > 100$ are regarded as strong, very strong and decisive evidence for H_1, respectively. We opted for evidence thresholds of $BF_{10} = 6$ (or 1/6) for concluding our analyses.

Chapter 4

V1 size predicts the contextual modulation strength in low- and high-level vision

Mehmet Umut Canoluk[1], Jolien P. Schuurmans[1], Matthew A. Bennett[2], Valerie Goffaux[1,2]

1. Research Institute for Psychological Science (IPSY), UCLouvain, Louvain-la-Neuve, Belgium,

2. Institute of Neuroscience (IoNS), UCLouvain, Louvain-la-Neuve, Belgium

KEYWORDS

Primary visual cortex, V1, Contextual modulation, face perception, holistic processing, individual differences

Abstract

The strength of contextual modulations at low- and high-stages of visual processing hierarchy shows great variability across individuals. The variability in the magnitude of contextual modulations at low-levels can be partially explained by the variance in the functionally defined size of V1, notably, human individuals with larger V1s exhibit smaller contextual modulation magnitude. However, especially in high-level contextual modulations seen with upright faces (so-called holistic processing), this variability is largely unexplained. In the current study, we addressed the variability in both low- and high-level contextual modulations that can be explained by V1 size. Using fMRI and psychophysics methods, we functionally measured the size of V1 and the magnitude of contextual modulations in upright and inverted facial feature matching, and grating contrast detection in 23 individuals. The interindividual variability of V1 size did not account for the contextual modulations seen in upright facial feature matching. This could be attributed to the dominant involvement of face-specialized regions in upright face processing, which may operate more autonomously from the low-level contextual processes present in V1. Variations in V1 size on the other hand seem to explain the contextual modulation magnitude in inverted facial feature matching partially. Inversion impedes the activation of face-specialized processing areas and allows for the measurement of high-level contextual modulations not specialized for faces. These high-level, non face-specialized contextual modulations exhibited a stronger functional connection with low-level mechanisms. Finally, we showed that variations in V1 size can predict the magnitude of orientation-dependent contextual modulation of contrast. These findings lend support to theories proposing the existence of specialized mechanisms for the dedicated processing of upright faces, which exhibit certain modular-like characteristics while also indicating that V1 size can partly account for the magnitude of contextual modulations in both low- and high-level, non-specialized mechanisms.

Introduction

Sensory perception of the world varies largely in the normal population. An instance of that variability is in susceptibility to visual contextual modulations. Psychophysical evidence for contextual effects can be found on various levels of visual processing hierarchy, ranging from aspects such as lightness (Adelson, 1999), orientation (Gibson & Radner, 1937; O'Toole & Wenderoth, 1977), contrast (Ejima & Takahashi, 1985; Chubb et al., 1989; Polat & Sagi, 1993), color (Webster et al., 2002) to face processing (Young et al., 1987; Tanaka & Farah, 1993). The susceptibility to such contextual effects can vary 2- to 4-fold across individuals (Coren & Porac, 1987; DeGutis et al., 2013; Konar et al., 2010a, 2010b; Cretenoud et al., 2021).

This large variability in contextual modulation (CM) magnitudes has been partially attributed to the size of V1. Across healthy young adults, cortical size of V1 varies up to 3.5-fold (Filimonoff, 1933; Stensaas et al., 1974; Andrews et al., 1997; Duncan & Boynton, 2003; Dougherty et al., 2003; Benson et al., 2022). The conventional scaling hypothesis, as in Ringo (1991) and Kaas (2000), posits that individuals with a larger V1 surface size would be less susceptible to contextual modulations, leading to improved perception of finer, local visual details. Conversely, individuals with a smaller V1 would experience stronger contextual modulations and poorer local processing. As such, increasing V1 size was associated with better contrast sensitivity (Himmelberg et al., 2022), smaller magnitude of contextual size illusions such as Ebbinghaus and Ponzo (Schwarzkopf et al., 2011; Schwarzkopf & Rees, 2013), and better local orientation sensitivity (Song et al., 2013). Individuals with smaller V1 size are more susceptible to contextual influence, have stronger visual contextual shape illusions and a stronger global modulation of orientation.

Face contextual modulations (or so-called "holistic processing") are a hallmark of face perception. They are considered a highly adaptive process that developed from the social necessity to distinguish between individual faces despite their high homogeneity (i.e., having the same basic features arranged in a consistent configuration, such as eyes, nose, and mouth in the same positions; e.g. Diamond & Carey, 1986; Leder & Bruce, 2000; McKone et al., 2007). They manifest themselves in various ways. For example, a facial feature is discriminated more accurately when presented within a full face than in isolation (or within a scrambled face; Sergent, 1984; Tanaka & Farah, 1993), and facial features become harder to identify when they are inserted in two different face contexts, due to their perceptual fusing with the context (Young et al., 1987; Hole, 1994). When the face is turned upside down (i.e., inverted), contextual modulations weaken drastically (Yin, 1969), possibly leading to an inability to spatially integrate information across features of a given face (Rossion & Gauthier, 2002; Rossion, 2009; Van Belle et al., 2010). Inversion disrupts the specialization of face-processing, which manifests as increased activity in non-face-specialized regions(Gilaie-Dotan et al., 2010; Taubert et

al., 2015). Inversion is a reversible manipulation in which the low-level properties of the stimulus are largely preserved.

The hierarchical nature of the visual system, and the fact that the contextual mechanisms specialized for the processing of faces are disrupted by inversion has pushed the idea that these mechanisms are high-level, detached from, and mainly unaffected by early visual regions (Gross et al., 1972; DiCarlo & Cox, 2007). However, they have been found to be sensitive to low-level properties such as stimulus size and position in the visual field (Levy et al., 2001; Schwarzlose et al., 2008; Arcaro et al., 2009; Carlson et al., 2011; Yue et al., 2011; Ross & Gauthier, 2015; Poltoratski et al., 2021). Furthermore, face contextual modulations are influenced by the orientation content (Dakin & Watt, 2009; Goffaux & Dakin, 2010; Goffaux & Greenwood, 2016) and spatial frequency information in the image (Boutet et al., 2003; Goffaux et al., 2005; Goffaux & Rossion, 2006). Building on the influence of low-level properties, as feature discriminability increases, the impact of context on the target feature becomes less pronounced, (Goffaux, 2012; Canoluk et al., 2023), akin to the spatial summation and lateral interaction effects in V1 neurons' responses to lower local contrast stimuli (Levitt & Lund, 1997; Sceniak et al., 1999; Angelucci et al., 2002; Nauhaus et al., 2009; Krause & Pack, 2014). In this study, drawing from the existing literature on the impact of V1 surface size on visual contextual modulations and through the joint application of fMRI and psychophysics, we aim to address a potential link between face contextual modulations and low-level mechanisms.

Joint examination of a low- and high-level task measuring the same phenomenon allows for quantifying the shared variation across them (Geisler & Chou, 1995; Peterzell, 2016), and investigation of whether one enables or constrains the other. In a previous study (Chapter 2), we addressed the link between low and high-level contextual modulations. We behaviorally measured the contextual modulations in grating contrast detection and upright and inverted face matching tasks in a group of individuals (Canoluk et al., 2023) and investigated the shared variance between these stimulus types from different processing levels. We observed no relationship between upright face matching and contrast detection, and a weak correlation between inverted face matching and contrast detection contextual modulation magnitudes. These findings bring support to the autonomy of face-specific contextual mechanisms from low-level ones, simultaneously suggesting that non-face-specific contextual mechanisms can be partially attributed to low-level processes. However, it should still be noted that observing correlations in the magnitudes of behaviorally measured contextual illusions has proven to be a challenge, both for low-level illusions (Song et al., 2011; Cretenoud, Francis, et al., 2020; Cretenoud, Grzeczkowski, et al., 2020; Samaey et al., 2020; Cretenoud et al., 2021) and measures of face holistic processing (Konar et al., 2010a, 2010b; Wang et al., 2012; Sunday et al., 2017; DeGutis et al., 2013; Ventura et al., 2022). Among other factors, correlations in behavioral measures can be contaminated by the task, the design, the definition of contextual conditions, the retinal position, or the choice of the dependent variable (e.g.; Green et al., 2016). The

registered report proposed in Chapter 3 aims to circumvent such factors by making the two tasks as similar to each other as possible, while investigating a contextual effect that is shared across low- and high-levels. While strict control of behavioral measurements may overcome the possible contaminations and reveal correlations, using a neuroimaging method has a potential of stripping such behavioral factors from the measurement, giving a more robust estimate of the contribution of V1-based mechanisms to face CM. Considering the impact of low-level visual properties on face contextual modulations and their shared dependence on local input strength, it seems essential to investigate the variance that is shared between these two levels. Therefore, in the current study, we investigate the relationships between V1 anatomy and contextual modulation magnitudes. Following Ringo (1991) and Kaas, (2000), we hypothesize that individuals with smaller V1s may exhibit stronger contextual modulation magnitudes, while those with larger V1s may demonstrate weaker magnitudes. To the best of our knowledge, V1 size has never been tested in conjunction with a contrast-dependent contextual modulation effect. Given their close proximity and their stronger connection to V1-based mechanisms, we anticipate that the magnitudes of contrast detection can be relatively strongly accounted for by the size of V1 in individuals. As for the facial feature matching task, we expect a similar relationship between contextual modulation magnitudes and V1 size. Considering the amount of overlap in the activated regions for upright and inverted faces, we hypothesize that the correlation between either orientation of face and V1 size would be similar, possibly with a more pronounced association when the faces are inverted, owing to their less-specialized manner of processing.

Methods

Participants

Twenty-three healthy subjects (age 26.57 ± 3.86, nine males) completed three sessions of psychophysical experiments, and one session of MRI scanning. The psychophysics data of five subjects were already collected in Canoluk et al. (2023). Eighteen new subjects were recruited and tested using the identical paradigms for the purpose of the current study. Additionally, the MRI data for nine of the subjects in this project comes from a different project which involved a total of 20 experimental runs divided over 3 scanning sessions (as described on Schuurmans et al., 2023). The first session included a T1-weighted anatomical scan and the second session a functional face localizer run. Identical retinotopy and localizer scans were collected in a single session on further 14 subjects that were only recruited for the purpose of this project. Below we omit the details of the runs collected for the other project.

Before the psychophysical experiments and scanning, we tested the acuity of the subjects with the Freiburg Visual Acuity and Contrast Test (FrACT), which revealed all subjects had normal or

corrected-to-normal vision (LogMar = -0.1 ± 0.12 decVA = 1.33 ± 0.33). All subjects but one reported being right-handed as measured with the Edinburgh handedness questionnaire (84.61 ± 39.62; Oldfield, 1971).

Functionally defined V1 size measurements

The MRI experiment stimuli were presented with PsychoPy v3.2.4 on an NNL LCD Monitor (32-inch, 1920 x 1080 pixels, 698.40 x 392.85 mm, refresh rate = 60 Hz) situated at the end of the scanner bore. Subjects viewed the stimuli via a mirror attached to the head-coil, at a viewing distance of 175cm.

Retinotopic mapping

Borders of V1 were delineated using the retinotopic maps acquired through functional runs. All functional runs included a 12 second baseline at the beginning and end. The retinotopic mapping stimuli were high contrast checkerboard patterns on a gray background. Stimuli subtended a radius of 6.4° of visual angle from fixation and reversed contrast polarity at a temporal frequency of ~4Hz. The grey background included a 'spider-web' pattern (see Fig. 1a). In all functional runs, subjects were instructed to fixate at a small red central cross at all times. The subjects' task was to indicate with a button press when the cross rotated by 45°.

The first run (lasting 280 secs) consisted in the simultaneous presentation of a smoothly counter-clockwise rotating wedge (45° width, 6 rotation cycles at a frequency of 42.67 secs) and a ring that expanded outward (the outer radius was 1.5 times the inner radius, 5 expansion cycles at a frequency of 51.333 secs). The size of checkerboards for this run logarithmically increased with eccentricity. In the next two runs, (each lasting 344 secs) we presented 16 smoothly sweeping bar stimuli (1.6° wide, sweep frequency of 20 secs; see Figure 1b.) which swept from one side of the screen to the other along eight equally spaced radial axes (once in one direction, once in the other). The size of the checkerboards in these runs was uniform (~0.53°).

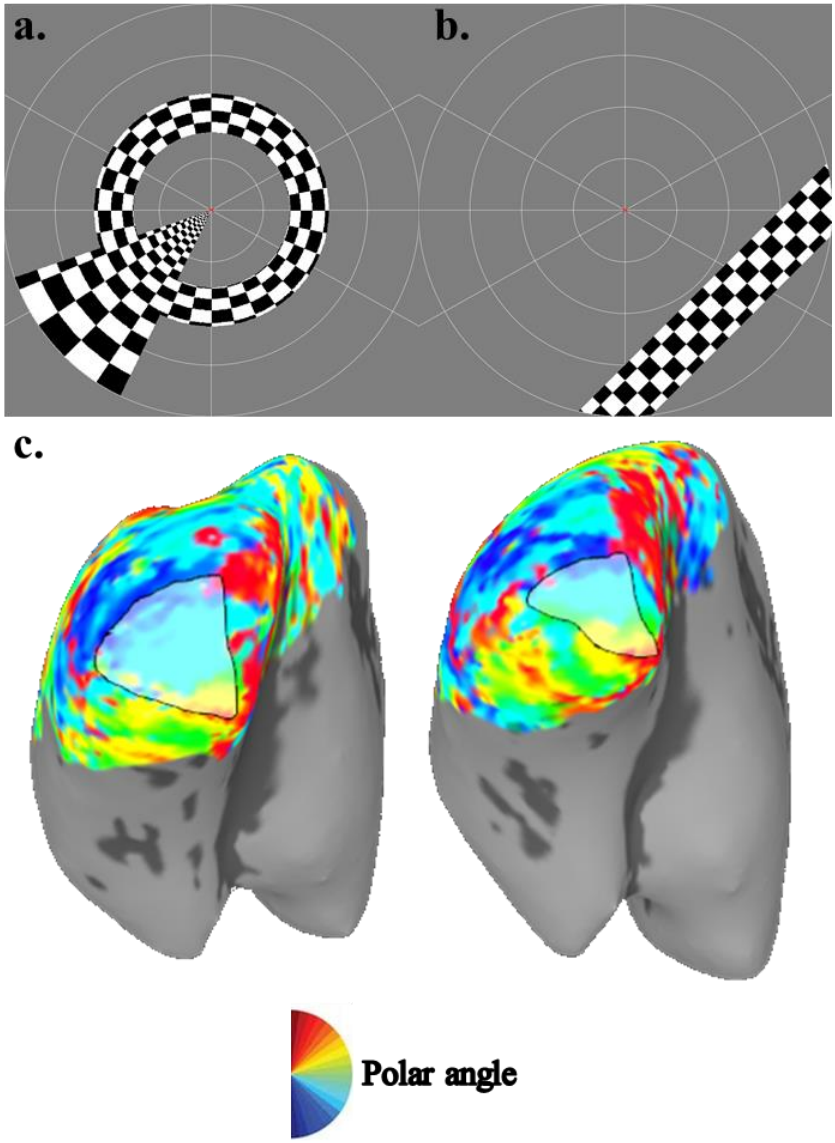

*Figure 1.**a.** The rotating wedge pattern used for retinotopic mapping. **b.** The sweeping bar pattern used for retinotopic mapping. **c.** Functionally defined V1 Areas of two subjects with biggest (left) and smallest (right) V1s. The thin black line represents the region belonging to V1, as delineated using the polar angle maps. The blue activation pattern represents the lower part of the visual field, and the yellow-red represents the upper part.*

MRI Data Acquisition

Subjects were scanned in a 3Tesla GE Signa Premier MRI scanner with a 48ch head coil, based at Cliniques Universitaires UCL Saint-Luc in Brussels. As anatomical references, whole-brain T1-weighted images were obtained during the first sessions (3D MP-RAGE, 1 x 1 x 1 mm, FOV = 256 mm, TI = 900ms, FA = 8°). Functional T2*-weighted GE echo-planar imaging was used to obtain the blood oxygen level-dependent (BOLD) signal as an indirect measurement of neural activity. Thirty-two 2.4-mm axial slices were acquired (2.4mm isotropic, FOV = 240mm, TR = 2000ms, TE = 30ms, FA = 90°).

Preprocessing

Functional and anatomical data were organized into the Brain Imaging Data Structure (Gorgolewski et al., 2016). Pre-processing of the data was carried out with fMRIPrep 20.1.1 (Esteban et al., 2018, 2019) which is based on Nipype 1.5.0 (Gorgolewski et al., 2011, 2018). To ensure reproducibility using the specific software versions for fMRIPrep and all its dependencies, it was executed via its Docker container (Merkel, 2014).

Anatomical data pre-processing

Each T1-weighted (T1w) volume was corrected for intensity non-uniformity (Tustison et al., 2010; N4BiasFieldCorrection), and skull-stripping was executed (antsBrainExtraction.shv, OASIS30ANTs template). Next, brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and grey-matter (GM) was performed on the brain-extracted T1w (using fast from FSL 5.0.9; Zhang et al., 2001). Finally, brain surfaces were reconstructed (using recon-all from FreeSurfer 6.0.1; Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical grey-matter of Mindboggle (Klein et al., 2017).

Functional data pre-processing

For each of the functional runs per subject (across all tasks and sessions), the following pre-processing was performed. First, to generate a functional reference, volumes with substantial T1w contrast derived from nonsteady states of the scanner (volumes at the beginning of EPI sequence) were identified, realigned, and averaged. After skull-stripping of the functional reference volume, head motion parameters with respect to the functional reference (transformation matrices, and six corresponding rotation and translation parameters) were estimated (Jenkinson et al., 2002; mcflirt - FSL 5.0.9). On average, the maximum movement was 1.38 ± 0.21 mm.

After correcting for slice timing (Cox & Hyde, 1997; 3dTshift from AFNI 20160207), the functional reference was co-registered to the T1w reference using boundary-based registration (bbregister, FreeSurfer; Greve & Fischl, 2009). Since the volumes were within the same subject, co-registration was configured with six degrees of freedom (i.e. 3 rotations and 3 translations).

Next, all functional data were resampled onto their original, native space by applying the transforms to correct for head-motion. Several confounding time series were calculated based on the functional data: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed with the absolute sum of relative motions (Power et al., 2014) and the relative root mean square displacement between affines (Jenkinson et al., 2002). FD and DVARS were calculated for each functional run, both using their implementations in Nipype (Power et al., 2014). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. The three global signals were extracted within the CSF, the WM, and the whole-brain masks.

Using FEAT (FMRI Expert Analysis Tool; Version 6.0; FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl) the functional data were smoothed in the space domain using a Gaussian kernel of FWHM 5mm. And high-pass temporal filtering was carried out using a Gaussian-weighted least-squares straight-line fitting (sigma=50.0s).

Measurement of pRFs

We created predicted time courses of 2D Gaussian pRF models, each model was centered at a grid location (with a spacing of 0.2 degrees of visual angle) and excluding those locations falling outside the circular stimulation area. At each grid location, we included 4 models with sigma's of either 0.05, 0.3, 0.65 and 0.95 visual degrees in line with the range of V1 sigma values found by Dumoulin & Wandell (2008). Predicted time courses were formed by convolving a neural timeseries with a (double gamma) hemodynamic response function. The neural timeseries were formed following an identical procedure to Dumoulin and Wandell (2008).

Before fitting, the data were spatially smoothed using MATLAB's smooth3 function with a Gaussian kernel size of 3x3x3 voxels. For each voxel, the mean of each run was removed and the data concatenated in the time dimension before being detrended using MATLAB's detrend function with an 8th degree polynomial. For every voxel in the posterior lobe, we computed the correlation between the observed time course and each of the predicted time course. The winning model was simply the one with the one that was correlated most strongly with the voxel's time course. We did not employ any further fine fitting procedure. Indeed, we observed a high level of agreement between the borders of V1/V2 as suggested by the functional maps and the V1/V2 border of the Freesurfer atlas labels fitted based on surface curvature (Smith & Muckli, 2010; Bordier et al., 2015; Petro et al., 2023).

Regions of interest

Similar to Schwarzkopf et al. (2011) the V1 region of interest was delineated manually: the border between V1 and V2 was defined by mirror reversals in the phase map, which correspond to the representation of the vertical and horizontal meridians. The portion of V1 representing up to 10 visual degrees (i.e. the full extent of the retinotopic stimulation) was defined by the band of maximal eccentricity values, which was confirmed to follow the drop off of model (r^2) (indicating poor model fits where cortex was too eccentric to be stimulated by the retinotopic mapping sequence). The surface area of the defined region was then determined by the SurfMeasures command provided by AFNI (Analysis of Functional NeuroImages; Cox, 1996).

Analyses

Contextual modulation magnitude was calculated using the same procedures described in the Analyses section of Chapter 2. In order to corroborate our results from Canoluk et al., (2023) here we also addressed whether contextual modulations for eye matching and grating contrast detection are functionally related by submitting contextual modulation magnitudes to a Bayesian correlation

analysis in the Supplementary of this chapter. Upon visual examination of the individual fits for the contrast detection task, the data from six subjects was discarded due to inadequate fitting .

