2 Memory for Moving Faces: The Interplay of Two Recognition Systems

Alice O'Toole and Dana Roark

The human face is a captivating stimulus, even when it is stationary. In motion, however, the face comes to life and offers us a myriad of information about the intent and personality of its owner. Through facial movements such as expressions we can gauge a person’s current state of mind. By perceiving the movements of the mouth as a friend speaks, a conversation becomes more intelligible in a noisy environment. Through the rigid movement of the head and the direction of eye gaze, we can follow another person’s focus of attention in a crowded room. The amount and diversity of social information that can be conveyed by a face ensures its place as a central focal object in any scene. Beyond the rich communication signals that we perceive in facial expressions, head orientation, eye gaze, and facial speech motions, it is also pertinent to ask whether the movements of a face help us to remember a person. The answer to this question can potentially advance our understanding of how the complex tasks we perform with faces, including those having to do with social interaction and memory, coexist in a neural processing network. It can also shed light on the computational processes we use to extract recognition cues from the steady stream of facial movements that carry social meanings.

Several years ago we proposed a psychological and neural framework for understanding the contribution of motion to face recognition (O’Toole, Roark, & Abdi, 2002; Roark, Barrett, Spence, Abdi, & O’Toole, 2003). At the time, there were surprisingly few studies on the role of motion in face recognition and so the data we used to support this framework were sparse and in some cases tentative. In recent years, however, interest in the perception and neural processing of dynamic human faces has expanded, bringing with it advances in our understanding of the role of motion in face recognition.

In this chapter we revisit and revise the psychological and neural framework we proposed previously (O’Toole et al., 2002). We begin with an overview of the original model and a sketch of how we arrived at the main hypotheses. Readers who are familiar with this model and its neural underpinnings can skip to the next section of this chapter. In that section we update this past perspective with new findings that
address the current status of psychological and neural hypotheses concerning the role of motion in face recognition. Finally, we discuss open questions that continue to challenge our understanding of how the dynamic information in faces affects the accuracy and robustness of their recognition.

A Psychological and Neural Framework for Recognizing Moving Faces: Past Perspectives

Before 2000, face recognition researchers rarely considered the role of motion in recognition. It is fair to say that with little or no data on the topic, many of us simply assumed that motion would benefit face recognition in a whole variety of ways. After all, a moving face is more engaging; it conveys contextual and emotional information that can be associated with identity; and it reveals the structure of the face more accurately than a static image. It is difficult to recall now, in the “You-tube” era, that a few short years ago the primary reason psychologists avoided the use of dynamic faces in perception and memory studies was the limited availability of digital video on the computers used in most psychology labs. This problem, combined with the lack of controlled video databases of faces, hindered research efforts on the perception and recognition of dynamic faces. As computing power increased and tools for manipulating digital video became standard issue on most computers, this state of affairs changed quickly. Research and interest in the topic burgeoned, with new data to consider and integrate appearing at an impressive rate.

The problem of stimulus availability was addressed partially in 2000 with a database designed to test computational face recognition algorithms competing in the Face Recognition Vendor Test 2002 (Phillips, Grother, Micheals, Blackburn, Tabassi, & Bone, 2003). This database consisted of a large collection of static and dynamic images of people taken over multiple sessions (O’Toole, Harms, Snow, Hurst, Pappas, & Abdi, 2005). In our first experiments using stimuli from this database, we tested whether face recognition accuracy could be improved when people learned a face from a high-resolution dynamic video rather than from a series of static images that approximated the amount of face information available from the video. Two prior studies returned a split decision on this question, with one showing an advantage for motion (Pike, Kemp, Towell, & Phillips, 1997) and a second showing no difference (Christie & Bruce, 1998). In our lab, we conducted several experiments with faces turning rigidly, expressing, and speaking, and found no advantage for motion information either at learning or at test. In several replications, motion neither benefited nor hindered recognition. It was as if the captivating motions of the face were completely irrelevant for recognition.

