CrossMark

What's in a "face file"? Feature binding with facial identity, emotion, and gaze direction

Daniel Fitousi[1]

Abstract A series of four experiments investigated the binding of facial (i.e., facial identity, emotion, and gaze direction) and non-facial (i.e., spatial location and response location) attributes. Evidence for the creation and retrieval of temporary memory face structures across perception and action has been adduced. These episodic structures—dubbed herein "face files"—consisted of both visuo–visuo and visuo–motor bindings. Feature binding was indicated by partial-repetition costs. That is repeating a combination of facial features or altering them altogether, led to faster responses than repeating or alternating only one of the features. Taken together, the results indicate that: (a) "face files" affect both action and perception mechanisms, (b) binding can take place with facial dimensions and is not restricted to low-level features (Hommel, Visual Cognition 5:183–216, 1998), and (c) the binding of facial and non-facial attributes is facilitated if the dimensions share common spatial or motor codes. The theoretical contributions of these results to "person construal" theories (Freeman, & Ambady, Psychological Science, 20(10), 1183–1188, 2011), as well as to face recognition models (Haxby, Hoffman, & Gobbini, Biological Psychiatry, 51(1), 59–67, 2000) are discussed.

Introduction

Faces are multidimensional visual stimuli that are capable of transmitting a great deal of information regarding a host of physical and social attributes. These attributes include (but are not limited to) the identity, sex, emotional expression, or gaze direction of the face. A fundamental question in the study of faces concerns the manner by which facial attributes are integrated into a unified phenomenal experience (Bruce, & Young, 1986; Haxby, Hoffman, & Gobbini, 2000; Fitousi, 2013; Fitousi, & Wenger, 2013; Young, & Yamane, 1992). Whereas extensive research has been conducted on the binding of simple features, such as color, shape, and spatial location (Treisman, 1996; Hommel, 1998), less effort has been invested in studying the binding of more complex dimensions, such as facial features. The present study sought to fill in this gap. It addressed the question of whether facial features (i.e., identity, emotion, and gaze direction), as well as non-facial features (i.e., spatial location and response location) are integrated in and across perception and action.

A novel hypothesis advanced in the present study postulates that people create, maintain, and retrieve transient memory structures of facial features. The quest for such "face files" in the present study has been inspired by the notions of "object files" (Kahneman, Triesman, & Gibbs, 1992), and "event files" (Hommel, 1998, 2004). These notions have been instrumental in the study of objects and attention (Gordon, & Irwin, 1996; Henderson, 1994; Hommel, 2005), but they have been rarely applied to faces. Harnessing concepts and methodologies from these literatures, the current investigation yielded consistent evidence for the existence of "face files"—transient memories of facial and non-facial features bindings. The results bear important implications for current theories of face and object recognition (Haxby et al., 2000), person construal (Freeman, & Ambady, 2011), and feature binding (Hommel, 1998; Treisman, 1996).

✉ Daniel Fitousi
danielfi@ariel.ac.il

[1] Department of Behavioral Sciences, Ariel University, Ariel, Israel

Springer

Faces and the binding problem

The primate brain codes the dimensions of perceptual objects in a distributed manner (Hubel, & Wiesel, 1977; Felleman, & Van Essen, 1991). In this process, elementary features, such as color, shape, and location, are represented in different feature maps in the visual cortex (Livingstone, & Hubel, 1987, 1988). A major challenge facing our perceptual systems is that of recombining the separate features into veridical representations of the viewed objects (Treisman, & Gelade, 1980). To accomplish this task, the primate brain should coordinate information from several independent and often temporally discordant sources. This formidable computational challenge has been often dubbed the binding problem (Singer, & Gray, 1995; Treisman, 1996; von der Malsburg, 1999). One notable example for the presence of a binding problem in perception is the finding of "illusory conjunctions" with color and shape (Triesman, & Schmidt, 1982).

Very much like objects, faces may pose binding problems to our visual system. This is because facial attributes are represented as separate codes in the brain. There is now ample evidence to suggest the involvement of a distributed network of brain areas that is responsible for the perception of specific facial dimensions (Haxby et al., 2002). For example, the processing of facial expression is governed by the amygdala (Breiter et al., 1996), whereas the processing of facial identity is held mainly in the fusiform area (FFA, Kanwisher, McDermott, Chun, 1997) and the superior temporal sulcus (STS, Haxby et al., 2000). Moreover, recordings in temporal cortex of nonhuman primates (Rolls, & Tovee, 1995; Sugase, Yamane, Ueno, & Kawano, 1999) support the existence of neuronal activity that is distributed across many neurons (Rogers, & McClelland, 2004; Spivey, & Dale, 2004).

