1 How does gaze to faces support face-to-face interaction? A review and perspective

2 Roy S. Hessels

3 Experimental Psychology, Helmholtz Institute, Utrecht University, Utrecht, the
4 Netherlands & Developmental Psychology, Utrecht University, Utrecht, the Netherlands

5 Author Note

6 Corresponding author:

7 Roy S. Hessels

8 E-mail: royhessels@gmail.com; r.s.hessels@uu.nl

9 Heidelberglaan 1, 3584CS Utrecht, the Netherlands

10
11 This paper has been accepted in this form for publication in Psychonomic Bulletin
12 & Review. Please refer to the journal's website for any updates.

13
14 This work was supported by the Consortium on Individual Development (CID).
15 CID is funded through the Gravitation program of the Dutch Ministry of Education,
16 Culture, and Science and the NWO (Grant No. 024.001.003). I am particularly grateful
17 to Ignace Hooze for extensive discussions and comments on the theoretical framework
18 here proposed. I am further grateful to Chantal Kemner, Gijs Holleman, Yentl de Kloe,
19 Niilo Valtakari, Katja Dindar, and two anonymous reviewers for valuable comments on
20 earlier versions of this paper.

Abstract

Gaze – where one looks, how long, and when – plays an essential part in human social behaviour. While many aspects of social gaze have been reviewed, there is no comprehensive review or theoretical framework that describes how gaze to faces supports face-to-face interaction. In this review, I address the following questions: (1) When does gaze need to be allocated to a particular region of a face in order to provide the relevant information for successful interaction; (2) How do humans look at other people, and faces in particular, regardless of whether gaze needs to be directed at a particular region to acquire the relevant visual information; (3) How does gaze support the regulation of interaction? The work reviewed spans psychophysical research, observational research and eye-tracking research in both lab-based and interactive contexts. Based on the literature overview, I sketch a framework for future research based on dynamic systems theory. The framework holds that gaze should be investigated in relation to sub-states of the interaction, encompassing sub-states of the interactors, the content of the interaction as well as the interactive context. The relevant sub-states for understanding gaze in interaction vary over different timescales from microgenesis to ontogenesis and phylogenesis. The framework has important implications for vision science, psychopathology, developmental science and social robotics.

Keywords: gaze, faces, facial features, social interaction, dynamic system theory

How does gaze to faces support face-to-face interaction? A review and perspective

Introduction

Understanding how, when and where gaze or visual attention is allocated in the visual world is an important goal in (vision) science, as it reveals fundamental insights into the organism-environment interaction. Throughout vision science's history, the dominant approach to attaining this goal has been to study the 'atomic' features that 'constitute' the visual world – edges, orientations, colours, and so forth (e.g. Marr, 1982) – and determine how they drive the allocation of visual attention and gaze (e.g. Treisman & Gelade, 1980; Itti & Koch, 2000). Humans, as objects in the world that can be looked at or attended, have generally been treated as a special case to the visual system. Yet, in a world so fundamentally social, it would seem that encountering humans is the norm, while encountering single 'features' – or perhaps a few features combined as in a single red tilted line in the visual field – are the exception.

In this paper, I address the question of how gaze supports, and is an integral part of, social behaviour. Specifically, how does gaze to faces and facial features support dyadic face-to-face interactions? I focus on gaze, not visual attention, as gaze can be measured continuously using eye-tracking technology, as opposed to (covert) visual attention which is generally inferred from differences in manual reaction times. Gaze is here defined as the act of directing the eyes toward a location in the visual world, i.e. I thus always consider gaze as being directed somewhere or to something¹. Moreover, one's gaze direction often is accessible to other humans. For example, one can judge where one's fellow commuter on the train is looking and use this information to either start, or refrain from starting, a conversation. In interaction, gaze can thus support visual information uptake, but also signal information to others.

Previous reviews have addressed the evolution of social gaze and its function (Emery, 2000), how sensitivity to the eyes of others emerges and facilitates social behaviour (Grossmann, 2017), the affective effects of eye contact (Hietanen, 2018), and

¹It is important to realise that one may have the feeling of staring 'into nothingness', yet this act may be perceived as a strong social signal by someone else.

how the neural correlates of gaze (or joint attention in particular) in social interaction can be studied (Pfeiffer et al., 2013), for example through the simulation of social interactions (Caruana et al., 2017). However, there is no review that integrates empirical evidence from multiple research fields on how gaze supports social interaction at the resolution of faces and facial features for (neuro-)cognitive research to build on. Therefore, I introduce a dynamic system approach to interaction in order to understand gaze to faces for the support of social interaction. That this is relevant for vision research stems from the fact that there is a growing appreciation for the hypothesis that the human visual system has evolved for a large part under social constraints, which means that vision may be more ‘social’ in nature than previously considered (Adams et al., 2011).

Apart from the importance for the understanding of social gaze, an integrative theoretical framework of gaze in social interaction has key implications for multiple research fields. First, atypical gaze to people is symptomatic of a number of psychopathologies including Autism Spectrum Disorder (Senju & Johnson, 2009; Guillon et al., 2014) and Social Anxiety disorder (Horley et al., 2003; Wieser et al., 2009). In both disorders, atypical gaze, such as difficulties in making eye contact, seems particularly evident in interactive settings (as extensively discussed in Hessels et al., 2018a). A theoretical framework of interactive gaze might shed new light on atypicalities of gaze in these disorders. Second, gaze in interaction is considered an important social learning mechanism for development (e.g. Mundy et al., 2007; Brooks & Meltzoff, 2008; Gredebäck et al., 2010). Understanding which factors play a role in interactive gaze is a requirement for developmental theories of social learning through gaze. Finally, applied fields such as social robotics may benefit from a model of gaze in interaction to simulate gaze for the improvement of human-robot interaction (see e.g. Raidt et al., 2007; Mutlu et al., 2009; Skantze et al., 2014; Ruhland et al., 2015; Bailly et al., 2018; Willemse & Wykowska, 2019, for current applications of gaze modelling in virtual agents and social robots).

Outline of this review

In order to give the reader a general idea of the framework that I aim to present and of the interactions (see Table 1 for important definitions) to which it applies, consider the following example. In panel A of Figure 1, two musicians are depicted who are learning to play a song together. Sheet music is placed on the table in front of them. The person on the left seems to be indicating a particular part of the score for the other person to attend, perhaps to point out which chord should be played next. By looking at the eyes of the other, he can verify that his fellow musician is indeed paying attention to the score. Thus, gaze to parts of the face of the other here serves information acquisition about the state of the world. The person on the right clearly needs to look at the score in order to understand which bar the other person is pointing towards. Yet, his gaze direction (towards the table) is observable by the other, and may signal to the other where his visual attention is directed. Thus, one's gaze also affords information, often in combination with head or body orientation. Of course there is more to social interaction than just gaze. Should the interaction continue, the person on the right might look back to the face of the other and verify whether he has understood correctly that he should play an E minor chord next. From the smile on the left person's face, he concludes that this is indeed the case.

This example should make it clear that there are at least two important aspects of gaze in face-to-face interaction. On the one hand, visual information is gathered by directing gaze to parts of the visual world. On the other hand, gaze direction may be observable by others, and may thus afford information as well.² The latter is particularly evident in face-to-face meetings including multiple people (such as in panel B of Figure 1), where gaze can guide the flow of the interaction. Additionally, the fact

²This fact has been well known for a long time. For example, Kendon (1967) writes: "we shall offer some suggestions as to the function of gaze-direction, both as an act of perception by which one interactant can monitor the behaviour of the other, and as an expressive sign and regulatory signal by which he may influence the behaviour of the other." (p. 24). In recent eye-tracking research, the use of photos and videos of faces has been predominant. In this part of the literature, the regulatory signal of gaze-direction may have been, perhaps, overlooked.

that gaze may also signal information is thought to be an important aspect of social learning (as in the example depicted in panel C of Figure 1).

The overarching question of this paper thus is how gaze to faces and facial features supports the face-to-face interactions just described. The following sub-questions can be identified. What visual information is extracted from faces? Does gaze need to be allocated to a particular facial feature to accomplish a given task relevant for interaction, and if so, when? Where do people look when they interact with others? When is gaze allocated to a particular location in the world to acquire visual information, and when to signal information? How is gaze driven by the content of the interaction, e.g. what is said (and done) in interaction? While the goal is to describe how gaze to faces supports face-to-face interaction, much of the relevant research has been conducted in non-interactive situations.

This review proceeds as follows. I first review the evidence in regard to the question when gaze *needs* to be allocated to a particular region of a face in order to ensure successful interaction. This part covers whether and when the visual system is data-limited (c.f. Norman & Bobrow, 1975)), i.e. when visual information is required in order for successful social interaction to ensue. Second, I review the face-scanning literature to ascertain how humans look at other people, and faces in particular, and whether gaze to faces is dependent on the content of that face, the task being carried out, and the characteristics of the observer and the context. In this part, I ask how humans gaze at other humans *regardless* of whether visual information is required or not. The studies covered in these first two sections mainly cover non-interactive settings, i.e. when the stimulus is not a live person, but a photo or video of a person. Note that for these sections, the default stimuli used are static faces (e.g. photographs). I will mention it explicitly when videos or a live person was used. Third, I review the observational literature on the role of gaze in regulating interaction. Fourth, I review the recent work that has combined eye-tracking technology and the study of interaction proper. Finally, I sketch the overall picture of gaze to faces in support of social interaction and propose a dynamic system approach to gaze in interaction for future

research to build on. I end with important outstanding questions for research on this topic.

Functional constraints of gaze for information acquisition from faces

Humans are foveated animals and use movements of the eyes, specifically saccades, to direct the most sensitive part of the retina (fovea) towards new locations in the visual world. During fixations (i.e. when the same location in the visual world is looked at), objects that appear in the periphery are represented at a lower spatial resolution while objects that appear in central vision (i.e. are projected to the central part of the retina) are represented at a higher spatial resolution. Thus, in order to perceive the visual world in detail, saccades are made continuously, usually at a rate of 3 to 4 per second to project new areas of the visual world to the fovea (see Hessels et al., 2018b, for a discussion on the definitions of fixation and saccades).

Studying gaze thus intuitively reveals something about the information-processing strategy used when interacting with the world (e.g Hooze & Erkelens, 1999; Land et al., 1999; Hayhoe, 2000; Over et al., 2007). However, gaze doesn't necessarily need to be directed at an object in the world in order to perceive it. For example, one need not look at a car directly to notice it coming towards one. In the context of face-to-face interaction, this question can be rephrased as follows: when does a location on the face (e.g. the mouth or eyes) of another need to be fixated, in order to acquire the relevant information which could ensure the continuation of a potential interaction? In the remainder of this section, I address this question with regard to (1) facial identity and emotional expression, which I assume are factors relevant to the establishment of interaction, and (2) the perception of speech and (3) the perception of another's gaze direction, which I assume are important building blocks for many dyadic, triadic, and multiparty interactions. Note that emotional expressions are relevant to the flow of the interaction as well, but in its dynamic nature rather than as a static expression (as they have often been used in eye-tracking research). I return to this point later.

Facial identity, emotional expressions and gaze

Facial identity has been an important area of study, particularly with regard to learning and recognising faces. The consensus in the literature is that the eye region is an important feature for learning face identities. For example, McKelvie (1976) has shown that masking the eyes of a face impairs face learning and recognition more than masking the mouth (see also Goldstein & Mackenberg, 1966). Sekiguchi (2011) has shown that a group that outperforms another in a facial-recognition task using videos of faces looked longer at the eyes and made more transitions between the eyes than the low-performing group. Caldara et al. (2005) furthermore reported that a patient with prosopagnosia (see e.g. Damasio et al., 1982) did not use information from the eyes to identify faces.

Eye-tracking studies have further investigated whether fixations to the eyes are necessary for both encoding and recognising faces. With regard to encoding, Henderson et al. (2005) reported that making saccades during the learning phase yields better recognition performance for faces than restricted viewing (i.e. not making saccades) and Laidlaw & Kingstone (2017) reported that fixations to the eyes were beneficial for facial encoding, whereas covert visual attention was not. With regard to recognition, Peterson & Eckstein (2012) showed that observers, under time restraints of 350 ms, fixate just below the eyes for the recognition of identity, emotion and sex, which was the optimal fixation location according to a Bayesian ideal observer model. This is corroborated by Hills et al. (2011), who showed that cueing the eyes improves facial recognition performance compared to cueing the mouth area and Royer et al. (2018), who showed that face-recognition performance was related to the use of visual information from the eye region. Hsiao & Cottrell (2008) reported that for facial identity recognition two fixations suffice: more fixations do not improve performance. Finally, reduced viewing time during face learning, but not face recognition, has been shown to impede performance (Arizpe et al., 2019).

The study of gaze during the viewing and identification of emotional expressions has likewise yielded crucial insights into the relation between gaze and information

acquisition from faces. Buchan et al. (2007), for example, reported that people generally fixate the eyes of videotaped faces more during an emotion-recognition task than during a speech-perception task. However, recognition of emotional expression is often already possible within 50 ms (Neath & Itier, 2014), and does not depend on which feature is fixated (see also Peterson & Eckstein, 2012, and the Section *Face scanning* below). In other words, it seems that the recognition of emotional expressions is not limited by having to fixate a specific region on the face. Indeed, Calvo et al. (2014) have shown that the recognition of emotional expressions in peripheral vision is possible. Performance in peripheral vision is best for happy faces, and is hardly impaired by showing only the mouth. However, in face-to-face interaction, it is unlikely that emotional expressions are constantly as pronounced as they are in many studies on the perception of emotional expressions. Emotional expressions in interaction are likely more subtle visually (see e.g. Jack & Schyns, 2015), and can likewise be derived from the context and, for example, speech content, acoustics (Banse & Scherer, 1996), intonation (Bänziger & Scherer, 2005), gaze direction (Kleck, 2005), and bodily movement (de Gelder, 2009). As a case in point, Vaidya et al. (2014) showed that fixation patterns predicted the correct categorisation of emotional expressions better for subtle expressions than for extreme expressions. In other words, gaze may be more important for categorising subtle emotional expressions as they occur in interaction than extreme expressions as often used in emotion-recognition experiments.

Speech perception and gaze

The perception of speech is one of the building blocks of face-to-face interaction. Although one may assume it is mainly an auditory affair, it has long been known that the availability of visual information from the face increases intelligibility of speech embedded in noise, such as white noise or multi-talker noise (e.g. Sumby & Pollack, 1954; Schwartz et al., 2004; Ma et al., 2009). The question then is what area of the face is important for the perception of speech, and whether gaze needs to be directed there in order to perceive it. Intuitively, the mouth is the main carrier of visual information

relevant to speech perception. However, movement of other facial regions is predictive of vocal-tract movements as well (Yehia et al., 1998). Lansing & McConkie (1999) have further shown that the upper face is more diagnostic for intonation patterns than for decisions about word segments or sentence stress.

With regard to gaze during speech perception, Vatikiotis-Bateson et al. (1998) have shown that the proportion of fixations to the mouth of videotaped faces increased from roughly 35 to 55% as noise (i.e. competing voices and party music) increased in intensity. Moreover, the number of transitions (i.e. saccades between relevant areas in the visual world) between the mouth and the eyes decreased. Buchan et al. (2007) showed that gaze was directed closer to the mouth of videotaped faces during speech perception than during emotion perception, and even closer to the mouth when multi-talker noise was added to the audio. Median fixation durations to the mouth were also longer under noise conditions compared to no-noise conditions. In slight contrast to the findings from Buchan et al. (2007) and Vatikiotis-Bateson et al. (1998), Buchan et al. (2008) showed that the number of fixations to the nose (not the mouth) of videotaped faces increased during speech perception under multi-speaker noise, and the number of fixation to the eyes and mouth decreased. However, fixation durations to the nose and mouth were higher when noise was present, and fixation durations to the eyes were lower. Yi et al. (2013) showed that when noise was absent, fixating anywhere within 10° of the mouth of a single videotaped talker was adequate for speech perception (the eye-to-mouth distance was approximately 5°). However, when noise in the audio and a distracting second talking face was presented, observers made many more saccades towards the mouth of the talking face than when noise was absent. Finally, developmental work by Lewkowicz & Hansen-Tift (2012) has shown that infants start looking more at the mouth of videotaped faces around 4-8 months of age, presumably to allow infants to pick up (redundant) audiovisual information for language learning.