These studies led us to set aside the preconceived bias that motion, as an ecologically valid and visually compelling stimulus, must necessarily make a useful contribution to face recognition. Although this made little sense psychologically, it seemed to
fit coherently into the consensus that was beginning to emerge about the way moving versus static faces and bodies were processed neurally.

**Evidence from Neuroscience**

A framework for understanding the neural organization of processing faces was suggested by Haxby, Hoffman, and Gobbini (2000) based on functional neuroimaging and neurophysiological studies. They proposed a distributed neural model in which the invariant information in faces, useful for identification, was processed separately from the changeable, motion-based information in faces that is useful for social communication. The invariant information includes the features and configural structure of a face, whereas the changeable information refers to facial expression, facial speech, eye gaze, and head movements. Haxby et al. (2000) proposed the lateral fusiform gyrus for processing the invariant information and the posterior superior temporal sulcus (pSTS) for processing changeable information in faces. The pSTS is also implicated in processing body motion.

The distributed network model pointed to the possibility that the limited effect of motion on face recognition might be a consequence of the neural organization of the brain areas responsible for processing faces. The distributed idea supports the main tenets of the role of the dorsal and ventral processing streams in vision with, at least, a partial dissociation of the low-resolution motion information from the higher-resolution color and form information. From the Haxby et al. model, it seemed possible, even likely, that the brain areas responsible for processing face identity in the inferior temporal cortex might not have much (direct) access to the areas of the brain that process facial motions for social interaction.

A second insight from the distributed network model was an appreciation of the fact that most facial motions function primarily as social communication signals (Allison, Puce, & MacCarthy, 2000). As such, they are likely to be processed preeminently for this purpose. Although facial motions might also contain some unique or identity-specific information about individual faces that can support recognition, this information seems secondary to the more important social communication information conveyed by motion.

**Fitting the Psychological Evidence into the Neural Framework**

A distributed neural network offered a framework for organizing the (albeit) limited data on recognition of people and faces from dynamic video. O’Toole et al. (2002) proposed two nonmutually exclusive ways in which motion might benefit recognition. The supplemental information hypothesis posits a representation of the characteristic facial motions or gestures of individual faces (“dynamic identity signatures”) in addition to the invariant structure of faces. We assumed that when both static and dynamic identity information is present, people rely on the static information
because it provides a more reliable marker of facial identity. The representation enhancement hypothesis posits that motion contributes to recognition by facilitating the perception of the three-dimensional structure of the face via standard visual structure-from-motion processes. Implicit in this hypothesis is the assumption that the benefit of motion transcends the benefit of seeing the multiple views and images embedded in a dynamic motion sequence.

At the time, there were two lines of evidence for the supplemental information hypothesis. The first came from clever experiments that pitted the shape of a face (from a three-dimensional laser scan of a head model), which could be manipulated with morphing, against characteristic facial motions projected onto heads that varied in shape (Hill & Johnston, 2001; Knappmeyer, Thornton, & Bülthoff, 2001; Knappmeyer, Thornton, & Bülthoff, 2003). These studies provided a prerequisite demonstration that the facial motion in dynamic identity signatures can bias an identification decision. The second line of evidence came from studies showing that recognition of famous faces was better with dynamic than with static presentations. This was demonstrated most compellingly when image quality was degraded (Knight & Johnston, 1997; Lander, Bruce, & Hill, 2001; Lander & Bruce, 2000). O’Toole et al. (2002) concluded that the role of motion in face identification depends on both image quality and face familiarity.

For the representation enhancement hypothesis, the empirical support was less compelling. The idea that facial motion could be perceptually useful for forming better representations of faces is consistent with a role for structure-from-motion processes in learning new faces. It seemed reasonable to hypothesize that motion could contribute, at least potentially, to the quality of the three-dimensional information perceptually available in faces, even if evidence for this was not entirely unambiguous.