Given the involvement of a highly distributed network in processing facial attributes, an acute binding problem may arise. Consider a situation in which you are presented with two facial identities with each conveying a different facial emotion (Jim happy, Dan sad). Your visual system must ensure that each identity is integrated with the correct emotion (Jim + happy, and Dan + sad). This is not a trivial task. Binding problems with faces may be even more difficult than with elementary low-level features. This is because faces, in addition to carrying invariant attributes (e.g., identity and gender), transmit a great deal of dynamic information, such as eye-gaze and emotional expressions. These attributes frequently change their physical appearance as well as their semantic meaning and thus require greater effort in maintaining accurate bindings.

A concrete example may be constructive here. Imagine you are standing in a crowded airport terminal, expecting your uncle to show up. You suddenly detect someone who is smiling at you. Then, you note that this "stranger" is approaching you. Finally, you understand that the man who is weeping on your shoulders is your uncle. Your uncle's face went through many feature changes in the course of a relatively short period of time. Still, you succeeded in maintaining a single coherent representation. How can this be accomplished? It is likely that some sort of binding mechanism has been operative. Evidence for such binding mechanism comes primarily from situations in which binding fails. In the well-known McGurk effect (McGurk, & MacDonald, 1976), the vocal sound produced by a face is erroneously integrated with the lips movements, such that the perceiver hears a different phoneme than that articulated.

Facial attributes and "person construal"

Cognitive psychologists have invested much effort in studying the perceptual mechanisms that govern face processing (Bruce, & Young, 1986; Burton, Bruce, & Johnston, 1990; Calder, & Young, 2005; Farah, Wilson, Drain, & Tanaka, 1998; Fitousi, & Wenger, 2013; Fitousi, 2015, 2016; Haxby et al., 2000). Social psychologists have also studied the implications of perceiving the faces of others. This work has come to be known as "person construal" (Fiske, & Neuberg, 1990; Freeman, & Ambady, 2011; Macrae, Bodenausen, & Milne, 1995). Person construal research investigates the lower level[1] perceptual mechanisms that produce social cognitive phenomena. A recent influential theory by Freeman and Ambady (2011) has proposed that perception of the social attributes in a face is a dynamic process that evolves over hundreds of milliseconds. In this model, perceptual processing of irrelevant social face attributes can partially activate other face attributes, including motor actions. Event-related potential (ERP) studies supported this conjecture, showing that the extraction of facial attributes (e.g., sex, race, and age) is immediately and concomitantly shared with the motor cortex (Freeman, Ambady, Midgley, & Holcomb, 2011).

Another source of support in the interactive theory of Freeman and Ambady (2011) comes from studies on response trajectories (Freeman, Pauker, Apfbelbaum, & Ambady, 2010; Freeman, & Ambady, 2009). In this type of studies, participants classify faces on a predefined facial attribute (e.g., age) by moving their hand toward one of

---

[1] Note that by "low-level", social psychologists refer to such attributes as gender, race, and age; these attributes are considered as "high-level" features by cognitive psychologists, who often study "low level" features such as color, shape, and orientation.

two labels on the screen. The faces also vary on an irrelevant dimension (e.g., gender). Participants' hand trajectories are often attracted to the label carrying the name of the irrelevant facial attribute (e.g., woman), indicating its abrupt online activation. These studies support the idea that face attributes interact with other face attributes at perceptual, cognitive, or motor levels. Freeman and Ambady's (2011) theory contributes valuable insights into the interaction of perceptual and motor aspects of face perception, but it is moot with respect to the binding mechanism that shapes the ultimate representation. What is needed is a broader theoretical framework that can shed light on the binding of facial and motor attributes. The following section proposes such a framework.