A classic example showing that visual information from the face can influence speech perception is the McGurk effect (McGurk & MacDonald, 1976): If an auditory and visual syllable do not concur, a different syllable altogether is perceived. Paré et al.

(2003) have shown that this effect diminishes slightly when looking at the hairline compared to the mouth, diminishes substantially when looking 10-20° away from the talker's mouth, and is negligible only at 60° eccentricity (the eye-to-mouth distance was approximately 5°). There is thus substantial influence of visual information from the face, and the mouth area in particular, that affects perception even when looking away from the face. In sum, it seems that the mouth is an important source of information for the perception of speech. Visual information from the mouth can be used for perception even when not looking at the face, although the mouth is looked at more and for longer durations when the conditions make it necessary (e.g. under high levels of ambient noise). When visual information is degraded, the mouth is looked at less again (Wilson et al., 2016).

Perception of looking direction and gaze

The perception of another's gaze direction can be considered as a second building block of face-to-face interaction, as it can reveal the locus of another's spatial attention. In fact, one's gaze direction can even automatically cue the spatial attention of others. Early studies on the perception of gaze direction have concluded that, under ideal conditions, humans are experts at perceiving one's looking direction. It has been estimated that humans are sensitive to sub-millimetre displacements of another person's iris at 1 - 2 meters observer-looker distance with a live looker (Gibson & Pick, 1963; Cline, 1967). Furthermore, this sensitivity to another person's gaze direction develops early in life (Symons et al., 1998). In a more recent study, Symons et al. (2004) reported that acuity for triadic gaze, i.e. gaze towards an object in between the observer and a live looker, was equally high (with threshold of around 30 seconds of arc), and is suggested to be limited by the ability to resolve changes in iris shifts of the looker. Yet, under less ideal conditions (e.g. when the looker does not face the observer directly but with a turned head), both the average error and standard deviation of observer judgements increased (Cline, 1967), although only the average error, not the standard deviation increased in Gibson & Pick (1963).

A number of studies have examined how perception of gaze direction relies on information beyond the eyes alone. Estimates of gaze direction have been shown to be biased by, for example, head orientation (Langton et al., 2004; Klutetz et al., 2009; Wollaston, 1824; Langton, 2000) and other cues (Langton et al., 2000). Many studies have since been conducted on the perception of gaze direction (e.g. Gamer & Hecht, 2007; Mareschal et al., 2013a,b), and one important conclusion that has been drawn from this work is that people have the tendency to believe that gaze is directed towards them (see also von Cranach & Ellgring, 1973, for a review of early studies on this topic).

One's gaze direction has also been shown to cue the spatial attention of other's automatically. The gaze direction of a face depicted in a photo, for example, can result in lower manual reaction times to targets that appear in the direction of the face's gaze direction, and higher reaction times to targets appearing in the opposite direction (Friesen & Kingstone, 1998). This effect is known as the 'gaze cueing' effect, and has been observed from adults to infants as young as 3 months (Hood et al., 1998). Although it has been suggested that reflexive cueing was unique to biologically relevant stimuli (e.g. faces and gaze direction), it has since been shown also to occur with non-predictive arrow cues, although this is perhaps subserved by different brain systems (Ristic et al., 2002). Regardless, gaze cueing is considered an important mechanism in social interaction. For in-depth reviews on the topic of gaze cueing, the reader is referred to other work (e.g. Frischen et al., 2007; Birmingham & Kingstone, 2009; Shepherd, 2010). For a model of the development of gaze following, see Triesch et al. (2006).

Again, the important question is whether perceiving one's gaze direction (or the gaze-cueing effect) requires fixation to the eyes. With regard to the perception of looking direction in general, Loomis et al. (2008) have reported that head orientation of a live person can be judged with high accuracy in peripheral vision (up to 90° eccentricity), when the head changes in orientation. When the head remains in a fixed position, judgements of its orientation were accurate from peripheral vision up to 45° eccentricity. With regard to the judgement of gaze direction from the eyes alone, these were accurate only within 8° eccentricity for a 84 cm observer-looker distance. For a 300

cm observer-looker distance, judgements of gaze direction from the eyes alone were accurate only within 4° eccentricity. To compare, the mean horizontal eccentricity encompassed by the eye region was 1.7° for the near condition (84 cm inter-person distance), and 0.5° for the far condition (300 cm inter-person distance). Florey et al. (2015) similarly reported that the perception of a looker's gaze direction from the periphery depends mostly on head orientation, not eye orientation. They concluded that the poorer resolution in the periphery is not the only cause of this dependence on head orientation, but other effects such as crowding (see e.g. Toet & Levi, 1992) and the expectation of how heads and eyes are oriented likely contribute. Furthermore, Palanica & Itier (2014) reported that discriminating direct from averted gaze within 150 ms is accurate within 3 to 6° of face eccentricity. To compare, the eye region subtended 2.5° horizontally by 0.5° vertically. With regard to the automatic cueing by gaze direction, Yokoyama & Takeda (2019) reported that a 2.3 by 2.3° schematic face could elicit gaze cueing effects when presented up to 5° above and below central fixation, but not 7.5° above or below.

It is important to realise that where one needs to look in order to perceive another's gaze direction depends on the accuracy with which another's gaze direction needs to be estimated. The work by Loomis et al. (2008), for example, exemplifies that making a judgement of whether another looks towards or away from oneself with head and eyes rotated is readily possible from peripheral vision. At the other extreme, making a judgement of whether another looks at one's eyes or mouth might not even be reliable under foveal scrutiny (see e.g. Chen, 2002). Obviously, within these two extremes, another's gaze direction may be useful in estimating that person's locus of spatial attention.

Interim summary

The allocation of gaze to multiple facial features is beneficial for encoding facial identity. However, recognising facial identity is near-optimal already within two fixations. The region just below the eyes appears optimal for recognising identity,

emotion and sex. These findings are likely relevant for establishing, not maintaining face-to-face interaction. For the maintenance of face-to-face interaction, the perception of speech and gaze direction are relevant. Gaze to the mouth can aid speech perception when conditions necessitate it (i.e. under high noise). The perception of gaze direction doesn't likely require gaze to be directed at the eyes, particularly if the orientation of the head co-varies with the gaze direction. However, a direct link between gaze position on a face (i.e. how far it is away from another's eyes) and the acuity of gaze-direction perception hasn't been shown. It is expected that an observer's gaze needs to be directed towards the eyes for more fine-grained judgements of gaze direction of the other. Finally, it seems relevant that future studies investigate data-limitations (i.e. when gaze is necessary to acquire specific visual information) of the kind described here in actual interactive settings.

Face scanning

In this section, I review the literature with regard to face scanning behaviour under less restrained conditions, for example during prolonged viewing of faces or when the observer is free to look around. I aim to review the evidence with regard to the follow questions: (1) what are the biases in gaze to faces and to what degree are these under volitional control, (2) how is gaze to faces dependent on the content of the face, (3) how is gaze to faces dependent on the task posed to the observer, and (4) how is gaze to faces dependent on characteristics of the observer? Note that the studies in this section have mainly been conducted in non-interactive settings. The (fewer) studies on gaze to faces in interaction proper are covered in a later section.

Biases in gaze to faces

The classic studies by Buswell (1935) and Yarbus (1967) were the first to suggest that people, faces, and eyes are preferentially looked at. This has since been corroborated by many studies (e.g. Birmingham et al., 2008a,b, as well as the many studies that follow). Interestingly, it appears that the bias for faces or eyes cannot be predicted by salience (as defined on the basis of stimulus features such as colour,

intensity and orientation; Itti & Koch, 2000) for faces (Nyström & Holmqvist, 2008) or eyes (Birmingham et al., 2009), but see Shen & Itti (2012) for an example of where salience of videotaped faces does have some predictive value. Amso et al. (2014) reported that salient faces were looked at slightly more often (71%) than non-salient faces (66%), but this is marginal (5%) compared to how often faces were looked at when not being salient.

The bias for looking at faces is already present at birth, as infants preferentially track faces compared to e.g. scrambled faces (Goren et al., 1975; Johnson et al., 1991), and preferentially make the first saccade to faces in complex displays (Gliga et al., 2009). The bias for looking at the eyes seems to develop in the first year after birth. Wilcox et al. (2013), for example, reported that 9-month-olds looked more at eyes than 3-4 month-olds for dynamic faces. Frank et al. (2009) further reported that the bias for looking at faces increased between 3 and 9 months of age, whereas gaze of 3-month-olds was best predicted by saliency (see also Leppänen, 2016). Humans are not the only animals with preferences for looking at conspecifics, faces and eyes. Chimpanzees have been shown to preferentially look at bodies and faces (Kano & Tomonaga, 2009), and rhesus monkeys to preferentially look at the eyes in faces (Guo et al., 2003). Chimpanzees, however, appear to gaze at both eyes and mouth and make saccades often between them (Kano & Tomonaga, 2010), more so than humans.

An important question is to what degree the bias for looking at faces is compulsory. In this regard, it has been shown that faces automatically attract attention (I discuss automatic attraction of gaze in the next paragraph) (Langton et al., 2008), although Pereira et al. (2019) state that this isn't always the case. Automatic attention-attraction by faces can, however, be overcome by top-down control of attention to support the goals of the observer (Bindemann et al., 2007), e.g. to attend something other than faces. Faces have also been shown to retain attention (Bindemann et al., 2005), already for 7-month-old infants (Peltola et al., 2018). Furthermore, the degree to which attention is maintained by faces is modulated by the emotional expression in the faces. For example, fearful faces have been shown to delay attentional

disengagement more than neutral, happy and control faces for infants (Peltola et al., 2008, 2009), and for high-anxious adults (Georgiou et al., 2005). Angry faces additionally maintained attention longer than happy faces and non-faces for 3-year-old children (Leppänen et al., 2018).

Apart from attracting and maintaining visual attention, several studies have also shown that the eyes automatically attract gaze. Laidlaw et al. (2012), for example, showed that when instructed to avoid the eyes, observers could not inhibit some fixations to the eyes. This was, however, possible for the mouth or for the eyes of inverted faces. Similarly, Itier et al. (2007) have reported that eyes always attracted gaze, even when the eye-region was not task-relevant. In another study, it was shown that although faces were preferentially fixated, the time of first fixating a face could still be lowered by giving a different task (i.e. spot people as fast as possible; End & Gamer, 2019).

Finally, a left-side bias in looking at faces has been reported in the literature and the use of information from that side in judging e.g. sex (Butler et al., 2005). A similar bias seems to occur in rhesus monkeys and dogs (Guo et al., 2009). Arizpe et al. (2012), have, however, cautioned that this left-side bias may partly be explained by the position of the initial fixation point.

Content-dependent gaze to faces

Gaze to moving faces, talking faces and faces making eye contact.

Apart from general biases and task-dependent gaze to faces, several studies have suggested that gaze to faces depends on what that face is doing, for example, talking, moving, making eye contact, etc.

As noted before, Buchan et al. (2007) and Buchan et al. (2008) have shown that gaze to videotaped faces is dependent on the intelligibility of speech, with longer fixations to the mouth and nose under noise conditions, shorter fixations to the eyes, and more fixations to the nose. An important question then is whether gaze is also directed more at the mouth when speech occurs and the conditions are favourable (i.e.

speech is intelligible). In a free-viewing experiment with videos of faces, V $\ddot{o}$ et al. (2012) showed that for audible talking faces, fixations occurred equally often to the eyes, nose and mouth. For muted videos of faces, fewer fixations to the mouth were observed. V $\ddot{o}$ et al. (2012) go on to show that gaze is dependent on the content and action of the face (audibility, eye contact, movement), with each its own facial region associated. For example, when the talking person in the video made eye contact (i.e. looked straight into the camera), the percentage of fixations to the eyes increased and the percentage of fixations to the mouth decreased. When the face in the video moved, the percentage of fixations to the nose increased. Similarly, Tenenbaum et al. (2013) reported that infants from 6-12 months of age (when language production starts to emerge) looked primarily at the mouth of a talking videotaped face (see also Frank et al., 2012), but that they looked more at the eyes of a smiling face than the eyes of a talking face. Lewkowicz & Hansen-Tift (2012) corroborated that information from the mouth is important for the development of language skills by showing that, for infants aged between 4-12 months, the youngest infants (4-6 months) primarily looked at the eyes, while older infants (8-12) looked more to the mouth, presumably to pick up (redundant) audiovisual information from the mouth. Importantly, infants aged 10 months fixated the mouth more (relative to the eyes) than the 12-month-olds. This latter ‘shift’ back towards the eyes did not occur for infants that grow up in a bilingual environment, suggesting that they exploit the audiovisual redundancy for learning language for a longer time (Pons et al., 2015). Foulsham et al. (2010) also showed that speech was a good predictor for which videotaped person was being looked at, although it co-depended on the social status of that speaker. I.e. speakers were looked at more often than non-speakers, but speakers with higher social status were looked at more than speakers with lower social status.

There is also contrasting evidence that suggests that the mouth need not always be looked at when speech occurs. While Foulsham et al. (2010) showed that speech was a good predictor of who was being looked at, observers predominantly looked at the eyes of the person. Moreover, Foulsham & Sanderson (2013) showed that this also occurred for videos from which the sound was removed. In another study, Scott et al.

(2018) showed observers three videos of an actor carrying out a monologue, manual actions (how to make a cup of tea) and misdirection (a magic trick ‘cups and balls’). They reported that faces were looked at most during monologues, but hands were looked at much more often during manual actions and misdirections in videos portrayed by an actor. Critically, hearing speech increased looking time to the face, but rather the eyes than the mouth. As noted before, however, information for speech recognition need not be confined to the mouth (Lansing & McConkie, 1999; Yehia et al., 1998). Finally, Scott et al. (2018) showed that eye contact by the actor (during manual activity and misdirection in particular) increased observers’ fixation time to the face.

Gaze to emotional faces. Multiple studies have investigated how gaze to faces is dependent on the emotional expression contained in the face, particularly for static emotional expressions. Green et al. (2003) asked observers to judge how the person they saw was feeling and showed that inter-fixation distances (saccadic amplitudes) were larger for angry and fearful facial expressions compared to non-threat related facial expressions. Furthermore, more and longer fixations to the facial features (eyes, nose, mouth) occurred for angry and fearful expressions. The authors interpret their findings as a ‘vigilant’ face-scanning style for threat-related expressions. Hunnius et al. (2011) reported that during a free-viewing experiment, dwell times and the percentage of fixations to the inner features (eyes, nose, mouth) were lower for threat-related (anger, fear) emotional expressions for both adults and infants. This was interpreted as a ‘vigilant’ face-scanning style, albeit a different manifestation than that observed by Green et al. (2003). The eyes of threat-related expressions were looked at less compared to happy, sad and neutral expressions only by the adults, not the infants. In other work, Eisenbarth & Alpers (2011) asked observers to look at faces and judge the emotional expression as positive or negative. They showed that across emotional expressions, the eyes were fixated most often and the longest. Fixations to the mouth were longer for happy expressions compared to sad and fearful expressions, and the eye-to-mouth index (higher values represent more looking at the eyes relative to the mouth) was lowest for happy faces, then angry faces, and then fearful, neutral and sad faces. Bombari et al.

(2013) showed that, during an emotion-recognition experiment, the eye region was looked at less for happy expressions, and the mouth looked at more for fear and happy expressions, compared to angry and sad facial expressions. Finally, Beaudry et al. (2014) reported that the mouth was fixated longer for happy facial expressions than for other expressions, and the eyes and brow region were fixated longer for sad emotional expressions. No other differences were observed between the emotional expressions.