To summarize, combining human face recognition data with the distributed network model, we proposed that processing the visual information from faces for recognition involves the interplay of two systems (O’Toole et al., 2002). The first system is equivalent to the one proposed by Haxby et al. (2000) in the ventral temporal cortex. It includes the lateral fusiform gyrus and associated structures (e.g., occipital face area, OFA) and processes the invariant information in faces. The second system processes facial movements and is the part of the distributed network useful for representing the changeable aspects of faces in the pSTS. O’Toole et al. (2002) amended the distributed model to specify the inclusion of both social communication signals and person-specific dynamic identity signatures in facial movements. We suggested that two caveats apply to the effective use of this secondary system. First, the face must be familiar (i.e., characteristic motions of the individual must be known) and second, the viewing conditions must be poor (i.e., otherwise the more reliable pictorial code will dominate recognition and the motion system will not be needed).
The familiarity caveat is relevant for understanding the well-established differences in processing capabilities for familiar and unfamiliar faces (Hancock, Bruce, & Burton, 2000). When we know a person well, a brief glance from a distance even under poor illumination, is often all that is required for recognition. For unfamiliar faces, changes in viewpoint, illumination, and resolution between learning and test all produce reliable decreases in recognition performance (see O’Toole, Jiang, Roark, & Abdi, 2006, for a review). We suggested that this secondary system might underlie the highly robust recognition performance that humans show in suboptimal viewing conditions for the faces they know best.

A second, more tentative amendment we made to the distributed model was the addition of structure-from-motion analyses that could proceed through the dorsal stream to the middle temporal (MT) and then back to the inferior temporal (IT) cortex as “motionless form.” We proposed possible neural mechanisms for this process and will update these presently.

In the next section we provide an updated account of the evidence for the supplemental information and the representation enhancement hypotheses. We also look at some studies that suggest a role for motion in recognition but that do not fit easily within the framework we outlined originally (O’Toole et al., 2002; Roark et al., 2003).

**“Backup Identity System” and Supplemental Motion Information: An Update**

Three lines of evidence now support the supplemental information hypothesis. The first adds to previous psychological studies of face recognition and further supports the beneficial effects of dynamic identity signatures for recognition. The second provides new evidence from studies indicating a benefit of motion-based identity codes in the efficiency of visually based “speech-reading” tasks. The third line of evidence comes from studies of prosopagnosics’ perceptions of moving faces.

**Psychological Studies of Dynamic Identity Signatures**

Lander and Chuang (2005) found a supportive role for motion when recognizing people in challenging viewing conditions. They replicated the results of earlier studies and expanded their inquiry to examine the types of motions needed to show the benefit, evaluating both rigid and nonrigid motions. Lander and Chuang found a recognition advantage for nonrigid motions (expressions and speech), but not for rigid motions (head nodding and shaking). Moreover, they found a motion advantage only when the facial movements were “distinctive.” They conclude that some familiar faces have characteristic motions that can help in identification by incorporating supplemental motion-based information about the face.

In a follow-up study, Lander, Chuang, and Wickam (2006) demonstrated human sensitivity to the “naturalness” of the dynamic identity signatures. Their results
showed that recognition of personally familiar faces was significantly better when the faces were shown smiling naturally than when they were shown smiling in an artificial way. Artificial “smile videos” were created by morphing from a neutral expression image to a smiling face image. Speeding up the motion of the natural smile impaired identification but did not impair recognition from the morphed artificial smile. Lander et al. conclude that characteristic movements of familiar faces are stored in memory. The study offers further support for a reasonably precise spatio-temporal code of characteristic face motions.

**Evidence from Facial Speech-Reading**

Several recent studies demonstrate that the supplemental information provided by speaker-specific facial speech movements can improve the accuracy of visually based “speech-reading.” Kaufmann and Schweinberger (2005), for example, implemented a speeded classification task in which participants were asked to distinguish between two vowel articulations across variations in speaker identity. Changes in facial identity slowed the participants’ ability to classify speech sounds from dynamic stimuli but did not affect classification performance from single static or multiple-static stimuli. Thus, individual differences in dynamic speech patterns can modulate facial speech processing. Kaufmann and Schweinberger (2005) conclude that the systems for processing facial identity and speech-reading are likely to overlap.