In order to investigate the relationship between V1 surface size and contextual modulations, we first collapsed the measured left and right hemisphere V1 surface areas of each subject. The collapsed surface values and contextual modulation magnitudes from each task were then submitted to separate Bayesian (Pearson) correlation analyses. We chose a Bayesian approach because of its ability to determine the strength of evidence presented by the data in supporting a particular hypothesis (H0 or H1), hence, its potential to interpret a finding, even when it is null. We computed Bayes Factors (BFs) of the correlation using BayesFactor package (Morey & Rouder, 2018) in R. A BF10 value indicates the support for H1 (e.g., correlation) over H0 (e.g., no correlation) given the data (conversely, BF01 would indicate support for H0 over H1, and can be calculated with 1/BF10). We followed Jeffreys' guidelines (Jarosz & Wiley, 2014; H. Jeffreys, 1998) to interpret the BF values: a BF10 value of 1 would mean equal amount of evidence for both H0 and H1, and therefore inconclusive. A BF10 value between 1 and 3 is considered anecdotal evidence in favor of H1 over H0, whereas a BF10 between 3 and 10 indicates substantial evidence for H1. BF10 > 10, BF10 > 30 and BF10 > 100 are regarded as strong, very strong and decisive evidence for H1, respectively.

Results

Contextual modulation magnitudes of the 23 subjects ranged from -1.08 to 2.02 (μ = -.14, σ = .75) in upright face matching, from -.47 to .60 (μ = -.09, σ = .26) in inverted face matching. After removing the subjects with poor fits in the contrast detection task the magnitudes in the contrast detection task ranged from -.38 to .21 (μ = -.04, σ = .26). It should be noted that these values are obtained through regression of regressions method (DeGutis et al., 2013), which uses the residuals of the regressed values, and thus standardized over the dataset. In this context, a zero value does not signify the absence of effect but rather represents a moderate level of contextual modulation magnitude. The comparisons of contextual modulation magnitudes across tasks are further examined in the Supplementary section of the current chapter.

The size of functionally defined V1 surface area (left and right hemispheres collapsed; in mm^2) ranged from 1400 to 2600 mm^2 (see Figure 4c.; μ = 1966.5, σ = 325.0), almost a two-fold variation within a sample of 23 subjects. We observed substantial evidence for a correlation of the V1 surface area across two hemispheres (r = .62, BF_{10} = 26.8). The contextual modulation magnitudes for each stimulus type (Upright Eye Matching, Inverted Eye Matching, Contrast Detection) were submitted to a Bayesian correlation with V1 surface area size from the MRI scans (see Figure 2). We found anecdotal evidence for the absence of a correlation between V1 surface area and Upright Face eye matching contextual modulation (r = -.22, BF_{10}= .91) suggesting the independence of the V1 size and

face-specialized complex image contextual modulations. However, there was anecdotal evidence in favor of a weak negative correlation between V1 surface area and Inverted eye matching contextual modulations (r = -.32, BF_{10} = 2.05). This suggests a link between V1 surface area and non-(face-)specialized contextual modulations. Lastly, we observed evidence for a (relatively stronger) negative correlation between V1 surface area and contrast detection contextual modulations (r = -.4, BF_{10} = 3.01). Similar correlations were observed across hemispheres when analyzed separately.

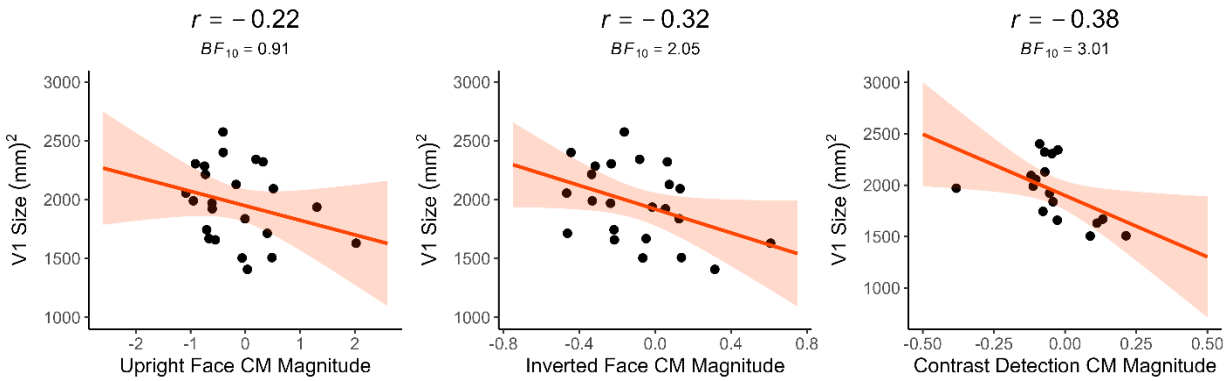

Figure 2. Correlations between V1 size and contextual modulation magnitudes of three stimulus types.

Discussion

Over recent years, there's been a growing interest in the influence of cortical anatomy and brain structure on visual perceptual functions (Kanai & Rees, 2011; Song et al., 2015; Himmelberg et al., 2022; Song et al., 2022; Genon et al., 2022). In particular, variations in the magnitude of visual contextual illusions have been associated with the functionally defined size of V1 (Schwarzkopf et al., 2011; Song et al., 2013). Although most past studies used simplistic stimuli, contextual modulations in face processing, or 'holistic processing', display key similarities to lower-level processing mechanisms. Moreover, the magnitude of these modulations varies widely across individuals, a phenomenon yet to be explained. Building upon the past research on the link between V1 anatomy and contextual illusions, the current study aimed to link the magnitude of contextual modulations in low-level (i.e., gratings) and high-level stimuli (i.e., upright and inverted faces) to the size of V1. Through means of fMRI and psychophysics, we investigated the amount of variation in low- and high-level contextual effects that is explained by V1 surface area. Low-level contextual effects were measured via a grating contrast detection task. For high-level contextual effects, upright and inverted facial feature matching tasks were used. Upright faces activate specific high-level regions, suggesting they might be processed by unique mechanisms specialized to the natural statistics of the upright face stimulus (George et al., 1999; Goffaux et al., 2016; Iidaka, 2014; Tong et al., 2000; Yue et al., 2011) . However, a connection to lower-level regions would indicate face contextual modulations are not

limited to these specialized areas but can also be partly explained by contextual mechanisms in early processing regions. Inverted faces activate mid- and high-level brain regions like general objects (Kanwisher et al., 1998; Gilaie-Dotan et al., 2010), and their relationship to early regions can shed light on how non-specialized complex images are contextually modulated. V1 size of 23 individuals and its link to the contextual modulation magnitudes measured in these three types of stimuli were investigated in the current study, in order to understand the impact of V1 size on the variations in the magnitudes of contextual modulations at these three levels of processing.

We observed no correlation between V1 size and upright face contextual modulation magnitudes. When the visual system aligns with the internal representation of an upright face, the face-specific areas become so dominant in their engagement that they might overshadow any processing that originates from lower-level regions. Indeed, this result aligns with the strong body of evidence showing that faces are "special", and are processed by a domain-specific module which almost works in isolation from other regions (Kanwisher, 2000; Downing et al., 2006). . However, there is evidence of a sustained engagement of V1 for the processing of complex stimuli such as faces. Using backward masking and neuroimaging, Schuurmans et al. (2023) recently found that high-level regions specialized for face recognition continually engage with low-level regions to enhance and refine the preliminary image representation over time. The engagement of high-level regions drastically reduced when the subjects were presented with contrast-negated stimuli. Despite being quite robust when defined functionally, V1 size is a static measure that might not capture the real-time dynamics of this interaction. Future research using transcranial magnetic stimulation (TMS) to virtually lesion V1 could better evaluate the relationship between low-level contextual modulations and face-specialized regions. It can be hypothesized that lesioning V1 right after the initial processing in order to disrupt recurrency may lead to changes in upright face processing, manifesting the nature of connections between low- and high-level face-specialized regions.

Interestingly, we found anecdotal evidence for a weak negative correlation between V1 size and inverted face contextual modulation magnitudes. Inverted faces have a less sparse activation profile compared to upright faces (Young & Yamane, 1992; Rolls & Tovee, 1995); they cause weaker engagement of face-specialized regions, with a broader spread of activation across mid-to-high level regions that are more responsive to other object categories (Kanwisher et al., 1998; Yovel & Kanwisher, 2005; Downing et al., 2006; Gilaie-Dotan et al., 2010; Goffaux et al., 2013; James et al., 2013; Taubert et al., 2015; Liu-Shuang et al., 2022). Inversion has also been shown to reduce the size of the perceptual integration window (Rossion, 2008, 2009; Van Belle et al., 2010), leading to a more localized processing of the facial features. This aligns with neuroimaging evidence that inversion reduces the size of the population receptive field (pRF) and overall visual field coverage of face-specialized regions such as inferior occipital gyrus (IOG-faces), posterior fusiform gyrus (pFus-faces/FFA-1), mid-fusiform gyrus (mFus-faces/FFA-2) and posterior end of the superior temporal

sulcus (pSTS-faces; Poltoratski et al., 2021). Furthermore, this negative correlation is a sign of a possible link between V1 size and complex image contextual modulations, therefore further research on such links is worth investigating with other object categories that do not have a specialized activation profile in high-level regions.

Finally, we found a relationship between the magnitude of orientation-dependent contextual modulation of contrast and V1 size. Orientation-dependent contextual modulations are considered to be processed mainly at lower-levels of the visual processing hierarchy, specifically due to their simplicity (with highly controlled gratings as opposed to complex images). To the best of our knowledge, this is the first time such a relationship is demonstrated between V1 anatomy and a contextual illusion that modulates the perceived contrast. Considering how past research mainly focused on the relationship between V1 and size/shape related illusions (e.g., (Kanai & Rees, 2011; Schwarzkopf et al., 2011; Song et al., 2011), this novel finding warrants further investigation into the interactions across low-level features. Behavioral and perceptual consequences of variations in the size of cortical regions was also investigated by Song et al. (2013). In a series of psychophysical experiments and measurements of surface sizes of V1, V2 and V3, they explored the relationship between the size of these regions and local and contextual processing of orientation, contrast and luminance. Although the conventional scaling hypothesis, as described by Ringo (1991) and Kaas (2000) seems to encompass any kind of spatial relationship between low-level visual properties, due to their varying nature in intracortical connectivity, visual properties might be impacted differently from surface size constraints. For example, neurons that have similar orientation preferences are more closely interconnected. However, specific V1 neurons do not show a preference for certain luminance gradients or contrast levels, so they do not have stronger connections based on those attributes. As such, Song et al. (2013) found that increased V1 size was indeed associated with a reduced contextual modulation for the processing of orientation, but not for the processing of luminance or contrast. They suggest that the impact of V1 size may be limited to specific features, such as orientation and size, and is not universally applicable to all visual properties. The low-level paradigm employed in our study focused on orientation-dependent contextual modulation of contrast. Although it primarily measured the magnitude of contextual effects on contrast, it heavily relied on orientation-specific mechanisms, which likely explains its correlation with V1 size.

In the psychophysics experiments (which employed the same design as the study in Chapter 2 in a partially new sample), we replicated several aspects of the past findings (results in the Supplementary). The detection and discrimination thresholds were higher when the subsequently displayed targets were presented within the same face context in the Upright Face Matching and when the target grating was presented within a larger grating of same orientation in the Contrast Detection tasks, demonstrating the impact of context on the processing of local elements at high and low-levels of the visual processing hierarchy. As expected, inversion drastically reduced the effect of context on

local feature processing. In contrast to our past study, however, we found evidence for a weak correlation between upright face and contrast detection contextual modulations. However, it must be noted that the sample sizes differed across studies (59 in the previous study, 17 in the current after removal of 6 subjects with poor fits). This finding is more likely to be a Type-II error due to the limited sample size. We replicated the correlations of inverted face contextual modulations to both contrast detection and upright face contextual modulations. Specifically, the relationship between inverted faces and contrast detection is in line with their connection to V1 size, and their relationship to upright faces further corroborates the amount of overlap between specialized and non-specialized face processing, as was evidenced in past research (Kanwisher et al., 1998; Yovel & Kanwisher, 2005; Taubert et al., 2015).

Previous research has indicated either no correlation or, at most, a weak and anecdotal correlation in the magnitude of contextual modulations measured through behavioral measures, even with large samples. This finding holds true regardless of whether the measurements are confined to the domain of basic stimuli (Cretenoud, Francis, et al., 2020, N > 300; Samaey et al., 2020, N = 52) or extend to more complex stimuli (Konar et al., 2010b, N = 48, Chamberlain et al., 2017, N > 105; Boutet et al., 2021, N > 200). In our case, the current sample of 17 participants is unlikely to have sufficient statistical power to make any inferences for the correlations across behavioral measures. However, studies relating V1 anatomy to behavioral estimations were more likely to find correlations despite smaller samples (Schwarzkopf et al., 2011, N=30; Schwarzkopf & Rees, 2013, N=26; Song et al., 2013, N=20), possibly due to the robustness of the measurements obtained through functionally defined surface size of V1 (Sereno et al., 1995; as opposed to measurements that are based purely on anatomical landmarks, which proves to be less likely to explain inter-individual differences as seen in Schwarzkopf & Rees (2013)). Our finding of a substantial evidence for correlation between V1 size and contrast detection task is therefore more likely to reflect a real effect despite a sample size of 17. Another important point to mention is that we resorted to a coarse fitting procedure in our pRF measurements. For every voxel in the posterior lobe, we computed the correlation between the observed time course and each of the predicted time course. The winning model was simply the one with the one that was correlated most strongly with the voxel's time course. While more sophisticated parameter selection methods exist (e.g., model fitting on cortical surface space with sequentially decreasing levels of smoothness applied to the data), we are confident that any noise introduced into parameter would not systematically impact the large-scale organization of the retinotopic maps produced from these model fits. Therefore, we did not see a necessity for a further fine fitting procedure. Moreover, even though this may have impacted the measurements, it is unlikely that people with larger/smaller V1s would be impacted differently by our V1 size definition procedure.

In conclusion, our study contributed to the literature showing that V1 size can explain the magnitude of size (Schwarzkopf et al., 2011; Schwarzkopf & Rees, 2013) or orientation (Song et al., 2013)

illusions, and showed that variations in V1 size can also predict the magnitudes of a contextual orientation-dependent contrast illusion. Furthermore, variations in V1 size are likely to predict the variations in the magnitude of contextual modulations in high-level, non-face-specialized tasks, as was evidenced by the correlations in inverted facial feature matching task. However, these size variations in V1 did not explain upright facial feature matching contextual modulations, possibly due to the strong engagement of face-specialized regions, which may work more independently from low-level contextual mechanisms implemented in V1. When inversion disrupted face-specialized processing, high-level, non-face-specialized contextual modulations were more functionally linked with low-level mechanisms. These results bring further support to theories claiming a special status for upright face processing, and face contextual modulations.

Supplementary

Figure S.1 demonstrates the z-scaled threshold values for tasks involving eye matching in both upright and inverted faces, as well as contrast detection. As anticipated, the thresholds for contrast detection increased when the grating context was isotropically oriented (same) compared to an orthogonally oriented (different) context. A similar trend was observed in the eye matching task: thresholds were increased when eyes were matched within a consistent (same) facial context as opposed to varying (different) facial contexts. Regarding the Isolated condition, its relative ranking compared to the Same and Different context conditions was consistent in the contrast detection and Upright eye matching tasks, but it differed in the Inverted eye matching task. The threshold difference between Same and Different contexts was larger in the contrast detection task than in eye matching tasks (HDI [2.56, 3.48]). In the eye matching task, the threshold difference between Same and Different context was non-negligible for Upright Faces (HDI [.85, 1.73]), while it was negligible for Inverted Faces (HDI [-.21, .51]), corroborating past evidence that inversion attenuates contextual influences on the encoding of features.

In different tasks, the ranking of the Same and Different context conditions remained fairly consistent. However, the Isolated condition's ranking showed more variation. For the contrast detection task, the thresholds for Isolated and Different were similar, with a slight edge for the Different context condition (HDI [-.35, -1.24]). This suggests that detecting local grating contrast is nearly as effective in an orthogonal context as when observed in isolation. Yet, for the upright eye matching task, the Isolated condition thresholds were more aligned with the Same context than the Different one (HDI [.17, .80] and [-.37, -1.25], respectively). For the Inverted eye matching, the variation between context condition thresholds was lesser, and negligible (HDI [-.18, -28], [-.15, .59] for the comparisons of Same-Isolated and Different-Isolated respectively). Inverted faces are processed more independently of their context compared to upright ones.

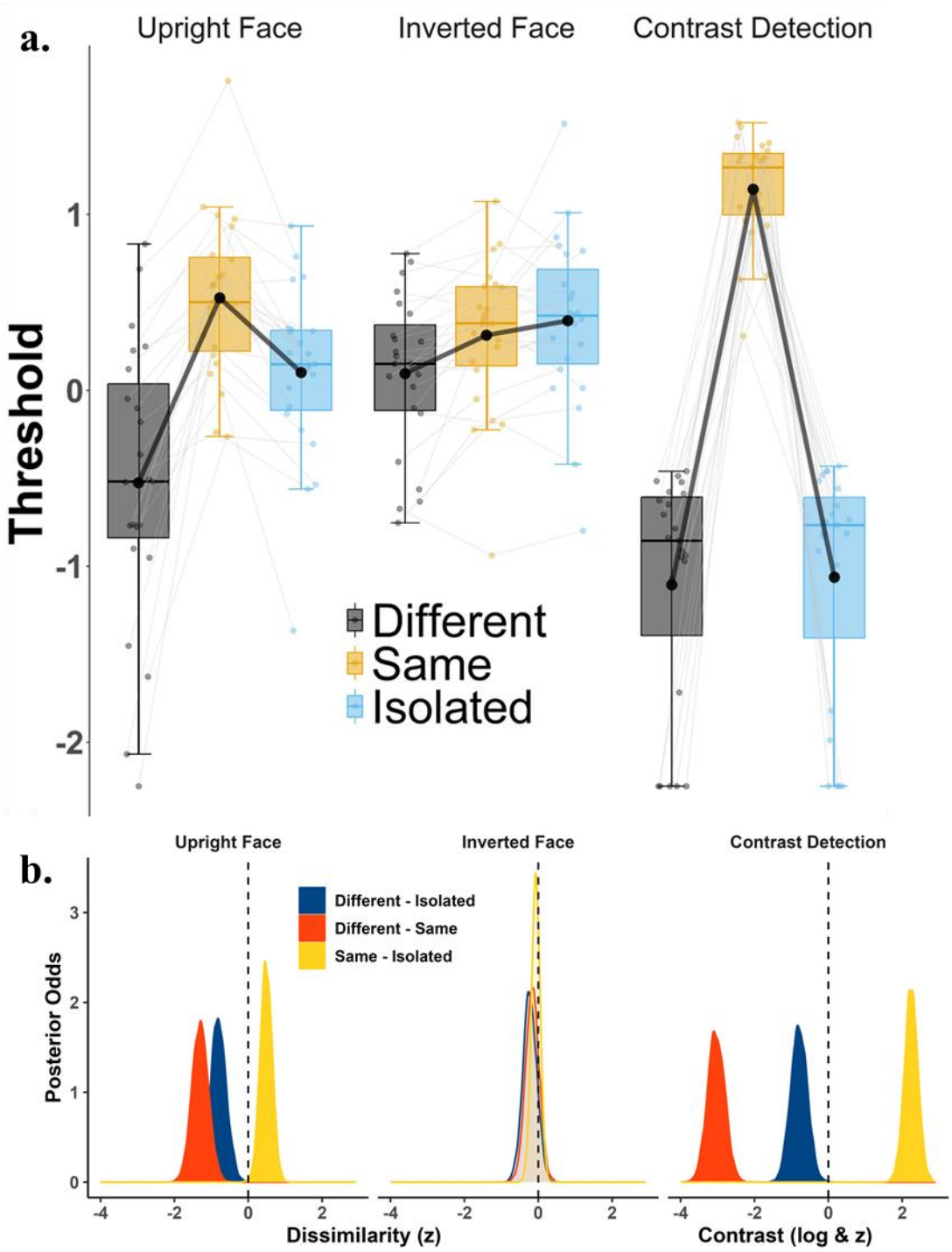

Figure S.1. ***a.*** *Z-scaled threshold estimates in eye matching (left) and contrast detection (right) tasks. Each dot represents the z-scaled threshold estimate for a given task and condition for a participant. The top, middle and bottom lines of boxes correspond to the first, second (median) and third quartiles (25th, 50th and 75th percentiles) respectively. The upper and lower whiskers extend until 1.5* Inter-Quartile Range. Faded grey lines connect the thresholds in respective conditions for a given participant. Bold black dots and lines display group means.* ***b****. Posterior odds of differences across the three tasks. The density plots display the posterior odds of the threshold differences. 95% interval of the distributions did not overlap with zero, therefore all differences were non-negligible in Upright Face and Contrast Detection tasks, but in Inverted Face, the difference between conditions were practically zero.*

The contextual modulation magnitudes for each stimulus type (Upright Eye Matching, Inverted Eye Matching, Contrast Detection) were submitted to a Bayesian correlation. Six participants' data in contrast detection task was considered outliers as per visual inspection of their psychometric fits and their threshold values, and were excluded from further analyses. Their data in other tasks were not outliers, and were used fully in the Upright Face and Inverted Face correlations, and were included in the V1 surface correlation analyses. We found anecdotal evidence for a correlation between Contrast Detection and Upright Face Matching (r = 0.3, BF_{10} = 1.48), anecdotal evidence for a correlation between Contrast Detection and Inverted Face Matching (r = 0.3, BF_{10} = 1.48), and decisive evidence for a correlation between Upright Face and Inverted Face Matching tasks (r = 0.59, BF_{10} = 99.6).