From "object files" to "event files"

A systematic analysis of feature binding with objects has been performed by Kahneman and Treisman (1984) and Kahneman et al., (1992). They have used a preview task in which a letter appears in a prime display, and then the same letter or different letters is presented in a probe display. Naming latencies for the probe letter were faster if the letter's identity was repeated and associated with the same object/location.[2] Kahneman et al. (1992) called this object-specific preview effect. According to these authors, the processing of a visual object leads to the creation of an "object file", an episodic representation of the object's identity and location that allows its identification in spite of spatiotemporal discontinuities.

Considerable progress in understanding "object files" has been made by Hommel (1998). He has advanced the theory in various creative ways (Hommel, 2004, 2005; Hommel, & Colzato, 2009). First, Hommel showed that priming effects can be documented even when an object's location is not repeated, but other of its features are (i.e., object-nonspecific repetition effects). Second, he demonstrated that "objects files" may consist of a subset (i.e., binary bindings) of their features, not necessarily the entire list of features, as argued by Kahneman et al. (1992). Third, object-nonspecific repetition effects represent a processing cost, rather than a benefit (Hommel, & Colzato, 2009). In particular, repeating two given features (e.g., a red square) or alternating the same features (e.g., a blue triangle) yields performance levels that are superior to those observed in conditions in which one of the features is repeated and the other is alternated (e.g., a red triangle). This pattern is called partial-repetition costs (Hommel, 2004, p. 496). Fourth, Hommel introduced the concept of action codes.

These are motor and response attributes that are distributed in the brain and are amenable to integration just like visual features-codes (Hommel, Müsseler, Aschersleben, & Prinz, 2001). When action codes integrate with feature codes, they create an "event file"—a mid-level representation or a pointer to a visuo–motor episodic trace (Hommel, 1998). For example, responding to a red object with your right hand may lead to the binding of the red color with the motor code associated with the right hand. Complete repetition or alternation of the features in this newly created combination would enjoy more efficient processing than partial repetitions.

The distributed coding of simple attributes, such as color, shape, and orientation, in the primate brain is well established (Livingstone, & Hubel, 1987). But are more complex attributes, such as facial dimensions coded in a distributed fashion? Haxby and his colleagues (Haxby et al., 2002; Hoffman, & Haxby, 2000) have presented evidence for the existence of a neural system in the human brain of separate localized regions. This system specializes in processing facial attributes. In this system, the ventral temporal cortex and the fusiform gyrus (Kanwisher et al., 1997) are responsible for the processing of invariant facial aspects, such as identity, whereas the superior temporal sulcus (STS) is responsible for the processing of variant attributes, such as eye gaze and emotion (Vuilleumier, Armony, Driver, & Dolan, 2001). The neuronal distributed model proposed by Haxby and his colleagues (Haxby et al., 2002; Hoffman, & Haxby, 2000) suggests that facial attributes are coded in separate brain areas. To date, no direct attempt has been made to study how these face codes are integrated with each other, or how they are bound with action codes (Hommel, 2000).

Overview of the present experiments

Using simple colored shapes, Hommel (1998) adduced consistent evidence for the presence of binding processes, supporting the existence of both visuo–visuo integrations (i.e., form and color, form and location, and color and location) and visuo–motor integrations (i.e., color and response location, form and response location). Hommel's (1998) methodology and results provide strong evidence for the existence of "object files" and "event files" with low-level features. The present study tested the hypothesis that similar "object files" and "event files" exist for face attributes. A recent study by Keizer, Colzato, and Hommel, (2008) documented integrations of faces with houses, motion, and manual response. The present study departs from the Keizer et al. study in an important way. In that study, the whole face served as the elementary unit of integration, whereas here, facial attributes (e.g., eye gaze

---

[2] In Kahneman's et al. (1992) study, no strict distinction has been postulated between location and object.

and expression) are the integration units, and the main question of interest concerns the binding of these attributes.

Five facial and non-facial attributes were elected for testing: facial identity, emotion (i.e., expression), eye-gaze direction, the face's spatial location, and the location of the manual response emitted toward the face. Subsets of these five attributes have been tested in a series of four experiments. The reason for choosing these attributes is that they represent the most important and studied face attributes (cf. Haxby et al., 2000, 2002). Another reason is that they encompass both variant (i.e., emotion and gaze direction) and invariant (i.e., identity) attributes (Haxby et al., 2000, 2002).

