ARTICLE IN PRESS

Consciousness and Cognition xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Perceptual averaging of facial expressions requires visual awareness and attention

Elric Elias*, Lauren Padama, Timothy D. Sweeny

Department of Psychology, University of Denver, United States

ARTICLE INFO

Keywords:
Face perception
Awareness, object substitution masking
Consciousness
Attention

ABSTRACT

Humans, as highly social animals, are regularly exposed to the faces of conspecifics—often more than one at a time. This feature of social living is important for understanding face perception, not just because it means that information from faces is available in bulk, but also because it changes the way individuals are perceived. For instance, when two faces are seen nearby one another, they tend to look like each other. This phenomenon of perceptual averaging is robust when both faces are seen and attended. But in everyday life, some faces may not receive the full benefit of attention, or they may not be visible at all. We evaluated whether perceptual averaging of relatively complex and simple information on faces, including facial expression and head orientation, can still occur even in these circumstances. In particular, we used object-substitution masking (OSM) and a dual-task designed to disrupt visual awareness and attention, respectively, during evaluations of briefly presented face pairs. Disruptions of awareness or attention eliminated averaging of facial expression, whereas orientation averaging persisted in spite of these challenges. These results demonstrate boundary conditions for the process of perceptual averaging. More generally, they provide insight into how the visual system processes multitudes of objects, both simple and complex, both with and without attention and awareness.

1. Introduction

People are sometimes seen one at a time. A friend's smile, for example, is easy to recognize across the room or in a video chat. In situations like these, sophisticated neural mechanisms support rapid analysis of a face's structure (Young & Yamane, 1992), identity (Hasselmo, Rolls, & Baylis, 1989), and emotion (Streit et al., 1999), sometimes even if that face's expression is not consciously seen (Adams, Gray, Garner, & Graf, 2010). Much of the psychological and vision research that focuses on the process of face perception is constrained to this solitary level of analysis. Yet humans are social creatures and people are often seen in the presence of others (Van Vugt & Kameda, 2012). Understanding how faces are encoded and then seen in these more complex circumstances is not simply a matter of extrapolating knowledge about how faces are perceived one at a time. Rather, the way people are seen depends on additional neural computations that take into account the presence of others nearby (Cosmides & Tooby, 2005). Work that highlights person-perception as it commonly occurs—in the context of other people—is crucial for a complete understanding of how faces are represented and then perceived by the visual system.

Toward this end, a great deal of attention has been focused on how visual information transcends individuals to the level of the collective (e.g., Alvarez, 2011; Whitney, Haberman, & Sweeny, 2014). For example, the process of ensemble coding allows an observer to appreciate the average emotion of an entire crowd at the expense of knowledge about its constituents (e.g., Elias, Dyer &

* Corresponding author at: University of Denver, Department of Psychology, 2155 S Race St, Frontier Hall, Denver, CO 80210, United States.
E-mail address: elric.elias@du.edu (E. Elias).

https://doi.org/10.1016/j.concog.2018.03.005
Received 25 September 2017; Received in revised form 27 December 2017; Accepted 13 March 2018
1053-8100/

Please cite this article as: Elias, E., Consciousness and Cognition (2018), https://doi.org/10.1016/j.concog.2018.03.005

Sweeny, 2017; Haberman & Whitney, 2007, 2009). Less attention has been focused on a separate but possibly related process—*perceptual averaging*—in which visual information is transferred *between* individuals, within the context of a crowd (Sweeny, Grabowecky, Paller, & Suzuki, 2009; Sweeny, Grabowecky, Kim, & Suzuki, 2011). In perceptual averaging, two spatially distinct faces can be made to appear similar to each other. For example, the anger of one face may appear to spread to a neutral face seen nearby. Although some details about the perceptual averaging of facial expression are clear—for example, it is likely to be, at least in part, a consequence of the large receptive fields of high-level ventral visual neurons—many questions remain about its mechanisms, its limitations, and when it can be expected to occur. The goal of the current investigation was to begin to bridge this gap by evaluating how perceptual averaging between multiple faces may depend on visual awareness and attention. In doing so, we hoped to better characterize the automaticity of face-specific perceptual averaging, determining whether it operates quickly, potentially based on feedforward visual representation characteristic of unconscious vision, or if it instead requires more in-depth and iterative representation associated with visual awareness (Pascual-Leone & Walsh, 2001; Silvanto, Cowey, Lavie, & Walsh, 2005; Juan & Walsh, 2003). Before discussing how perceptual averaging of facial expressions may depend on visual awareness, we first describe the phenomenon more generally from both neurophysiological and psychophysical perspectives.