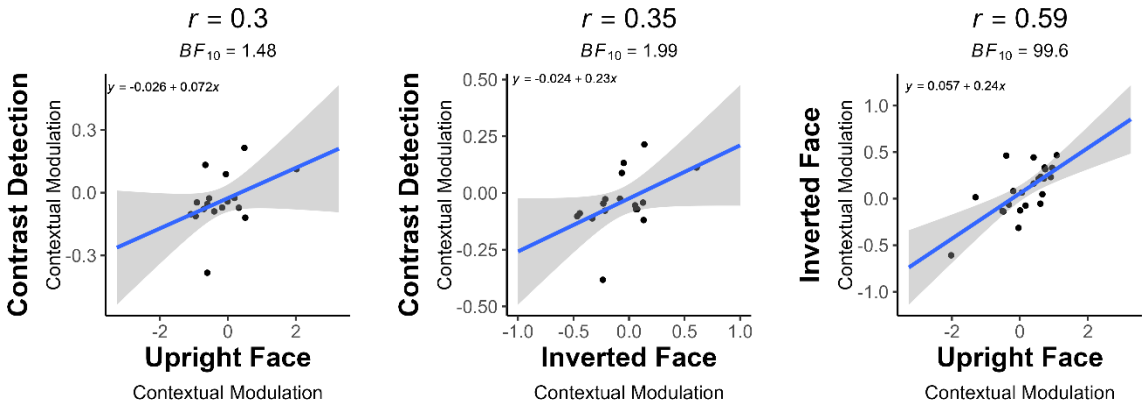

Figure S.2. Correlations of contextual modulation magnitudes across tasks.

The contextual modulation magnitudes for each stimulus type (Upright Eye Matching, Inverted Eye Matching, Contrast Detection) were put in a Bayesian correlation with V1 surface area size (in mm^2) in each hemisphere (see Figure S.3). We found anecdotal evidence for the absence of a correlation between both left and right V1 surface area and Upright Face eye matching contextual modulation (r = -.19, BF_{10}= .76 and r = -.21, BF_{10}= .84, respectively) further suggesting the independence of the V1 size and face-specialized complex image contextual modulations. There was anecdotal evidence in favor of a weak negative correlation between both left and right V1 surface area and Inverted eye matching contextual modulations (r = -.25, BF_{10} = 1.15 for left), with stronger evidence for a correlation in right V1 (r = -.33, BF_{10}= 2.18 for right V1). This suggests a link between V1 surface area and non-(face-)specialized complex image contextual modulations. Lastly, we observed evidence for a (relatively stronger) negative correlation between both left and right V1 surface area and contrast detection contextual modulations (r = -.35, BF_{10} = 1.95 for left), again, with stronger evidence for a correlation in the right V1 (r = -.38, BF_{10}= 2.76 for right V1).

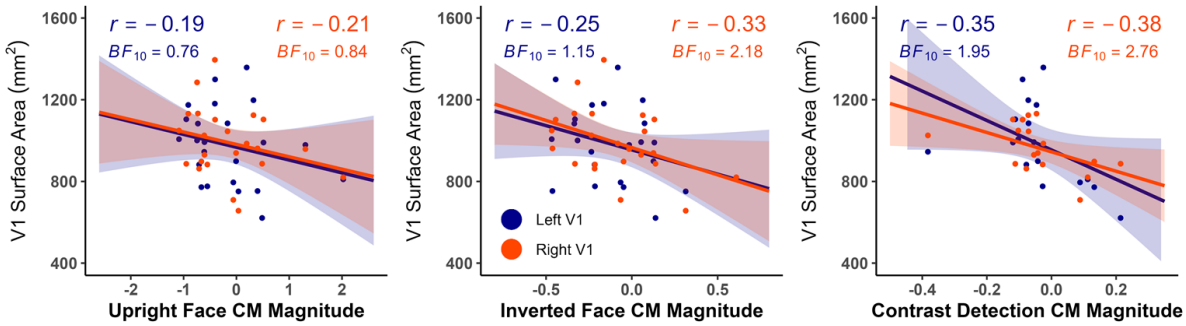

Figure S.3 Correlations between Left and Right V1 size and contextual modulation magnitudes of three stimulus types.

Bayes Factor Robustness Analyses

In Bayesian statistics, the interpretability of outcomes and the definitions of H0 and H1 are directly linked to, and relies significantly on the selection of prior distributions. Greater information availability allows for more precise specification of "informed" priors. Nevertheless, when informed priors significantly differ from default priors, it becomes increasingly essential to provide justifications. In such situations, Bayes Factor Robustness checks can be utilized to evaluate the degree to which the results are influenced by the particular choice of prior distribution. In the current study, we opted for a stretched beta distribution with a width of 0.33. Traditionally, a stretched beta prior distribution with a width of 1 is used as a non-informative prior. In this distribution, any parameter value over the range of correlations between -1 and 1 are equally likely. However, it has been shown that this choice of priors can produce Bayes factors that can be biased towards null models (Vanpaemel, 2010; Quintana & Williams, 2018). In order to justify our specification of priors and the robustness of our results to different prior specifications, below we provide the prior/posterior distributions and Bayes Factor Robustness checks (Figure S.3) for the three main correlation analyses we ran using contextual modulation magnitudes from different tasks.

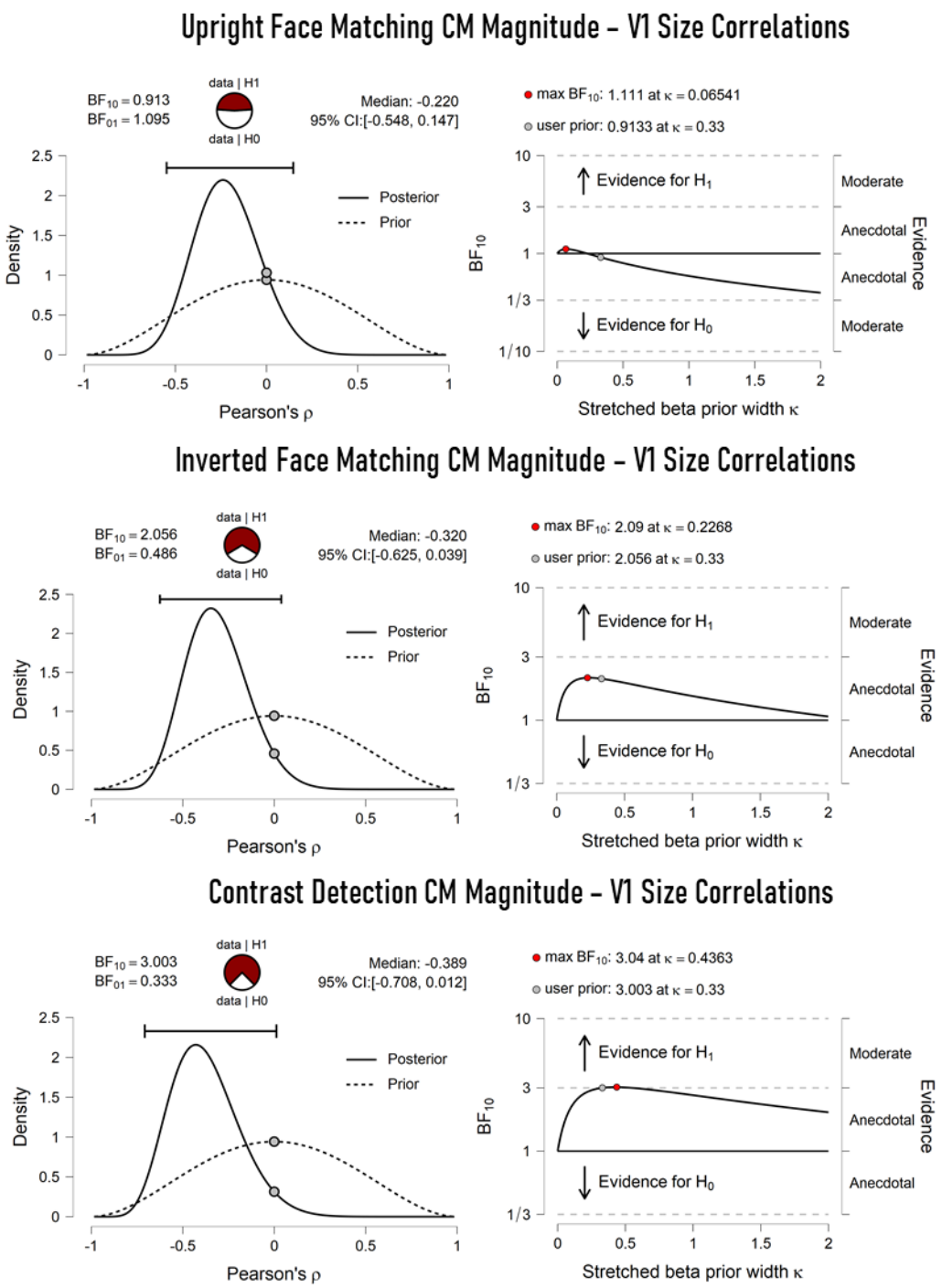

Figure S.3. Prior/posterior distributions and Bayes Factor Robustness checks for CM magnitude and V1 size correlation analyses. *Left column displays the selected prior distribution with dashed lines. Combining this prior distribution with the observed data forms the posterior distribution (solid lines), and the Bayes factor is the ratio between the marginal likelihoods of the null model (no correlation) and the alternative model (correlation), given the data. The pie charts illustrate the ratios for H1 and H0. No specific minimum value for "correlation" is employed. Right column displays Bayes Factor robustness checks. The figures display the changes in* BF_{01} *value as a function*

of specified beta prior width. As can be seen in all three figures, regardless of the informativeness of the specified prior, the results are in the same direction in most cases.

Chapter 5

General Discussion

Summary of findings

The primary objective of this thesis was to delve into the contextual modulations at both the lower and higher stages of the visual processing hierarchy. Holistic processing, a high-level contextual modulation mechanism, is a hallmark of face processing, and has been proposed to mark the recruitment of face-specialized regions. It can be easily disrupted by inversion, an image manipulation that largely preserves image properties processed at primary stages of vision. The preservation of low-level properties with inversion directed the traditional focus to high-level mechanisms to account for face holistic processing. Due to this focus, holistic processing has been considered independent from low-level mechanisms and its connection with the early regions was not adequately explored. In the present thesis, we examined their relationship to early regions through behavioral and neuroimaging methods. Contextual modulations at both of these distinct stages demonstrate a reliance on local input intensity when it comes to contextual modulations. This shared fundamental principle provided a refined lens for our combined studies. Inversion was observed to reduce the strength of face contextual modulations, decrease the response of face-specific brain regions, and lead to brain activation patterns more similar to non-face object categories. This makes inversion a valuable tool for exploring contextual modulations in advanced, non-specialized processing stages. Therefore, by testing inverted faces alongside upright faces, we were also able to investigate the interactions between low level regions and non-specialized, high-level regions. The interactions between these regions were investigated through an inter-individual differences framework throughout our studies.

Contextual modulations in the primary stages of visual processing are influenced by the strength of local input. Similarly, in high-level stages of facial processing, the discriminability of a facial feature dictates the extent of the face context's impact on that feature. Despite these aligning observations, the emergence of high-level contextual modulations from primary mechanisms remains ambiguous, largely due to limited empirical research exploring the functional relationship between them. In Chapter 2, we behaviorally explored their relationship with low-level contextual modulations in psychophysical tests. We measured the shared variance in the magnitude and profile of contextual modulations in upright and inverted face matching and grating contrast detection. The variance in the magnitude of upright face contextual modulations was not explained by low-level mechanisms, which

points to their independence from early regions. Conversely, the variance in the inverted face contextual modulation magnitudes was partially explained by these mechanisms. This indicates a closer association with the early visual regions. Despite the difference in their connection to low-level mechanisms, we found that contextual effects in inverted faces account for 37% of the variability in upright face contextual modulations. This finding resonates with the studies that show brain reactions to inverted faces engage similar regions as those activated by upright faces. However, the significant variability in upright face contextual modulation that persists even after considering the variability in inverted face contextual modulations underscores the unique functional nature of the contextual mechanisms triggered by this visual category. The measurement of shared variance in the magnitude of contextual modulations was to study their functional dependence. In order to examine whether they operate with similar principles across different levels of the hierarchy, we also analyzed their profile. Contextual modulation profiles showed strong correlations across tasks. This reinstates that these distinct mechanisms indeed share functional similarities despite their independence.

There were a few points in the experiments employed in Chapter 2 that may have obscured an existing relationship. Among other things, the tasks differed in their definition of local input strength, and foveal position. Namely, the local input was related to shape similarity in the facial feature matching tasks, and to oriented contrast in the grating contrast detection task. This disparity in local input strength may indeed account for the absence of such a relationship, as contextual illusions with comparable local input characteristics tend to have a greater tendency to exhibit a correlation (Grzeczkowski et al., 2017). Furthermore, the eccentricity in which the local input was presented at differed across tasks, with one shown foveally and the other peripherally. Contextual effects in the fovea and periphery are based on distinct, non-intersecting mechanisms (Xing & Heeger, 2000, 2001). In Chapter 3 we proposed a registered report that refines the behavioral methods from Chapter 2 to enhance measurement sensitivity. In the proposed study, we aligned both how the local input strength was defined, by using contrast as the independent variable, and the eccentricity of the targets, which resulted in foveal presentation for both tasks. Therefore, the examination of the same relationship using these alignments will present a more comprehensive picture of the mechanisms underlying face contextual modulations.

The receptive field coverage of feedback connections to V1 expands progressively as the distance from the region providing the feedback increases (Angelucci et al., 2002). These connections operate at an exceptionally high speed, surpassing even the velocity of horizontal connections within V1, allowing them to exert an influence on intra-V1 horizontal connections (Girard et al., 2001). Our contrast detection task primarily depended on near-surround contextual modulations. Considering the distance of the regions that are engaged for inverted and upright faces from V1, it is noteworthy that the correlation observed between contextual modulations in inverted faces and contrast detection might be attributed to the activation of more similar near- and far surround effects. In contrast, the separation of

upright faces from early regions may be attributed to their greater dependence on the far-surround and their recruitment primarily from high-level regions with expansive receptive fields (see Figure 1). Indeed, this account is highly speculative, and requires deeper investigation. A possible approach to address and distinguish these factors in the future would be to examine them in tandem with low-level contextual effects that exclusively depend on near and far surround modulations, or by manipulating the distance between the local input and its surround in faces.

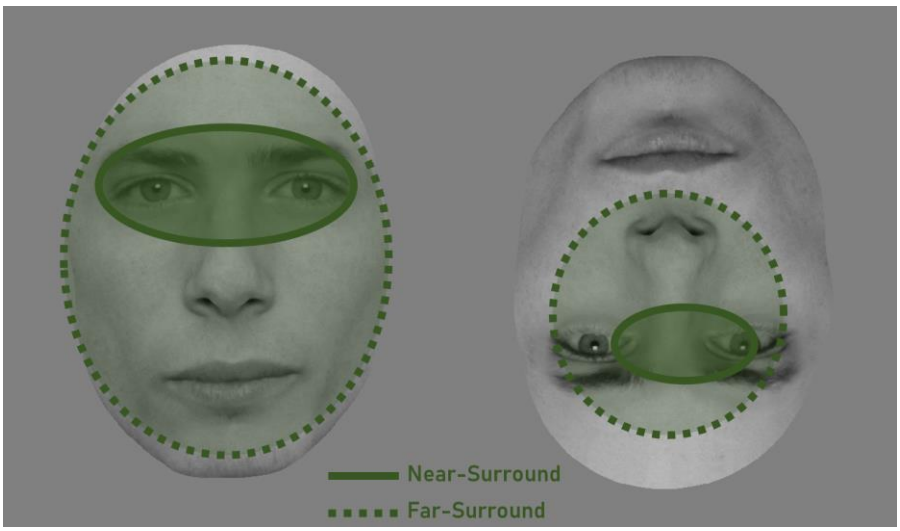

Figure 1. Possible sizes of near- and far-surround modulations elicited by upright and inverted faces. The feedback connections resulting from regions activated by upright faces lead to near- and far-surround modulations that extend over more distal regions due to their relative distance from V1. In contrast, inverted faces cover a smaller field in terms of these modulations.

Finally, the variability in the magnitude of contextual modulations at low-levels can be attributed in part to differences in the functional size of V1; specifically, larger V1s tend to have reduced contextual modulation magnitudes. Yet, for high-level contextual modulations observed in upright faces, the reasons for such variability remain largely unexplained. Chapter 4 honed in on the influence of V1 anatomy on the magnitude of both low- and high-level contextual modulations. Studies investigating correlations between contextual modulations only through behavioral methodology rarely found correlations across various kinds of illusions or contextual effects (e.g., Konar et al., 2010; Cretenoud et al., 2021; Cretenoud, Francis, et al., 2020; Cretenoud, Grzeczkowski, et al., 2020; Boutet et al., 2021; DeGutis et al., 2013). However, several contextual effects have been found to be related to the size of V1 (e.g., Schwarzkopf et al., 2011; Schwarzkopf & Rees, 2013; Song et al., 2013). Leaning on the literature on the spatial integration principles in V1 (Ringo, 1991; Kaas, 2000), we investigated the amount of variance in upright/inverted face and grating contrast detection contextual modulation magnitudes that can be explained by the size of V1 in a group of individuals. V1 size did not impact the magnitude of contextual modulations in upright faces. As was also apparent from the behavioral findings of Chapter 2, it appears that when the visual input matches the internal model of an upright face, face-specific regions activate so dominantly that they might drown

out the influence from the lower-level areas, making their correlation negligible. This further suggests that upright faces may indeed be processed by specialized mechanisms. The association with V1 size was more evident in inverted face contextual modulations, suggesting a potential relationship between V1 size and complex image contextual modulations. Thus, exploring this connection with other object categories, especially those without a specialized activation pattern in high-level regions warrants further research. Considering the pronounced relation of our low-level task to early processing and V1 compared to both upright and inverted faces, it likely had minimal engagement with higher-level processes, and the connection was more apparent.

In essence, this thesis sought to explore the interactions of low-level contextual mechanisms, as well as the anatomy of early visual regions with contextual modulations observed in face-specialized and non-specialized high-level regions. Subsequent sections will further elaborate on the methodological approaches employed in this thesis, such as the inter-individual differences framework, the advantages of using Hierarchical Models and the regression method used to obtain more sensitive estimates. Furthermore, it will aim to examine the reasons behind the absence of correlations between low-level and high-level, face-specialized contextual mechanisms. The intent is to amalgamate the findings, offer a critical assessment of the methods used, and position this research within the wider field of face holistic processing and vision science in general.

Leveraging inter-individual differences to understand underlying processes

Research in human cognition and neuroscience often regard variations between individuals as noise, and usually aim to eliminate them by averaging data across a group of participants. However, if these variations persistently appear across diverse tests, they signify individual traits that might be indicative of distinctions in brain function (see Figure 2; Kanai & Rees, 2011; Peterzell, 2016). Investigating these distinct brain functions across individuals provides an invaluable method to understand the mechanisms and factors underlying an observed phenomenon.

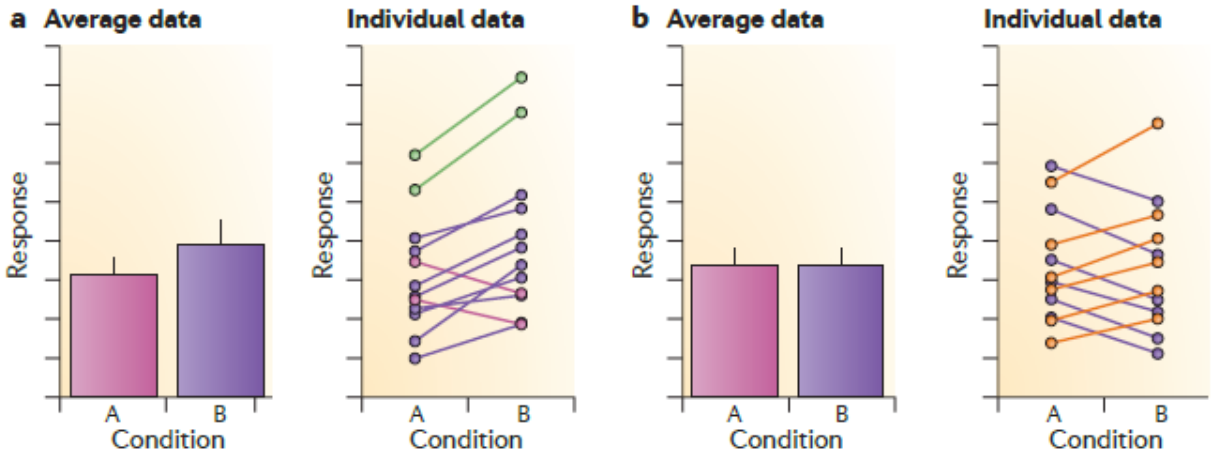

Figure 2. Two examples of average and individual data. ***a.*** *Example dataset; while the average data shows a difference between condition A and B, individual data shows that some participants showed opposite trends (pink lines), and some showed much higher difference (green lines).* ***b.*** *The mean responses between two conditions do not show a difference, however, when individual data is plotted, two distinct patterns become evident in the data, showing opposite trends. These distinct patterns inside the data can be used as a source of information. Figure from Kanai & Rees, (2011)*

In an inspiring paper, Geisler & Chou, (1995) explained how this approach can in fact be leveraged to assess the role of low-level factors in complex tasks. By jointly investigating the same phenomenon observed in the low-level and high-level stages, one can assess how much of the variance in the high-level task can be predicted by the variations in the low-level task. Determining the relative impact of the low-level factors is especially crucial in the study of face processing, because most of the inter-individual differences studies in this literature focused on high-level explanations for the underlying mechanisms (except for a few, e.g. Barton et al., 2004). Recently, the individual differences approach has been applied in the field of face processing (Wang et al., 2012; DeGutis et al., 2013; Yovel et al., 2014; Wilmer, 2017). These studies have been mainly used to address whether various measures of holistic processing can be used to predict face identity recognition abilities, or whether paradigms measuring holistic processing really reflect the same mechanism. Early regions, such as V1 and V2 are proposed to recurrently contribute to the perception of complex stimuli (e.g., Bar, 2003, 2004; Schuurmans et al., 2023). Furthermore, V1 anatomy has been linked to the contextual modulations seen in basic images (Schwarzkopf et al., 2011). The relative distance in the hierarchy between processing stages in the high-level, face-specialized regions and low-level regions possibly brings more noise in the data than investigating an effect within a single processing stage. For that reason, in the present thesis, we did not only seek for the origins of high-level contextual modulations in the lower-levels of the visual hierarchy, but while doing so, we aimed to attain the highest statistical sensitivity and control as we could exploit in individual difference studies by using hierarchical modeling.