A word is in order regarding the non-facial attribute of spatial location. A-priori, it seems likely that faces are individuated via their identity (John's face). However, there is also the possibility that faces are individuated through their location in space. Interestingly, spatial location has not been considered as a consequential variable in face recognition studies, although it has been attributed a fundamental role in tagging an addressing "object files" (Kahneman, & Treisman, 1984; Kahneman et al., 1992; Wolfe, & Bennett, 1997). Hommel has documented partial-repetition costs for combinations of location and response, location and form, but not for combinations of location and color (Hommel, 1998). It is, therefore, crucial to see whether spatial location is critical to the individuation of faces, or for the integration of facial features into an "object file" or "face file."

The paradigm deployed throughout the present experiments is similar to that used by Hommel (1998, 2004, see also Zmigrod, de Sonneville, Colzato, Swaab, & Hommel, 2013). It is a variation on the original preview method developed by Kahneman et al. (1992). Each trial consisted of a sequence of displays, starting with a cue to response, followed by a face (S1), and replaced by a blank. The blank was then substituted by another face stimulus (S2). Response to the first face, S1, is termed R1, and response to the second face, S2, is called R2. Figure 1 shows a schematic illustration of displays and timings in the experiments. On a trial, each one of the features could be either repeated or alternated from S1 to S2. Similarly, the response feature (i.e., left- vs right-hand response) could be repeated, alternated, or neutral from R1 to R2. The neutral condition means that no response was required in R1. This condition can help decide whether repetition was beneficial or alternation was harmful for performance. The execution of R2 was performed according to the relevant dimension for response (e.g., identity) in the given experiment. The target dimension for response was varied across experiments.

In the present experiments, each facial dimension could take one of two values. Thus, facial identity could belong
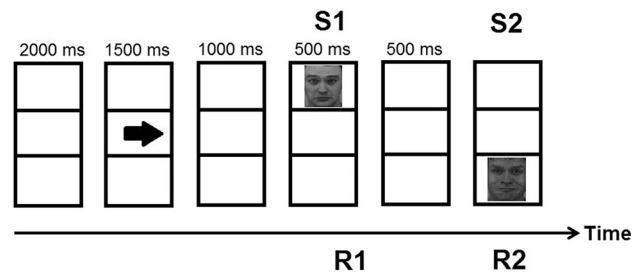

Fig. 1 Illustration of the displays used in Experiments 1–4 following Hommel (1998). R1 and R2 are responses to the first stimulus (S1) and second stimulus (S2)

to either person A or person B (Experiments 1 and 2); similarly, facial emotion could take one of two possible values—sad vs angry in Experiments 1 and 2, or frightened vs angry in Experiments 3 and 4; eye-gaze direction was either averted to the left or to the right (in Experiments 3 and 4), and the spatial location of the face was either on the top or bottom of the screen (Experiments 1–4).

Three effects of major theoretical significance may emerge in this priming setup (Hommel, 1998, 2004). The first is a main effect of stimulus or response feature repetition. Perceivers may benefit from the repetition of facial identity S1 (e.g., Jim) in S2 (e.g., Jim), or due to the repetition of R1 response to S1 (e.g., right-hand key) in S2 (e.g., right-hand key). In that case, perceivers may respond faster to the probe in the identity-repeated condition than in the identity-alternated condition (Burton, Kelly, & Bruce, 1998; Ellis, Young, Flude, & Hay, 1987). This type of effect does not imply integration of features, but it indicates feature priming in short-term memory.

A second type of effect is called partial-repetition costs (Hommel, 2004) and is due to repetition or alteration of combinations of features from S1 to S2. To better understand how this effect is measured, consider the following three types of trials: (1) complete repetitions are trials in which the two features of the stimulus in S1 (e.g., Jim + happy) are repeated in S2 (e.g., Jim + happy), (2) complete alternations are trials in which the two features in S1 (e.g., Jim + happy) are replaced by two different features in S2 (e.g., David + sad), and (3) partial repetitions are trials in which one of the features in S1 is repeated in S2, whereas the other feature is alternated (e.g., Jim+ happy in S1 and David + happy in S2). Partial-repetition costs (Hommel, 2004) are recorded when performance in the partial-repetition trials is worse than that in the complete repetition or complete alternation trials. The presence of such costs entails the formation of an "object file" consisting of a pairwise binding trace of the two pertinent features (Hommel, 1998).