In a related study, Lander, Hill, Kamachi, and Vatikiotis-Bateson (2007) found that speaker-specific mannerisms enhanced speech-reading ability. Participants matched silent video clips of faces to audio voice recordings using unfamiliar faces and voices as stimuli. In one experiment, the prosody of speech segments was varied in clips that were otherwise identical in content (e.g., participants heard the statement “I’m going to the library in the city” or the question “I’m going to the library in the city?”). Participants were less accurate at matching the face and the voice when the prosody of the audio clip did not match the video clip, or vice versa. Of note, Lander and her colleagues also showed that participants were most successful matching faces to voices when speech cues came in the form of naturalistic, conversational speech. Even relatively minor variations in speaker mannerisms (e.g., unnatural enunciation, hurried speech) inhibited the participants’ ability to correctly match faces with voices.

Familiarity with a speaker also seems to play a role in speech-reading ability. Lander and Davies (2008) found that as participants’ experience with a speaker increases through exposure to video clips of the speakers reciting letters and telling stories, so does speech-reading accuracy. The mediating role of familiarity in the use of motion information from a face is consistent with the proposals we made previously (O’Toole et al., 2002) for a secondary face identity system in the pSTS. It is also consistent with the suggestion that the face identity code in this dorsal backup
system is more robust than the representation in the ventral stream. Lander and Davies (2008) conclude that familiarity with a person’s idiosyncratic speaking gestures can be coupled with speech-specific articulation movements to facilitate speech-reading.

**Evidence for the Supplemental Motion Backup System from Prosopagnosia**

The possible existence of a recognition backup system that processes dynamic identity signatures in the pSTS makes an intriguing prediction about face recognition skills in prosopagnosics. Specifically, it suggests that face recognition could be partially spared in prosopagnosics when a face is presented in motion. The rationale behind this prediction is based on the anatomical separation of the ventral temporal face areas and the pSTS. Damage to the part of the system that processes invariant features of faces would not necessarily affect the areas in the pSTS that process dynamic identity signatures. Two studies have addressed this question with prosopagnosics of different kinds. In the first study, Lander, Humphreys, and Bruce (2004) tested a stroke patient who suffered a relatively broad pattern of bilateral lesion damage throughout ventral-occipital regions, including the lingual and fusiform gyri. “HJA,” who is “profoundly prosopagnosic” (Lander et al., 2004), suffers also from a range of other nonface-specific neuropsychological deficits (see Humphreys and Riddoch, 1987 for a review) that include object agnosias, reading difficulties, and achromatopsia. Despite these widespread visual perception difficulties, HJA is able to perform a number of visual tasks involving face and body motion. For example, he is able to categorize lip movements accurately (Campbell, 1992). He also reports relying on voice and gait information for recognizing people (Lander et al., 2004).

For present purposes, Lander et al. (2004) tested HJA on several tasks of face recognition with moving faces. HJA was significantly better at matching the identity of moving faces than matching the identity of static faces. This pattern of results was opposite to that found for control subjects. However, HJA was not able to use face motion to explicitly recognize faces and was no better at learning names for moving faces than for control faces. Thus, although the study suggests that HJA is able to make use of motion information in ways not easy for control subjects, it does not offer strong evidence for a secondary identity backup system. Given the extensive nature of the lesion damage in HJA, however, the result is not inconsistent with the hypothesis of the backup system.

The prediction that motion-based face recognition could be spared in prosopagnosics was examined further by Steede, Tree, and Hole (2007). They tested a developmental prosopagnosic (“CS”) who has a purer face recognition deficit than HJA. CS has no difficulties with visual and object processing, but has profound recognition difficulties for both familiar and unfamiliar faces. Steede et al. tested CS with
dynamic faces and found that he was able to discriminate between dynamic identities. He was also able to learn the names assigned to individuals based only on their idiosyncratic facial movements. This learning reached performance levels comparable to those of control subjects. These results support the posited dissociation between the mechanisms involved in recognizing faces from static versus dynamic information. A cautionary note on concluding this too firmly, based on these results, is that CS is a developmental (congenital) rather than an acquired prosopagnosic. Thus it is possible that some aspects of his face recognition system have been organized developmentally to compensate for his difficulties with static face recognition. More work of this sort is needed to test patients with relatively pure versions of acquired prosopagnosia.