As a potential explanation of the different gaze distributions to emotional expressions, Eisenbarth & Alpers (2011) proposed that regions that are most characteristic of an emotional expression are looked at. If one considers the diagnostic information (see Smith et al., 2005) of seven facial expressions (happy, surprised, fearful, angry, disgusted, sad, and neutral), it seems that for the happy expressions this claim holds, although it is less clear for the other emotional expressions. A potential problem with interpreting these studies in terms of information-usage is that either there is no task (i.e. free-viewing, see also Tatler et al., 2011), or gaze to the face is not the bottleneck for the task. With regard to the latter, it has been shown that emotion recognition can already be done in 50 ms (e.g. Neath & Itier, 2014), so how informative is the gaze about information-usage during prolonged viewing? In contrast to the studies described in the Section *Functional constraints of gaze for information acquisition from faces*, here the necessity of gaze location is more difficult to relate to task performance. It may be expected that during prolonged viewing, recognition of the emotional expressions has already been achieved and that gaze is (partly) determined by whatever social consequences an emotion may have. Clearly, describing face-scanning behaviour as ‘vigilant’ seems to suggest so. Indeed, Becker & Detweiler-Bedell (2009), showed that when multiple faces were presented in a free-viewing experiment, fearful and angry faces were avoided already from 300 ms after stimulus onset, suggesting that any threat-related information was processed rapidly in peripheral vision and consequently avoided.

Furthermore, the content of a face, such as the emotional expression, during interaction is dynamic and not static as in many of the studies described in this section.

Moreover, it is likely more nuanced and tied closely to other aspects of the interaction such as speech (e.g. intonation). Dynamic aspects of emotional expressions can aid their recognition, particularly when the expressions are subtle or when visual information is degraded (e.g. low spatial resolution). For a review on this topic, see Krumhuber et al. (2013). Jack & Schyns (2015, 2017) have also discussed in-depth that the human face contains a lot of potential information that is transmitted for social communication, and outline how to potentially study the dynamics of it. I am not aware of any studies available at the time of writing that have investigated gaze to the dynamic emotional expressions in e.g. social interaction and how it depends on the diagnostic information for an expression at each point in time. Blais et al. (2017), however, reported that fixation distributions to emotional expressions were different for dynamic as compared to static expressions, with fewer fixations made to the main facial features (i.e. eyes, mouth) for dynamic expressions. However, face stimuli were only presented for 500 ms with the emotional expression unfolding in this time period, yielding only 2 fixations on average to compare (with the first one likely on the centre of the face due to the position of the fixation cross prior to the face).

Task-related gaze to faces

Already since the work of Yarbus (1967), it has been known that the task given to a person may affect gaze to faces. Since then, gaze has often been interpreted as a means of extracting visual information from the world for the task at hand. Here, I briefly outline the differences in gaze to faces that have been observed for different tasks. Walker-Smith et al. (1977) have shown that during face learning and recognition gaze is confined to the internal features of the face (eyes, nose, mouth). This holds both for when faces are presented sequentially and side-by-side. Similarly, Luria & Strauss (1978) have shown that the eyes, nose, and mouth are looked at most often during face learning and recognition, and Henderson et al. (2005) noted that most time was spent looking at the eyes during face learning. During face recognition, they reported that gaze was more restricted (primarily to the eyes and nose) than during face learning.

Williams & Henderson (2007) furthermore reported that the eyes, nose and mouth were looked at most (and the eyes in particular) during face learning and recognition for both upright and inverted faces.

A common theory from the early days of face-scanning research was the scanpath theory (Noton & Stark, 1971), which held that a face that was learned by fixating features in a certain order would be recognised by following that same order. Walker-Smith et al. (1977) have shown that this model does not hold, as scanpaths shown during face learning are not repeated during face recognition (see also Henderson et al., 2005). Walker-Smith et al. (1977) proposed a model in which the first fixation provides the gestalt of the face. Subsequent fixations to different facial features are used to flesh out the face-percept. In order to compare faces, the same feature must be fixated in both faces.

With regard to other tasks, Nguyen et al. (2009) have shown that the eye region was looked at most when judging age and fatigue. Cheeks were looked at more for the less-tired faces than for the more tired faces. Eye brows and the glabella were looked at more for the older half of faces compared to the younger half. In a similar study, Kwart et al. (2012) had observers judge the age and attractiveness of faces. They showed that the eyes and nose were looked at most of the time, with very little difference in the distribution of gaze between the two tasks. Buchan et al. (2007) had observers judge either emotion or speech of videotaped faces and found that observers looked more often and longer at the eyes when judging emotion. Finally, Lansing & McConkie (1999) reported that observers looked more often and longer at the upper face when forming judgements about intonation and more at mid and lower face when forming judgements about sentence stress or segmentation, which mimics the diagnostic information: The upper face was more diagnostic for intonation patterns than for decisions about word segments or sentence stress.

Observer-dependent gaze to faces

Idiosyncratic face-scanning patterns. A particularly interesting observation that was reported by Walker-Smith et al. (1977) in their early work on gaze during face learning and recognition was that their 3 subjects showed very different scan patterns. Recently, a number of studies have corroborated and extended these findings substantially. Peterson & Eckstein (2013), for example, had observers perform a face-identification task under three conditions: (1) free-viewing a 350 ms presented face, (2) free-viewing a 1500 ms presented face, and (3) a fixed fixation location somewhere on the face with the face presented for 200 ms. Observers showed large inter-individual differences in their preferred fixation locations during the free-viewing conditions, the location of which was highly correlated between the 350 and 1500 ms duration conditions. In other words, some observers preferred to fixate the nose while other preferred to fixate the eyes. Interestingly, restricting fixation location to the eyes for ‘nose-lookers’ degraded face-identification performance, whereas restricting fixation location to the nose degraded face-identification performance for the ‘eye-lookers’. Thus, Peterson & Eckstein (2013) concluded that face scanning patterns are idiosyncratic, and reflect observer-specific optimal viewing locations for task performance.

In subsequent work, Mehoudar et al. (2014) have shown that idiosyncratic face scanning patterns were stable over a period of 18 months, and were not predictive of face-recognition performance. Kanan et al. (2015) have additionally shown that observers not only have idiosyncratic face scanning patterns, but also that they have task-specific idiosyncratic face-scanning patterns (e.g. for judging age or for judging attractiveness). Inferring task from a face-scanning pattern was accurate for eye-tracking data from an individual, but not when inferring task based on eye-tracking data from multiple other observers. Arizpe et al. (2017) have further reported that the idiosyncratic face-scanning patterns of multiple observers could be clustered into 4 groups, respectively having a fixation-density peak over the left eye, right eye, nasion, or nose-philtrum-upper lip regions. Face-recognition performance did not differ between the groups and face-scanning patterns were equally distinct for inverted faces. Finally, it seems that idiosyncratic face-scanning patterns are hereditary to a degree. Constantino

et al. (2017) have shown that the proportion of time spent looking at the eyes and mouth was correlated by 0.91 between monozygotic twin toddlers, and only by 0.35 for dizygotic twins. Even spatiotemporal characteristics of gaze to faces, such as when saccades were made and in which direction, seemed to have a hereditary component.

Sex-dependent gaze to faces. Several studies have indicated that males and females differ in how they look at faces. In early observational work with live people, it has been reported that females tend to look more at an interviewer than males regardless of the sex of the interviewer (Exline et al., 1965). In recent eye-tracking work using videos, Shen & Itti (2012) have reported that fixation durations to faces, bodies and people were longer for male observers than for female observers. Moreover, males were more likely to look at the mouth, and less likely to look at the eyes, than females. Coutrot et al. (2016) corroborated and extended some of these findings. They showed that fixation durations to faces were longer, saccade amplitudes shorter, and overall dispersion smaller for male observers than for female observers. Furthermore, a largest left-side bias was observed for female observers looking at faces of females. Note that these differences are based on a large eye-tracking data set of 405 participants, looking at 40 videos each.

Cross-cultural differences in gaze to faces. Cross-cultural differences in face perception and gaze to faces have been a long-standing area of research. Differences between cultures have been observed for gaze during face learning and recognition, emotion discrimination and free-viewing. Blais et al. (2008), for example, have reported that east-asian (EA) observers looked more at the nose and less at the eyes compared to western-caucasian (WC) observers during face learning, face recognition and judgement of race. Furthermore, EA observers were better at recognition of EA faces, and WC observers of WC faces. The authors suggested that not looking at the eyes for the EA observers may be a gaze-avoidant strategy, as eye contact can be considered rude in some EA cultures. Jack et al. (2009) showed that during an emotion-discrimination task, WC observers distributed their fixations across the facial features (eyes, nose, mouth), whereas EA observers focused mostly on the eyes (cf. Blais et al., 2008, during

face learning and recognition). Furthermore, Jack et al. (2009) reported that EA observers, but not WC observers, exhibited a deficit in categorising fearful and disgusted facial expressions, perhaps due to the fact that the eyes were mostly fixated, which do not contain diagnostic information for e.g. disgust (Smith et al., 2005). Jack et al. (2009) thus questioned the suggestion by Blais et al. (2008) that EA observers actively avoided looking into the eyes. Moreover, even if EA observers were to look more at the nose than at the eyes (as Blais et al., 2008, suggest), it is unlikely that this is a gaze-avoidance strategy, as observers tend not to be able to distinguish whether they're being looked in the nose or eyes (e.g. Chen, 2002; Gamer et al., 2011) and assume they're being looked at under uncertainty (e.g. Mareschal et al., 2013b).

In a study directly aimed at investigating information use by EA and WC observers during face learning and recognition, Caldara et al. (2010) showed observers faces of which a 2, 5 or 8° gaussian aperture was visible around the fixation point. WC observers fixated the eyes and partially the mouth for all aperture sizes. EA observers, however, fixated the eye region for the 2 and 5° aperture, and partially the mouth for the 5° aperture, but fixated mainly the central region of the face (i.e. the nose) for the 8° aperture. The authors conclude that EA and WC observers rely on the same information for learning and recognising faces when under visual constraints, but show different biases when no visual constraints are in place. In a particularly comprehensive set of experiments, Or et al. (2015) showed that both Asian and Caucasian observers' first fixation during a face-identification task were directed, on average, just below the eyes, which has been shown to be optimal in terms of information acquisition for identity, sex and emotion recognition (Peterson & Eckstein, 2012). Fixations were shifted slightly more to the left for Caucasian observers compared to Asian observers, however (approximately 8.1% of the interocular distance). For the remaining fixations during the 1500 and 5000-ms presentation, no substantial differences in fixation patterns between groups were observed. Greater variability was observed within groups than between groups, and a forced-fixation experiment showed that performance was optimal for idiosyncratic preferred fixation locations (see the Section *Idiosyncratic*

661 *face-scanning patterns*).

662 In a free-viewing experiment, Senju et al. (2013) showed that cross-cultural
663 differences were already evident for young children. Japanese children aged 1-7 years
664 looked more at the eyes and less at the mouth of videotaped faces than British children
665 of the same age. Moreover, (Gobel et al., 2017) reported that EA observers only looked
666 more at the nose and less at the eyes than WC observers when the gaze direction of the
667 videotaped talking face being looked at was direct (as if towards the observer), not
668 when the face's gaze was averted slightly (as if talking to another person). The authors
669 concluded that cross-cultural differences in gaze to faces need to be considered within
670 the interpersonal context in which gaze is measured.

671 Thus far I have considered only the cross-cultural differences in gaze to faces from
672 the perspective of the observers. However, multiple studies have reported an 'own-race'
673 effect, in that higher recognition performance has been observed for observers viewing
674 faces from their own race compared faces from another race. With regard to how people
675 scan own-race and other-race faces, a number of studies have been conducted. Fu et al.
676 (2012), for example, reported that Chinese observers spent more time looking at the
677 eyes and less time to the nose and mouth of Caucasian faces than of Chinese faces.
678 Wheeler et al. (2011) furthermore reported that older Caucasian infants (within a range
679 of 6-10 months of age) looked more at the eyes and less at the mouth of own-race faces
680 than younger infants, whereas this difference was not observed for other-race faces (see
681 also Xiao et al., 2013, for more in-depth findings). Liu et al. (2011) have finally
682 reported that older Asian infants (within a range of 4-9 months of age) tended to look
683 less at the internal features (eyes, nose, mouth) for other-race faces than younger
684 infants, which was not observed for own-race faces. Arizpe et al. (2016), however,
685 argued that differences in gaze to own-race and other-race faces are subtle at best, and
686 are dependent on the exact analysis used. When area-of-interest analyses are used,
687 subtle differences emerge, yet these are not found with spatial density maps (a method
688 that does not make a-priori specifications of where differences are expected to arise).

Interim summary

The studies reviewed in this section have revealed the following. When observers are unrestrained in where they can look or for how long they can look, other people are preferentially fixated over objects, faces over bodies and eyes over other facial features. However, exactly where one looks on the face of another is dependent on a multitude of factors. What the face does – e.g. whether it moves, talks, expresses emotion, or looks directly toward the observer – modulates gaze to the face and seems to attract gaze to the information source (e.g. the mouth for speech), although the evidence is not always clear-cut. Furthermore, the task being carried out by the observer affects gaze to the face, although intra-individual differences in task-specific face-scanning patterns are potentially as large as inter-individual differences. Small sex differences in gaze behaviour have been observed, as have cross-cultural differences, depending both on the observer and the person observed. Although cross-cultural differences have been observed in children and adults, and across multiple studies, the differences may be only in initial fixations or dependent on the interpersonal context. Finally, and particularly important, face-scanning patterns are highly idiosyncratic, and are, at least in part, under genetic control (i.e. hereditary).

Social context and the dual function of gaze

The studies described so far have highlighted how gaze is allocated to faces from a purely information-acquisition perspective, or have described general biases. Over the last years, a large number of researchers have argued that traditional laboratory studies of social attention or social gaze (i.e. gaze to people, faces, and so forth) have misrepresented how gaze may operate in ‘real world’ situations (e.g. Smilek et al., 2006; Kingstone et al., 2008; Kingstone, 2009; Risko et al., 2016; Cole et al., 2016; Hayward et al., 2017). This critique is particularly concerned with the fact that in interactive situations, one’s gaze direction is available to others too, and there may be social consequences to where one looks. The fact that the contrast between the human iris and sclera is large means that it can easily be distinguished from afar, and this high

contrast has been suggested to have had a facilitatory effect on the evolution of communicative and cooperative behaviours (Kobayashi & Kohshima, 1997).

What is of particular importance, is that gaze to faces appears to be sensitive to the particular social context (e.g. Risko & Kingstone, 2011; Richardson et al., 2012). Foulsham et al. (2010), for example, had participants look at a video of three people making a decision. Not only did the speaker role (i.e. who spoke at what point in time) predict gaze to that person, but participants also tended to look more at the eyes, face and body of people with higher social status than those of lower social status. Similarly, Gobel et al. (2015) reported that gaze to faces depended on the social rank of that person. They reported that the eye-to-mouth ratio of participants was higher when looking at videotaped people of lower social rank, but lower for people of higher social rank when participants believed the other person would look back at them (at a later point in time – their video was said to be recorded and shown later), compared to when participants believed there was no possibility for the other to look back. The authors argued that the inter-personal difference in social rank predicted gaze to facial features (eyes vs. mouth). These two studies show that interpersonal context may affect gaze to faces, and particularly when the other person is (believed to be) live.

In more direct investigations of the effects of ‘live’ presence of another person, Laidlaw et al. (2011) showed that participants would hardly look at a confederate in a waiting room, while they would often look at a video stream of a confederate placed in a waiting room. The authors argued that the potential for social interaction here led people to avoid looking at the confederate (see also Gregory & Antolin, 2019; Cañigüeral & Hamilton, 2019, who report similar findings). In other work, Foulsham et al. (2011) had participants walk around campus wearing an eye tracker, or look at a video of someone walking around campus. While pedestrians were looked at often in both situations, the timing of it showed subtle differences between the video and live conditions. When participants actually walked around campus, other pedestrians were looked at less at a close distance than when watching the video in the lab. Finally, Laidlaw et al. (2016) showed that people on the street tended to look more often at a

746 confederate carrying out a public action (saying hi and waving) than a private action
747 (saying hi on the phone), and concluded that covert visual attention must have been
748 necessary to assess the intention of the confederate, before gaze was either directed to
749 that person or not. These studies show that general biases for looking at other people,
750 faces and eyes do not necessarily generalise to all contexts.