A third type of result is due to the repetition or alteration of feature–response combinations. The repetition or

alteration of a specific combination of stimulus–response features conjunction in S1–R1 (e.g., Jim + left key in S1) may be facilitated if completely repeated in S2–R2 (e.g., Jim + left key in S2) or completely alternated (e.g., David + right key in S2), relative to a condition where only one of the features is repeated and the other is alternated (e.g., Jim + right key in S2). Partial-repetition costs with response–stimulus features indicate the formation of an "event file" (Hommel, 1998, 2004). In the theoretical context of the present study, this type of effect may speak to the integration of response codes with facial attributes.

Experiment 1

Faces in Experiment 1 varied on four dimensions: identity, emotion, spatial location, and response location. The relevant dimension for response was facial identity. A central goal of the experiment has been to examine whether facial identity plays a crucial role in the formation of "face files". Mitroff, Scholl, and Noles, (2007) have shown that the response to facial identity was speeded if identity reappeared in a previously presented object irrespective of the object's location. The results by Mitroff et al. (2007) suggest the involvement of episodic tokens in the formation of "object files". It is highly likely that facial identity is an important feature in the formation of "face files", allowing a coherent representation when a face undergoes spatiotemporal discontinuities. However, the Mitroff et al. study has not been designed to probe identity binding with other facial attributes of perceptual and conceptual variability (e.g., emotion).

Spatial location is another feature that might be operative in the formation of "face files", serving the visual system as an anchor or pointer toward the perceived face. This is a plausible idea, since spatial tagging mechanisms, such as inhibition of return (i.e., IOR, Posner, & Cohen, 1984), have been shown to affect the detection of faces (Tipper, Weaver, Jerreat, & Burak, 1994). If this hypothesis is correct, partial-repetition costs are expected with spatial location. Another prediction follows from Hommel's work (1998, 2004) on the binding of visual and action codes. Hommel found that the task-relevant feature is often highly likely to be bound with the response code. It is therefore predicted that facial identity, which serves here as the relevant feature, will be integrated with response code. Finally, Kahneman and Treisman (1984, see also Kahneman et al. 1992), have argued that the creation of "object files" is exhaustive, in the sense that it requires the binding of all constituent features. If such an exhaustive process occurs with faces, full-repetition costs with all four dimensions are expected. This would be indicted by a four-way interaction with identity × emotion × location × response.

Method

Participants

Twenty young volunteers from Ariel University took part in this experiment. These were young male and female undergraduate students (aged 20–28) who participated in partial fulfillment of course credit. All reported normal or corrected-to-normal vision, normal hearing, and unencumbered use of their two hands.

Apparatus and stimuli

The experiment was controlled by a desktop computer. Viewing distance subtended 76 cm from the computer screen. The stimuli consisted of three $3.16° × 2.7°$ black square outlines arranged vertically from the top to the bottom (see Fig. 1). Four facial identities were deployed. These consisted of two females and two males. Two separate sets of faces were constructed for the male and female faces (see Fig. 2). Each set of images was created by crossing two unfamiliar facial identities (person A and person B) with two facial expressions (sad and angry). The face images were downloaded with permission from the Karolinska directed emotional face (KDEF) database (Lundqvist, Flykt, & Ohman, 1998). The images were altered with the free GIMP software. Each face image subtended $1.88° × 2.33°$. The faces were equated for size, brightness, and overall shape. The face stimuli were presented as gray-scale images over a gray or black frame (see Fig. 2).

Each face could appear either in the upper box or in the lower box (see Fig. 1). A middle box, at the center of screen, was used for presenting the cue for response (R1). Response cues were full black arrows which were pointing to the right, left, or both directions (when no response in R1 was needed). Responses were made by pressing the left ("z") or right ("m") keys on a QWERTY keyboard.

Procedure and design

The procedure and design were similar to those reported by Hommel (1998). Each experimental trial started with an arrow cue for 1500 ms. Participants withheld their response (R1) to the first stimulus (S1) if the arrow was bidirectional. Participants made a response (R1) to S1 according to the cue if the arrow was pointing only in one direction (left or right). A leftward pointing arrow required a left-hand-key response and a rightward pointing arrow required a right-hand-key response. Participants were informed that there would be no systematic relationship between S1 and R1, so that they should execute the pre-cued response at the onset of S1 while ignoring the irrelevant dimension of S1. A second response (R2) was always