Hierarchical Modeling: a method to obtain higher reliability and sensitivity in the study of individual differences

In the present thesis, we leveraged the sensitivity and high reliability provided by hierarchical models. Employing these models enabled the fitting of trial-by-trial data from each individual subject in a single, large model, preserving the richness and informativeness of the data at the individual trial level instead of aggregating it across multiple trials, which would have diminished its informativeness by condensing it into a single data point (as demonstrated in Efron & Morris, 1977). . Moreover, in our hierarchical model we entered the effect introduced by an individual itself as random effect factor. Doing so, the threshold estimates in the model was not only informed by the complete dataset, but also when the thresholds of an individual were being estimated from one task, the estimates were informed by the estimates of the same individual in other tasks, painting a more informed picture of the individual's behavior. Therefore, using this model provided a more sensitive and robust measure compared to simple linear models. Considering the difficulties past studies had in observing correlations in inter-individual differences studies even with within-level contextual effects (e.g. in various low-level size illusions as in Grzeczkowski et al., (2017), or in high-level, measures of holistic face processing as in Konar et al., (2010a) or Boutet et al., (2021)), we believe that employing such a sensitive measure provides a richer metric of the effect being investigated, and has a potential to reveal correlations previously undetected. The correlation we observed across processing levels, particularly in contrast detection and inverted face matching tasks underscores this potential.

Commonly, research examining individual differences aims at determining correlation between performance on diverse tasks, seeking shared cognitive/perceptual processes. Yet, the results are often underwhelming, with low correlations even for seemingly related tasks. Researchers who study individual differences rely strongly on scores derived from experimental tasks (e.g., accuracy, RT, d'). Many of the tasks used with this framework (e.g., Stroop task) produce reliable results. In experimental context, a "reliable" effect is one that can be consistently replicated in varying experimental conditions, observed in the majority of study participants, and yields steady effect sizes. For instance, in the "Many labs 3" project (Ebersole et al., 2016), which assessed the reproducibility of effects across various labs using identical procedures, the Stroop effect was reproduced in 100% of the attempts, a success rate markedly higher than most other effects tested. In fact, the reliability in some of such tasks is so pronounced that they are considered nearly universal (MacLeod, 1991). Reliability plays a pivotal role in individual differences research as it imposes a constraint on the correlation that can be detected between two measures when their reliability is in question (Nunnally Jr, 1970; Spearman, 1904). Beyond their reliability, these tasks are expertly designed to spotlight distinct cognitive functions, ensure a commendable degree of internal validity, and the extensive body of literature surrounding many of these tasks add to their appeal, in both their implementation and

interpretation of results. Measures of holistic processing, especially the complete version of the composite task has been shown to have good validity. However, the reliability of these measures are reportedly low (Richler et al., 2014; Ross et al., 2015). Various holistic processing measures, such as the part-whole task, composite task, and face inversion effect, are thought to access somewhat distinct yet overlapping mechanisms (Boutet et al., 2021). For instance, the part-whole effect relates to the processing of an overarching face template (Leder & Bruce, 2000; Maurer et al., 2002; Mondloch et al., 1999), whereas the composite effect concerns the depiction of spatial relationships between facial features (Andrews et al., 2010; Meinhardt-Injac et al., 2014; Richler & Gauthier, 2014). These tasks, generically assumed to measure holistic processing, has low correlation among each other. The part-whole task shows no correlation to the partial version of the composite task (Rezlescu et al., 2017; Wang et al., 2012), a moderate correlation to the complete version of the composite task (DeGutis et al., 2013), and a weak correlation to the face inversion effect.

There are two possible explanations for these low correlations. The first is that the issue is fundamentally substantive; the observed low correlations in these studies with large samples might indicate that the tasks are tapping into genuinely distinct mechanisms. The alternative account suggests that the absence of correlation in between these tasks is primarily statistical in nature. Strong correlations only manifest when tasks exhibit high reliability, and low correlations across tasks with low reliability cannot be meaningfully interpreted (Rouder & Haaf, 2019). Another important statistical factor that may contribute to this is the portability of a given test If a test has the portability characteristic, the inherent population value of the tool remains unchanged regardless of the sample size chosen by the researcher. When investigating individual differences, it is common practice to consolidate an individual's data into a singular performance score, such as d’ or averages of accuracy or RT. While it's easy to assume that this averaging method is sufficient and any discrepancies will be balanced out, evidence suggests otherwise (Rouder & Haaf, 2019). Averaging might significantly weaken effect size and correlation measures. Hierarchical modeling offers an alternative analytical method for measuring task performance. This method examines how individuals’ performances vary across tasks in a more informed and comprehensive way by accounting for variations in performance on a trial-by-trial basis.

The metric for effect sizes within a given task and correlations between tasks are heavily influenced by design decisions, specifically the number of trials assigned per individual for each task (Green et al., 2016). Consequently, when researchers opt for varying numbers of trials per task, which is often the case, they may in fact end up examining distinct mechanisms (Rouder & Haaf, 2019). One problem especially in inter-individual difference studies is how the portability of a given task is underestimated. For example, if a tool is considered to have a test-retest reliability of .70, this value is accepted as a consistent truth across all samples. However, the reliability of a test is generally heavily dependent on sample size and number of trials (Hedge et al., 2018). This dependence points to low

portability. The challenge of portability can be effectively addressed through the utilization of established hierarchical statistical models (Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). Even in situations where study participants undergo a restricted number of trials, and their data may be subject to trial-specific noise, hierarchical models provide a viable solution. These models enable us to generate predictions for scenarios involving a substantially larger number of trials. In such simulations, our estimates of variability remain uninfluenced by the noise associated with individual trials. Consequently, the effect sizes and correlations derived under these conditions can be regarded as robust and dependable metrics, irrespective of the initial trial quantity.

Another advantage of hierarchical models is known as "shrinkage" (Efron & Morris, 1977). Due to their (hierarchical) nature, they recognize that individual data points (i.e., trials in this case) do not exist in isolation; they are a part of a larger structure (e.g. all trials belonging to the same individual subject, or all trials belonging to the same task). Therefore, individual estimates are informed by both their own data, and the data of the group they belong to. Shrinkage refers to the phenomenon in which estimates in a hierarchical model are "pulled" or "shrunk" towards a group mean or overall mean. In other words, rather than taking each individual estimate at face value, shrinkage introduces an adjustment based on the overall group behavior. This function provides a way to stabilize the estimates by drawing them closer to the group mean (see Figure 3). Especially in the case of outliers, they cause these values to not introduce drastic changes in the group estimate. Since the outlier data is informed by the overall group, it becomes more likely to reflect real life results, making it less likely to be completely removed. By borrowing strength from the overall group, shrinkage can prevent overfitting to noisy individual-level data. As individual estimates converge towards a group mean, the accuracy of predictions typically improves. This phenomenon is particularly pronounced in scenarios with limited data, where a single aggregate value is more prone to contain elevated levels of inherent noise (Moors et al., 2020; Rouder & Haaf, 2019; Samaey et al., 2020).

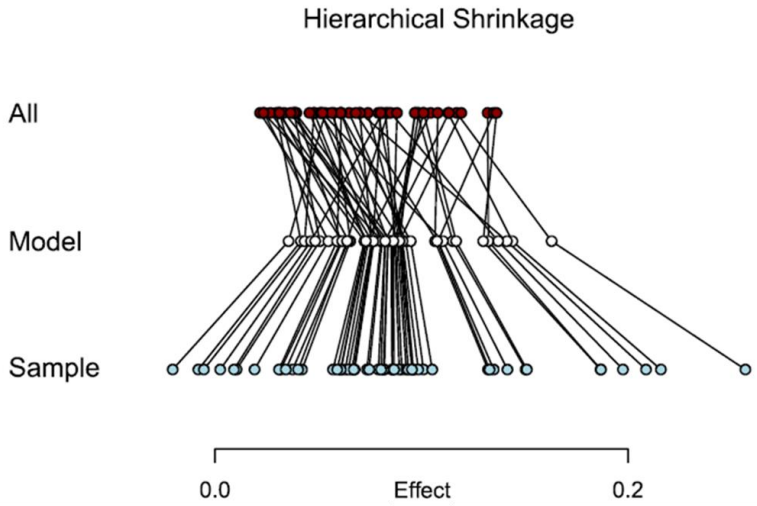

Figure 3. Hierarchical shrinkage in practice. Lower row displays sample estimates from a single block, the central row displays estimates derived from the hierarchical model, and the upper row displays sample estimates from all blocks combined. The hierarchical model's shrinkage reduces variability so that it aligns with that observed in much larger samples. Figure from Rouder & Haaf, (2019)

Using regression to estimate the contextual effects

Typically, a common metric for determining the magnitude of holistic processing or contextual modulations involves subtracting the dependent variable (e.g., accuracy, reaction time, d', or thresholds in our case) of the control condition (i.e., "wholes" in the part-whole task, "misaligned" or "congruent" condition in the composite task) from the condition of interest (i.e., "parts" in the part-whole task, "aligned" or "incongruent" condition depending on the kind of composite task used). Research examining the shared variance among different measures of holistic processing using this approach yielded inconsistent results. There's an ongoing debate regarding whether these varied measures truly assess holistic processing (Richler et al., 2012; Boutet et al., 2021), or if they can predict general face recognition capabilities (Wang et al., 2012; DeGutis et al., 2013). While assessing arithmetic differences in the dependent variable across conditions serves as a legitimate means to approximate these effects, simple subtraction methods have been highlighted as suboptimal for investigating contextual effects (DeGutis et al., 2013).

The subtraction is mainly applied to strip the influence of the condition of interest from the control condition as much as possible, aiming to capture the pure impact of the manipulation introduced by the condition of interest. Ideally, in a meticulously designed experiment and stimulus presentation, the estimated effect should strongly correlate with the condition of interest, while showing no correlation with the control condition. However, when assessing the shared variance between the condition of interest or control condition and the difference scores using correlations, it is evident that simple subtraction does not entirely eliminate these influences from either condition. With subtraction, while

difference scores do correlate with the condition of interest, they also retain some variance from the control condition, indicating an incomplete separation from the control's influence (see Figure 4). In the regression method, the control condition is regressed out of the condition of interest, and the residual values of this regression are used as the estimate of the effect. As a result, the residual values correlate strongly with the condition of interest, while their correlation with the control condition is practically zero. Therefore, this method presents a cleaner separation of the effect that is more sensitive to the manipulation.

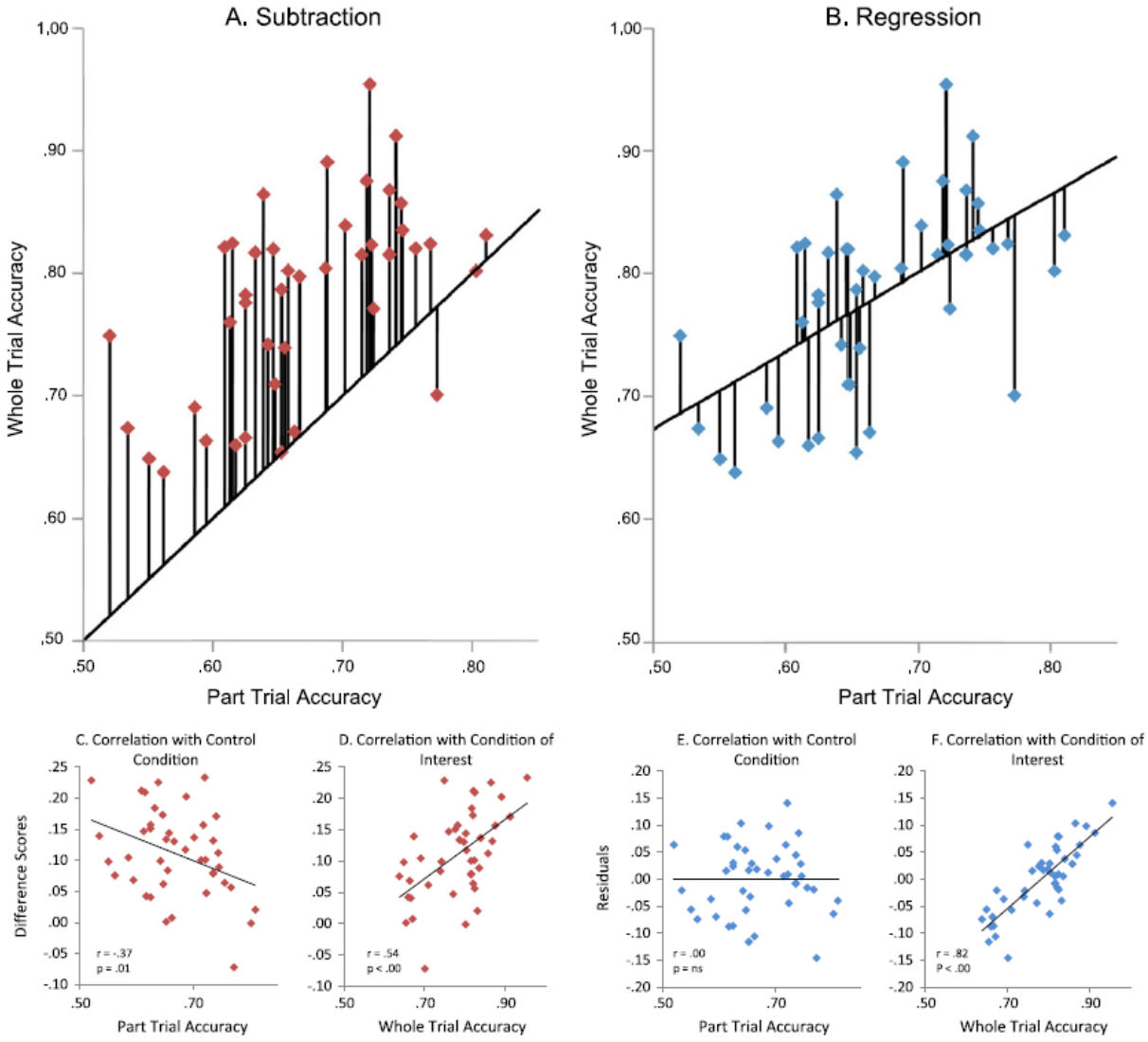

Figure 4. Estimates of holistic processing for part-whole task calculated using subtraction and regression methods. ***A.*** *Simple subtraction method, where the differences in the dependent variable across conditions (i.e., vertical black lines) are used as an estimate of the effect. Small plots below (left half) with red dots display the correlation of difference scores with the control condition and condition of interest. Although the scores correlate with condition of interest, there is still some variance shared with the control condition and difference scores.* ***B.*** *Regression method, where the control condition is regressed from the condition of interest, and the residuals (i.e. vertical black lines) are used as an estimate of the effect. Small plots below (right half) with blue dots display the correlation of residuals with the control condition and condition of interest. Residuals share no variance with the control condition while showing a strong correlation with condition of*

interest, demonstrating a cleaner separation of the effects introduced by the manipulation. Figure from DeGutis et al., (2013)

Most studies investigating different measures of holistic processing or their relationship to general face processing abilities using an inter-individual differences framework used subtraction method to estimate the effects (Konar et al., 2010a, 2010b; Wang et al., 2012; Konar et al., 2013; Ross et al., 2015; Sunday et al., 2017). The weak correlations, or the absence thereof can be attributed to the subtraction method which may not have been sensitive enough. The studies that used the regression method for the same purpose however were more likely to observe correlations across these measures, possibly due to better isolation of face-specific processing (DeGutis et al., 2013; Rezlescu et al., 2017; Boutet et al., 2021). In this thesis, we calculated all context-specific effects using the regression method as opposed to subtraction. Despite using this approach, we did not observe a correlation between contrast detection contextual modulations and upright faces, nor between V1 size and upright faces. However, the regression method, alongside the hierarchical modeling, may have increased the chance to observe the correlation between inverted face and contrast detection contextual modulations, as well as with V1 size. By adopting these two methods, we positioned ourselves in optimal conditions where overlooking an existing correlation would be highly improbable.

Specialization and modularity of upright face processing

Even with the strength of our methodology, one particular phenomenon persisted in our research: upright face contextual modulations share no variance with low-level mechanisms. This arises when the shared variance is determined either through its behavioral indicators (i.e., contrast detection contextual modulations) or by its correlation to V1 size. In a way, this is unsurprising and aligns with the strong body of literature that attributes a special status to their processing. However, a first question appears in terms of our statistical power and sensitivity. Can we confidently conclude based on our chosen methodology and sample size? Given the associations we observed across other tasks, notably between inverted face and contrast detection contextual modulations, it appears that our measurements were sufficiently sensitive to discern such correlations between low- and high-level contextual mechanisms. Due to the hierarchical nature of visual processing, the processes at primary vision should still resonate at higher levels. While the contextual modulations for inverted faces seem to still carry some of the signal propagated from the lower-levels of the visual hierarchy, the engagement of contextual mechanisms specialized for the processing of upright faces may have overshadowed, or detached this signal inherited from V1. Another possibility is that contextual modulations for upright faces are so specialized and so crucial that they are resistant and protected from the variations in the lower levels of the visual hierarchy that may constrain their involvement in any way. This brings up the question of modularity in upright face processing. For a process to meet the criteria of a 'module' within the Fodorian Modularity of Mind theory, it must conform to specific principles. Among these, the most

crucial are rapid and obligatory processing, the presence of dedicated neural pathways, distinctive processing characteristics, and the concept of informational encapsulation. The literature on face processing has demonstrated that upright faces do indeed undergo rapid processing (Bentin et al., 1996), possibly via a dedicated neural circuitry (Sergent et al., 1992; Kanwisher et al., 1997; Haxby et al., 2000; Tong et al., 2000). Our findings provide additional insight into the principles of informational encapsulation and distinctive processing characteristics of upright face processing. The separation of upright face contextual modulations from the lower levels of the visual processing hierarchy suggests that they exhibit informational encapsulation to a certain extent. They appear to be impervious to the constraints imposed by these early processing stages, whereas inverted faces show a partial reliance on them. However, upright and inverted face contextual modulations have at least partially overlapping variance (i.e., 37% of the variance is shared between the two), which challenges a complete informational encapsulation account, calling for a more nuanced explanation for the nature of this process. As such, it is important to recognize that these principles have faced criticism and debate due to their restrictive nature and oversimplification of the complexity of the human mind (Carruthers, 2005; Prinz, 2005). As with many other processes in the visual system, and considering its hierarchical nature, it appears unlikely that a visual process can be completely encapsulated. Hence, cognitive processes may be more interconnected than initially suggested by Fodor's theory. Still, it is crucial not to overlook the substantial portion of variance that remains unexplained (63%) in upright face processing, as this supports another Fodorian principle - the distinctive processing characteristics of upright faces. This variance indicates that upright faces undergo different processes compared to inverted faces and other complex images, emphasizing the strength of specialized processes that engage when the visual system aligns with the template of an upright face. Therefore, although upright face processing may not be as isolated as Fodor's theory proposes, it still exhibits a significant degree of specialization. These findings highlight the intricate and multifaceted nature of visual processing, which does not entirely conform to the strict principles of modularity.

Concluding remarks

The study of vision science aims to understand the complexities of the visual system, and how they manifest themselves in human behavior and perception. By examining low-level and high-level mechanisms in isolation, many studies in sensation, perception, and cognition have often overlooked the interplay between these levels. Consequently, they offer limited understanding of the relative significance of each mechanism in predicting outcomes in intricate real-world perception. Specifically, the models estimating the principles for contextual modulations are built using basic stimuli, but the estimations of these models with naturalistic stimuli is poor and yet to be understood. In order to truly

evaluate the accuracy of these models, they should be tested with natural stimuli like faces. The present thesis contributed to this aim by jointly studying visual contextual modulations at low- and high-level stages of visual processing hierarchy, using basic stimuli, high-level, non face-specialized stimuli and face-specialized stimuli. Although the high-level, face-specialized stimuli seems to be working independently of the early regions, we found that non face-specialized high-level stimuli indeed gets influenced by the constraints presented by lower-levels of the processing hierarchy. These findings pose new questions in the literature such as whether contextual modulations at low-levels would influence the contextual modulations of other complex images that are not handled by dedicated mechanisms, and further adds to the large literature on whether faces are truly processed in a special manner or not. It is important to highlight that the distinctions between the contributions of low- and high-level processes may be more intricate than proposed by the standard hierarchical model. Given the pervasive impact of low-level biases on high-level processes, what is typically considered to be low- or high-level may not be that isolated in the end. The modulation of neural signals by low-level features not only challenges such dichotomy but also reveals an interplay, wherein these features are more likely to overlap and interact with those linked to category-specific processes. Moving forward, a more fruitful approach may involve concentrating on comprehending the collective impact of various properties on a specific process, rather than rigidly classifying them as either low or high level. Consequently, the focus shifts from understanding how low-level stages influence high-level processing to exploring the neural representations in a more comprehensive manner, posing new and compelling research avenues for the future.