751 I do not aim to reiterate the ‘lab vs. the real world’ discussion, as this has often
752 been framed, nor the call for interactive paradigms. The interested reader is referred to
753 Kingstone et al. (2008) for a good starting point on this topic. For in-depth
754 comparisons of methodology across different levels of ‘situational complexity’ (i.e. from
755 watching static faces to full-fledged live interaction) see e.g. Risko et al. (2012) and
756 Pfeiffer et al. (2013). My aim is to integrate the available evidence from multiple
757 research fields to tackle the real problem of describing, understanding and predicting
758 gaze in social face-to-face interactions. These studies covered above make a number of
759 points clear: (1) gaze may be sensitive to many social factors that are not considered
760 from a purely information-acquisition perspective of gaze, but require an
761 information-signalling perspective of gaze, and (2) evidence on gaze in non-interactive
762 settings may not necessarily generalise to interactive settings. The question then
763 beckons how gaze operates in interaction? There are at least two strands of research to
764 help answer this question. First, there is a large observational literature on gaze in
765 interaction. Second, more recent studies – partly in response to the critique on research
766 using static pictures outlined in this paragraph – have used eye trackers to study gaze
767 in interaction. I review these strands of research below.

768 **Observational studies of gaze in interaction**

769 In stark contrast to the biases reported in the eye-tracking literature for looking at
770 people and faces, many social interactions that occur throughout a day can be
771 characterised by ‘civil inattention’. This phenomenon, described by Goffman (1966, p.
772 83-85), often occurs when two strangers meet and consists of a brief exchange of looks,
773 followed by ignoring each other as a form of courtesy (cf. Laidlaw et al., 2011). In other

words, people tend not to look at each other in such situations. As an example of this phenomenon, Cary (1978) reported that participants placed in a waiting room almost always gave an initial look to each other. When no initial look took place, it was unlikely that conversation would ensue between the participants. When an additional exchange of looks occurred, conversation was more likely to follow. In social interactions, gaze may thus serve to refrain from, or initiate, conversation. Many early observational studies have subsequently investigated how gaze may regulate interaction, of which I give a brief overview. The observational research described here is characterised by multiple people interacting in real life, while they are observed or recorded. Gaze is then scored in real time or subsequently from the video recordings and carefully annotated, often supplemented with annotations of e.g. speech or gestures.

Probably one of the most important studies on gaze in interaction was conducted by Kendon (1967), who showed that the time spent looking at the face of another during interaction varies heavily (between 28% and over 70%, cf. the Section *Idiosyncratic face-scanning patterns*), both during speaking and listening, and that the number of changes of gaze-direction was highly correlated between partners in a dyad. Kendon further showed that gaze was directed more often towards the other at the end of one's utterance, which was suggested to be to determine which action might be taken next, e.g. to give up the floor or to continue speaking. Gaze also tended to be directed away from the conversational partner when beginning an utterance, which was suggested to be to actively shut out the other and focus on what one wants to say. Some of these findings are summed as follows (p refers to one of the interactants): *"In withdrawing his gaze, p is able to concentrate on the organization of the utterance, and at the same time, by looking away he signals his intention to continue to hold the floor, and thereby forestall any attempt at action from his interlocutor. In looking up, which we have seen that he does briefly at phrase endings, and for a longer time at the ends of his utterances, he can at once check on how his interlocutor is responding to what he is saying, and signal to him that he is looking for some response from him."* (p. 42).

Allen & Guy (1977) tested the hypothesis of Kendon (1967) that looking away

from the other is causally related to reducing mental load, by investigating the relation of looks away from the conversational partner with the content of the speech. They found that when words relating to mental process (believe, guess, imagine, know, etc.) or judgement (bad, every, good, some, etc.) were spoken, looks away tended to occur more often than without such words. Furthermore, Beattie (1981) had participants either look freely or fixate the interviewer. While continuous looking at the interviewer did not affect speech speed or fluency, more hesitations ('ehm') and false starts (starting a sentence and restarting just briefly afterwards) occurred, suggesting that looking at the other indeed interferes with the production of spontaneous speech. This is known as the *cognitive interference hypothesis*.

Observational studies have further shown that gaze depends on e.g. the content of the conversation (i.e. personal or innocuous questions; Exline et al., 1965), on personality characteristics (Libby & Yaklevich, 1973), on inter-personal intimacy (Argyle & Dean, 1965; Patterson, 1976), and competition versus cooperation between the interlocutors (Foddy, 1978). For example, Foddy (1978) reported that cooperative negotiation resulted in longer bouts of looking at each other than competitive negotiation, although the frequency was the same across both negotiations. The authors suggested that frequency is related to the monitoring/checking function, while length is related to affiliative functions (cf. Jarick & Kingstone, 2015, for more recent work on this topic). Kleinke (1986) summarises multiple studies on this topic, stating that gaze can be used to exert social control during persuasion or for asserting dominance through prolonged gaze to the face of the other: "*People generally get along better and communicate more effectively when they look at each other. One exception is in bargaining interactions where cooperation can be undermined when gaze is used for expressing dominance and threat*" (p. 84).

As noted, the brief review I give of the observational literature is necessarily non-exhaustive. Most of the early research on gaze and eye contact in social interaction was reviewed by Argyle (e.g. 1972) and particularly Kleinke (1986), the latter organising the available evidence within the framework of Patterson (1982) on

nonverbal exchange. For a detailed overview, the reader is encouraged to read Kleinke's review. One of the essential points of his work, however, is that *"gaze synchronization and the operation of gaze in turn taking are less reliable than previously believed because they depend on the context and motives of the interactants"* (p. 81), which means that gaze cannot be fully understood as a regulator of interaction without understanding how personal and contextual factors contribute to gaze to faces, as has already been established above for the role of gaze in information acquisition.

As Bavelas et al. (2002) pointed out, the review of (Kleinke, 1986) was the last major review on observational research on gaze, with few new studies to (re-)define the field afterwards. In the years after 2000, a number of relevant studies have been conducted on this topic however. For example, in a study on how (non-)verbal communication aids understanding, Clark & Krych (2004) reported that looks to the face of a person giving instructions occurred when a conflict needed to be resolved. Hanna & Brennan (2007) furthermore showed that the gaze direction of someone giving instruction was rapidly used to disambiguate which object was referred to when the instruction could refer to multiple objects. These studies attest to the fact that information from gaze can be rapidly used depending on the contextual needs of the person in interaction.

The field of conversation analysis is another example which has continued to investigate the role of gaze as an important interactional resource. Apart from the role of gaze in the initiation and participation in interaction, and in the regulation of interaction, gaze is also considered to form independent actions in this field: e.g. to appeal for assistance (e.g. Kidwell, 2009). Kidwell (2005), for example, describes how children differentiate different types of looking from their caregiver in order to prolong or change their ongoing behaviour. Stivers & Rossano (2010) investigated how responses in conversation are elicited by extensively annotating conversations. They reported that a response was evoked from a conversational partner based on, among others, gaze, interrogative prosody (e.g. rising pitch at the end of a sentence) and lexico-morphosyntax (word- and sentence-formation). Stivers et al. (2009) have

furthermore shown that gaze towards another is a near-universal facilitator (across 9/10 investigated languages) of a speeded response from the conversational partner. For further research on this topic, the reader is referred to Rossano (2013).

Interim summary

Gaze plays an important role in initiating and regulating interaction. The initiation of conversation tends to be preceded by one's gaze being directed towards the conversational partner, and the timing of when gaze is directed towards or away from the conversational partner plays an important role in the turn-taking behaviour during interaction. Looking toward a conversational partner can be used to give up the turn, whereas looking away can be used to reduce load while thinking about what to say next. Finally, gaze is but one of multiple cues (e.g. prosody) that aid the regulation of interaction.

Eye tracking in interaction

The observational studies noted above have often been criticised for being subjective in how gaze is coded, whereas eye-tracking has been hailed as the objective counterpart. Early studies have estimated the validity of analysing gaze in interaction from observation to be around 70-80% for the best recording techniques (Beattie & Bogle, 1982). See also Kleinke (1986) in this regard, who noted that eye and face gaze cannot be reliably and validly distinguished by observational techniques. This is evident in observational research, which is all restricted to whether one looks towards a face or not. Whether one looks at the eyes, nose or mouth is not reliably established from observation. This is, however, an important distinction with regard to the studies described in the Sections *Functional constraints of gaze for information acquisition from faces* and *Face scanning*, where eyes, nose and mouth are considered as regions that may carry distinctive information useful for ensuring successful interaction. Eye-tracking studies have provided some remedy to these concerns: gaze direction can be objectively measured, although not all eye trackers are good enough to establish gaze to facial features in interactive settings (see e.g. Niehorster et al., 2020, for a

discussion). Furthermore, eye-tracking in interaction can be quite challenging (e.g. Clark & Gergle, 2011; Brône & Oben, 2018). In this section, I review the eye-tracking studies that have investigated (some aspect of) gaze in face-to-face interaction.

A number of eye-tracking studies in interaction have corroborated reports from the observational literature. For example, Freeth et al. (2013) reported that participants wearing eye-tracking glasses looked less at the face of the interviewer and more to the background when answering questions than when being asked a question. Furthermore, participants looked more at the face of the interviewer when she made eye contact with the participant than when she averted her gaze. Ho et al. (2015) had two participants fitted with wearable eye trackers and had them play games (20 Questions, Heads Up) in which turn-taking behaviour occurred. They showed that gaze to the other person preceded the talking of the other (by about 400 ms on average), and gaze was averted when one started talking up to around 700 ms on average after talking started. Holler & Kendrick (2015) furthermore had three people engage in interaction while wearing eye trackers and showed that the unaddressed interactant shifted their gaze from one speaker to the next speaker around (and often prior to) the end of the first speaker's turn (see also Hirvenkari et al., 2013; Casillas & Frank, 2017, for comparable research in non-interactive settings). Broz et al. (2012) showed that the amount spent looking at each other (mutual gaze) of a dyad during face-to-face conversation was correlated positively with the combined level of agreeableness and how well the participants knew each other. Finally, Mihoub et al. (2015) showed that gaze to faces in interaction depended on the interpersonal context, i.e. colleagues versus students. These studies combined show that, as has been previously established in the observational literature, gaze is important in regulating turn-taking behaviour in interaction, and is related to contextual characteristics (e.g. personality, familiarity, interpersonal context) .

Important innovations in multiple disciplines are beginning to appear. For example, Auer (2018) conducted a study on the role of gaze in regulating triadic conversation and showed that gaze serves both addressee selection and next-speaker selection separately. When speaking, the speaker's gaze was distributed across both

918 conversational partners, but the speaker's gaze was directed to one partner specifically
919 at the end of a turn to offer up the floor. The next speaker would then either start their
920 turn, give a small reply to signal the current speaker to continue, or gaze at the third
921 party to hand on the turn. However, it turned out that these contingencies were weak
922 and that speakers could easily self-select as the next speaker by simply starting to talk
923 at the end of a turn without having been 'offered the floor'. In another study using eye
924 tracking to build on early observational research, Jehoul et al. (2017) investigated the
925 relation between gazes away from a speaker and 'fillers' such as "uh" or "um" in dyadic
926 conversation. They showed that one particular filler ("um") was more associated with
927 looks away from the conversational partner than another filler ("uh"), highlighting the
928 multimodal nature of communication. In recent developmental work, Yu & Smith
929 (2016) showed that infants' sustained gaze (or sustained overt attention) to an object
930 was prolonged after their parent also looked at that object, implicating joint attention
931 in the development of sustained attention.

932 Macdonald & Tatler (2013, 2018) have conducted interesting studies on the role of
933 gaze during cooperative behaviour, and particularly in relation to instructions.
934 Macdonald & Tatler (2013) had participants wear eye-tracking glasses while building a
935 block model at the guidance of an instructor. When the instructions were ambiguous
936 and gaze cues were available from the instructor to resolve the ambiguity, participants
937 fixated the instructor's face more than when such gaze cues were not available or when
938 the instructions were unambiguous. Gazing at the face to resolve the ambiguity of
939 instructions predicted increased performance in picking up the right block for the next
940 move. The authors concluded that gaze cues were used only when necessary to
941 disambiguate other information. Macdonald & Tatler (2018), on the other hand, had
942 dyads make a cake together. Half of the dyads were given specific roles (one chef and
943 one gatherer), the other dyads were not. Participants spent very little time looking at
944 each other, but did look at each other often when receiving instructions. When roles
945 were given, moments of looking at each other were longer, and shared gaze (looking at
946 the same object) occurred faster (regardless of who initiated the first look to the

object). In another set of studies, Gullberg & Holmqvist (1999, 2006) investigated how gestures (as a nonverbal source of information that may support verbal information and a means for communicating) are fixated in face-to-face communication. One participant was therefore fitted with a wearable eye tracker and engaged in interaction. Gestures were fixated more often when they occurred peripheral compared to central and when the speaker fixated the gesture too. Note, however, that gestures were fixated on less than 10% of the cases, while gaze was directed at the face for most of the time. This occurs even in sign language, where gaze is also directed at the face most of the time (> 80%) (Muir & Richardson, 2005; Emmorey et al., 2009). Regardless, these studies combined show that gaze is attuned to the interactive context.

Two eye-tracking studies in interaction have paid particular attention to idiosyncratic scan patterns (see the Section *Idiosyncratic face-scanning patterns*). Peterson et al. (2016) investigated whether idiosyncratic biases also occurred during interaction. First, participants completed a face-identification task in the lab, based on which they were classified as upper looker, middle looker or lower looker in faces. Participants were then fitted with a wearable eye tracker and had to walk around campus. All fixations were then classified as being on the face or not using a crowdsourced group of raters (using Amazon Mechanical Turk). Similarly, the position of the upper lip (as a central feature in the face) was determined by a crowdsourced group of raters. The relative location of the first fixation on the face (i.e. where it occurred between the eyes and mouth) was highly correlated across the lab- and wearable eye-tracking experiment. This suggests that idiosyncratic face scanning patterns exist for interactive settings as well, not just for looking at static pictures of faces. Similarly, Rogers et al. (2018) had dyads engage in conversation while wearing eye-tracking glasses. They reported large inter-individual differences in whether the eyes, nose or mouth were preferentially looked at.

Recently, a series of studies on gaze to facial features during face-to-face interaction has been conducted by Hessels et al. (2017, 2018a, 2019). Hessels et al. (2017) used a video-based interaction setup with half-silvered mirrors that allows one to

both look directly into an invisible camera and at the eyes of the other, while their eye movements are recorded with remote eye trackers. They had dyads look at each other for 5 minutes and reported that participants spent most of the time looking at each other's eyes, followed by the nose and mouth. Interestingly, the time spent looking at each other's eyes was highly correlated across dyads (cf. Kendon, 1967, who reports a similar correlation for looking at the face across dyads). In a second experiment, a confederate either stared into the eyes of the other or looked around the face, although this did not affect the gaze of the other participant. Using the same setup, Hessels et al. (2018a) showed that looking at the eyes was correlated to traits of social anxiety and autism spectrum disorder in a student population. Moreover, paired gaze states (e.g. 'eye contact' or one-way averted gaze) were highly, but differentially, correlated to social anxiety and autistic traits. Higher combined traits of social anxiety predicted shorter periods of two-way and one-way eye gaze, and a higher frequency of one-way eye gaze (corroborating a hyper-vigilant scanning style). Higher combined autistic traits, on the other hand, predicted a lower total time in two-way, but a higher total time in one-way eye gaze (corroborating a gaze avoidance scanning style). See Vabalas & Freeth (2016), however, who find no relation between social anxiety or autistic traits and distribution of gaze to the face in a student sample in a wearable eye-tracking interview-setting. Finally, Hessels et al. (2019) reported that the eyes, nose and mouth of a confederate were fixated more often and for longer total durations when the participant was listening than while speaking and that this did not depend on whether the confederate himself was looking away or towards the participant. Interestingly, a gaze shift toward and away from the participant by the confederate caused a difference in the distribution of gaze over the facial features of the participants, which was found not to be due to stimulus factors in a second experiment. The authors concluded that the confederate's gaze shift away from the participant acted as a gaze guide, whereas the gaze shift toward the participant caused participants to distribute their gaze more over the facial features, in relation to the participant's subtask of monitoring when to start speaking. I.e. a gaze shift away from the participant by the confederate likely meant that the

participant didn't need to start speaking, whereas a gaze shift towards the participant might have signalled this.

Interim summary

Eye-tracking studies of gaze in interaction have corroborated findings from both the face-scanning literature and the observational literature. Findings that corroborate the face-scanning literature include the bias for looking at the eyes when one looks at the face of another and idiosyncratic face-scanning patterns. Findings that corroborate the observational literature include the relation between looking toward or away from the conversational partner and the production of speech, as well as patterns of gaze at turn start and end, and the relation to personality or interpersonal context. Several eye-tracking studies have also provided critical extensions, which include the finding that a gaze shift may guide another person's gaze related to the task of monitoring when to start speaking, as well as the rapid use of gaze cues during cooperative behaviours, and the relation between joint gaze to an object and attentional development.