Fig. 2 a Four male face images and **b** four female face images used in Experiment 1. Each set of the faces was created by crossing two levels of unfamiliar facial identity (*A* and *B*) with two levels of facial emotion (angry and sad)
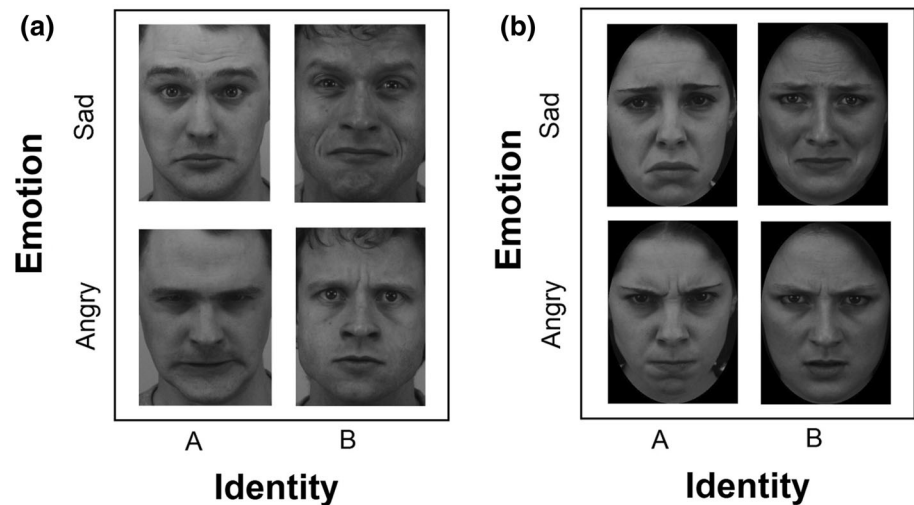
a binary-choice reaction to the second stimulus (S2). The critical stimulus dimension in S2 was facial identity. Half of the participants responded to "identity A" with a right-hand response ("m") and to "identity B" with a left-hand response ("z"), while the other half responded with the reverse assignment. To be able extend the validity of the results beyond a certain identity and gender, one group of participants (*n* = 12) was presented with the male images (Fig. 2a), whereas the other group of participants (*n* = 8) was presented with the female images (Fig. 2b).

Figure 1 shows a typical sequence of events in a trial. Each trial began with an arrow cue presented for 1500 ms followed by a blank interval for 500 ms. Then, S1 face appeared for 500 ms and R1 was expected. S1 was then replaced by another blank interval for 500 ms followed by S2. At this stage, R2 was expected. S2 remained on the screen for 2500 ms or until response. An inter-trial interval of 2500 ms preceded the presentation of a new response cue. A block consisted of the factorial combination of S2 identity (person A vs person B), R1 response (left vs right vs both), emotion (sad vs angry), location (top vs bottom box), and R2 response (left vs right), the possible relationships between S1 and S2 (i.e., repetition vs alternation) regarding identity, emotion and location, and the three possible relationships between R1 and R2 (repetition, alternation, or single response). Each experimental block consisted of 192 trials. The experiment consisted of three blocks of trials. The order of trials in each block was chosen randomly by the computer. A 1 min break was allowed between the blocks.

Results

Trials in which RTs were incorrect, longer than 1900 ms, or shorter than 150 ms were removed from the analysis. These amounted to 8.6 % of the total number of trials.

Mean RTs and mean proportion of errors were calculated for each possible level of stimuli and responses in the two tasks (R1 and R2). A five-way ANOVA with stimulus set (male, female), response (repeated, alternated), emotion (repeated, alternated), identity (repeated, alternated), and location (repeated, alternated) as factors was performed on mean RTs. Because the effect of stimulus set was far from significance, the data were collapsed to a four-way ANOVA. Table 1 reports those mean RTs along with the error rates (see also Table 5 in the Appendix for an exhaustive list of the ANOVA effects). A significant main effect of emotion [$F(1, 19) = 12.47$, MSE = 24,055, $p < 0.005$] revealed that repeating facial expression led to faster responses (802 ms) than alternating it (819 ms). Most importantly, the response × identity interaction [$F(1, 19) = 43.43$, MSE = 18,4767, $p < 0.00001$] indicated partial-repetition costs due to bindings of the response feature with the task-relevant facial feature of identity (see Fig. 3a). Responses were faster when both