References

Adelson, E. H. (1999). *Lightness perception and lightness illusions*. https://api.semanticscholar.org/CorpusID:5247134

Agrillo, C., Gori, S., & Beran, M. J. (2015). Do rhesus monkeys (Macaca mulatta) perceive illusory motion? *Animal Cognition*, *18*(4), 895–910. https://doi.org/10.1007/s10071-015-0860-6

Agrochao, M., Tanaka, R., Salazar-Gatzimas, E., & Clark, D. A. (2020). Mechanism for analogous illusory motion perception in flies and humans. *Proceedings of the National Academy of Sciences*, *117*(37), 23044–23053. https://doi.org/10.1073/pnas.2002937117

Albonico, A., & Barton, J. J. S. (2019). Progress in perceptual research: The case of prosopagnosia. *F1000Research*, *8*, 765. https://doi.org/10.12688/f1000research.18492.1

Alitto, H. J., & Usrey, W. M. (2008). Origin and Dynamics of Extraclassical Suppression in the Lateral Geniculate Nucleus of the Macaque Monkey. *Neuron*, *57*(1), 135–146. https://doi.org/10.1016/j.neuron.2007.11.019

Allison, T. (1999). Electrophysiological Studies of Human Face Perception. I: Potentials Generated in Occipitotemporal Cortex by Face and Non-face Stimuli. *Cerebral Cortex*, *9*(5), 415–430. https://doi.org/10.1093/cercor/9.5.415

Allman, J., Miezin, F., & McGuinness, E. (1985). Stimulus Specific Responses from Beyond the Classical Receptive Field: Neurophysiological Mechanisms for Local-Global Comparisons in Visual Neurons. *Annual Review of Neuroscience*, *8*(1), 407–430. https://doi.org/10.1146/annurev.ne.08.030185.002203

Anaki, D., Nica, E. I., & Moscovitch, M. (2011). Automatic Aspects in Face Perception: Evidence From Mandatory Processing of Distractor Facial Components. *Experimental Psychology*, *58*(1), 4–18. https://doi.org/10.1027/1618-3169/a000061

Anderson, J. C., & Martin, K. A. C. (2009). The Synaptic Connections between Cortical Areas V1 and V2 in Macaque Monkey. *The Journal of Neuroscience*, *29*(36), 11283–11293. https://doi.org/10.1523/JNEUROSCI.5757-08.2009

Andrews, T. J., Davies-Thompson, J., Kingstone, A., & Young, A. W. (2010). Internal and external features of the face are represented holistically in face-selective regions of visual cortex. *Journal of Neuroscience*, *30*(9), 3544–3552. https://doi.org/10.1523/JNEUROSCI.4863-09.2010

Andrews, T. J., Halpern, S. D., & Purves, D. (1997). Correlated Size Variations in Human Visual Cortex, Lateral Geniculate Nucleus, and Optic Tract. *The Journal of Neuroscience*, *17*(8), 2859–2868. https://doi.org/10.1523/JNEUROSCI.17-08-02859.1997

Angelucci, A., Bijanzadeh, M., Nurminen, L., Federer, F., Merlin, S., & Bressloff, P. C. (2017). Circuits and Mechanisms for Surround Modulation in Visual Cortex. *Annual Review of Neuroscience*, *40*(1), 425–451. https://doi.org/10.1146/annurev-neuro-072116-031418

Angelucci, A., Levitt, J. B., Walton, E. J. S., Hupé, J.-M., Bullier, J., & Lund, J. S. (2002). Circuits for local and global signal integration in primary visual cortex. *Journal of Neuroscience*, *22*(19), 8633–8646. https://doi.org/10.1523/JNEUROSCI.22-19-08633.2002

Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic Organization of Human Ventral Visual Cortex. *The Journal of Neuroscience*, *29*(34), 10638–10652. https://doi.org/10.1523/JNEUROSCI.2807-09.2009

Atick, J. J., & Redlich, A. N. (1990). Towards a Theory of Early Visual Processing. *Neural Computation*, *2*(3), 308–320. https://doi.org/10.1162/neco.1990.2.3.308

Bååth, R., Seno, T., & Kitaoka, A. (2014). Cats and Illusory Motion. *Psychology*, *05*(09), 1131–1134. https://doi.org/10.4236/psych.2014.59125

Bach, M. (1996). The Freiburg Visual Acuity Test—Automatic Measurement of Visual Acuity: *Optometry and Vision Science*, *73*(1), 49–53. https://doi.org/10.1097/00006324-199601000-00008

Bakken, T. E., Roddey, J. C., Djurovic, S., Akshoomoff, N., Amaral, D. G., Bloss, C. S., Casey, B. J., Chang, L., Ernst, T. M., Gruen, J. R., Jernigan, T. L., Kaufmann, W. E., Kenet, T., Kennedy, D. N., Kuperman, J. M., Murray, S. S., Sowell, E. R., Rimol, L. M., Mattingsdal, M., … Pediatric Imaging, Neurocognition, and Genetics Study. (2012). Association of common genetic variants in GPCPD1 with scaling of visual cortical surface area in humans.

Proceedings of the National Academy of Sciences, *109*(10), 3985–3990. https://doi.org/10.1073/pnas.1105829109

Bar, M. (2003). A Cortical Mechanism for Triggering Top-Down Facilitation in Visual Object Recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609. https://doi.org/10.1162/089892903321662976

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629. https://doi.org/10.1038/nrn1476

Bar, M., & Aminoff, E. (2003). Cortical Analysis of Visual Context. *Neuron*, *38*(2), 347–358. https://doi.org/10.1016/S0896-6273(03)00167-3

Bartlett, J. C., & Searcy, J. (1993). Inversion and Configuration of Faces. *Cognitive Psychology*, *25*(3), 281–316. https://doi.org/10.1006/cogp.1993.1007

Barton, J. J. S. (2008). Structure and function in acquired prosopagnosia: Lessons from a series of 10 patients with brain damage. *Journal of Neuropsychology*, *2*(1), 197–225. https://doi.org/10.1348/174866407X214172

Barton, J. J. S., Cherkasova, M. V., Press, D. Z., Intriligator, J. M., & O'Connor, M. (2004). Perceptual Functions in Prosopagnosia. *Perception*, *33*(8), 939–956. https://doi.org/10.1068/p5243

Barton, J. J. S., Press, D. Z., Keenan, J. P., & O'Connor, M. (2002). Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology*, *58*(1), 71–78. https://doi.org/10.1212/WNL.58.1.71

Behrmann, M., & Avidan, G. (2005). Congenital prosopagnosia: Face-blind from birth. *Trends in Cognitive Sciences*, *9*(4), 180–187. https://doi.org/10.1016/j.tics.2005.02.011

Belanova, E., Davis, J. P., & Thompson, T. (2021). The part-whole effect in super-recognisers and typical-range-ability controls. *Vision Research*, *187*, 75–84. https://doi.org/10.1016/j.visres.2021.06.004

Benson, N. C., Yoon, J. M. D., Forenzo, D., Engel, S. A., Kay, K. N., & Winawer, J. (2022). Variability of the Surface Area of the V1, V2, and V3 Maps in a Large Sample of Human

Observers. *The Journal of Neuroscience*, *42*(46), 8629–8646. https://doi.org/10.1523/JNEUROSCI.0690-21.2022

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of Cognitive Neuroscience*, *8*(6), 551–565. https://doi.org/10.1162/jocn.1996.8.6.551

Benton, A. L., & Van Allen, M. W. (1968). Impairment in Facial Recognition in Patients with Cerebral Disease. *Cortex*, *4*(4), 344-IN1. https://doi.org/10.1016/S0010-9452(68)80018-8

Berman, R. A., & Wurtz, R. H. (2011). Signals Conveyed in the Pulvinar Pathway from Superior Colliculus to Cortical Area MT. *The Journal of Neuroscience*, *31*(2), 373–384. https://doi.org/10.1523/JNEUROSCI.4738-10.2011

Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *352*(1358), 1203–1219. https://doi.org/10.1098/rstb.1997.0103

Blakemore, C., & Tobin, ElisabethA. (1972). Lateral inhibition between orientation detectors in the cat's visual cortex. *Experimental Brain Research*, *15*(4). https://doi.org/10.1007/BF00234129

Bodamer, J. (1947). Die Prosop-Agnosie: Die Agnosie des Physiognomieerkennens. *Archiv f*[illegible]*r Psychiatrie und Nervenkrankheiten Vereinigt mit Zeitschrift f*[illegible]*r die Gesamte Neurologie und Psychiatrie*, *179*(1–2), 6–53. https://doi.org/10.1007/BF00352849

Bonin, V., Mante, V., & Carandini, M. (2005). The Suppressive Field of Neurons in Lateral Geniculate Nucleus. *The Journal of Neuroscience*, *25*(47), 10844–10856. https://doi.org/10.1523/JNEUROSCI.3562-05.2005

Bordier, C., Hupé, J.-M., & Dojat, M. (2015). Quantitative evaluation of fMRI retinotopic maps, from V1 to V4, for cognitive experiments. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00277

Boutet, I., Collin, C., & Faubert, J. (2003). Configural face encoding and spatial frequency information. *Perception & Psychophysics*, *65*(7), 1078–1093. https://doi.org/10.3758/BF03194835

Boutet, I., Nelson, E. A., Watier, N., Cousineau, D., Béland, S., & Collin, C. A. (2021). Different measures of holistic face processing tap into distinct but partially overlapping mechanisms. *Attention, Perception, & Psychophysics*, *83*(7), 2905–2923. https://doi.org/10.3758/s13414-021-02337-7

Brants, M., Wagemans, J., & Op De Beeck, H. P. (2011). Activation of Fusiform Face Area by Greebles Is Related to Face Similarity but Not Expertise. *Journal of Cognitive Neuroscience*, *23*(12), 3949–3958. https://doi.org/10.1162/jocn_a_00072

Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience*, *7*(8), 880–886. https://doi.org/10.1038/nn1278

Bruce, V. (1988). *Recognising Faces* (1st ed.). Routledge. https://www.taylorfrancis.com/books/9781315471808

Bruyer, R., & Velge, V. (1981). [Unilateral cerebral lesion and disturbance of face perception. Specificity of the deficit? (Author's transl)]. *Acta Neurologica Belgica*, *81*(6), 321–332.

Bürkner, P.-C. (2017). **brms**: An *R* Package for Bayesian Multilevel Models Using *Stan*. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395. https://doi.org/10.32614/RJ-2018-017

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology*, *66*(8), 1467–1485. https://doi.org/10.1080/17470218.2013.800125

Busigny, T., & Rossion, B. (2011). Holistic processing impairment can be restricted to faces in acquired prosopagnosia: Evidence from the global/local Navon effect: Navon effect and prosopagnosia. *Journal of Neuropsychology*, *5*(1), 1–14. https://doi.org/10.1348/174866410X500116

Busigny, T., So Jeong, C., & Barton, J. J. S. (2012). Holistic face processing induces perceptual shifts in face perception. *Journal of Vision*, *12*(9), 639–639. https://doi.org/10.1167/12.9.639

Byosiere, S.-E., Feng, L. C., Woodhead, J. K., Rutter, N. J., Chouinard, P. A., Howell, T. J., & Bennett, P. C. (2017). Visual perception in domestic dogs: Susceptibility to the Ebbinghaus–

Titchener and Delboeuf illusions. *Animal Cognition*, *20*(3), 435–448. https://doi.org/10.1007/s10071-016-1067-1

Cabeza, R., & Kato, T. (2000). Features are Also Important: Contributions of Featural and Configural Processing to Face Recognition. *Psychological Science*, *11*(5), 429–433. https://doi.org/10.1111/1467-9280.00283

Cadieu, C., Kouh, M., Pasupathy, A., Connor, C. E., Riesenhuber, M., & Poggio, T. (2007). A Model of V4 Shape Selectivity and Invariance. *Journal of Neurophysiology*, *98*(3), 1733–1750. https://doi.org/10.1152/jn.01265.2006

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 527–551. https://doi.org/10.1037/0096-1523.26.2.527

Canoluk, M. U., Moors, P., & Goffaux, V. (2023). Contributions of low- and high-level contextual mechanisms to human face perception. *PLOS ONE*, *18*(5), e0285255. https://doi.org/10.1371/journal.pone.0285255

Carandini, M. (2005). Do We Know What the Early Visual System Does? *Journal of Neuroscience*, *25*(46), 10577–10597. https://doi.org/10.1523/JNEUROSCI.3726-05.2005

Carbon, C.-C., & Leder, H. (2005). When Feature Information Comes First! Early Processing of Inverted Faces. *Perception*, *34*(9), 1117–1134. https://doi.org/10.1068/p5192

Carey, S., & Diamond, R. (1977). From Piecemeal to Configurational Representation of Faces. *Science*, *195*(4275), 312–314. https://doi.org/10.1126/science.831281

Carey, S., Diamond, R., & Woods, B. (1980). Development of face recognition: A maturational component? *Developmental Psychology*, *16*(4), 257–269. https://doi.org/10.1037/0012-1649.16.4.257

Carlson, T., Hogendoorn, H., Fonteijn, H., & Verstraten, F. A. J. (2011). Spatial coding and invariance in object-selective cortex. *Cortex*, *47*(1), 14–22. https://doi.org/10.1016/j.cortex.2009.08.015

Carruthers, P. (2005). The case for massively modular models of mind. *Contemporary Debates in Cognitive Science, Ed. R. Stainton*, 205–225.

Cassia, V. M., Picozzi, M., Kuefner, D., Bricolo, E., & Turati, C. (2009). Holistic processing for faces and cars in preschool-aged children and adults: Evidence from the composite effect. *Developmental Science*, *12*(2), 236–248. https://doi.org/10.1111/j.1467-7687.2008.00765.x

Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *Journal of Neurophysiology*, *88*(5), 2530–2546. https://doi.org/10.1152/jn.00692.2001

Chalupa, L., Williams, R., & Hughes, M. (1983). Visual response properties in the tectorecipient zone of the cat's lateral posterior-pulvinar complex: A comparison with the superior colliculus. *The Journal of Neuroscience*, *3*(12), 2587–2596. https://doi.org/10.1523/JNEUROSCI.03-12-02587.1983

Chamberlain, R., Van der Hallen, R., Huygelier, H., Van de Cruys, S., & Wagemans, J. (2017). Local-global processing bias is not a unitary individual difference in visual processing. *Vision Research*, *141*, 247–257. https://doi.org/10.1016/j.visres.2017.01.008

Chao, L. L., Martin, A., & Haxby, J. V. (1999). Are face-responsive regions selective only for faces?: *NeuroReport*, *10*(14), 2945–2950. https://doi.org/10.1097/00001756-199909290-00013

Cheung, O. S., Richler, J. J., Palmeri, T. J., & Gauthier, I. (2008). Revisiting the role of spatial frequencies in the holistic processing of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1327–1336. https://doi.org/10.1037/a0011752

Cheung, O. S., Richler, J. J., Phillips, W. S., & Gauthier, I. (2011). Does temporal integration of face parts reflect holistic processing? *Psychonomic Bulletin & Review*, *18*(3), 476–483. https://doi.org/10.3758/s13423-011-0051-7

Chubb, C., Sperling, G., & Solomon, J. A. (1989). Texture interactions determine perceived contrast. *Proceedings of the National Academy of Sciences*, *86*(23), 9631–9635. https://doi.org/10.1073/pnas.86.23.9631

Civile, C., McLaren, R., & McLaren, I. P. L. (2016). The Face Inversion Effect: Roles of First and Second-Order Configural Information. *The American Journal of Psychology*, *129*(1), 23–35. https://doi.org/10.5406/amerjpsyc.129.1.0023

Collin, C. A., Liu, C. H., Troje, N. F., McMullen, P. A., & Chaudhuri, A. (2004). Face Recognition Is Affected by Similarity in Spatial Frequency Range to a Greater Degree Than Within-Category Object Recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(5), 975–987. https://doi.org/10.1037/0096-1523.30.5.975

Collins, J. A., & Olson, I. R. (2014). Beyond the FFA: The role of the ventral anterior temporal lobes in face processing. *Neuropsychologia*, *61*, 65–79. https://doi.org/10.1016/j.neuropsychologia.2014.06.005

Collishaw, S. M., & Hole, G. J. (2000). Featural and Configurational Processes in the Recognition of Faces of Different Familiarity. *Perception*, *29*(8), 893–909. https://doi.org/10.1068/p2949

Coren, S., & Porac, C. (1987). Individual differences in visual-geometric illusions: Predictions from measures of spatial cognitive abilities. *Perception & Psychophysics*, *41*(3), 211–219. https://doi.org/10.3758/BF03208220

Costen, N. P., Parker, D. M., & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception & Psychophysics*, *58*(4), 602–612. https://doi.org/10.3758/BF03213093

Cox, R. W. (1996). AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Computers and Biomedical Research*, *29*(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, *10*(4–5), 171–178. https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L

Cretenoud, A. F., Francis, G., & Herzog, M. H. (2020). When illusions merge. *Journal of Vision*, *20*(8), 12. https://doi.org/10.1167/jov.20.8.12

Cretenoud, A. F., Grzeczkowski, L., Bertamini, M., & Herzog, M. H. (2020). Individual differences in the Müller-Lyer and Ponzo illusions are stable across different contexts. *Journal of Vision*, *20*(6), 4. https://doi.org/10.1167/jov.20.6.4

Cretenoud, A. F., Grzeczkowski, L., Kunchulia, M., & Herzog, M. H. (2021). Individual differences in the perception of visual illusions are stable across eyes, time, and measurement methods. *Journal of Vision*, *21*(5), 26. https://doi.org/10.1167/jov.21.5.26

Curby, K. M., Entenman, R. J., & Fleming, J. T. (2016). Holistic face perception is modulated by experience-dependent perceptual grouping. *Attention, Perception, & Psychophysics*, *78*(5), 1392–1404. https://doi.org/10.3758/s13414-016-1077-8

Curby, K. M., & Moerel, D. (2019). Behind the face of holistic perception: Holistic processing of Gestalt stimuli and faces recruit overlapping perceptual mechanisms. *Attention, Perception, & Psychophysics*, *81*(8), 2873–2880. https://doi.org/10.3758/s13414-019-01749-w

Dakin, S., & Watt, R. J. (2009). Biological "bar codes" in human faces. *Journal of Vision*, *9*(4), 1–10. https://doi.org/doi:10.1167/9.4.2

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Damasio, A. R., Damasio, H., & Van Hoesen, G. W. (1982). Prosopagnosia: Anatomic basis and behavioral mechanisms. *Neurology*, *32*(4), 331–331. https://doi.org/10.1212/WNL.32.4.331

Davidoff, J., & Donnelly, N. (1990). Object superiority: A comparison of complete and part probes. *Acta Psychologica*, *73*(3), 225–243. https://doi.org/10.1016/0001-6918(90)90024-A

de Gelder, B., Frissen, I., Barton, J. J. S., & Hadjikhani, N. (2003). A modulatory role for facial expressions in prosopagnosia. *Proceedings of the National Academy of Sciences*, *100*(22), 13105–13110. https://doi.org/10.1073/pnas.1735530100

De Heering, A., Turati, C., Rossion, B., Bulf, H., Goffaux, V., & Simion, F. (2008). Newborns' face recognition is based on spatial frequencies below 0.5 cycles per degree. *Cognition*, *106*(1), 444–454. https://doi.org/10.1016/j.cognition.2006.12.012

DeGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition*, *126*(1), 87–100. https://doi.org/10.1016/j.cognition.2012.09.004

Desimone, R. (1991). Face-Selective Cells in the Temporal Cortex of Monkeys. *Journal of Cognitive Neuroscience*, *3*(1), 1–8. https://doi.org/10.1162/jocn.1991.3.1.1

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*(2), 107–117. https://doi.org/10.1037/0096-3445.115.2.107

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, *11*(8), 333–341. https://doi.org/10.1016/j.tics.2007.06.010

Dobbins, A., Zucker, S. W., & Cynader, M. S. (1987). Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature*, *329*(6138), 438–441. https://doi.org/10.1038/329438a0

Dong, D. W., & Atick, J. J. (1995). Statistics of natural time-varying images. *Network: Computation in Neural Systems*, *6*(3), 345–358. https://doi.org/10.1088/0954-898X_6_3_003

Dougherty, R. F., Koch, V. M., Brewer, A. A., Fischer, B., Modersitzki, J., & Wandell, B. A. (2003). Visual field representations and locations of visual areas V1/2/3 in human visual cortex. *Journal of Vision*, *3*(10), 1. https://doi.org/10.1167/3.10.1

Downing, P. E., Chan, A. W.-Y., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain Specificity in Visual Cortex. *Cerebral Cortex*, *16*(10), 1453–1461. https://doi.org/10.1093/cercor/bhj086

Duchaine, B. C., & Weidenfeld, A. (2003). An evaluation of two commonly used tests of unfamiliar face recognition. *Neuropsychologia*, *41*(6), 713–720. https://doi.org/10.1016/S0028-3932(02)00222-1

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, *39*(2), 647–660. https://doi.org/10.1016/j.neuroimage.2007.09.034

Duncan, J., Gosselin, F., Cobarro, C., Dugas, G., Blais, C., & Fiset, D. (2017). Orientations for the successful categorization of facial expressions and their link with facial features. *Journal of Vision*, *17*(14), 7. https://doi.org/10.1167/17.14.7

Duncan, R. O., & Boynton, G. M. (2003). Cortical Magnification within Human Primary Visual Cortex Correlates with Acuity Thresholds. *Neuron*, *38*(4), 659–671. https://doi.org/10.1016/S0896-6273(03)00265-4

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., Baranski, E., Bernstein, M. J., Bonfiglio, D. B. V., Boucher, L., Brown, E. R., Budiman, N. I., Cairo, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J. G., … Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology, 67*, 68–82. https://doi.org/10.1016/j.jesp.2015.10.012

Efron, B., & Morris, C. (1977). Stein's Paradox in Statistics. *Scientific American, 236*(5), 119–127. https://doi.org/10.1038/scientificamerican0577-119

Eimer, M. (2000). Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clinical Neurophysiology*, *111*(4), 694–705. https://doi.org/10.1016/S1388-2457(99)00285-0

Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia*, *45*(1), 15–31. https://doi.org/10.1016/j.neuropsychologia.2006.04.022

Ejima, Y. & Takahashi. (1985). Apparent contrast of a sinusoidal grating in the simultaneous presence of peripheral gratings. *Vision Research, 25*(9), 1223–1232. https://doi.org/10.1016/0042-6989(85)90036-7

Epstein, R. A., Higgins, J. S., Parker, W., Aguirre, G. K., & Cooperman, S. (2006). Cortical correlates of face and scene inversion: A comparison. *Neuropsychologia*, *44*(7), 1145–1158. https://doi.org/10.1016/j.neuropsychologia.2005.10.009

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., & Isik, A. I. (2018). *fMRIPrep. Software.* [Computer software]. Zenodo.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Fang, F., Boyaci, H., Kersten, D., & Murray, S. O. (2008). Attention-Dependent Representation of a Size Illusion in Human V1. *Current Biology*, *18*(21), 1707–1712. https://doi.org/10.1016/j.cub.2008.09.025

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 628–634. https://doi.org/10.1037/0096-1523.21.3.628

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka. (1998). What is "special" about face perception? *Psychological Review*, *105*(3), 482–498. https://doi.org/10.1037/0033-295X.105.3.482

Filimonoff, I. (1933). *Uber die variabilitä t der großhirnrindenstruktur. Mitteilung III. Regio occipitalis bei den höheren und niederen affen*. *44*, 1–96.