In summary, these three lines of evidence combined offer solid support for the supplemental information hypothesis.

**Representation Enhancement: An Update**

The clearest way to demonstrate a role for the representation enhancement hypothesis is to show that faces learned when they are in motion can be recognized more accurately than faces learned from a static image or set of static images that equate the “face” information (e.g., from extra views). If motion promotes a more accurate representation of the three-dimensional structure of a face, then learning a face from a moving stimulus should benefit later recognition. This advantage assumes that face representations incorporate information about the three-dimensional face structure that is ultimately useful for recognition. The benefit of motion in this case should be clear when testing with either a static or a moving image of the face—i.e., the benefit is a consequence of a better, more flexible face representation. At first glance, it seems reasonable to assume that the face representation we are talking about is in the ventral stream. In other words, this representation encodes the invariant feature-based aspects of faces rather than the idiosyncratic dynamic identity signatures. Thus it seems likely that it would be part of the system that represents static facial structure. We will qualify and question this assumption shortly.

For present purposes, to date there is still quite limited evidence to support the beneficial use of structure-from-motion analyses for face recognition. This lack of support is undoubtedly related to findings from the behavioral and neural literatures that suggest view-based rather than object-centered representations of faces, especially for unfamiliar faces. In particular, several functional neuroimaging studies have examined this question using the functional magnetic resonance adaptation (fMR-A) paradigm (cf. Grill-Spector, Kushnir, Hendler, Edelman, Itzchak, & Malach, 1999). fMR-A makes use of the ubiquitous finding that brain response decreases with repeated presentations of the “same” stimulus. The fusiform face area (FFA;
Kanwisher, McDermott, & Chun, 1997) and other face-selective regions in the ventral temporal cortex show adaptation for face identity, but a release from adaptation when the viewpoint of a face is altered (e.g., Andrews & Ewbank, 2004; Pourtois, Schwartz, Seghier, Lazeyras, & Vuilleumier, 2005). This suggests a view-based neural representation of unfamiliar faces in the ventral temporal cortex. [Although see Jiang, Blanz, and O’Toole (2009) for evidence of three-dimensional information contributing to codes for familiar faces.]

In the psychological literature, Lander and Bruce (2003) further investigated the role of motion in learning new faces. They show first that learning a face from either rigid (head nodding or shaking) or nonrigid (talking, expressions) motion produced better recognition than learning a face from only a single static image. However, the learning advantage they found for rigid motion could be accounted for by the different angles of view that the subjects experienced during the video. For nonrigid motions, the advantage could not be explained by the multiple sequences experienced in the video. Lander and Bruce suggest that this advantage may be due to the increased social attention elicited by nonrigid facial movements. This is because nonrigid facial motions (talking and expressing) may be more socially engaging than rigid facial ones (nodding and shaking). Nevertheless, the study opens up the possibility that structure-from-motion may benefit face learning, at least for some nonrigid facial motions. Before firmly concluding this, however, additional controls over the potential differences in the attention appeal of rigid and nonrigid motions are needed to eliminate this factor as an explanation of the results.

Bonner, Burton, and Bruce (2003) examined the role of motion and familiarization in learning new faces. Previous work by Ellis, Shepherd, and Davies (1979) showed that internal features tend to dominate when matching familiar faces, whereas external features are relied upon more for unfamiliar faces. Bonner et al. examined the time course of the shift from external to internal features over the course of several days. Also, based on the hypothesis that motion is relatively more important for recognizing familiar faces, they looked at the differences in face learning as a function of whether the faces were learned from static images or a video. The videos they used featured slow rigid rotations of the head, whereas the static presentations showed extracted still images that covered a range of the poses seen in the video. They found improvement over the course of 3 days in matching the internal features of the faces, up to the level achieved with the external features for the initial match period. Thus the internal feature matches continued to improve with familiarity but the external matches remained constant. Notably, they found no role for motion in promoting learning of the faces. This is consistent with a minimal contribution of motion for perceptual enhancement.