A perspective

In the Section *Functional constraints of gaze for information acquisition from faces*, I have identified when gaze may need to be directed at specific areas of another's face for acquiring the relevant information (e.g. speech, gaze direction) in order to ensure successful interaction. In the Section *Face scanning*, I have identified the biases in gaze to faces and how they are modulated by the content of the face and observer-characteristics. In the Sections *Observational studies of gaze in interaction & Eye tracking in interaction*, I have identified how gaze to faces may regulate social interaction. The studies reviewed here stem from different disciplines and different methodological backgrounds (psychophysical research, observational research, eye-tracking research) with various topics of research (emotion, conversation, interpersonal synchrony, social interaction, etc.). In what follows, I sketch a perspective in order to guide future research on the topic of gaze to faces in social interaction. The goals of this final section are (1) to summarise and organise the relevant factors that

might predict gaze to faces in social interaction, (2) to facilitate the development of future studies on this topic across the breadth of the disciplines involved, and (3) to suggest a way how future studies might describe their findings on gaze in the context of multimodal interaction. It should be noted up front that most studies described above have been designed to maximise the effect of one parameter of interest (e.g. task, context, facial expression) on gaze to faces. In a way, researchers have been working on the ‘atomic’ features of social interaction that might drive gaze. An important question is how conclusions from these studies generalise to the complexity of face-to-face interaction and its situational variance. For example, studies on gaze to emotional faces have mostly featured static pictures with prototypical expressions. Yet, in interaction, emotional expressions are likely much more nuanced. They are not static images, but moving faces bound to bodies that likely carry multiple redundant sources of information (intonation, body posture, etc.). In interaction, this “*varied bouquet of ... cues*” (cf. Koenderink et al., 2000, p. 69) is available to the observer (or better: interactor). It has been well-established that the world is full of redundancy for humans to exploit in guiding their behaviour (e.g. Brunswik, 1955).

I propose that one method that may be particularly helpful in guiding future research on gaze in face-to-face interaction is dynamic systems theory (see e.g. Smith & Thelen, 2003), which, as Beer (2000) explains in the context of cognitive science, focuses on how a process or behaviour unfolds over time and how the unfolding is shaped by various influences. This approach contrasts with, for example, a computational perspective which might focus on how behaviour is causally determined by a set of information-processing mechanisms – i.e. a linear A causes B approach with a set of computations in between. A dynamical approach to (aspects of) human interaction is not new per se. Similar approaches have been proposed and utilised, particularly in research on alignment and synchrony in interpersonal interaction and conversations (see e.g. Fusaroli & Tylén, 2012; Dale et al., 2013; Paxton & Dale, 2013; Fusaroli & Tylén, 2016). Such approaches have, however, not been commonly suggested or utilised in e.g. psychophysical research on the role of gaze to faces. However, the tenets of a dynamic

system approach can be applied to many aspects of this multidisciplinary research topic. In line with what previous researchers have suggested, a dynamic system approach seems to me particularly suited for the study of social interactions, as interactions unfold over time and stimulus and response are hard to disentangle. An analogy to acoustic resonance might help clarify this latter point. I assume that the reader is familiar with the phenomenon of audio feedback, which, for example, occurs when a microphone and speakers are not configured well, resulting in a continuous sound (often loud and high-pitched). While one can describe how this sound develops over time, claiming the microphone or speaker to be the single cause is illogical. It depends on the configuration of the entire system. In social interaction, an analogous phenomenon may occur when two people are forced to stare at each other. At some point they may burst out in laughter. Now, what is the cause of this outburst? Perhaps it begins with a slight twitch of one's corner of the mouth, yet a sequence of causes and effects leading up to the laughter can hardly be ascertained. Thus, the emphasis of the present framework is on changes over time in aspects of the interaction and relations between them.

A dynamic system approach to gaze in face-to-face interaction

I propose that face-to-face interaction can be considered as a dynamic system with multiple sub-states that vary at different timescales and that gaze can be considered as one of these time-varying sub-states. I hold that the state of the entire interaction can be characterised by a number of interactor-specific sub-states (task, drive or affect, idiosyncrasy), sub-states relevant to the content of the interaction itself (gaze direction, facial expression, communicative reference, speaker status), and sub-states relevant to the context in which interaction occurs (cultural and interpersonal context). A schematic overview of these sub-states is given in Figure 2, panel A. It is important to note that, in this framework, the interaction is not considered as two dynamic systems (i.e. the interactors), but one dynamic system which includes two interactors, a context in which the interaction takes place, and the content of the interaction itself (cf. De Jaegher et al., 2010; Dale & Spivey, 2018, for similar perspectives on interaction).

In order to grasp how gaze direction is related to the state of the interaction, it is necessary to understand the time-varying characteristics of the other sub-states and at what timescale these sub-states vary. Panel B in Figure 2 depicts how the sub-states differ in the timescale at which they vary, across microgenetic time (i.e. the timescale at which awareness arises), ontogenetic time (or developmental time) and phylogenetic time (or evolutionary time). At the shortest timescale, facial expressions such as emotional expressions vary as well as facial movements that correlate with speech patterns. One's gaze direction may (1) correlate with one's own facial expressions and (2) correlate with another's facial expressions, such that one's gaze is directed to the facial features (eyes, nose, mouth, eyebrows, etc.) of another in a predictable fashion based on that person's facial expression. At a slightly longer timescale, what I term 'communicative reference' may vary (e.g. Stukenbrock, 2018). These are, for example, gaze cues or gestures (or both, see Yu & Smith, 2013) which may predict when gaze is directed towards relevant objects, i.e. those that coincide with an interactor's locus of spatial attention, or those that are the topic of conversation. Speaker status then refers to which of the interactors is speaking and varies with the turn-taking behaviour of the interaction. Importantly, gaze direction may be correlated not only to who speaks at what point in time, but also to the time since changes of turn, the time to upcoming turn changes, and the stage of speaking (e.g. thinking of what to say versus saying it) (Bailly et al., 2010). Note that 'speaker status' does not suggest that one of the interactors is active, while the other is passive. Both can be actively engaged in the interaction and acquire information from or signal information to each other. It has merely been termed 'speaker status' to indicate whether one is primarily speaking or not, which has been shown to predict where one might look (see above).

A particularly relevant aspect of the interaction to consider is the task carried out by an interactor. As has been pointed out in previous research (e.g. Land & Furneaux, 1997; Land et al., 1999; Hayhoe, 2000), task is an important predictor of gaze in most daily settings, for example during driving, making tea or sandwiches, reading music or playing sports. As Macdonald & Tatler (2013, 2018) and Hessels et al. (2019) have

shown, task is an important predictor of gaze during social interaction as well. At a slightly longer timescale still, drive and affect are expected to vary. This is a rather broad category that relates to long-term emotional states (moods), personality characteristics and e.g. drives to establish, guide, maintain, or dominate a particular interaction. In other words, they are behavioural predictors of gaze that are not task- or context-specific by definition. At the ontogenetic timescale sub-states belonging to the context in which interaction occurs are expected to vary: the interpersonal and cultural context. Finally, idiosyncratic face-scanning patterns are expected to vary at the phylogenetic timescale, and have been suggested to operate as a form of biological niche construction (Constantino et al., 2017).

According to this framework, gaze behaviour of two interactors in face-to-face interaction should be investigated as a dynamic phenomenon developing over multiple timescales and investigated in relation to the time-varying characteristics of the other sub-states that compose the entire state of the interaction. One working hypothesis is that sub-states of the interaction generally correlate with gaze to faces and facial features at shorter timescales than personal and contextual sub-states which correlate with gaze to faces and facial features over longer timescales including ontogenesis and phylogenesis.

There are at least three important implications of the framework I propose for current research on the role of gaze in face-to-face interactions:

1. Perhaps self-evident, it is paramount that gaze is investigated in actual interactive contexts.
2. Gaze should not be treated as an isolated phenomenon, but as one aspect of the interaction, which is multimodal by nature.
3. The time-dependency of gaze to faces in interaction should be emphasised, by relating gaze to the time-varying characteristics of the other sub-states of the interaction. This is in contrast to the dominant approach of reporting aggregate measures of gaze averaged over large parts of an interaction.

Note that the sub-states that compose the state of the interaction as described here should not be considered as logical solids. These sub-states are mere descriptors based on the literature reviewed above, but may be more fluid than initially described. For example, the boundary between ‘task’ and ‘drive’ may not be clearcut, and changes in task may change one’s drive and vice versa. Note also that there is no reference in the framework to the medium through which the interaction takes place and the quality of the information thus transmitted (e.g. a noisy communication channel). Although it may be an important aspect to consider for e.g. video-mediated interactions, I assume the quality of the information is not generally problematic in live interaction, nor is it a characteristic of the interactors, interactional context, or content of the interaction proper.

A criticism one might raise is why the emphasis should be on gaze, as the state of the interaction is composed of many sub-states, of which gaze is but one. As noted in the Introduction, there has been a lot of emphasis on the importance of gaze in social behaviour in the literature (e.g. Emery, 2000; Grossmann, 2017; Pfeiffer et al., 2013), which was the starting point for the present review. Within the framework here proposed, gaze is considered to be one of multiple sub-states, not necessarily more important than the rest. As such, gaze is perhaps considered to be less ‘fundamental’ to social interaction than initially conceived at the start of the review. One advantage of gaze, however, is that it can be measured with a high signal-to-noise ratio using state-of-the-art eye trackers, which makes it an attractive aspect of the interaction to measure. This does not hold for all other aspects of the interaction, such as for example facial expressions, which are difficult to measure reliably using automated techniques (see e.g. Baltrušaitis et al., 2015).

Towards application of the framework. How might one design new experiments on the role of gaze in face-to-face interaction based on the proposed framework? While investigating gaze within the context of the entire multimodal interaction is theoretically attractive, it might not be practically feasible. To make the framework empirically tractable, one will have to make simplifying assumptions

depending on the specific research question.

Say that one is interested in the relation between gaze, facial expression, and speaker status. One could engage dyads into a conversation and operationalise the sub-states of the interaction as follows: Gaze is operationalised for each interactor using an area-of-interest method and is assigned one of the following labels: partner eyes, partner mouth, partner body, partner hands, away from partner. Facial expression is annotated manually for each interactor from a video recording and can take the following states: neutral, laughing, frowning. Speaker status is determined from the audio recording and takes the following two states: interactor 1 speaks, interactor 2 speaks, both speak, no one speaks. One may assume that the conversation is too brief for the task, interpersonal context or cultural context sub-states of the interaction to change meaningfully (although the states itself may be important to conceptualise and consider). From here it is possible to cross-correlate the changes in sub-states of the interaction over time, or to investigate the transition probabilities across combined speaker-expression-gaze states (i.e. as a simplified state of the interaction). In a similar vein, one might be interested in the relation between gaze and the interpersonal context between e.g. a parent and her child. Obviously, the timescale at which one describes changes in the gaze sub-state with respect to changes in the interpersonal context is different from that of the previous example, but the manner of description and analysis may be quite similar. Based on such analyses, one may begin to uncover how gaze supports face-to-face interaction from moment-to-moment, how gaze may stand in for other ostensive signals, or which patterns of gaze are typically elicited in certain interactive contexts.

Concluding remarks & outstanding questions

In this paper, I have reviewed the literature on how gaze can support social interaction, and in particular dyadic face-to-face interactions. I briefly summarise the conclusions of the review and outline a number of fruitful avenues for future research.

Maintaining face-to-face interaction builds on, among others, the perception of

speech and gaze direction. Gaze to the mouth can aid speech perception, while gaze to the eyes likely aids the perception of gaze direction for fine-grained, but not crude judgements of gaze direction. When participants are unrestrained in where they can look, or for how long they can look, there is a bias for fixating other people, faces and eyes. Gaze to faces is, however, modulated by what the face does (talking, expressing emotion, making eye contact), and seems to attract gaze to the source of information. The participant's task furthermore affects gaze to faces, although intra-individual differences in task-specific face-scanning patterns are large. Face-scanning patterns are further sex- and culturally-dependent, highly idiosyncratic and partly under genetic control.

Gaze plays an important role in initiating and regulating interaction, for example in initiating conversation and turn-taking behaviour. Giving up one's turn often includes a look towards the conversational partner, whereas load may be reduced by looking away. Finally, gaze seems to be tightly interwoven with other cues such as linguistic cues for the regulation of interaction. A substantial proportion of eye-tracking studies in interaction have corroborated observational findings on the initiation and regulation of interaction, as well as idiosyncratic face-scanning patterns from the non-interactive literature. These findings from non-interactive settings thus generalise to interactive situations. More recent eye-tracking studies have begun providing critical extensions of the observational literature.

I have sketched a dynamic system approach to interaction which may provide the basis for future research on the role of gaze in face-to-face interaction. Gaze can thus be investigated in relation to sub-states of the interaction, encompassing aspects of the interactors, the content of the interaction itself as well as the interactive context, with each sub-state varying over different timescales from microgenesis to ontogenesis and phylogenesis. A number of important questions for future research are as follows:

1. Not all evidence comes from interaction proper. Do all findings of gaze to faces hold for interactive situations as well? If so, what is their relative contribution?

An essential factor in making these questions tractable is that eye-tracking setups

capable of producing high spatial and temporal resolution eye-tracking data are used.

2. The dynamic system approach I propose is merely a framework at present. What would be minimally needed to predict gaze in interaction given the other sub-states of the interaction? Computational approaches may be particularly useful in this regard.

3. How do idiosyncratic gaze patterns come to develop? Are they mainly under genetic control (Constantino et al., 2017), or are they modulated through interactions with other people?

4. How are tasks or drives represented for the control of gaze in interaction? When does a task become a drive or vice versa? How wide or narrow are representations of tasks or drives (cf. “*suchbild*” in Koenderink, 2014)?

Open practices statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study. The article was not pre-registered. A preprint has been submitted to: <https://osf.io/8zta5>.

References

1249

1250 Adams, R. B., Ambady, N., Nakayama, K., & Shimojo, S. (2011). *The science of social*
1251 *vision*. Oxford University Press, Oxford.

1252 Allen, D. E. & Guy, R. F. (1977). Ocular breaks and verbal output. *Sociometry*, 40(1):
1253 90–96.

1254 Amso, D., Haas, S., & Markant, J. (2014). An Eye Tracking Investigation of
1255 Developmental Change in Bottom-up Attention Orienting to Faces in Cluttered
1256 Natural Scenes. *PLOS One*, 9(1):e85701.

1257 Argyle, M. (1972). Eye-contact and the direction of gaze. In *The psychology of*
1258 *interpersonal behavior*. Harmondsworth, Middlesex.

1259 Argyle, M. & Dean, J. (1965). Eye-contact, distance and affiliation. *Sociometry*, 28(3):
1260 289–304.

1261 Arizpe, J., Kravitz, D. J., Yovel, G., & Baker, C. I. (2012). Start Position Strongly
1262 Influences Fixation Patterns during Face Processing: Difficulties with Eye Movements
1263 as a Measure of Information Use. *PLOS One*, 7(2):e31106.

1264 Arizpe, J., Kravitz, D. J., Walsh, V., Yovel, G., & Baker, C. I. (2016). Differences in
1265 Looking at Own- and Other-Race Faces Are Subtle and Analysis-Dependent: An
1266 Account of Discrepant Reports. *PLOS One*, 11(2):e0148253.

1267 Arizpe, J., Walsh, V., Yovel, G., & Baker, C. I. (2017). The categories, frequencies, and
1268 stability of idiosyncratic eye-movement patterns to faces. *Vision Research*, 141:
1269 191–203.

1270 Arizpe, J. M., Noles, D. L., Tsao, J. W., & Chan, A. W. Y. (2019). Eye Movement
1271 Dynamics Differ between Encoding and Recognition of Faces. *Vision*, 3(9):1–31.

1272 Auer, P. (2018). Gaze, addresse selection and turn-taking in three-party interaction. In
1273 Brône, G. & Oben, B., editors, *Eye-tracking in interaction*. John Benjamins
1274 Publishing Company, Amsterdam / Philadelphia.