Table 1 Means reaction times (RT) in ms and percentage of error (PE) for R2 in Experiment 1 for conditions of repetition and alteration in S1 and S2 and in R1 and R2

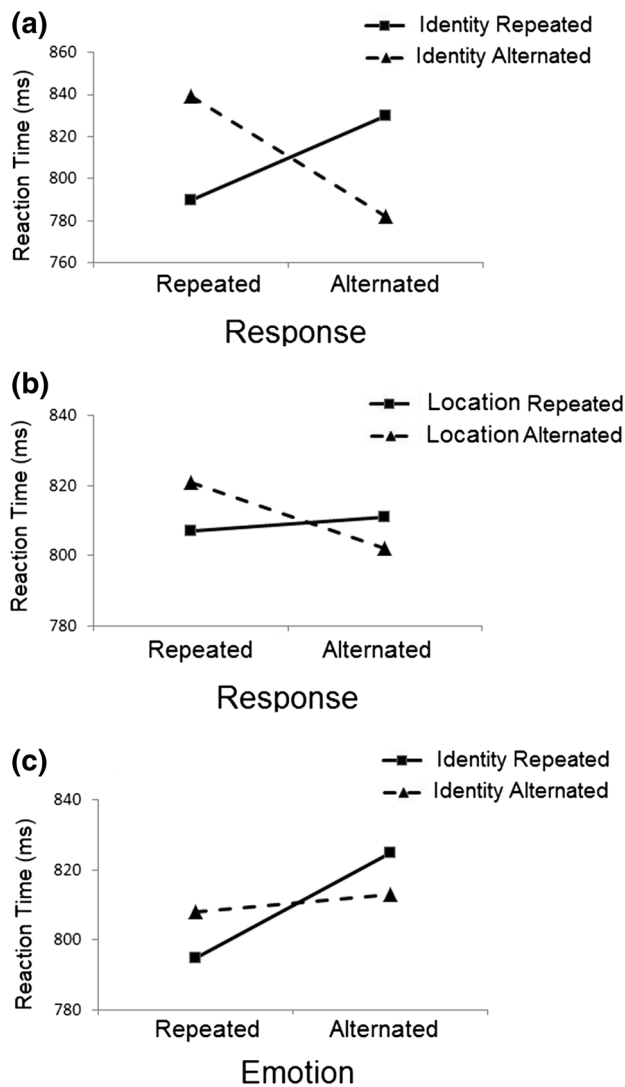| Face dimension repeated	Response					
	Repeated		Alternated		Single	
	RT	PE	RT	PE	RT	PE
Neither	847	0.9	779	4.4	805	2.7
Identity (I)	827	0.9	849	3.3	835	3.2
Emotion (E)	833	1.0	767	2.6	796	3.2
Location (L)	825	1.2	799	3.2	800	3.1
IE	773	2.8	820	1.7	777	2.7
IL	791	4.8	834	3.0	823	4.4
EL	847	2.0	791	3.0	826	1.6
IEL	769	4.3	817	2.3	780	1.6

Fig. 3 Feature binding in Experiment 1. **a** Facial identity and response. **b** Spatial location of the face and response. **c** Facial emotion and facial identity

interaction [$F (1, 19) = 4.6$, $MSE = 11{,}712$, $p < 0.05$], indicated the binding of identity and emotion (see Fig. 3c), and thus the creation of "object files". Faster responses were recorded when both facial identity and emotion repeated together (795 and 813 ms) than when only one of them repeated and the other alternated (808 and 825 ms). Two-tailed t tests verified that the benefits and costs associated with all these pairwise bindings were significantly different from zero (all $ps < 0.05$).

Discussion

Experiments 1 underscored partial-repetition costs with both facial and non-facial attributes, adducing consistent evidence for the formation and retrieval of both "object files" and "event files" with facial attributes. These episodic structures are dubbed herein "face files". The current patterns extend those observed with color-shape objects (Hommel, 1998). They show that: (a) binding can take place with subsets of features rather than the entire list of features (Kahneman et al., 1992) and (b) integration of response-stimulus features can occur with task-relevant as well as with task-irrelevant stimulus features (Hommel, 2004). The results support the hypothesis that high-level social and motor categories conveyed by faces are abstracted, extracted, and become available to perception and action. The results are commensurate with Freeman and Ambady's (2011) interactive model, according to which social aspects of a face interact with each other as well as with motor codes.