Before leaving this review of the representation enhancement hypothesis for learning new faces, it is worth noting that the results of studies with adults may not
generalize to learning faces in infancy. Otsuka, Konishi, Kanazawa, Yamaguchi, Abdi, and O’Toole (2009) compared 3–4-month-olds’ recognition of previously unfamiliar faces learned from moving or static displays. The videos used in the study portrayed natural dynamic facial expressions. Infants viewing the moving condition recognized faces more efficiently than infants viewing the static condition, requiring shorter familiarization times even when different images of a face were used in the familiarization and test phases. Furthermore, the presentation of multiple static images of a face could not account for the motion benefit. Facial motion, therefore, promotes young infants’ ability to learn previously unfamiliar faces. In combination with the literature on an adult’s processing of moving faces, the results of Otsuka et al. suggest a distinction between developmental and postdevelopmental learning in structure-from-motion contributions to building representations for new faces.

**Does Motion Contribute to Ventral Face Representations?**

From a broad-brush point of view, the bulk of the literature on visual neuroscience points to anatomically and functionally distinct pathways for processing high-resolution color or form information and for processing motion-based form. In our previous review (O’Toole et al., 2002), we discussed some speculative neural support for the possibility that motion information could contribute to face representations in the inferotemporal cortex. These neural links are obviously necessary if structure-from-motion processes are to enhance the quality of face representations in the traditional face-selective areas of the ventral temporal (VT) cortex. In O’Toole et al. (2002), we suggested the following data in support of motion-based contributions to the ventral cortex face representation. First, we noted that neurons in the primate IT, which are sensitive to particular forms, respond invariantly to form even when it is specified by pure motion-induced contrasts (Sary, Vogels, & Orban, 1993). Second, lesion studies have indicated that form discrimination mechanisms in the IT can make use of input from the motion-processing system (Britten, Newsome, & Saunders, 1992). Third, both the neurophysiological (Sary et al., 1993) and lesion studies (Britten et al., 1992) suggest known connections from the MT to the IT via V4 (Maunsell and Van Essen, 1983; Ungerleider & Desimone, 1986) as a plausible basis for their findings. We also noted in O’Toole et al. (2002) that psychological demonstrations of the usefulness of structure-from-motion for face recognition have been tentative and so strictly speaking there is no psychologically compelling reason to establish a mechanism for the process.

At present, the neural possibilities for contact between dorsal and ventral representations remain well established, but there are not enough results, at present, to require an immediate exploration of these links. However, there have been interesting developments in understanding the more general problem of recognizing people in
motion, particularly from point-light walker displays (Grossman & Blake, 2003; Giese & Poggio, 2003). These studies suggest a role for both ventral and dorsal pathways in the task. Giese and Poggio (2003) caution, however, that there are still open questions and unresolved paradoxes in the data currently available.

For present purposes, we have wondered recently if one problem in making sense of the data concerns the assumption we made originally that structure-from-motion must somehow feed back information to the ventral face representations (O'Toole et al., 2002). As noted, this assumption was based on the rationale that structure is an invariant property of faces. Based on the more recent findings in the perception of moving bodies, we tentatively suggest that some aspects of facial structure might also become part of the pSTS representation of identity. This representation of face structure would be at least partially independent of the specific facial motions used to establish it, but would nonetheless need a moving face to activate it. In other words, we hypothesize that the dorsal stream pSTS identity representation might include, not only idiosyncratic facial gestures, but also a rough representation of the facial shape independent of these idiosyncratic motions.