- 1275 Bailly, G., Raidt, S., & Elisei, F. (2010). Gaze, conversational agents and face-to-face
1276 communication. *Speech Communication*, 52(6):598–612.
- 1277 Bailly, G., Mihoub, A., Wolf, C., & Elisei, F. (2018). Gaze and face-to-face interaction.
1278 In Brône, G. & Oben, B., editors, *Eye-tracking in interaction*. John Benjamins
1279 Publishing Company, Amsterdam / Philadelphia.
- 1280 Baltrušaitis, T., Mahmoud, M., & Robinson, P. (2015). Cross-dataset learning and
1281 person-specific normalisation for automatic Action Unit detection. *11th IEEE*
1282 *International Conference and Workshops on Automatic Face and Gesture Recognition*,
1283 6:1–6.
- 1284 Banse, R. & Scherer, K. R. (1996). Acoustic Profiles in Vocal Emotion Expression.
1285 *Journal of Personality and Social Psychology*, 70(3):614–636.
- 1286 Bänziger, T. & Scherer, K. R. (2005). The role of intonation in emotional expressions.
1287 *Speech Communication*, 46(3-4):252–267.
- 1288 Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener Responses as a Collaborative
1289 Process: The Role of Gaze. *Journal of Communication*, 52(3):566–580.
- 1290 Beattie, G. W. (1981). A further investigation of the cognitive interference hypothesis of
1291 gaze patterns during conversation. *British Journal of Social Psychology*, 20:243–248.
- 1292 Beattie, G. W. & Bogle, G. (1982). The reliability and validity of different
1293 video-recording techniques used for analysing gaze in dyadic interaction. *British*
1294 *Journal of Social Psychology*, 21:31–34.
- 1295 Beaudry, O., Roy-Charland, A., Perron, M., Cormier, I., & Tapp, R. (2014). Featural
1296 processing in recognition of emotional facial expressions. *Cognition & Emotion*, 28(3):
1297 416–432.
- 1298 Becker, M. W. & Detweiler-Bedell, B. (2009). Short article: Early detection and
1299 avoidance of threatening faces during passive viewing. *Quarterly Journal of*
1300 *Experimental Psychology*, 62(7):1257–1264.

- 1301 Beer, R. D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive*
1302 *Sciences*, 4(3):91–99.
- 1303 Bindemann, M., Burton, A. M., Hooge, I. T. C., Jenkins, R., & de Haan, E. H. F.
1304 (2005). Faces retain attention. *Psychological Bulletin & Review*, 12(6):1048–1053.
- 1305 Bindemann, M., Burton, A. M., Langton, S. R. H., Schweinberger, S. R., & Doherty,
1306 M. J. (2007). The control of attention to faces. *Journal of Vision*, 7(10):1–8.
- 1307 Birmingham, E. & Kingstone, A. (2009). Human Social Attention. *Annals of the New*
1308 *York Academy of Sciences*, 1156(1):118–140.
- 1309 Birmingham, E., Bischof, W. F., & Kingstone, A. (2008a). Social attention and
1310 real-world scenes: The roles of action, competition and social content. *The Quarterly*
1311 *Journal of Experimental Psychology*, 61(7):986–998.
- 1312 Birmingham, E., Bischof, W. F., & Kingstone, A. (2008b). Gaze selection in complex
1313 social scenes. *Visual Cognition*, 16(2-3):341–355.
- 1314 Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Saliency does not account for
1315 fixations to eyes within social scenes. *Vision Research*, 49(24):2992–3000.
- 1316 Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture Shapes
1317 How We Look at Faces. *PLOS One*, 3(8):e3022.
- 1318 Blais, C., Fiset, D., Roy, C., Saumure Régimbald, C., & Gosselin, F. (2017). Eye
1319 Fixation Patterns for Categorizing Static and Dynamic Facial Expressions. *Emotion*,
1320 17(7):1107–1119.
- 1321 Bombari, D., Schmid, P. C., Schmid Mast, M., Birri, S., Mast, F. W., & Lobmaier, J. S.
1322 (2013). Emotion Recognition: The Role of Featural and Configural Face Information.
1323 *Quarterly Journal of Experimental Psychology*, 66(12):2426–2442.
- 1324 Brône, G. & Oben, B., editors. (2018). *Eye-tracking in interaction*. Studies on the role
1325 of eye gaze in dialogue. John Benjamins Publishing Company, Amsterdam /
1326 Philadelphia.

- 1327 Brooks, R. & Meltzoff, A. N. (2008). Infant gaze following and pointing predict
1328 accelerated vocabulary growth through two years of age: a longitudinal, growth curve
1329 modeling study. *Journal of Child Language*, 35(1):207–220.
- 1330 Broz, F., Lehmann, H., Nehaniv, C. L., & Dautenhahn, K. (2012). Mutual Gaze,
1331 Personality, and Familiarity: Dual Eye-Tracking During Conversation. *2012 IEEE*
1332 *RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive*
1333 *Communication*.
- 1334 Brunswik, E. (1955). Representative design and probabilistic theory in a functional
1335 psychology. *Psychological Review*, 62(3):193–217.
- 1336 Buchan, J. N., Paré, M., & Munhall, K. G. (2007). Spatial statistics of gaze fixations
1337 during dynamic face processing. *Social Neuroscience*, 2(1):1–13.
- 1338 Buchan, J. N., Paré, M., & Munhall, K. G. (2008). The effect of varying talker identity
1339 and listening conditions on gaze behavior during audiovisual speech perception. *Brain*
1340 *Research*, 1242:162–171.
- 1341 Buswell, G. T. (1935). *How people look at pictures*. A study of the psychology of
1342 perception in art. The University of Chicago Press, Chicago, Illinois.
- 1343 Butler, S., Gilchrist, I. D., Burt, D. M., Perrett, D. I., & Jones, E. (2005). Are the
1344 perceptual biases found in chimeric face processing reflected in eye-movement
1345 patterns? *Neuropsychologia*, 43:52–59.
- 1346 Caldara, R., Schyns, P., Mayer, E., Smith, M. L., Gosselin, F., & Rossion, B. (2005).
1347 Does Prosopagnosia Take the Eyes Out of Face Representations? Evidence for a
1348 Defect in Representing Diagnostic Facial Information following Brain Damage.
1349 *Journal of Cognitive Neuroscience*, 17(10):1652–1666.
- 1350 Caldara, R., Zhou, X., & Miellet, S. (2010). Putting Culture Under the ‘Spotlight’
1351 Reveals Universal Information Use for Face Recognition. *PLOS One*, 5(3):e9708.

- 1352 Calvo, M. G., Fernández-Martín, A., & Nummenmaa, L. (2014). Facial expression
1353 recognition in peripheral versus central vision: role of the eyes and the mouth.
1354 *Psychological Research*, 78(2):180–195.
- 1355 Cañigüeral, R. & Hamilton, A. F. C. (2019). Being watched: Effects of an audience on
1356 eye gaze and prosocial behaviour. *Acta Psychologica*, 195:50–63.
- 1357 Caruana, N., McArthur, G., Woolgar, A., & Brock, J. (2017). Simulating social
1358 interactions for the experimental investigation of joint attention. *Neuroscience &*
1359 *Biobehavioral Reviews*, 74(Part A):115–125.
- 1360 Cary, M. S. (1978). The role of gaze in the initiation of conversation. *Social Psychology*,
1361 41(3):269–271.
- 1362 Casillas, M. & Frank, M. C. (2017). The development of children’s ability to track and
1363 predict turn structure in conversation. *Journal of memory and language*, 92(C):
1364 234–253.
- 1365 Chen, M. (2002). Leveraging the Asymmetric Sensitivity of Eye Contact for
1366 Videoconference. *CHI '02 Proceedings of the SIGCHI Conference on Human Factors*
1367 *in Computing Systems*, 4(1):49–56.
- 1368 Clark, A. T. & Gergle, D. (2011). Mobile Dual Eye-Tracking Methods: Challenges and
1369 Opportunities. *DUET Dual Eye Tracking workshop at ECSCW 2011*, pages 1–7.
- 1370 Clark, H. H. & Krych, M. A. (2004). Speaking while monitoring addressees for
1371 understanding. *Journal of memory and language*, 50(1):62–81.
- 1372 Cline, M. G. (1967). The Perception of Where a Person Is Looking. *The American*
1373 *Journal of Psychology*, 80(1):41–50.
- 1374 Cole, G. G., Skarratt, P. A., & Kuhn, G. (2016). Real person interaction in visual
1375 attention research. *European Psychologist*, 21(2):141–149.
- 1376 Constantino, J. N., Kennon-McGill, S., Weichselbaum, C., Marrus, N., Haider, A.,
1377 Glowinski, A. L., Gillespie, S., Klaiman, C., Klin, A., & Jones, W. (2017). Infant

- 1378 viewing of social scenes is under genetic control and is atypical in autism. *Nature*, 547
1379 (7663):340–344.
- 1380 Coutrot, A., Binetti, N., Harrison, C., Mareschal, I., & Johnston, A. (2016). Face
1381 exploration dynamics differentiate men and women. *Journal of Vision*, 16(14):1–19.
- 1382 Dale, R. & Spivey, M. J. (2018). Weaving oneself into others. In Brône, G. & Oben, B.,
1383 editors, *Eye-tracking in interaction*. John Benjamins Publishing Company,
1384 Amsterdam / Philadelphia.
- 1385 Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2013). The
1386 Self-Organization of Human Interaction. In *Psychology of learning and motivation*,
1387 pages 43–95. Elsevier.
- 1388 Damasio, A. R., Damasio, H., & van Hoesen, G. W. (1982). Prosopagnosia: anatomic
1389 basis and behavioral mechanisms. *Neurology*, 32:331–341.
- 1390 de Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in
1391 affective neuroscience. *Philosophical Transactions of the Royal Society B: Biological*
1392 *Sciences*, 364(1535):3475–3484.
- 1393 De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute
1394 social cognition? *Trends in Cognitive Sciences*, 14(10):441–447.
- 1395 Eisenbarth, H. & Alpers, G. W. (2011). Happy mouth and sad eyes: Scanning
1396 emotional facial expressions. *Emotion*, 11(4):860–865.
- 1397 Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of
1398 social gaze. *Neuroscience & Biobehavioral Reviews*, 24:581–604.
- 1399 Emmorey, K., Thompson, R., & Colvin, R. (2009). Eye Gaze During Comprehension of
1400 American Sign Language by Native and Beginning Signers. *Journal of Deaf Studies*
1401 *and Deaf Education*, 14(2):237–243.
- 1402 End, A. & Gamer, M. (2019). Task instructions can accelerate the early preference for
1403 social features in naturalistic scenes. *Royal Society Open Science*, 6(180596):1–19.

- 1404 Exline, R., Gray, D., & Schuette, D. (1965). Visual behavior in a dyad as affected by
1405 interview content and sex of respondent. *Journal of Personality and Social*
1406 *Psychology*, 1(3):201–209.
- 1407 Florey, J., Clifford, C. W. G., Dakin, S. C., & Mareschal, I. (2015). Peripheral
1408 processing of gaze. *Journal of Experimental Psychology: Human Perception and*
1409 *Performance*, 41(4):1084–1094.
- 1410 Foddy, M. (1978). Patterns of gaze in cooperative and competitive negotiation. *Human*
1411 *Relations*, 31(11):925–938.
- 1412 Foulsham, T. & Sanderson, L. A. (2013). Look who's talking? Sound changes gaze
1413 behaviour in a dynamic social scene. *Visual Cognition*, 21(7):922–944.
- 1414 Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze
1415 allocation in a dynamic situation: Effects of social status and speaking. *Cognition*,
1416 117(3):319–331.
- 1417 Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze
1418 allocation in the lab and the natural environment. *Vision Research*, 51(17):1920–1931.
- 1419 Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to
1420 faces during the first year. *Cognition*, 110(2):160–170.
- 1421 Frank, M. C., Vul, E., & Saxe, R. (2012). Measuring the Development of Social
1422 Attention Using Free-Viewing. *Infancy*, 17(4):355–375.
- 1423 Freeth, M., Foulsham, T., & Kingstone, A. (2013). What Affects Social Attention?
1424 Social Presence, Eye Contact and Autistic Traits. *PLOS One*, 8(1):e53286.
- 1425 Friesen, C. K. & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered
1426 by nonpredictive gaze. *Psychonomic Bulletin & Review*, 5(3):490–495.
- 1427 Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual
1428 attention, social cognition, and individual differences. *Psychological Bulletin*, 133(4):
1429 694–724.

- 1430 Fu, G., Hu, C. S., Wang, Q., Quinn, P. C., & Lee, K. (2012). Adults Scan Own- and
1431 Other-Race Faces Differently. *PLOS One*, 7(6):e37688.
- 1432 Fusaroli, R. & Tylén, K. (2012). Carving language for social coordination: A dynamical
1433 approach. *Interaction Studies*, 13(1):103–124.
- 1434 Fusaroli, R. & Tylén, K. (2016). Investigating Conversational Dynamics: Interactive
1435 Alignment, Interpersonal Synergy, and Collective Task Performance. *Cognitive*
1436 *Development*, 40(1):145–171.
- 1437 Gamer, M. & Hecht, H. (2007). Are you looking at me? Measuring the cone of gaze.
1438 *Journal of Experimental Psychology: Human Perception and Performance*, 33(3):
1439 705–715.
- 1440 Gamer, M., Hecht, H., Seipp, N., & Hiller, W. (2011). Who is looking at me? The cone
1441 of gaze widens in social phobia. *Cognition & Emotion*, 25(4):756–764.
- 1442 Georgiou, G., Bleakley, C., Hayward, J., Russo, R., Dutton, K., Eltiti, S., & Fox, E.
1443 (2005). Focusing on fear: attentional disengagement from emotional faces. *Visual*
1444 *Cognition*, 12(1):145–158.
- 1445 Gibson, J. J. & Pick, A. D. (1963). Perception of Another Person's Looking Behavior.
1446 *The American Journal of Psychology*, 76(3):386–394.
- 1447 Gliga, T., Elsabbagh, M., Andravizou, A., & Johnson, M. (2009). Faces Attract Infants'
1448 Attention in Complex Displays. *Infancy*, 14(5):550–562.
- 1449 Gobel, M. S., Kim, H. S., & Richardson, D. C. (2015). The dual function of social gaze.
1450 *Cognition*, 136:359–364.
- 1451 Gobel, M. S., Chen, A., & Richardson, D. C. (2017). How Different Cultures Look at
1452 Faces Depends on the Interpersonal Context. *Canadian Journal of Experimental*
1453 *Psychology/Revue canadienne de psychologie expérimentale*, 71(3):258–264.
- 1454 Goffman, E. (1966). *Behavior in public places*. Notes on the social organization of
1455 gatherings. The Free Press, New York.

- 1456 Goldstein, A. G. & Mackenberg, E. J. (1966). Recognition of human faces from isolated
1457 facial features: A developmental study. *Psychonomic Science*, 84:149–150.
- 1458 Goren, C. C., Sarty, M., & Wu, P. Y. K. (1975). Visual following and pattern
1459 discrimination of face-like stimuli by newborn infants. *Pediatrics*, 56(4):544–549.
- 1460 Gredebäck, G., Fikke, L., & Melinder, A. (2010). The development of joint visual
1461 attention: a longitudinal study of gaze following during interactions with mothers and
1462 strangers. *Developmental Science*, 13(6):839–848.
- 1463 Green, M., Williams, L., & Davidson, D. (2003). In the face of danger: Specific viewing
1464 strategies for facial expressions of threat? *Cognition & Emotion*, 17(5):779–786.
- 1465 Gregory, N. & Antolin, J. (2019). Does Social Presence or the Potential for Interaction
1466 reduce Social Gaze in Online Social Scenarios? Introducing the "Live Lab" paradigm.
1467 . *The Quarterly Journal of Experimental Psychology*, 72(4):779–791.
- 1468 Grossmann, T. (2017). The Eyes as Windows Into Other Minds. *Perspectives on*
1469 *Psychological Science*, 12(1):107–121.
- 1470 Guillon, Q., Hadjikhani, N., Baduel, S., & Rogé, B. (2014). Visual social attention in
1471 autism spectrum disorder: Insights from eye tracking studies. *Neuroscience &*
1472 *Biobehavioral Reviews*, 42:279–297.
- 1473 Gullberg, M. & Holmqvist, K. (1999). Keeping an eye on gestures: Visual perception of
1474 gestures in face-to-face communication. *Pragmatics & Cognition*, 7(1):35–63.
- 1475 Gullberg, M. & Holmqvist, K. (2006). What speakers do and what addressees look at:
1476 Visual attention to gestures in human interaction live and on video. *Pragmatics &*
1477 *Cognition*, 14(1):53–82.
- 1478 Guo, K., Robertson, R. G., Mahmoodi, S., Tadmor, Y., & Young, M. P. (2003). How
1479 do monkeys view faces?—a study of eye movements. *Experimental Brain Research*,
1480 150(3):363–374.