Fiorentini, A., Maffei, L., & Sandini, G. (1983). The Role of High Spatial Frequencies in Face Perception. *Perception*, *12*(2), 195–201. https://doi.org/10.1068/p120195

Fitzpatrick, D. (2000). Seeing beyond the receptive field in primary visual cortex. *Current Opinion in Neurobiology*, *10*(4), 438–443. https://doi.org/10.1016/S0959-4388(00)00113-6

Fleuret, F., Li, T., Dubout, C., Wampler, E. K., Yantis, S., & Geman, D. (2011). Comparing machines and humans on a visual categorization test. *Proceedings of the National Academy of Sciences*, *108*(43), 17621–17625. https://doi.org/10.1073/pnas.1109168108

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.

Foster, C., Bülthoff, I., Bartels, A., & Zhao, M. (2021). Investigating holistic face processing within and outside of face-responsive brain regions. *NeuroImage*, *226*, 117565. https://doi.org/10.1016/j.neuroimage.2020.117565

Fox, C. J., Iaria, G., & Barton, J. J. S. (2009). Defining the face processing network: Optimization of the functional localizer in fMRI. *Human Brain Mapping*, *30*(5), 1637–1651. https://doi.org/10.1002/hbm.20630

Gao, C., Conte, S., Richards, J. E., Xie, W., & Hanayik, T. (2019). The neural sources of N170: Understanding timing of activation in face-selective areas. *Psychophysiology*, *56*(6), e13336. https://doi.org/10.1111/psyp.13336

García, A. M., & Ibáñez, A. (Eds.). (2023). *The Routledge handbook of semiosis and the brain*. Routledge.

Gaspar, C., Sekuler, A. B., & Bennett, P. J. (2008). Spatial frequency tuning of upright and inverted face identification. *Vision Research*, *48*(28), 2817–2826. https://doi.org/10.1016/j.visres.2008.09.015

Gauthier, I., & Bukach, C. (2007). Should we reject the expertise hypothesis? *Cognition*, *103*(2), 322–330. https://doi.org/10.1016/j.cognition.2006.05.003

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

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nature Neuroscience*, *2*(6), 568–573. https://doi.org/10.1038/9224

Geisler, W. S., & Chou, K.-L. (1995). Separation of low-level and high-level factors in complex tasks: Visual Search. *Psychological Review*, *102*(2), 356–378. https://doi.org/10.1037/0033-295x.102.2.356

Genon, S., Eickhoff, S. B., & Kharabian, S. (2022). Linking interindividual variability in brain structure to behaviour. *Nature Reviews Neuroscience*, *23*(5), 307–318. https://doi.org/10.1038/s41583-022-00584-7

George, N., Dolan, R. J., Fink, G. R., Baylis, G. C., Russell, C., & Driver, J. (1999). Contrast polarity and face recognition in the human fusiform gyrus. *Nature Neuroscience*, *2*(6), 574–580. https://doi.org/10.1038/9230

Gibson, J. J., & Radner, M. (1937). Adaptation, after-effect and contrast in the perception of tilted lines. I. Quantitative studies. *Journal of Experimental Psychology*, *20*(5), 453–467. https://doi.org/10.1037/h0059826

Gilaie-Dotan, S., Gelbard-Sagiv, H., & Malach, R. (2010). Perceptual shape sensitivity to upright and inverted faces is reflected in neuronal adaptation. *NeuroImage*, *50*(2), 383–395. https://doi.org/10.1016/j.neuroimage.2009.12.077

Gilbert, C. D., & Wiesel, T. N. (1990). The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vision Research*, *30*(11), 1689–1701. https://doi.org/10.1016/0042-6989(90)90153-C

Girard, P., Hupé, J. M., & Bullier, J. (2001). Feedforward and Feedback Connections Between Areas V1 and V2 of the Monkey Have Similar Rapid Conduction Velocities. *Journal of Neurophysiology*, *85*(3), 1328–1331. https://doi.org/10.1152/jn.2001.85.3.1328

Goffaux, V. (2008). The horizontal and vertical relations in upright faces are transmitted by different spatial frequency ranges. *Acta Psychologica*, *128*(1), 119–126. https://doi.org/10.1016/j.actpsy.2007.11.005

Goffaux, V. (2009). Spatial interactions in upright and inverted faces: Re-exploration of spatial scale influence. *Vision Research*, *49*(7), 774–781. https://doi.org/10.1016/j.visres.2009.02.009

Goffaux, V. (2012). The discriminability of local cues determines the strength of holistic face processing. *Vision Research*, *64*, 17–22. https://doi.org/10.1016/j.visres.2012.04.022

Goffaux, V., & Dakin, S. (2010). Horizontal information drives the behavioral signatures of face processing. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00143

Goffaux, V., Duecker, F., Hausfeld, L., Schiltz, C., & Goebel, R. (2016). Horizontal tuning for faces originates in high-level Fusiform Face Area. *Neuropsychologia*, *81*, 1–11. https://doi.org/10.1016/j.neuropsychologia.2015.12.004

Goffaux, V., Gauthier, I., & Rossion, B. (2003). Spatial scale contribution to early visual differences between face and object processing. *Cognitive Brain Research*, *16*(3), 416–424. https://doi.org/10.1016/S0926-6410(03)00056-9

Goffaux, V., & Greenwood, J. A. (2016). The orientation selectivity of face identification. *Scientific Reports*, *6*(1), 34204. https://doi.org/10.1038/srep34204

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The Respective Role of Low and High Spatial Frequencies in Supporting Configural and Featural Processing of Faces. *Perception*, *34*(1), 77–86. https://doi.org/10.1068/p5370

Goffaux, V., & Rossion, B. (2006). Faces are “spatial”—Holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(4), 1023–1039. https://doi.org/10.1037/0096-1523.32.4.1023

Goffaux, V., & Rossion, B. (2007). Face inversion disproportionately impairs the perception of vertical but not horizontal relations between features. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 995–1002. https://doi.org/10.1037/0096-1523.33.4.995

Goffaux, V., Rossion, B., Sorger, B., Schiltz, C., & Goebel, R. (2009). Face inversion disrupts the perception of vertical relations between features in the right human occipito-temporal cortex. *Journal of Neuropsychology*, *3*(1), 45–67. https://doi.org/10.1348/174866408X292670

Goffaux, V., Schiltz, C., Mur, M., & Goebel, R. (2013). Local Discriminability Determines the Strength of Holistic Processing for Faces in the Fusiform Face Area. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00604

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research*, *39*(21), 3537–3560. https://doi.org/10.1016/S0042-6989(99)00080-2

Goldberg, M. E., & Wurtz, R. H. (1972). Activity of superior colliculus in behaving monkey. I. Visual receptive fields of single neurons. *Journal of Neurophysiology*, *35*(4), 542–559. https://doi.org/10.1152/jn.1972.35.4.542

Gonchar, Y., & Burkhalter, A. (2003). Distinct GABAergic Targets of Feedforward and Feedback Connections Between Lower and Higher Areas of Rat Visual Cortex. *The Journal of Neuroscience*, *23*(34), 10904–10912. https://doi.org/10.1523/JNEUROSCI.23-34-10904.2003

Goren, C. C., Sarty, M., & Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, *56*(4), 544–549.

Gorgolewski, K., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of

neuroimaging experiments. *Scientific Data*, *3*(1), 160044. https://doi.org/10.1038/sdata.2016.44

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Frontiers in Neuroinformatics*, *5*. https://doi.org/10.3389/fninf.2011.00013

Gorgolewski, K., Nichols, T., Kennedy, D. N., Poline, J.-B., & Poldrack, R. A. (2018). Making replication prestigious. *Behavioral and Brain Sciences*, *41*, e131. https://doi.org/10.1017/S0140525X18000663

Gori, S., Agrillo, C., Dadda, M., & Bisazza, A. (2014). Do Fish Perceive Illusory Motion? *Scientific Reports*, *4*(1), 6443. https://doi.org/10.1038/srep06443

Green, S. B., Yang, Y., Alt, M., Brinkley, S., Gray, S., Hogan, T., & Cowan, N. (2016). Use of internal consistency coefficients for estimating reliability of experimental task scores. *Psychonomic Bulletin & Review*, *23*(3), 750–763. https://doi.org/10.3758/s13423-015-0968-3

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Grill-Spector, K., Weiner, K. S., Kay, K., & Gomez, J. (2017). The Functional Neuroanatomy of Human Face Perception. *Annual Review of Vision Science*, *3*(1), 167–196. https://doi.org/10.1146/annurev-vision-102016-061214

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

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology*, *35*(1), 96–111. https://doi.org/10.1152/jn.1972.35.1.96

Grzeczkowski, L., Clarke, A. M., Francis, G., Mast, F. W., & Herzog, M. H. (2017). About individual differences in vision. *Vision Research*, *141*, 282–292. https://doi.org/10.1016/j.visres.2016.10.006

Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General*, *144*(2), 432–446. https://doi.org/10.1037/xge0000053

Haines, N., Kvam, P. D., Irving, L. H., Smith, C., Beauchaine, T. P., Pitt, M. A., Ahn, W.-Y., & Turner, B. (2020). *Theoretically Informed Generative Models Can Advance the Psychological and Brain Sciences: Lessons from the Reliability Paradox* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/xr7y3

Halit, H., De Haan, M., Schyns, P. G., & Johnson, M. H. (2006). Is high-spatial frequency information used in the early stages of face detection? *Brain Research*, *1117*(1), 154–161. https://doi.org/10.1016/j.brainres.2006.07.059

Hansen, T., & Neumann, H. (2008). A recurrent model of contour integration in primary visual cortex. *Journal of Vision*, *8*(8), 8–8. https://doi.org/10.1167/8.8.8

Harris, A., & Aguirre, G. K. (2008). The effects of parts, wholes, and familiarity on face-selective responses in MEG. *Journal of Vision*, *8*(10), 4–4. https://doi.org/10.1167/8.10.4

Hashemi, A., Pachai, M. V., Bennett, P. J., & Sekuler, A. B. (2019). The role of horizontal facial structure on the N170 and N250. *Vision Research*, *157*, 12–23. https://doi.org/10.1016/j.visres.2018.02.006

Haxby, Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. https://doi.org/10.1016/S1364-6613(00)01482-0

Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., & Martin, A. (1999a). The Effect of Face Inversion on Activity in Human Neural Systems for Face and Object Perception. *Neuron*, *22*(1), 189–199. https://doi.org/10.1016/S0896-6273(00)80690-X

Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., & Martin, A. (1999b). The Effect of Face Inversion on Activity in Human Neural Systems for Face and Object Perception. *Neuron*, *22*(1), 189–199. https://doi.org/10.1016/S0896-6273(00)80690-X

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

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

Hole, G. J. (1994). Configurational Factors in the Perception of Unfamiliar Faces. *Perception*, *23*(1), 65–74. https://doi.org/10.1068/p230065

Hole, G. J., George, P. A., & Dunsmore, V. (1999). Evidence for Holistic Processing of Faces Viewed as Photographic Negatives. *Perception*, *28*(3), 341–359. https://doi.org/10.1068/p2622

Howell, D. C. (2013). *Statistical methods for psychology* (8th ed). Wadsworth Cengage Learning.

Huang, X., Albright, T. D., & Stoner, G. R. (2007). Adaptive Surround Modulation in Cortical Area MT. *Neuron*, *53*(5), 761–770. https://doi.org/10.1016/j.neuron.2007.01.032

Hubel, D. H., & Wiesel, T. N. (1965). Receptive Fields and Functioinal Architecture in Two Nonstriate Visual Areas (18 and 19) of the Cat. *Journal of Neurophysiology*, *28*(2), 229–289. https://doi.org/10.1152/jn.1965.28.2.229

Iidaka, T. (2014). Role of the fusiform gyrus and superior temporal sulcus in face perception and recognition: An empirical review: Neuroimaging of face recognition. *Japanese Psychological Research*, *56*(1), 33–45. https://doi.org/10.1111/jpr.12018

Jack, R. E., & Schyns, P. G. (2017). Toward a Social Psychophysics of Face Communication. *Annual Review of Psychology*, *68*(1), 269–297. https://doi.org/10.1146/annurev-psych-010416-044242

Jacobs, C., Petras, K., Moors, P., & Goffaux, V. (2020). Contrast versus identity encoding in the face image follow distinct orientation selectivity profiles. *PLOS ONE*, *15*(3), e0229185. https://doi.org/10.1371/journal.pone.0229185

Jacques, C., & Rossion, B. (2006). The Speed of Individual Face Categorization. *Psychological Science*, *17*(6), 485–492. https://doi.org/10.1111/j.1467-9280.2006.01733.x

Jacques, C., & Rossion, B. (2009). The initial representation of individual faces in the right occipito-temporal cortex is holistic: Electrophysiological evidence from the composite face illusion. *Journal of Vision*, *9*(6), 8–8. https://doi.org/10.1167/9.6.8

Jacques, C., & Rossion, B. (2010). Misaligning face halves increases and delays the N170 specifically for upright faces: Implications for the nature of early face representations. *Brain Research*, *1318*, 96–109. https://doi.org/10.1016/j.brainres.2009.12.070

James, T. W., Arcurio, L. R., & Gold, J. M. (2013). Inversion Effects in Face-selective Cortex with Combinations of Face Parts. *Journal of Cognitive Neuroscience*, *25*(3), 455–464. https://doi.org/10.1162/jocn_a_00312

Jarosz, A. F., & Wiley, J. (2014). What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors. *The Journal of Problem Solving*, *7*(1). https://doi.org/10.7771/1932-6246.1167

Jeffreys, D. A. (1996). Evoked Potential Studies of Face and Object Processing. *Visual Cognition*, *3*(1), 1–38. https://doi.org/10.1080/713756729

Jeffreys, H. (1998). *Theory of probability* (3rd ed). Clarendon Press ; Oxford University Press.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Johnson, M. H., & De Haan, M. (2011). *Developmental cognitive neuroscience: An introduction* (3rd ed). Wiley-Blackwell.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1–2), 1–19. https://doi.org/10.1016/0010-0277(91)90045-6

Johnson, M. H., Farroni, T., Brockbank, M., & Simion, F. (2000). Preferential orienting to faces in 4-month-olds: Analysis of temporal-nasal visual field differences. *Developmental Science*, *3*(1), 41–45. https://doi.org/10.1111/1467-7687.00097

Kaas, J. H. (2000). Why is Brain Size so Important: Design Problems and Solutions as Neocortex Gets Bigger or Smaller. *Brain and Mind*, *1*(1), 7–23. https://doi.org/10.1023/A:1010028405318

Kamps, F. S., Morris, E. J., & Dilks, D. D. (2019). A face is more than just the eyes, nose, and mouth: fMRI evidence that face-selective cortex represents external features. *NeuroImage*, *184*, 90–100. https://doi.org/10.1016/j.neuroimage.2018.09.027

Kanai, R., & Rees, G. (2011). The structural basis of inter-individual differences in human behaviour and cognition. *Nature Reviews Neuroscience*, *12*(4), 231–242. https://doi.org/10.1038/nrn3000

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*(8), 759–763. https://doi.org/10.1038/77664

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *The Journal of Neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Kanwisher, N., Stanley, D., & Harris, A. (1999). The fusiform face area is selective for faces not animals: *NeuroReport*, *10*(1), 183–187. https://doi.org/10.1097/00001756-199901180-00035

Kanwisher, N., Tong, F., & Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition*, *68*(1), B1–B11. https://doi.org/10.1016/S0010-0277(98)00035-3

Kimchi, R., & Amishav, R. (2010). Faces as perceptual wholes: The interplay between component and configural properties in face processing. *Visual Cognition*, *18*(7), 1034–1062. https://doi.org/10.1080/13506281003619986

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Chaibub Neto, E., & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, *13*(2), e1005350. https://doi.org/10.1371/journal.pcbi.1005350

Knierim, J. J., & van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology*, *67*(4), 961–980. https://doi.org/10.1152/jn.1992.67.4.961

Knoblauch, K., & Maloney, L. T. (2012). *Modeling psychophysical data in R*. Springer.

Knudsen, E. I., & Konishi, M. (1978). A Neural Map of Auditory Space in the Owl. *Science*, *200*(4343), 795–797. https://doi.org/10.1126/science.644324

Konar, Y., Bennett, P. J., & Sekuler, A. B. (2010a). Holistic Processing Is Not Correlated With Face-Identification Accuracy. *Psychological Science*, *21*(1), 38–43. https://doi.org/10.1177/0956797609356508

Konar, Y., Bennett, P. J., & Sekuler, A. B. (2013). Effects of aging on face identification and holistic face processing. *Vision Research*, *88*, 38–46. https://doi.org/10.1016/j.visres.2013.06.003

Konar, Y., Bennett, P., & Sekuler, A. (2010b). Face identification and the evaluation of holistic indexes: CFE and the whole-part task. *Journal of Vision*, *10*(7), 678–678. https://doi.org/10.1167/10.7.678

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, *39*(16), 2729–2737. https://doi.org/10.1016/S0042-6989(98)00285-5

Krause, M. R., & Pack, C. C. (2014). Contextual modulation and stimulus selectivity in extrastriate cortex. *Vision Research*, *104*, 36–46. https://doi.org/10.1016/j.visres.2014.10.006

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

Krieger, G., & Zetzsche, C. (1996). Nonlinear image operators for the evaluation of local intrinsic dimensionality. *IEEE Transactions on Image Processing*, *5*(6), 1026–1042. https://doi.org/10.1109/83.503917

Kruschke, J. K. (2018). Rejecting or Accepting Parameter Values in Bayesian Estimation. *Advances in Methods and Practices in Psychological Science*, *1*(2), 270–280. https://doi.org/10.1177/2515245918771304

Kruschke, J. K., & Liddell, T. M. (2018a). Bayesian data analysis for newcomers. *Psychonomic Bulletin & Review*, *25*(1), 155–177. https://doi.org/10.3758/s13423-017-1272-1

Kruschke, J. K., & Liddell, T. M. (2018b). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206. https://doi.org/10.3758/s13423-016-1221-4

Kuefner, D. (2010). Early visually evoked electrophysiological responses over the human brain (P1, N170) show stable patterns of face-sensitivity from 4 years to adulthood. *Frontiers in Human Neuroscience*, *3*. https://doi.org/10.3389/neuro.09.067.2009

Kurz, S. (2019). *Statistical Rethinking with brms, ggplot2, and the tidyverse.* (Retrieved from osf.io/97t6w).

Laguesse, R., Dormal, G., Biervoye, A., Kuefner, D., & Rossion, B. (2012). Extensive visual training in adulthood significantly reduces the face inversion effect. *Journal of Vision*, *12*(10), 14–14. https://doi.org/10.1167/12.10.14

Lamme, V. A. F. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *The Journal of Neuroscience*, *15*(2), 1605–1615. https://doi.org/10.1523/JNEUROSCI.15-02-01605.1995

Lamme, V. A. F., Supèr, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology*, *8*(4), 529–535. https://doi.org/10.1016/S0959-4388(98)80042-1

Landis, T., Cummings, J. L., Christen, L., Bogen, J. E., & Imhof, H.-G. (1986). Are Unilateral Right Posterior Cerebral Lesions Sufficient to Cause Prosopagnosia? Clinical and Radiological Findings in Six Additional Patients. *Cortex*, *22*(2), 243–252. https://doi.org/10.1016/S0010-9452(86)80048-X

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2004). Impairment in Holistic Face Processing Following Early Visual Deprivation. *Psychological Science*, *15*(11), 762–768. https://doi.org/10.1111/j.0956-7976.2004.00753.x

Leder, H., & Bruce, V. (2000). When Inverted Faces are Recognized: The Role of Configural Information in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A*, *53*(2), 513–536. https://doi.org/10.1080/713755889

Leehey, S., Carey, S., Diamond, R., & Cahn, A. (1978). Upright and Inverted Faces: The Right Hemisphere Knows the Difference. *Cortex*, *14*(3), 411–419. https://doi.org/10.1016/S0010-9452(78)80067-7

Levick, W. R., Cleland, B. G., & Dubin, M. W. (1972). Lateral geniculate neurons of cat: Retinal inputs and physiology. *Investigative Ophthalmology*, *11*(5), 302–311.