Evidence for this hypothesis can be found in two studies that suggest that the beneficial contribution from the motion system for learning new faces may be limited to tasks that include dynamic information both at learning and at test (Lander & Davies, 2007; Roark, O'Toole, & Abdi, 2006). First, from work in our lab, Roark et al. (2006) familiarized participants with previously unknown people using surveillancelike, whole-body images (gait videos) and then tested recognition using either close-up videos of the faces talking and expressing or a single, static image. The results showed that recognition of the people learned from the gait videos was more accurate when the test images were dynamic than when they were static. Similarly, when we reversed the stimuli so that participants learned faces either from the close-up still images or the close-up video faces and were tested using the gait videos, we found that recognition from the gait videos was more accurate when participants had learned the faces from the dynamic videos. Taken together, this pattern of results indicates that it is easier to obtain a motion advantage in a recognition task with unfamiliar faces when a moving image is present both at learning and at test. Furthermore, it is indicative of a system in which “motion-motion” matches across the learning and test trials are more useful for memory than either “static-motion” or “motion-static” mismatches across the learning and test trials.

Lander and Davies (2007) found a strikingly similar result. In their study, participants learned faces from either a moving sequence of the person speaking and smiling or a static freeze frame selected from the video sequence. At test, participants viewed either moving or static images. This was a different moving sequence or static image than that presented during the learning phase. Lander and Davies found that there was an advantage for recognizing a face in motion only when participants had
learned the faces from dynamic images. This result adds further support to the idea that motion is most helpful when participants have access to it during both learning and test times.

It should be noted that the results of both Lander and Davies (2007) and Roark et al. (2006) indicate that it is not a prerequisite that identical motions be present at learning and at test in order to obtain the motion advantage; both studies included different motions across the learning and test trials. Rather, it seems sufficient merely to activate the motion system across the learn-test transfer to see gains in recognition accuracy. Returning to the representation enhancement hypothesis, the motion-motion benefit may reflect the efficiency of having to access only a single channel (i.e., the dorsal motion system) when bridging between two moving images. Conversely, when dynamic information is present only at learning but not at test (or only at test but not at learning), cross-access between the motion and static information streams is required for successful recognition.

This motion-motion advantage must be put into perspective, however, with work on the effect of moving “primes” on face perception. Thornton and Kourtzi (2002) implemented a face-matching task with unfamiliar faces in which participants briefly viewed either moving or static images of faces and then had to identify whether the prime image matched the identity of a static face presented immediately afterward. The participants’ responses were faster following the dynamic primes than following the static primes. Pilz, Thornton, and Bülthoff (2006) observed a similar advantage for moving primes, with the additional finding that the benefit of moving primes extends across prime-target viewpoint changes. Pilz et al. also reported that moving primes led to faster identity matching in a delayed visual search task. Neither of these studies, however, included dynamic stimuli during the match trials, so it is difficult to tie these results directly to those of Roark et al. (2006) and Lander and Davies (2007), where motion was most useful when it was available from both the learning and the test stimuli. Interpretation of the motion-motion match hypothesis in the context of priming studies is a topic that is clearly ripe for additional work.

In conclusion, there is a more general need for studies that can clarify the extent to which motion can act as carrier or conduit for dorsal representations of face identity that have both moving and stationary components.

Summary

The faces we encounter in daily life are nearly always in motion. These motions convey social information and also can carry information about the identity of a person in the form of dynamic identity signatures. There is solid evidence that these motions comprise part of the human neural code for identifying faces and that they can be used for recognition, especially when the viewing conditions are suboptimal and
when the people to be recognized are familiar. Intriguing questions remain about the sparing of these systems in classic cases of acquired prosopagnosia. Moreover, an understanding of dorsal face and body representations, established through experience with dynamic stimuli, might be important for computational models of face recognition aimed at robust performance across viewpoint changes (cf. Giese & Poggio, 2003 and this volume). There is little evidence for the representation enhancement hypothesis having a major role in face recognition. Again, this raises basic questions about the extent to which ventral and dorsal face representations are independent. There is still a great deal of work to be done in bridging the gap between the well-studied face representations in the ventral stream and less well-understood face representations in the dorsal stream.

Acknowledgments

Thanks are due to Technical Support Working Group/United States Department of Defense for funding A. O'Toole during the preparation of this chapter.

References