- 1481 Guo, K., Meints, K., Hall, C., Hall, S., & Mills, D. (2009). Left gaze bias in humans,
1482 rhesus monkeys and domestic dogs. *Animal cognition*, 12(3):409–418.
- 1483 Hanna, J. E. & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring
1484 expressions early during face-to-face conversation. *Journal of memory and language*,
1485 57(4):596–615.
- 1486 Hayhoe, M. (2000). Vision Using Routines: A Functional Account of Vision. *Visual*
1487 *Cognition*, 7(1-3):43–64.
- 1488 Hayward, D. A., Voorhies, W., Morris, J. L., Capozzi, F., & Ristic, J. (2017). Staring
1489 Reality in the Face: A Comparison of Social Attention Across Laboratory and Real
1490 World Measures Suggests Little Common Ground. *Canadian Journal of Experimental*
1491 *Psychology/Revue canadienne de psychologie expérimentale*, 71(3):212–225.
- 1492 Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional
1493 during face learning. *Memory & Cognition*, 33(1):98–106.
- 1494 Hessels, R. S., Cornelissen, T. H. W., Hooge, I. T. C., & Kemner, C. (2017). Gaze
1495 Behavior to Faces During Dyadic Interaction. *Canadian Journal of Experimental*
1496 *Psychology/Revue canadienne de psychologie expérimentale*, 71(3):226–242.
- 1497 Hessels, R. S., Holleman, G. A., Cornelissen, T. H. W., Hooge, I. T. C., & Kemner, C.
1498 (2018a). Eye contact takes two – autistic and social anxiety traits predict gaze
1499 behavior in dyadic interaction. *Journal of Experimental Psychopathology*,
1500 (April-June):1–17.
- 1501 Hessels, R. S., Niehorster, D. C., Nyström, M., Andersson, R., & Hooge, I. T. C.
1502 (2018b). Is the eye-movement field confused about fixations and saccades? A survey
1503 among 124 researchers. *Royal Society Open Science*, 5(180502):1–23.
- 1504 Hessels, R. S., Holleman, G. A., Kingstone, A., Hooge, I. T. C., & Kemner, C. (2019).
1505 Gaze allocation in face-to-face communication is affected primarily by task structure
1506 and social context, not stimulus-driven factors. *Cognition*, 184:28–43.

- 1507 Hietanen, J. K. (2018). Affective Eye Contact: An Integrative Review. *Frontiers in*
1508 *Psychology*, 9:1–15.
- 1509 Hills, P. J., Ross, D. A., & Lewis, M. B. (2011). Attention misplaced: The role of
1510 diagnostic features in the face-inversion effect. *Journal of Experimental Psychology:*
1511 *Human Perception and Performance*, 37(5):1396–1406.
- 1512 Hirvenkari, L., Ruusuvuori, J., Saarinen, V.-M., Kivioja, M., Peräkylä, A., & Hari, R.
1513 (2013). Influence of Turn-Taking in a Two-Person Conversation on the Gaze of a
1514 Viewer. *PLOS One*, 8(8):1–6.
- 1515 Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and Listening with the Eyes:
1516 Gaze Signaling during Dyadic Interactions. *PLOS One*, 10(8):e0136905.
- 1517 Holler, J. & Kendrick, K. H. (2015). Unaddressed participants' gaze in multi-person
1518 interaction: optimizing reciprocity. *Frontiers in Psychology*, 6(98):1–14.
- 1519 Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult's Eyes Trigger Shifts of Visual
1520 Attention in Human Infants. *Psychological Science*, 9(2):131–134.
- 1521 Hooge, I. T. C. & Erkelens, C. J. (1999). Peripheral vision and oculomotor control
1522 during visual search. *Vision Research*, 39:1567–1575.
- 1523 Horley, K., Williams, L. M., Gonsalvez, C., & Gordon, E. (2003). Social phobics do not
1524 see eye to eye: A visual scanpath study of emotional expression processing. *Journal*
1525 *of anxiety disorders*, 17:33–44.
- 1526 Hsiao, J. H.-w. & Cottrell, G. (2008). Two fixations suffice in face recognition.
1527 *Psychological Science*, 19(10):998–1006.
- 1528 Hunnius, S., de Wit, T. C. J., Vries, S., & von Hofsten, C. (2011). Facing threat:
1529 Infants' and adults' visual scanning of faces with neutral, happy, sad, angry, and
1530 fearful emotional expressions. *Cognition & Emotion*, 25(2):193–205.

- 1531 Itier, R. J., Villate, C., & Ryan, J. D. (2007). Eyes always attract attention but gaze
1532 orienting is task-dependent: Evidence from eye movement monitoring.
1533 *Neuropsychologia*, 45(5):1019–1028.
- 1534 Itti, L. & Koch, C. (2000). A saliency-based search mechanism for overt and covert
1535 shifts of visual attention. *Vision Research*, 40:1489–1506.
- 1536 Jack, R. E. & Schyns, P. G. (2015). The Human Face as a Dynamic Tool for Social
1537 Communication. *Current Biology*, 25(14):R621–R634.
- 1538 Jack, R. E. & Schyns, P. G. (2017). Toward a Social Psychophysics of Face
1539 Communication. *Annual Review of Psychology*, 68(1):269–297.
- 1540 Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural
1541 Confusions Show that Facial Expressions Are Not Universal. *Current Biology*, 19(18):
1542 1543–1548.
- 1543 Jarick, M. & Kingstone, A. (2015). The duality of gaze: Eyes extract and signal social
1544 information during sustained cooperative and competitive dyadic gaze. *Frontiers in*
1545 *Psychology*, 6(1423):1–7.
- 1546 Jehoul, A., Brône, G., & Feyaerts, K. (2017). Gaze patterns and fillers: Empirical data
1547 on the difference between Dutch 'euh' and 'euhm'. *Proceedings of the th European*
1548 *and th Nordic Symposium on Multimodal Communication MMSYM*, (141):43–50.
- 1549 Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential
1550 tracking of face-like stimuli and its subsequent decline. *Cognition*, 40(1-2):1–19.
- 1551 Kanan, C., Bseiso, D. N. F., Ray, N. A., Hsiao, J. H., & Cottrell, G. W. (2015).
1552 Humans have idiosyncratic and task-specific scanpaths for judging faces. *Vision*
1553 *Research*, 108:67–76.
- 1554 Kano, F. & Tomonaga, M. (2009). How chimpanzees look at pictures: a comparative
1555 eye-tracking study. *Proceedings of the Royal Society B: Biological Sciences*, 276
1556 (1664):1949–1955.

- 1557 Kano, F. & Tomonaga, M. (2010). Face scanning in chimpanzees and humans:
1558 continuity and discontinuity. *Animal behaviour*, 79(1):227–235.
- 1559 Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta*
1560 *Psychologica*, 26:22–63.
- 1561 Kidwell, M. (2005). Gaze as Social Control: How Very Young Children Differentiate
1562 "The Look" From a "Mere Look" by Their Adult Caregivers. *Research on Language &*
1563 *Social Interaction*, 38(4):417–449.
- 1564 Kidwell, M. (2009). Gaze Shift as an Interactional Resource for Very Young Children.
1565 *Discourse Processes*, 46(2-3):145–160.
- 1566 Kingstone, A. (2009). Taking a real look at social attention. *Current Opinion in*
1567 *Neurobiology*, 19:52–56.
- 1568 Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive Ethology: A new
1569 approach for studying human cognition. *British Journal of Psychology*, 99(3):317–340.
- 1570 Kleck, R. E. (2005). Effects of Direct and Averted Gaze on the Perception of Facially
1571 Communicated Emotion. *Emotion*, 5(1):3–11.
- 1572 Kleinke, C. L. (1986). Gaze and Eye Contact: A Research Review. *Psychological*
1573 *Bulletin*, 100(1):78–100.
- 1574 Kluttz, N. L., Mayes, B. R., West, R. W., & Kerby, D. S. (2009). The effect of head
1575 turn on the perception of gaze. *Vision Research*, 49(15):1979–1993.
- 1576 Kobayashi, H. & Kohshima, S. (1997). Unique morphology of the human eye. *Nature*,
1577 387:767–768.
- 1578 Koenderink, J. (2014). The All Seeing Eye? *Perception*, 43(1):1–6.
- 1579 Koenderink, J. J., van Doorn, A. J., & Lappin, J. S. (2000). Direct Measurement of the
1580 Curvature of Visual Space. *Perception*, 29(1):69–79.

- 1581 Krumhuber, E. G., Kappas, A., & Manstead, A. S. R. (2013). Effects of Dynamic
1582 Aspects of Facial Expressions: A Review. *Emotion Review*, 5(1):41–46.
- 1583 Kwart, D. G., Foulsham, T., & Kingstone, A. (2012). Age and Beauty are in the Eye of
1584 the Beholder. *Perception*, 41(8):925–938.
- 1585 Laidlaw, K. E. W., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social
1586 interactions are important to social attention. *Proceedings of the National Academy
1587 of Sciences*, 108(14):5548–5553.
- 1588 Laidlaw, K. E. W. & Kingstone, A. (2017). Fixations to the eyes aids in facial
1589 encoding; covertly attending to the eyes does not. *Acta Psychologica*, 173(C):55–65.
- 1590 Laidlaw, K. E. W., Risko, E. F., & Kingstone, A. (2012). A new look at social
1591 attention: Orienting to the eyes is not (entirely) under volitional control. *Journal of
1592 Experimental Psychology: Human Perception and Performance*, 38(5):1132–1143.
- 1593 Laidlaw, K. E. W., Rothwell, A., & Kingstone, A. (2016). Camouflaged attention:
1594 covert attention is critical to social communication in natural settings. *Evolution and
1595 Human Behavior*, 37(6):449–455.
- 1596 Land, M., Mennie, N., & Rusted, J. (1999). The Roles of Vision and Eye Movements in
1597 the Control of Activities of Daily Living. *Perception*, 28:1311–1328.
- 1598 Land, M. F. & Furneaux, S. (1997). The knowledge base of the oculomotor system.
1599 *Philosophical Transactions of the Royal Society B: Biological Sciences*, 352:1231–1239.
- 1600 Langton, S. R. H. (2000). The mutual influence of gaze and head orientation in the
1601 analysis of social attention direction. *The Quarterly Journal of Experimental
1602 Psychology*, 53(3):825–845.
- 1603 Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the
1604 direction of social attention. *Trends in Cognitive Sciences*, 4(2):50–59.

- 1605 Langton, S. R. H., Honeyman, H., & Tessler, E. (2004). The influence of head contour
1606 and nose angle on the perception of eye-gaze direction. *Perception & Psychophysics*,
1607 66(5):752–771.
- 1608 Langton, S. R. H., Law, A. S., Burton, A. M., & Schweinberger, S. R. (2008). Attention
1609 capture by faces. *Cognition*, 107(1):330–342.
- 1610 Lansing, C. R. & McConkie, G. W. (1999). Attention to Facial Regions in Segmental
1611 and Prosodic Visual Speech Perception Tasks. *Journal of Speech, Language and*
1612 *Hearing Research*, 42:526–539.
- 1613 Leppänen, J. M. (2016). Using Eye Tracking to Understand Infants’ Attentional Bias
1614 for Faces. *Child Development Perspectives*.
- 1615 Leppänen, J. M., Cataldo, J. K., Bosquet Enlow, M., & Nelson, C. A. (2018). Early
1616 development of attention to threat-related facial expressions. *PLOS One*, 13(5):1–13.
- 1617 Lewkowicz, D. J. & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the
1618 mouth of a talking face when learning speech. *Proceedings of the National Academy*
1619 *of Sciences*, 109(5):1431–1436.
- 1620 Libby, W. L. & Yaklevich, D. (1973). Personality determinants of eye contact and
1621 direction of gaze aversion. *Journal of Personality and Social Psychology*, 27(2):
1622 197–206.
- 1623 Liu, S., Quinn, P. C., Wheeler, A., Xiao, N., Ge, L., & Lee, K. (2011). Similarity and
1624 difference in the processing of same- and other-race faces as revealed by eye tracking
1625 in 4- to 9-month-olds. *Journal of Experimental Child Psychology*, 108(1):180–189.
- 1626 Loomis, J. M., Kelly, J. W., Pusch, M., Bailenson, J. N., & Beall, A. C. (2008).
1627 Psychophysics of perceiving eye-gaze and head direction with peripheral
1628 vision: Implications for the dynamics of eye-gaze behavior. *Perception*, 37(9):
1629 1443–1457.

- 1630 Luria, S. M. & Strauss, M. S. (1978). Comparison of Eye Movements over Faces in
1631 Photographic Positives and Negatives. *Perception*, 7:349–358.
- 1632 Ma, W. J., Zhou, X., Ross, L. A., Foxe, J. J., & Parra, L. C. (2009). Lip-Reading Aids
1633 Word Recognition Most in Moderate Noise: A Bayesian Explanation Using
1634 High-Dimensional Feature Space. *PLOS One*, 4(3):1–14.
- 1635 Macdonald, R. G. & Tatler, B. W. (2013). Do as eye say: Gaze cueing and language in
1636 a real-world social interaction. *Journal of Vision*, 13(4):1–12.
- 1637 Macdonald, R. G. & Tatler, B. W. (2018). Gaze in a real-world social interaction: A
1638 dual eye-tracking study. *Quarterly Journal of Experimental Psychology*, 71(10):
1639 2162–2173.
- 1640 Mareschal, I., Calder, A. J., Dadds, M. R., & Clifford, C. W. G. (2013a). Gaze
1641 categorization under uncertainty: Psychophysics and modeling. *Journal of Vision*, 13
1642 (5):18–18.
- 1643 Mareschal, I., Calder, A. J., & Clifford, C. W. G. (2013b). Humans Have an
1644 Expectation That Gaze Is Directed Toward Them. *Current Biology*, 23(8):717–721.
- 1645 Marr, D. (1982). *Vision: A computational investigation into the human representation*
1646 *and processing of visual information*. W. H. Freeman and Company, San Francisco.
- 1647 McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264:
1648 746–748.
- 1649 McKelvie, S. J. (1976). The Role of Eyes and Mouth in the Memory of a Face. *The*
1650 *American Journal of Psychology*, 89(2):311–323.
- 1651 Mehoudar, E., Arizpe, J., Baker, C. I., & Yovel, G. (2014). Faces in the eye of the
1652 beholder: Unique and stable eye scanning patterns of individual observers. *Journal of*
1653 *Vision*, 14(7):1–11.

- Mihoub, A., Bailly, G., Wolf, C., & Elisei, F. (2015). Learning multimodal behavioral models for face-to-face social interaction. *Journal on Multimodal User Interfaces*, 9(3):195–210.
- Muir, L. J. & Richardson, I. E. G. (2005). Perception of Sign Language and Its Application to Visual Communications for Deaf People. *Journal of Deaf Studies and Deaf Education*, 10(4):390–401.
- Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., & Parlade, M. V. (2007). Individual Differences and the Development of Joint Attention in Infancy. *Child Development*, 28(3):938–954.
- Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H., & Hagita, N. Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *the 4th ACM/IEEE international conference*, pages 61–69, New York, New York, USA, (2009). ACM Press.
- Neath, K. N. & Itier, R. J. (2014). Facial expression discrimination varies with presentation time but not with fixation on features: A backward masking study using eye- tracking. *Cognition & Emotion*, 28(1):115–131.
- Nguyen, H. T., Isaacowitz, D. M., & Rubin, P. A. D. (2009). Age- and Fatigue-related Markers of Human Faces: An Eye-Tracking Study. *OPHTHA*, 116(2):355–360.
- Niehorster, D. C., Santini, T., Hessels, R. S., Hooge, I. T. C., Kasneci, E., & Nyström, M. (2020). The impact of slippage on the data quality of head-worn eye trackers. *Behavior Research Methods*, pages 1–21.
- Norman, D. A. & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, 7:44–64.
- Noton, D. & Stark, L. (1971). Scanpaths in Eye Movements during Pattern Perception. *Science*, 171(3968):308–311.