Although the encoding and recognition of facial expressions of emotion is a distributed process (e.g., Pallett & Meng, 2013; Streit et al., 1999), a key stage of this analysis occurs in temporal visual cortex (e.g., Hasselmo et al., 1989). Neurons in high-level areas of the ventral visual stream like inferotemporal cortex (IT) are known to have very large receptive fields (e.g., Boussaoud, Desimone, & Ungerleider, 1991; Chelazzi, Duncan, Miller, & Desimone, 1998; Desimone & Gross, 1979; Niemeier, Goltz, Kuchinad, Tweed, & Vilis, 2004; Op De Beeck & Vogles, 2000). These large receptive fields, which often take up the entire contralateral visual hemifield, provide a notable computational benefit—they allow cells to respond consistently to a complex pattern, like a single face, despite dramatic changes to its location. However, this position-invariant coding introduces a cost of poor spatial resolution, which can be problematic when multiple objects or faces appear within a neuron's receptive field. Overcoming this tradeoff is often as simple as engaging selective attention, which can suppress the representation of the unattended object and allow such a neuron to respond as if only one object were in its receptive field (Chelazzi et al., 1998). But of course, selective attention cannot always be engaged because the locations of important objects are not always certain, attention must sometimes be distributed, and perception can often be fleeting. In these kinds of circumstances, ventral visual neurons are unable to resolve the multiple objects within their receptive fields, and they respond as if they "average" across those patterns (e.g., Kastner et al., 2001; Miller, Gochin, & Gross, 1993; Rolls & Tovee, 1995; Sato, 1989; Zoccolan, Cox, & Di Carlo, 2005). For example, when a pattern that elicits a strong response from a particular neuron (at least when seen in isolation, e.g., 30 spikes/s) is seen nearby a pattern that does little to push the neuron past its spontaneous firing rate (e.g., 10 spikes/s), this unit will produce an intermediate response (e.g., 20 spikes/s). This neural averaging does not occur when selective attention is engaged on one object from a pair, nor does it occur when the two objects appear in different receptive fields of ventral visual neurons (i.e., one on either side of the vertical meridian). Crucially for the current investigation, this process of neural averaging has a striking perceptual consequence.

When a face with a surprised expression is seen nearby a second face with a happy (or angry) expression, for example, the affect from each face appears to spread to the other (Sweeny et al., 2009, 2011). In this case, the happy face would appear less happy next to a surprised face, and vice versa. Crucially, this perceptual averaging occurs most strongly when a pair of faces is seen within a visual hemifield compared to when the faces appear in separate visual hemifields, illustrating that perceptual averaging can be constrained by the spatial selectivity (or lack thereof) of high-level visual neurons. Furthermore, the fact that averaging, both neural and perceptual, occurs when selective attention is *not* engaged on any individual face suggests that this process may operate as if by default, a natural consequence of the visual system's architecture that occurs any time information about multiple objects is present yet unbound by selection. Here, we put this notion of default averaging to a more stringent test, examining whether or not it would occur even when one face from a pair is not visible. This is important to examine, since lapses in visual awareness can be reasonably expected to occur in cluttered and rapidly changing visual scenes where multiple people are so often encountered. Moreover, faces, especially those with emotional expressions, are known to engage rapid but coarse visual processing even when they are not visible (for reviews, see Axelrod, Bar, & Rees, 2015; Faivre, Berthet, & Koudier, 2012). Many masking techniques could have allowed us to manipulate visual awareness in our investigation, but one in particular was best suited for this investigation—object-substitution masking (OSM).

Object-substitution masking (OSM) is a unique and powerful tool for disrupting perception (Enns, 2004; Enns & Di Lollo, 1997). In most OSM paradigms, a target object is shown for a brief amount of time, typically flanked by four masking dots. When these dots persist after the target disappears, even for just a fraction of a second, they can disrupt discrimination of that object's features and even eliminate awareness of its presence (although these effects can be independent; Gellatly, Pilling, Cole, & Skarratt, 2006). Although recent evidence suggests that OSM is not special in terms of its interaction with spatial attention (Argyropoulos, Gellatly, Pilling, & Carter, 2013; Filmer, Mattingley, & Dux, 2014, 2015; Goodhew & Edwards, 2016; Pilling, Gellatly, Argyropoulos, & Skarratt, 2014), it can still be differentiated from other types of masking in terms of its time course (Enns, 2004). More importantly, OSM is known to have a relatively late stage of interruption (Chakravarthi & Cavanagh, 2009), and relative to other types of visual masking (e.g., sandwich masking; Harris, Wu & Woldorff, 2011, backward masking; Woodman & Luck, 2003), we expected OSM to be the least likely to interfere with visual processing of facial expression over and above disrupting visual awareness. Below, we provide some background on the potential mechanisms of OSM in order to place our design and hypotheses in a deeper theoretical context.

Both early and more recent accounts of OSM (e.g., Di Lollo, 2014; Enns & Di Lollo, 1997; Enns, 2004) highlight the way it appears to selectively disrupt the kind of re-entrant communication between higher- and lower-level visual areas (e.g., extrastriate areas and V1) that appears to be necessary for visual awareness of objects (Lamme, Super, & Spekreijse, 1998; Pascual-Leon & Walsh, 2001; Silvanto, Cowey, Lavie, & Walsh, 2005). More specifically, these accounts of OSM propose that when information about a masked