Levine, S. C., Banich, M. T., & Koch-Weser, M. P. (1988). Face recognition: A general or specific right hemisphere capacity? *Brain and Cognition*, *8*(3), 303–325. https://doi.org/10.1016/0278-2626(88)90057-7

Levitt, J. B., & Lund, J. S. (1997). Contrast dependence of contextual effects in primate visual cortex. *Nature*, *387*(6628), 73–76. https://doi.org/10.1038/387073a0

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, C.-Y., & Li, W. (1994). Extensive integration field beyond the classical receptive field of cat's striate cortical neurons—Classification and tuning properties. *Vision Research*, *34*(18), 2337–2355. https://doi.org/10.1016/0042-6989(94)90280-1

Liu-Shuang, J., Ales, J. M., Rossion, B., & Norcia, A. M. (2015). The effect of contrast polarity reversal on face detection: Evidence of perceptual asymmetry from sweep VEP. *Vision Research*, *108*, 8–19. https://doi.org/10.1016/j.visres.2015.01.001

Liu-Shuang, J., Yang, Y.-F., Rossion, B., & Goffaux, V. (2022). Natural Contrast Statistics Facilitate Human Face Categorization. *Eneuro*, *9*(5), ENEURO.0420-21.2022. https://doi.org/10.1523/ENEURO.0420-21.2022

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163

Maffei, L., & Fiorentini, A. (1976). The unresponsive regions of visual cortical receptive fields. *Vision Research*, *16*(10), 1131-IN5. https://doi.org/10.1016/0042-6989(76)90253-4

Mannion, D. J., Donkin, C., & Whitford, T. J. (2017). No apparent influence of psychometrically-defined schizotypy on orientation-dependent contextual modulation of visual contrast detection. *PeerJ*, *5*, e2921. https://doi.org/10.7717/peerj.2921

Marozzi, M. (2007). Some remarks about the number of permutations one should consider to perform a permutation test [Application/octet-stream]. *Statistica; Vol 64*, No 1 (2004); 193201. https://doi.org/10.6092/ISSN.1973-2201/32

Marrocco, R., McClurkin, J., & Young, R. (1982). Spatial summation and conduction latency classification of cells of the lateral geniculate nucleus of macaques. *The Journal of Neuroscience*, *2*(9), 1275–1291. https://doi.org/10.1523/JNEUROSCI.02-09-01275.1982

Mattson, A., Levin, H., & Grafman, J. (2000). A Case of Prosopagnosia Following Moderate Closed Head Injury with Left Hemisphere Focal Lesion. *Cortex*, *36*(1), 125–137. https://doi.org/10.1016/S0010-9452(08)70841-4

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255–260. https://doi.org/10.1016/S1364-6613(02)01903-4

Mazard, A., Schiltz, C., & Rossion, B. (2006). Recovery from adaptation to facial identity is larger for upright than inverted faces in the human occipito-temporal cortex. *Neuropsychologia*, *44*(6), 912–922. https://doi.org/10.1016/j.neuropsychologia.2005.08.015

McCaffery, J. M., Robertson, D. J., Young, A. W., & Burton, A. M. (2018). Individual differences in face identity processing. *Cognitive Research: Principles and Implications*, *3*(1), 21. https://doi.org/10.1186/s41235-018-0112-9

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-Specific Processing in the Human Fusiform Gyrus. *Journal of Cognitive Neuroscience*, *9*(5), 605–610. https://doi.org/10.1162/jocn.1997.9.5.605

McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC Press/Taylor & Francis Group.

McIlwain, J. T. (1964). RECEPTIVE FIELDS OF OPTIC TRACT AXONS AND LATERAL GENICULATE CELLS: PERIPHERAL EXTENT AND BARBITURATE SENSITIVITY. *Journal of Neurophysiology*, *27*(6), 1154–1173. https://doi.org/10.1152/jn.1964.27.6.1154

McKone, E. (2009). Holistic processing for faces operates over a wide range of sizes but is strongest at identification rather than conversational distances. *Vision Research*, *49*(2), 268–283. https://doi.org/10.1016/j.visres.2008.10.020

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences*, *11*(1), 8–15. https://doi.org/10.1016/j.tics.2006.11.002

McKone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. *Psychonomic Bulletin & Review*, *16*(5), 778–797. https://doi.org/10.3758/PBR.16.5.778

Meadows, J. C. (1974). The anatomical basis of prosopagnosia. *Journal of Neurology, Neurosurgery & Psychiatry*, *37*(5), 489–501. https://doi.org/10.1136/jnnp.37.5.489

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, *34*(4), 865–876. https://doi.org/10.3758/BF03193433

Meinhardt-Injac, B., Persike, M., & Meinhardt, G. (2014). Holistic processing and reliance on global viewing strategies in older adults' face perception. *Acta Psychologica*, *151*, 155–163. https://doi.org/10.1016/j.actpsy.2014.06.001

Merkel, D. (2014). Docker: Lightweight Linux Containers for Consistent Development and Deployment. *Linux Journal*, *239*(2).

Michel, C., Rossion, B., Han, J., Chung, C.-S., & Caldara, R. (2006). Holistic Processing Is Finely Tuned for Faces of One's Own Race. *Psychological Science*, *17*(7), 608–615. https://doi.org/10.1111/j.1467-9280.2006.01752.x

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural Face Processing Develops more Slowly than Featural Face Processing. *Perception*, *31*(5), 553–566. https://doi.org/10.1068/p3339

Mondloch, C. J., Lewis, T. L., Budreau, D. R., Maurer, D., Dannemiller, J. L., Stephens, B. R., & Kleiner-Gathercoal, K. A. (1999). Face Perception During Early Infancy. *Psychological Science*, *10*(5), 419–422. https://doi.org/10.1111/1467-9280.00179

Moors, P., Costa, T. L., & Wagemans, J. (2020). Configural superiority for varying contrast levels. *Attention, Perception, & Psychophysics*, *82*(3), 1355–1367. https://doi.org/10.3758/s13414-019-01917-y

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes Factors for Common Designs.* [Computer software]. https://CRAN.R-project.org/package=BayesFactor

Morris, J. S., Friston, K. J., & Dolan, R. J. (1997). Neural responses to salient visual stimuli. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *264*(1382), 769–775. https://doi.org/10.1098/rspb.1997.0109

Moscatelli, A., Mezzetti, M., & Lacquaniti, F. (2012). Modeling psychophysical data at the population-level: The generalized linear mixed model. *Journal of Vision*, *12*(11), 26–26. https://doi.org/10.1167/12.11.26

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What Is Special about Face Recognition? Nineteen Experiments on a Person with Visual Object Agnosia and Dyslexia but Normal Face Recognition. *Journal of Cognitive Neuroscience*, *9*(5), 555–604. https://doi.org/10.1162/jocn.1997.9.5.555

Murphy, J., Gray, K. L. H., & Cook, R. (2017). The composite face illusion. *Psychonomic Bulletin & Review*, *24*(2), 245–261. https://doi.org/10.3758/s13423-016-1131-5

Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research*, *39*(23), 3824–3833. https://doi.org/10.1016/S0042-6989(99)00096-6

Nauhaus, I., Busse, L., Carandini, M., & Ringach, D. L. (2009). Stimulus contrast modulates functional connectivity in visual cortex. *Nature Neuroscience*, *12*(1), 70–76. https://doi.org/10.1038/nn.2232

Nelson, J. I., & Frost, B. J. (1978). Orientation-selective inhibition from beyond the classic visual receptive field. *Brain Research*, *139*(2), 359–365. https://doi.org/10.1016/0006-8993(78)90937-x

Nishimura, M., Rutherford, M. D., & Maurer, D. (2008). Converging evidence of configural processing of faces in high-functioning adults with autism spectrum disorders. *Visual Cognition*, *16*(7), 859–891. https://doi.org/10.1080/13506280701538514

Nolt, M. J., Kumbhani, R. D., & Palmer, L. A. (2004). Contrast-Dependent Spatial Summation in the Lateral Geniculate Nucleus and Retina of the Cat. *Journal of Neurophysiology*, *92*(3), 1708–1717. https://doi.org/10.1152/jn.00176.2004

Nunnally Jr, J. C. (1970). *Introduction to psychological measurement.*

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Olsen, S. R., & Wilson, R. I. (2008). Lateral presynaptic inhibition mediates gain control in an olfactory circuit. *Nature*, *452*(7190), 956–960. https://doi.org/10.1038/nature06864

Olshausen, B., & Field, D. (2004). Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, *14*(4), 481–487. https://doi.org/10.1016/j.conb.2004.07.007

Oruc, I., & Barton, J. J. S. (2010). Critical frequencies in the perception of letters, faces, and novel shapes: Evidence for limited scale invariance for faces. *Journal of Vision*, *10*(12), 20–20. https://doi.org/10.1167/10.12.20

O'Toole, B., & Wenderoth, P. (1977). The tilt illusion: Repulsion and attraction effects in the oblique meridian. *Vision Research*, *17*(3), 367–374. https://doi.org/10.1016/0042-6989(77)90025-6

Pachai, M. V., Bennett, P. J., & Sekuler, A. B. (2018). The Bandwidth of Diagnostic Horizontal Structure for Face Identification. *Perception*, *47*(4), 397–413. https://doi.org/10.1177/0301006618754479

Pack, C. C., Hunter, J. N., & Born, R. T. (2005). Contrast Dependence of Suppressive Influences in Cortical Area MT of Alert Macaque. *Journal of Neurophysiology*, *93*(3), 1809–1815. https://doi.org/10.1152/jn.00629.2004

Palermo, R., Willis, M. L., Rivolta, D., McKone, E., Wilson, C. E., & Calder, A. J. (2011). Impaired holistic coding of facial expression and facial identity in congenital prosopagnosia. *Neuropsychologia*, *49*(5), 1226–1235. https://doi.org/10.1016/j.neuropsychologia.2011.02.021

Palmer, S. E. (2002). *Vision science: Photons to phenomenology* (3. printing). MIT Press.

Palmer, T. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*(5), 519–526. https://doi.org/10.3758/BF03197524

Passarotti, A. M., Smith, J., DeLano, M., & Huang, J. (2007). Developmental differences in the neural bases of the face inversion effect show progressive tuning of face-selective regions to the

upright orientation. *NeuroImage*, *34*(4), 1708–1722. https://doi.org/10.1016/j.neuroimage.2006.07.045

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peli, E., Lee, E., Trempe, C. L., & Buzney, S. (1994). Image enhancement for the visually impaired: The effects of enhancement on face recognition. *Journal of the Optical Society of America A*, *11*(7), 1929. https://doi.org/10.1364/JOSAA.11.001929

Peters, J. C., Goebel, R., & Goffaux, V. (2018). From coarse to fine: Interactive feature processing precedes local feature analysis in human face perception. *Biological Psychology*, *138*, 1–10. https://doi.org/10.1016/j.biopsycho.2018.07.009

Peterzell, D. H. (2016). Discovering Sensory Processes Using Individual Differences: A Review and Factor Analytic Manifesto. *Electronic Imaging*, *28*(16), 1–11. https://doi.org/10.2352/ISSN.2470-1173.2016.16.HVEI-112

Petras, K., ten Oever, S., Dalal, S. S., & Goffaux, V. (2021). Information redundancy across spatial scales modulates early visual cortical processing. *NeuroImage*, *244*, 118613. https://doi.org/10.1016/j.neuroimage.2021.118613

Petras, K., ten Oever, S., Jacobs, C., & Goffaux, V. (2019). Coarse-to-fine information integration in human vision. *NeuroImage*, *186*, 103–112. https://doi.org/10.1016/j.neuroimage.2018.10.086

Petro, L. S., Smith, F. W., Abbatecola, C., & Muckli, L. (2023). The Spatial Precision of Contextual Feedback Signals in Human V1. *Biology*, *12*(7), 1022. https://doi.org/10.3390/biology12071022

Petrov, Y., & McKee, S. P. (2006). The effect of spatial configuration on surround suppression of contrast sensitivity. *Journal of Vision*, *6*(3), 4. https://doi.org/10.1167/6.3.4

Phillips, & O'Toole, A. J. (2014). Comparison of human and computer performance across face recognition experiments. *Image and Vision Computing*, *32*(1), 74–85. https://doi.org/10.1016/j.imavis.2013.12.002

Phillips, R. J., & Rawles, R. E. (1979). Recognition of Upright and Inverted Faces: A Correlational Study. *Perception*, *8*(5), 577–583. https://doi.org/10.1068/p080577

Pinsk, M. A., Arcaro, M., Weiner, K. S., Kalkus, J. F., Inati, S. J., Gross, C. G., & Kastner, S. (2009). Neural Representations of Faces and Body Parts in Macaque and Human Cortex: A Comparative fMRI Study. *Journal of Neurophysiology*, *101*(5), 2581–2600. https://doi.org/10.1152/jn.91198.2008

Pitcher, D., Ianni, G., & Ungerleider, L. G. (2019). A functional dissociation of face-, body- and scene-selective brain areas based on their response to moving and static stimuli. *Scientific Reports*, *9*(1), 8242. https://doi.org/10.1038/s41598-019-44663-9

Polat, U., & Sagi, D. (1993). Lateral interactions between spatial channels: Suppression and facilitation revealed by lateral masking experiments. *Vision Research*, *33*(7), 993–999. https://doi.org/10.1016/0042-6989(93)90081-7

Polat, U., & Sagi, D. (1994). Spatial interactions in human vision: From near to farvia experience-dependent cascades of connections. *Proceedings of the National Academy of Sciences*, *91*(4), 1206–1209. https://doi.org/10.1073/pnas.91.4.1206

Poltoratski, S., Kay, K., Finzi, D., & Grill-Spector, K. (2021). Holistic face recognition is an emergent phenomenon of spatial processing in face-selective regions. *Nature Communications*, *12*(1), 4745. https://doi.org/10.1038/s41467-021-24806-1

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*, 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048

Prinz, J. J. (2005). *Is the mind really modular*. https://api.semanticscholar.org/CorpusID:6991934

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential Sensitivity of Human Visual Cortex to Faces, Letterstrings, and Textures: A Functional Magnetic Resonance Imaging Study. *The Journal of Neuroscience*, *16*(16), 5205–5215. https://doi.org/10.1523/JNEUROSCI.16-16-05205.1996

Puce, A., Allison, T., Gore, J. C., & McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. *Journal of Neurophysiology*, *74*(3), 1192–1199. https://doi.org/10.1152/jn.1995.74.3.1192

Quintana, D. S., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. *BMC Psychiatry*, *18*(1), 178. https://doi.org/10.1186/s12888-018-1761-4

R Core Team. (2020). *R: A language and environment for statistical computing*. https://www.R-project.org/

Rajimehr, R., Young, J. C., & Tootell, R. B. H. (2009). An anterior temporal face patch in human cortex, predicted by macaque maps. *Proceedings of the National Academy of Sciences*, *106*(6), 1995–2000. https://doi.org/10.1073/pnas.0807304106

Ramon, M., Busigny, T., & Rossion, B. (2010). Impaired holistic processing of unfamiliar individual faces in acquired prosopagnosia. *Neuropsychologia*, *48*(4), 933–944. https://doi.org/10.1016/j.neuropsychologia.2009.11.014

Ramon, M., Dricot, L., & Rossion, B. (2010). Personally familiar faces are perceived categorically in face-selective regions other than the fusiform face area: Response properties of face-preferential regions. *European Journal of Neuroscience*, *32*(9), 1587–1598. https://doi.org/10.1111/j.1460-9568.2010.07405.x

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). sage.

Regaiolli, B., Rizzo, A., Ottolini, G., Miletto Petrazzini, M. E., Spiezio, C., & Agrillo, C. (2019). Motion Illusions as Environmental Enrichment for Zoo Animals: A Preliminary Investigation on Lions (Panthera leo). *Frontiers in Psychology*, *10*, 2220. https://doi.org/10.3389/fpsyg.2019.02220

Renzi, E. D. (1986). Prosopagnosia in two patients with CT scan evidence of damage confined to the right hemisphere. *Neuropsychologia*, *24*(3), 385–389. https://doi.org/10.1016/0028-3932(86)90023-0

Rezlescu, C., Susilo, T., Wilmer, J. B., & Caramazza, A. (2017). The inversion, part-whole, and composite effects reflect distinct perceptual mechanisms with varied relationships to face recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(12), 1961–1973. https://doi.org/10.1037/xhp0000400

Rhodes, G. (1988). Looking at Faces: First-Order and Second-Order Features as Determinants of Facial Appearance. *Perception*, *17*(1), 43–63. https://doi.org/10.1068/p170043

Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces? *Cognition*, *47*(1), 25–57. https://doi.org/10.1016/0010-0277(93)90061-Y

Richardson, H., Taylor, J., Kane-Grade, F., Powell, L., Bosquet Enlow, M., & Nelson, C. A. (2021). Preferential responses to faces in superior temporal and medial prefrontal cortex in three-year-old children. *Developmental Cognitive Neuroscience*, *50*, 100984. https://doi.org/10.1016/j.dcn.2021.100984

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011a). Beliefs alter holistic face processing ... If response bias is not taken into account. *Journal of Vision*, *11*(13), 17–17. https://doi.org/10.1167/11.13.17

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011b). Holistic processing predicts face recognition. *Psychological Science*, *22*(4), 464–471. https://doi.org/10.1177/0956797611401753

Richler, J. J., Cheung, O. S., Wong, A. C. N., & Gauthier, I. (2009). Does response interference contribute to face composite effects? *Psychonomic Bulletin & Review*, *16*(2), 258–263. https://doi.org/10.3758/PBR.16.2.258

Richler, J. J., Floyd, R. J., & Gauthier, I. (2014). The Vanderbilt Holistic Face Processing Test: A short and reliable measure of holistic face processing. *Journal of Vision*, *14*(11), 10–10. https://doi.org/10.1167/14.11.10

Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychological Bulletin*, *140*(5), 1281–1302. https://doi.org/10.1037/a0037004

Richler, J. J., Gauthier, I., Wenger, M. J., & Palmeri, T. J. (2008). Holistic processing of faces: Perceptual and decisional components. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 328–342. https://doi.org/10.1037/0278-7393.34.2.328

Richler, J. J., Mack, M. L., Gauthier, I., & Palmeri, T. J. (2009). Holistic processing of faces happens at a glance. *Vision Research*, *49*(23), 2856–2861. https://doi.org/10.1016/j.visres.2009.08.025

Richler, J. J., Palmeri, T. J., & Gauthier, I. (2012). Meanings, Mechanisms, and Measures of Holistic Processing. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00553

Richler, J. J., Tanaka, J. W., Brown, D. D., & Gauthier, I. (2008). Why does selective attention to parts fail in face processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1356–1368. https://doi.org/10.1037/a0013080

Richler, J. J., Wong, Y. K., & Gauthier, I. (2011). Perceptual Expertise as a Shift From Strategic Interference to Automatic Holistic Processing. *Current Directions in Psychological Science*, *20*(2), 129–134. https://doi.org/10.1177/0963721411402472

Ringo, J. L. (1991). Neuronal Interconnection as a Function of Brain Size. *Brain, Behavior and Evolution*, *38*(1), 1–6. https://doi.org/10.1159/000114375

Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition*, *103*(1), 34–79. https://doi.org/10.1016/j.cognition.2006.02.008

Rock, I. (1974). The Perception of Disoriented Figures. *Scientific American*, *230*(1), 78–85. https://doi.org/10.1038/scientificamerican0174-78

Rolls, E. T., & Tovee, M. J. (1995). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology*, *73*(2), 713–726. https://doi.org/10.1152/jn.1995.73.2.713

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Ross, D. A., & Gauthier, I. (2015). Holistic processing in the composite task depends on face size. *Visual Cognition*, *23*(5), 533–545. https://doi.org/10.1080/13506285.2015.1049678

Ross, D. A., Richler, J. J., & Gauthier, I. (2015). Reliability of composite-task measurements of holistic face processing. *Behavior Research Methods*, *47*(3), 736–743. https://doi.org/10.3758/s13428-014-0497-4

Rossion, B. (1999). Spatio-temporal localization of the face inversion effect: An event-related potentials study. *Biological Psychology*, *50*(3), 173–189. https://doi.org/10.1016/S0301-0511(99)00013-7

Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica*, *128*(2), 274–289. https://doi.org/10.1016/j.actpsy.2008.02.003

Rossion, B. (2009). Distinguishing the cause and consequence of face inversion: The perceptual field hypothesis. *Acta Psychologica*, *132*(3), 300–312. https://doi.org/10.1016/j.actpsy.2009.08.002

Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition*, *21*(2), 139–253. https://doi.org/10.1080/13506285.2013.772929

Rossion, B., & Boremanse, A. (2008). Nonlinear relationship between holistic processing of individual faces and picture-plane rotation: Evidence from the face composite illusion. *Journal of Vision*, *8*(4), 3. https://doi.org/10.1167/8.4.3

Rossion, B., & Gauthier, I. (2002). How Does the Brain Process Upright and Inverted Faces? *Behavioral and Cognitive Neuroscience Reviews*, *1*(1), 63–75. https://doi.org/10.1177/1534582302001001004

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: An electrophysiological account of face-specific processes in the human brain. *NeuroReport*, *11*(1), 69–72. https://doi.org/10.1097/00001756-200001170-00014

Rossion, B., Hanseeuw, B., & Dricot, L. (2012). Defining face perception areas in the human brain: A large-scale factorial fMRI face localizer analysis. *Brain and Cognition*, *79*(2), 138–157. https://doi.org/10.1016/j.bandc.2012.01.001

Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *NeuroImage*, *39*(4), 1959–1979. https://doi.org/10.1016/j.neuroimage.2007.10.011

Rossion, B., & Michel, C. (2018). Normative accuracy and response time data for the computerized Benton Facial Recognition Test (BFRT-c). *Behavior Research Methods*, *50*(6), 2442–2460. https://doi.org/10.3758/s13428-018-1023-x

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, *26*(2), 452–467. https://doi.org/10.3758/s13423-018-1558-y

Roux-Sibilon, A., Peyrin, C., Greenwood, J. A., & Goffaux, V. (2023). Radial bias in face identification. *Proceedings of the Royal Society B: Biological Sciences*, *290*(2001), 20231118. https://doi.org/10.1098/rspb.2023.1118

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252–257. https://doi.org/10.3758/PBR.16.2.252

Sachdev, R. N. S., Krause, M. R., & Mazer, J. A. (2012). Surround suppression and sparse coding in visual and barrel cortices. *Frontiers in Neural Circuits*, *6*. https://doi.org/10.3389/fncir.2012.00043

Samaey, C., Wagemans, J., & Moors, P. (2020). Individual differences in processing orientation and proximity as emergent features. *Vision Research*, *169*, 12–24. https://doi.org/10.1016/j.visres.2020.02.002

Scapinello, K. F., & Yarmey, A. D. (1970). The role of familiarity and orientation in immediate and delayed recognition of pictorial stimuli. *Psychonomic Science*, *21*(6), 329–330. https://doi.org/10.3758/BF03335807

Sceniak, M. P., Chatterjee, S., & Callaway, E. M. (2006). Visual Spatial Summation in Macaque Geniculocortical Afferents. *Journal of Neurophysiology*, *96*(6), 3474–3484. https://doi.org/10.1152/jn.00734.2006

Sceniak, M. P., Hawken, M. J., & Shapley, R. (2001). Visual Spatial Characterization of Macaque V1 Neurons. *Journal of Neurophysiology*, *85*(5), 1873–1887. https://doi.org/10.1152/jn.2001.85.5.1873

Sceniak, M. P., Ringach, D. L., Hawken, M. J., & Shapley, R. (1999). Contrast's effect on spatial summation by macaque V1 neurons. *Nature Neuroscience*, *2*(8), 733–739. https://doi.org/10.1038/11197

Schiltz, C., Dricot, L., Goebel, R., & Rossion, B. (2010). Holistic perception of individual faces in the right middle fusiform gyrus as evidenced by the composite face illusion. *Journal of Vision*, *10*(2), 1–16. https://doi.org/10.1167/10.2.25

Schiltz, C., & Rossion, B. (2006). Faces are represented holistically in the human occipito-temporal cortex. *NeuroImage*, *32*, 1385–1394. https://doi.org/10.1016/j.neuroimage.2006.05.037

Schloss, K. B., Fortenbaugh, F. C., & Palmer, S. E. (2014). The configural shape illusion. *Journal of Vision*, *14*(8), 23–23. https://doi.org/10.1167/14.8.23

Schobert, A.-K., Corradi-Dell'Acqua, C., Frühholz, S., van der Zwaag, W., & Vuilleumier, P. (2018). Functional organization of face processing in the human superior temporal sulcus: A 7T high-resolution fMRI study. *Social Cognitive and Affective Neuroscience*, *13*(1), 102–113. https://doi.org/10.1093/scan/nsx119

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, *25*(1), 128–142. https://doi.org/10.3758/s13423-017-1230-y

Schuurmans, J. P., Bennett, M. A., Petras, K., & Goffaux, V. (2023). Backward masking reveals coarse-to-fine dynamics in human V1. *NeuroImage*, *274*, 120139. https://doi.org/10.1016/j.neuroimage.2023.120139

Schwartz, O., Hsu, A., & Dayan, P. (2007). Space and time in visual context. *Nature Reviews Neuroscience*, *8*, 522–535. https://doi.org/10.1038/nrn2155

Schwarzkopf, D. S., & Rees, G. (2013). Subjective Size Perception Depends on Central Visual Cortical Magnification in Human V1. *PLoS ONE*, *8*(3), e60550. https://doi.org/10.1371/journal.pone.0060550

Schwarzkopf, D. S., Song, C., & Rees, G. (2011). The surface area of human V1 predicts the subjective experience of object size. *Nature Neuroscience*, *14*(1), 28–30. https://doi.org/10.1038/nn.2706

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

Self, M. W., Lorteije, J. A. M., Vangeneugden, J., van Beest, E. H., Grigore, M. E., Levelt, C. N., Heimel, J. A., & Roelfsema, P. R. (2014). Orientation-Tuned Surround Suppression in Mouse

Visual Cortex. *Journal of Neuroscience*, *34*(28), 9290–9304. https://doi.org/10.1523/JNEUROSCI.5051-13.2014

Sengpiel, F., Sen, A., & Blakemore, C. (1997). Characteristics of surround inhibition in cat area 17: *Experimental Brain Research*, *116*(2), 216–228. https://doi.org/10.1007/PL00005751

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R., & Tootell, R. B. H. (1995). Borders of Multiple Visual Areas in Humans Revealed by Functional Magnetic Resonance Imaging. *Science*, *268*(5212), 889–893. https://doi.org/10.1126/science.7754376

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology*, *75*(2), 221–242. https://doi.org/10.1111/j.2044-8295.1984.tb01895.x

Sergent, J., Ohta, S., & Macdonald, B. (1992). FUNCTIONAL NEUROANATOMY OF FACE AND OBJECT PROCESSING: A POSITRON EMISSION TOMOGRAPHY STUDY. *Brain*, *115*(1), 15–36. https://doi.org/10.1093/brain/115.1.15

Shaqiri, A., Roinishvili, M., Grzeczkowski, L., Chkonia, E., Pilz, K., Mohr, C., Brand, A., Kunchulia, M., & Herzog, M. H. (2018). Sex-related differences in vision are heterogeneous. *Scientific Reports*, *8*(1), 7521. https://doi.org/10.1038/s41598-018-25298-8

Shushruth, S., Ichida, J. M., Levitt, J. B., & Angelucci, A. (2009). Comparison of Spatial Summation Properties of Neurons in Macaque V1 and V2. *Journal of Neurophysiology*, *102*(4), 2069–2083. https://doi.org/10.1152/jn.00512.2009

Sillito, A. M., Grieve, K. L., Jones, H. E., Cudeiro, J., & Davls, J. (1995). Visual cortical mechanisms detecting focal orientation discontinuities. *Nature*, *378*(6556), 492–496. https://doi.org/10.1038/378492a0

Simion, F., & Giorgio, E. D. (2015). Face perception and processing in early infancy: Inborn predispositions and developmental changes. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00969

Simion, F., Macchi Cassia, V., Turati, C., & Valenza. (2003). Non-specific perceptual biases at the origins of face processing. In *The Development of Face Processing in Infancy and Early Childhood* (pp. 13–25).

Simion, F., Macchi Cassia, V., Turati, C., & Valenza, E. (2001). The origins of face perception: Specific versus non-specific mechanisms. *Infant and Child Development*, *10*(1–2), 59–65. https://doi.org/10.1002/icd.247

Simion, F., Turati, C., Valenza, E., & Leo, I. (2006). The emergence of cognitive specialization in infancy: The case of face preference. *Attention and Performance, XXI*, 189–208.

Smith, F. W., & Muckli, L. (2010). Nonstimulated early visual areas carry information about surrounding context. *Proceedings of the National Academy of Sciences*, *107*(46), 20099–20103. https://doi.org/10.1073/pnas.1000233107

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed). Sage.

Solomon, S. G. (2006). Suppressive Surrounds and Contrast Gain in Magnocellular-Pathway Retinal Ganglion Cells of Macaque. *Journal of Neuroscience*, *26*(34), 8715–8726. https://doi.org/10.1523/JNEUROSCI.0821-06.2006

Solomon, S. G., White, A. J. R., & Martin, P. R. (2002). Extraclassical Receptive Field Properties of Parvocellular, Magnocellular, and Koniocellular Cells in the Primate Lateral Geniculate Nucleus. *The Journal of Neuroscience*, *22*(1), 338–349. https://doi.org/10.1523/JNEUROSCI.22-01-00338.2002

Song, C., Sandberg, K., Rutiku, R., & Kanai, R. (2022). Linking human behaviour to brain structure: Further challenges and possible solutions. *Nature Reviews Neuroscience*, *23*(8), 517–518. https://doi.org/10.1038/s41583-022-00614-4

Song, C., Schwarzkopf, D. S., Kanai, R., & Rees, G. (2015). Neural Population Tuning Links Visual Cortical Anatomy to Human Visual Perception. *Neuron*, *85*(3), 641–656. https://doi.org/10.1016/j.neuron.2014.12.041

Song, C., Schwarzkopf, D. S., & Rees, G. (2011). Interocular induction of illusory size perception. *BMC Neuroscience*, *12*(1), 27. https://doi.org/10.1186/1471-2202-12-27

Song, C., Schwarzkopf, D. S., & Rees, G. (2013). Variability in visual cortex size reflects tradeoff between local orientation sensitivity and global orientation modulation. *Nature Communications*, *4*(1), 2201. https://doi.org/10.1038/ncomms3201

Sowden, P. T., & Schyns, P. G. (2006). Channel surfing in the visual brain. *Trends in Cognitive Sciences*, *10*(12), 538–545. https://doi.org/10.1016/j.tics.2006.10.007

Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, *15*(1), 72. https://doi.org/10.2307/1412159

Stan Development Team. (2023). *RStan: The R interface to Stan*. https://mc-stan.org/

Stensaas, S. S., Eddington, D. K., & Dobelle, W. H. (1974). The topography and variability of the primary visual cortex in man. *Journal of Neurosurgery*, *40*(6), 747–755. https://doi.org/10.3171/jns.1974.40.6.0747

Sterling, P., & Wickelgren, B. G. (1969). Visual receptive fields in the superior colliculus of the cat. *Journal of Neurophysiology*, *32*(1), 1–15. https://doi.org/10.1152/jn.1969.32.1.1

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Sunday, M. A., Richler, J. J., & Gauthier, I. (2017). Limited evidence of individual differences in holistic processing in different versions of the part-whole paradigm. *Attention, Perception, & Psychophysics*, *79*(5), 1453–1465. https://doi.org/10.3758/s13414-017-1311-z

Sutter, M. L., Schreiner, C. E., McLean, M., O'connor, K. N., & Loftus, W. C. (1999). Organization of Inhibitory Frequency Receptive Fields in Cat Primary Auditory Cortex. *Journal of Neurophysiology*, *82*(5), 2358–2371. https://doi.org/10.1152/jn.1999.82.5.2358

Tanaka, & Farah, M. J. (1993). Parts and Wholes in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A*, *46*(2), 225–245. https://doi.org/10.1080/14640749308401045

Tanaka, & Gauthier, I. (1997). Expertise in Object and Face Recognition. In *Psychology of Learning and Motivation* (Vol. 36, pp. 83–125). Elsevier. https://doi.org/10.1016/S0079-7421(08)60282-0

Tanaka, J. W., Kaiser, M. D., Butler, S., & Le Grand, R. (2012). Mixed emotions: Holistic and analytic perception of facial expressions. *Cognition & Emotion*, *26*(6), 961–977. https://doi.org/10.1080/02699931.2011.630933

Tarr, M. J., & Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, *3*(8), 764–769. https://doi.org/10.1038/77666

Taubert, J., Apthorp, D., Aagten-Murphy, D., & Alais, D. (2011). The role of holistic processing in face perception: Evidence from the face inversion effect. *Vision Research*, *51*(11), 1273–1278. https://doi.org/10.1016/j.visres.2011.04.002

Taubert, J., Van Belle, G., Vanduffel, W., Rossion, B., & Vogels, R. (2015). The effect of face inversion for neurons inside and outside fMRI-defined face-selective cortical regions. *Journal of Neurophysiology*, *113*(5), 1644–1655. https://doi.org/10.1152/jn.00700.2014

Thompson, P., & Wilson, J. (2012). Why Do Most Faces Look Thinner Upside Down? *I-Perception*, *3*(10), 765–774. https://doi.org/10.1068/i0554

Todorov, A. (2012). The role of the amygdala in face perception and evaluation. *Motivation and Emotion*, *36*(1), 16–26. https://doi.org/10.1007/s11031-011-9238-5

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (2000). RESPONSE PROPERTIES OF THE HUMAN FUSIFORM FACE AREA. *Cognitive Neuropsychology*, *17*(1–3), 257–280. https://doi.org/10.1080/026432900380607

Toyama, J. S. (1977). *THE EFFECT OF ORIENTATION ON THE RECOGNITION OF FACES: A REPLY TO YIN.*

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of Face Perception. *Annual Review of Neuroscience*, *31*(1), 411–437. https://doi.org/10.1146/annurev.neuro.30.051606.094238

Tsao, D. Y., Moeller, S., & Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences*, *105*(49), 19514–19519. https://doi.org/10.1073/pnas.0809662105

Turati, C. (2004). Why Faces Are Not Special to Newborns: An Alternative Account of the Face Preference. *Current Directions in Psychological Science*, *13*(1), 5–8. https://doi.org/10.1111/j.0963-7214.2004.01301002.x

Tustison, N. J., Avants, B. B., Cook, P. A., Yuanjie Zheng, Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology*, *79*(4), 471–491. https://doi.org/10.1111/j.2044-8295.1988.tb02747.x

Valentine, T., & Bruce, V. (1986). Recognizing familiar faces: The role of distinctiveness and familiarity. *Canadian Journal of Psychology / Revue Canadienne de Psychologie*, *40*(3), 300–305. https://doi.org/10.1037/h0080101

Valenza, E., Simion, F., Cassia, V. M., & Umiltà, C. (1996). Face preference at birth. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 892–903. https://doi.org/10.1037/0096-1523.22.4.892

Van Belle, G., De Graef, P., Verfaillie, K., Rossion, B., & Lefevre, P. (2010). Face inversion impairs holistic perception: Evidence from gaze-contingent stimulation. *Journal of Vision*, *10*(5), 10–10. https://doi.org/10.1167/10.5.10

Van Belle, G., Lefèvre, P., & Rossion, B. (2015). Face inversion and acquired prosopagnosia reduce the size of the perceptual field of view. *Cognition*, *136*, 403–408. https://doi.org/10.1016/j.cognition.2014.11.037

Van Geert, E., Moors, P., Haaf, J., & Wagemans, J. (2022). Same stimulus, same temporal context, different percept? Individual differences in hysteresis and adaptation when perceiving multistable dot lattices. *I-Perception*, *13*(4), 204166952211093. https://doi.org/10.1177/20416695221109300

Vanpaemel, W. (2010). Prior sensitivity in theory testing: An apologia for the Bayes factor. *Journal of Mathematical Psychology*, *54*(6), 491–498. https://doi.org/10.1016/j.jmp.2010.07.003

Vega-Bermudez, F., & Johnson, K. O. (1999). SA1 and RA Receptive Fields, Response Variability, and Population Responses Mapped with a Probe Array. *Journal of Neurophysiology*, *81*(6), 2701–2710. https://doi.org/10.1152/jn.1999.81.6.2701

Ventura, P., Ngan, V., Pereira, A., Cruz, F., Guerreiro, J. C., Rosário, V., Delgado, J., Faustino, B., Barros, M., Domingues, M., & Wong, A. (2022). The relation between holistic processing as measured by three composite tasks and face processing: A latent variable modeling approach. *Attention, Perception, & Psychophysics*, *84*(7), 2319–2334. https://doi.org/10.3758/s13414-022-02543-x

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2017). General and specific factors in the processing of faces. *Vision Research*, *141*, 217–227. https://doi.org/10.1016/j.visres.2016.12.014

Versavel, M., Orban, G. A., & Lagae, L. (1990). Responses of visual cortical neurons to curved stimuli and chevrons. *Vision Research*, *30*(2), 235–248. https://doi.org/10.1016/0042-6989(90)90039-N

Vinje, W. E., & Gallant, J. L. (2000). Sparse Coding and Decorrelation in Primary Visual Cortex During Natural Vision. *Science*, *287*(5456), 1273–1276. https://doi.org/10.1126/science.287.5456.1273

Vinje, W. E., & Gallant, J. L. (2002). Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *Journal of Neuroscience*, *22*(7), 2904–2915. https://doi.org/10.1523/JNEUROSCI.22-07-02904.2002

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature Neuroscience*, *6*(6), 624–631. https://doi.org/10.1038/nn1057

Wagner, D. A. (1977). Ontogeny of the Ponzo Illusion: Effects of Age, Schooling, and Environment. *International Journal of Psychology*, *12*(3), 161–176. https://doi.org/10.1080/00207597708247386

Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012). Individual Differences in Holistic Processing Predict Face Recognition Ability. *Psychological Science*, *23*(2), 169–177. https://doi.org/10.1177/0956797611420575

Webster, M. A., Malkoc, G., Bilson, A. C., & Webster, S. M. (2002). Color contrast and contextual influences on color appearance. *Journal of Vision*, *2*(6), 7. https://doi.org/10.1167/2.6.7

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–1313. https://doi.org/10.3758/BF03194544

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (2nd ed. 2016). Springer International Publishing : Imprint: Springer. https://doi.org/10.1007/978-3-319-24277-4

Wiese, H., Kachel, U., & Schweinberger, S. R. (2013). Holistic face processing of own- and other-age faces in young and older adults: ERP evidence from the composite face task. *NeuroImage*, *74*, 306–317. https://doi.org/10.1016/j.neuroimage.2013.02.051

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. https://doi.org/10.3758/BRM.42.3.671

Williams, N. R., Willenbockel, V., & Gauthier, I. (2009). Sensitivity to spatial frequency and orientation content is not specific to face perception. *Vision Research*, *49*(19), 2353–2362. https://doi.org/10.1016/j.visres.2009.06.019

Wilmer, J. B. (2017). Individual Differences in Face Recognition: A Decade of Discovery. *Current Directions in Psychological Science*, *26*(3), 225–230. https://doi.org/10.1177/0963721417710693

Wilson, H. R., & Richards, W. A. (1992). Curvature and separation discrimination at texture boundaries. *Journal of the Optical Society of America A*, *9*(10), 1653. https://doi.org/10.1364/JOSAA.9.001653

Xing, J., & Heeger, D. J. (2000). Center-surround interactions in foveal and peripheral vision. *Vision Research*, *40*(22), 3065–3072. https://doi.org/10.1016/S0042-6989(00)00152-8

Xing, J., & Heeger, D. J. (2001). Measurement and modeling of center-surround suppression and enhancement. *Vision Research*, *41*(5), 571–583. https://doi.org/10.1016/S0042-6989(00)00270-4

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141–145. https://doi.org/10.1037/h0027474

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configural information in face perception. *Perception*, *16*(6), 747–759. https://doi.org/10.1068/p160747

Young, & Yamane, S. (1992). Sparse Population Coding of Faces in the Inferotemporal Cortex. *Science*, *256*(5061), 1327–1331. https://doi.org/10.1126/science.1598577

Yovel, G., & Duchaine, B. (2006). Specialized Face Perception Mechanisms Extract Both Part and Spacing Information: Evidence from Developmental Prosopagnosia. *Journal of Cognitive Neuroscience*, *18*(4), 580–593. https://doi.org/10.1162/jocn.2006.18.4.580

Yovel, G., & Kanwisher, N. (2005). The Neural Basis of the Behavioral Face-Inversion Effect. *Current Biology*, *15*(24), 2256–2262. https://doi.org/10.1016/j.cub.2005.10.072

Yovel, G., Wilmer, J. B., & Duchaine, B. (2014). What can individual differences reveal about face processing? *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00562

Yssaad-Fesselier, R., & Knoblauch, K. (2006). Modeling psychometric functions in R. *Behavior Research Methods*, *38*(1), 28–41. https://doi.org/10.3758/BF03192747

Yue, X., Cassidy, B. S., Devaney, K. J., Holt, D. J., & Tootell, R. B. H. (2011). Lower-Level Stimulus Features Strongly Influence Responses in the Fusiform Face Area. *Cerebral Cortex*, *21*(1), 35–47. https://doi.org/10.1093/cercor/bhq050

Zachariou, V., Nikas, C. V., Safiullah, Z. N., Gotts, S. J., & Ungerleider, L. G. (2016). Spatial Mechanisms within the Dorsal Visual Pathway Contribute to the Configural Processing of Faces. *Cerebral Cortex*, cercor;bhw224v1. https://doi.org/10.1093/cercor/bhw224

Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J. P., Chang, W.-C., Jenvay, S., Miyamichi, K., Luo, L., & Dan, Y. (2014). Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing. *Science (New York, N.Y.)*, *345*(6197), 660–665. https://doi.org/10.1126/science.1254126

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57. https://doi.org/10.1109/42.906424

Zhao, M., Bülthoff, H. H., & Bülthoff, I. (2016). Beyond Faces and Expertise: Facelike Holistic Processing of Nonface Objects in the Absence of Expertise. *Psychological Science*, *27*(2), 213–222. https://doi.org/10.1177/0956797615617779

Zhou, G., Liu, J., Xiao, N. G., Wu, S. J., Li, H., & Lee, K. (2018). The Fusiform Face Area Plays a Greater Role in Holistic Processing for Own-Race Faces Than Other-Race Faces. *Frontiers in Human Neuroscience*, *12*, 220. https://doi.org/10.3389/fnhum.2018.00220

Zhu, Q., Li, X., Chow, K., & Liu, J. (2009). The Part Task of the Part-Spacing Paradigm Is Not a Pure Measurement of Part-Based Information of Faces. *PLoS ONE*, *4*(7), e6239. https://doi.org/10.1371/journal.pone.0006239