- 1679 Nyström, M. & Holmqvist, K. (2008). Semantic Override of Low-level Features in Image
1680 Viewing – Both Initially and Overall. *Journal of Eye Movement Research*, 2(2):1–11.
- 1681 Or, C. C. F., Peterson, M. F., & Eckstein, M. P. (2015). Initial eye movements during
1682 face identification are optimal and similar across cultures. *Journal of Vision*, 15(13):
1683 1–25.
- 1684 Over, E. A. B., Hooge, I. T. C., Vlaskamp, B. N. S., & Erkelens, C. J. (2007).
1685 Coarse-to-fine eye movement strategy in visual search. *Vision Research*, 47(17):
1686 2272–2280.
- 1687 Palanica, A. & Itier, R. J. (2014). Effects of peripheral eccentricity and head
1688 orientation on gaze discrimination. *Visual Cognition*, 22(9-10):1216–1232.
- 1689 Paré, M., Richler, R. C., & ten Hove, M. (2003). Gaze behavior in audiovisual speech
1690 perception: The influence of ocular fixations on the McGurk effect. *Perception &*
1691 *Psychophysics*, 65(4):553–567.
- 1692 Patterson, M. L. (1976). An Arousal Model of Interpersonal Intimacy. *Psychological*
1693 *Review*, 83(3):235–245.
- 1694 Patterson, M. L. (1982). A Sequential Functional Model of Nonverbal Exchange.
1695 *Psychological Review*, 89(3):231–249.
- 1696 Paxton, A. & Dale, R. (2013). Multimodal Networks of Interpersonal Interaction and
1697 Conversational Contexts. *Proceedings of the Annual Meeting of the Cognitive Science*
1698 *Society*, 35:1121–1126.
- 1699 Peltola, M. J., Leppänen, J. M., & Vogel-Farley, V. K. (2009). Fearful faces but not
1700 fearful eyes alone delay attention disengagement in 7-month-old infants. *Emotion*, 9
1701 (4):560–565.
- 1702 Peltola, M. J., Leppänen, J. M., Palokangas, T., & Hietanen, J. K. (2008). Fearful faces
1703 modulate looking duration and attention disengagement in 7-month-old infants.
1704 *Developmental Science*, 11(1):60–68.

- 1705 Peltola, M. J., Yrttiaho, S., & Leppänen, J. M. (2018). Infants' attention bias to faces
1706 as an early marker of social development. *Developmental Science*, 9:e12687–14.
- 1707 Pereira, E. J., Birmingham, E., & Ristic, J. (2019). The eyes do not have it after all?
1708 Attention is not automatically biased towards faces and eyes. *Psychological Research*.
- 1709 Peterson, M. F. & Eckstein, M. P. (2012). Looking just below the eyes is optimal across
1710 face recognition tasks. *Proceedings of the National Academy of Sciences*, pages
1711 E3314–E3323.
- 1712 Peterson, M. F. & Eckstein, M. P. (2013). Individual Differences in Eye Movements
1713 During Face Identification Reflect Observer-Specific Optimal Points of Fixation.
1714 *Psychological Science*, 24(7):1216–1225.
- 1715 Peterson, M. F., Lin, J., Zaun, I., & Kanwisher, N. (2016). Individual differences in
1716 face-looking behavior generalize from the lab to the world. *Journal of Vision*, 16(7):
1717 12–18.
- 1718 Pfeiffer, U. J., Vogeley, K., & Schilbach, L. (2013). From gaze cueing to dual
1719 eye-tracking: Novel approaches to investigate the neural correlates of gaze in social
1720 interaction. *Neuroscience & Biobehavioral Reviews*, 37:2516–2528.
- 1721 Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism Modulates Infants'
1722 Selective Attention to the Mouth of a Talking Face. *Psychological Science*, 26(4):
1723 490–498.
- 1724 Raidt, S., Bailly, G., & Elisei, F. (2007). Analyzing and modeling gaze during
1725 face-to-face interaction. *7th International Conference on Intelligent Virtual Agents*
1726 (IVA).
- 1727 Richardson, D. C., Street, C. N. H., Tan, J. Y. M., Kirkham, N. Z., Hoover, M. A., &
1728 Cavanaugh, A. G. (2012). Joint perception: gaze and social context. *Frontiers in*
1729 *Human Neuroscience*, pages 1–8.

- 1730 Risko, E. F. & Kingstone, A. (2011). Eyes wide shut: implied social presence, eye
1731 tracking and attention. *Attention, Perception & Psychophysics*, 73(2):291–296.
- 1732 Risko, E. F., Laidlaw, K. E. W., Freeth, M., Foulsham, T., & Kingstone, A. (2012).
1733 Social attention with real versus reel stimuli: toward an empirical approach to
1734 concerns about ecological validity. *Frontiers in Human Neuroscience*, 6(1):143.
- 1735 Risko, E. F., Richardson, D. C., & Kingstone, A. (2016). Breaking the Fourth Wall of
1736 Cognitive Science. *Current Directions in Psychological Science*, 25(1):70–74.
- 1737 Ristic, J., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how
1738 you look at it. *Psychonomic Bulletin & Review*, 9(3):507–513.
- 1739 Rogers, S. L., Speelman, C. P., Guidetti, O., & Longmuir, M. (2018). Using dual eye
1740 tracking to uncover personal gaze patterns during social interaction. *Scientific*
1741 *Reports*, pages 1–9.
- 1742 Rossano, F. (2013). Gaze in Conversation. In *The handbook of conversation analysis*,
1743 pages 1–22.
- 1744 Royer, J., Blais, C., Charbonneau, I., Déry, K., Tardif, J., Duchaine, B., Gosselin, F., &
1745 Fiset, D. (2018). Greater reliance on the eye region predicts better face recognition
1746 ability. *Cognition*, 181:12–20.
- 1747 Ruhland, K., Peters, C. E., Andrist, S., Badler, J. B., Badler, N. I., Gleicher, M.,
1748 Mutlu, B., & McDonnell, R. (2015). A Review of Eye Gaze in Virtual Agents, Social
1749 Robotics and HCI: Behaviour Generation, User Interaction and Perception.
1750 *Computer Graphics Forum*, 34(6):299–326.
- 1751 Schwartz, J.-L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better:
1752 evidence for early audio-visual interactions in speech identification. *Cognition*, 93(2):
1753 B69–B78.
- 1754 Scott, H., Batten, J. P., & Kuhn, G. (2018). Why are you looking at me? It's because

- 1755 I'm talking, but mostly because I'm staring or not doing much. *Attention, Perception*
1756 *& Psychophysics*, pages 1–10.
- 1757 Sekiguchi, T. (2011). Individual differences in face memory and eye fixation patterns
1758 during face learning. *Acta Psychologica*, 137(1):1–9.
- 1759 Senju, A. & Johnson, M. H. (2009). Atypical eye contact in autism: models,
1760 mechanisms and development. *Neuroscience & Biobehavioral Reviews*, 33:1204–1214.
- 1761 Senju, A., Verneti, A., Kikuchi, Y., Akechi, H., & Hasegawa, T. (2013). Cultural
1762 Modulation of Face and Gaze Scanning in Young Children. *PLOS One*, 8(8):e74017.
- 1763 Shen, J. & Itti, L. (2012). Top-down influences on visual attention during listening are
1764 modulated by observer sex. *Vision Research*, 65(C):62–76.
- 1765 Shepherd. (2010). Following gaze: gaze-following behavior as a window into social
1766 cognition. *Frontiers in Integrative Neuroscience*.
- 1767 Skantze, G., Hjalmarsson, A., & Oertel, C. (2014). Turn-taking, feedback and joint
1768 attention in situated human-robot interaction. *Speech Communication*, 65(C):50–66.
- 1769 Smilek, D., Birmingham, E., Cameron, D., Bischof, W., & Kingstone, A. (2006).
1770 Cognitive Ethology and exploring attention in real-world scenes. *Brain Research*,
1771 1080:101–119.
- 1772 Smith, L. B. & Thelen, E. (2003). Development as a dynamic system. *Trends in*
1773 *Cognitive Sciences*, 7(8):343–348.
- 1774 Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and
1775 Decoding Facial Expressions. *Psychological Science*, 16(3):184–189.
- 1776 Stivers, T. & Rossano, F. (2010). Mobilizing Response. *Research on Language & Social*
1777 *Interaction*, 43(1):3–31.
- 1778 Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T.,
1779 Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009).

- 1780 Universals and cultural variation in turn-taking in conversation. *Proceedings of the*
1781 *National Academy of Sciences*, 106(26):10587–10592.
- 1782 Stukenbrock, A. (2018). Mobile dual eye-tracking in face-to-face interaction. In Brône,
1783 G. & Oben, B., editors, *Eye-tracking in interaction*. John Benjamins Publishing
1784 Company, Amsterdam / Philadelphia.
- 1785 Sumbly, W. H. & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in
1786 Noise. *The Journal of the Acoustical Society of America*, 26(2):212–215.
- 1787 Symons, L. A., Hains, S. M. J., & Muir, D. W. (1998). Look at me: five-month-old
1788 infants' sensitivity to very small deviations in eye-gaze during social interactions.
1789 *Infant Behavior and Development*, 21(3):531–536.
- 1790 Symons, L. A., Lee, K., Cedrone, C. C., & Nishimura, M. (2004). What are you looking
1791 at? Acuity for triadic eye gaze. *J Gen Psychology*, 131(4):451–469.
- 1792 Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in
1793 natural vision: Reinterpreting salience. *Journal of Vision*, 11(5):1–23.
- 1794 Tenenbaum, E. J., Shah, R. J., Sobel, D. M., Malle, B. F., & Morgan, J. L. (2013).
1795 Increased Focus on the Mouth Among Infants in the First Year of Life: A
1796 Longitudinal Eye-Tracking Study. *Infancy*, 18(4):534–553.
- 1797 Toet, A. & Levi, D. M. (1992). The Two-Dimensional Shape of Spatial Interaction
1798 Zones in the Parafovea. *Vision Research*, 32(7):1349–1357.
- 1799 Treisman, A. M. & Gelade, G. (1980). A feature-integration theory of attention.
1800 *Cognitive Psychology*, 12:97–136.
- 1801 Triesch, J., Teuscher, C., Deák, G. O., & Carlson, E. (2006). Gaze following: why (not)
1802 learn it. *Developmental Science*, 9(2):125–157.
- 1803 Vabalas, A. & Freeth, M. (2016). Brief Report: Patterns of Eye Movements in Face to
1804 Face Conversation are Associated with Autistic Traits: Evidence from a Student
1805 Sample. *Journal of Autism and Developmental Disorders*, 46(1):305–314.

- Vaidya, A. R., Jin, C., & Fellows, L. K. (2014). Eye spy: The predictive value of fixation patterns in detecting subtle and extreme emotions from faces. *Cognition*, 133(2):443–456.
- Vatikiotis-Bateson, E., Eigsti, I.-M., Yano, S., & Munhall, K. G. (1998). Eye movement of perceivers during audiovisual speech perception. *Perception & Psychophysics*, 60(6):926–940.
- Võ, M. L. H., Smith, T. J., Mital, P. K., & Henderson, J. M. (2012). Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *Journal of Vision*, 12(13):1–14.
- von Cranach, M. & Ellgring, J. H. (1973). Problems in the Recognition of Gaze Direction. *Social communication and movement: Studies of interaction and expression in man and chimpanzee*, 4:419.
- Walker-Smith, G. J., Gale, A. G., & Findlay, J. M. (1977). Eye movement strategies involved in face perception. *Perception*, 6:313–326.
- Wheeler, A., Anzures, G., Quinn, P. C., Pascalis, O., Omrin, D. S., & Lee, K. (2011). Caucasian Infants Scan Own- and Other-Race Faces Differently. *PLOS One*, 6(4): e18621.
- Wieser, M. J., Pauli, P., Alpers, G. W., & Mühlberger, A. (2009). Is eye to eye contact really threatening and avoided in social anxiety?—An eye-tracking and psychophysiology study. *Journal of anxiety disorders*, 23:93–103.
- Wilcox, T., Stubbs, J. A., Wheeler, L., & Alexander, G. M. (2013). Infants’ scanning of dynamic faces during the first year. *Infant Behavior and Development*, 36(4):513–516.
- Willemse, C. & Wykowska, A. (2019). In natural interaction with embodied robots, we prefer it when they follow our gaze: a gaze-contingent mobile eyetracking study. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1771):1–7.

- 1831 Williams, C. C. & Henderson, J. M. (2007). The face inversion effect is not a
1832 consequence of aberrant eye movements. *Memory & Cognition*, 35(8):1977–1985.
- 1833 Wilson, A. H., Alsius, A., Paré, M., & Munhall, K. G. (2016). Spatial Frequency
1834 Requirements and Gaze Strategy in Visual-Only and Audiovisual Speech Perception.
1835 *Journal of Speech, Language and Hearing Research*, 59(4):601–615.
- 1836 Wollaston, W. H. (1824). On the Apparent Direction of Eyes in a Portrait.
1837 *Philosophical Transactions of the Royal Society of London*, 114:247–256.
- 1838 Xiao, W. S., Xiao, N. G., Quinn, P. C., Anzures, G., & Lee, K. (2013). Development of
1839 face scanning for own- and other-race faces in infancy. *International Journal of*
1840 *Behavioral Development*, 37(2):100–105.
- 1841 Yarbus, A. L. (1967). *Eye movements and vision*. Plenum Press, New York.
- 1842 Yehia, H., Rubin, P., & Vatikiotis-Bateson, E. (1998). Quantitative association of
1843 vocal-tract and facial behavior. *Speech Communication*, 26:23–43.
- 1844 Yi, A., Wong, W., & Eizenman, M. (2013). Gaze Patterns and Audiovisual Speech
1845 Enhancement. *Journal of Speech, Language and Hearing Research*, 56(2):471–480.
- 1846 Yokoyama, T. & Takeda, Y. (2019). Gaze Cuing Effects in Peripheral Vision. *Frontiers*
1847 *in Psychology*, 10:1–9.
- 1848 Yu, C. & Smith, L. B. (2013). Joint Attention without Gaze Following: Human Infants
1849 and Their Parents Coordinate Visual Attention to Objects through Eye-Hand
1850 Coordination. *PLOS One*, 8(11):1–10.
- 1851 Yu, C. & Smith, L. B. (2016). The Social Origins of Sustained Attention in
1852 One-Year-Old Human Infants. *Current Biology*, 26:1235–1240.

Table 1

Important definitions.

Concept	Definition
Stimulus	Content presented to an observer in an experiment, e.g. image or video
Observer	Person observing a set of stimuli
Participant	Person engaged in, or believing to be engaged in or part of, a social situation
Interactor	An agent involved in interaction
Interaction	Reciprocal action or influence between two or more interactors

Figure 1. Example face-to-face interactions in which gaze plays an important role. (a) Two musicians learning a song for guitar and mandolin together. Notice how the left person can infer the spatial locus of the right person's visual attention from his gaze direction. (b) A meeting among co-workers. Gaze direction is often an important regulator of the flow of conversation in such meetings as a key resource for turn allocation. (c) An infant engaged in play with her parent and a third person. Following a parent's gaze direction is thought to be an important learning mechanism. Picture *a* courtesy of Gijs Holleman, pictures *b* & *c* courtesy of Ivar Pel and the YOUth study at Utrecht University.

Figure 2. Overview of the framework. (a) Schematic drawing of the relevant aspects of face-to-face interactions considered as sub-states of a dynamic system of interaction. These sub-states are organised as those belonging to the individual interactors (task, drive and affect, idiosyncratic scan patterns), the content of the interaction itself (gaze direction, facial expression, communicative reference, speaker status) and the broader context in which the interaction takes place (interpersonal context, cultural context). (b) Overview of the timescale at which the sub-states of the dynamic system that describes face-to-face interaction are expected to vary, from microgenetic, to ontogenetic and phylogenetic timescales.