Note that spatial location interacted with the response, but not with any of the other facial features; while the task-relevant attribute (e.g., identity) was bound with the response feature and with the facial attribute of emotion. This might be because identity served as the task-relevant dimension. An alternative explanation is that facial identity serves as a quintessential facial dimension in the individuation of a face. According to this account, identity should be automatically bound with response, as well as with other facial features. In addition, this should hold true even when identity is not the relevant dimension for the task at hand. A plausible hypothesis is, therefore, that it is facial identity and not spatial location that maintains the retrieval of integrated face attributes. A central goal of Experiment 2 has been to decide between these two hypotheses. In Experiment 2 facial emotion was made the relevant dimension for response. If the former hypothesis is correct, it is expected that facial identity would not be integrated with response. If the latter hypothesis is correct, it is expected that facial identity would be integrated with the response feature, as well as with other features.

identity and response features repeated or alternated (790 and 782 ms) than when only one of them repeated and the other alternated (830 and 839 ms). It is important to emphasize that interpreting binding effects strictly requires focusing on the interaction as such. The main effects, whether significant or insignificant, are irrelevant to the interpretation of the binding effect. A response × location [$F (1, 19) = 5.41$, $MSE = 10{,}586$, $p < 0.05$] reflected the binding of spatial location with response (see Fig. 3b). Responses were faster when both location and response features repeated or alternated (807 and 802 ms) than when only one of them repeated and the other alternated (811 and 821 ms). In addition to the creation of these "event files", which reflected a visuo–motor binding, a significant identity × emotion

Experiment 2

Experiment 2 was identical to Experiment 1 in terms of design, procedure and stimuli, except for the fact that emotion served as the relevant dimension for response. Participants were asked to ignore the facial identity as well as other irrelevant features.

Method

Participants

A new group of eleven young volunteers from Ariel University took part in Experiment 2. These were male and female undergraduate students who participated in partial fulfillment of course credit. All reported normal or corrected-to-normal vision, normal hearing, and unencumbered use of their two hands.

Apparatus and stimuli

Apparatus and stimuli were identical to those reported in Experiment 1 with the female set of face stimuli.

Procedure and design

The procedure and design were identical to those reported in Experiment 1. The only difference between the two experiments was that in the current experiment participants responded to the feature of facial emotion rather than to that of facial identity of the target's face. Participants indicated whether the face was sad or angry by pressing one of two response keys. Response assignment was balanced across observers.

Results

Trials in which RTs were incorrect, longer than 1900 ms, or shorter than 150 ms were removed from the analysis. These amounted to 4.6 % of the total number of trials. Mean RTs and mean proportion of errors were calculated for each possible level of stimuli and responses in the two tasks (R1 and R2). Table 2 reports those mean RTs along with the error rates. A four-way ANOVA with response (repeated, alternated), emotion (repeated, alternated), identity (repeated, alternated), and location (repeated, alternated) as factors was performed on mean RTs. The full list of ANOVA effects is presented in Table 6 at the Appendix. A marginally significant main effect of spatial location [F (1, 10) = 4.37, MSE = 11,641, p = 0.06] revealed that repeating that feature led to faster responses (822 ms) than alternating it (835 ms).

Table 2 Means reaction times (RT) in ms and percentage of error (PE) for R2 in Experiment 2 for conditions of repetition and alteration in S1 and S2 and in R1 and R2

| Face dimension repeated | Response | | | | | |
| | Repeated | | Alternated | | Single | |
	RT	PE	RT	PE	RT	PE
Neither	956	1.2	804	5.2	786	4.6
Identity (I)	852	1.9	833	2.9	795	2.9
Emotion (E)	843	4.2	854	4.9	865	9.5
Location (L)	811	3.8	799	4.8	792	6.8
IE	792	5.4	859	6.3	784	6.2
IL	806	2.0	797	4.1	783	6.8
EL	844	4.8	849	6.7	873	5.0
IEL	765	3.7	861	3.9	809	2.2

Most importantly, the response × emotion interaction [F (1, 10) = 18.12, MSE = 50,889, p < 0.005] indicated partial-repetition costs due to bindings of the response feature with the task-relevant facial feature of emotion (see Fig. 4a). Responses were faster when both emotion and response features repeated or alternated (816 and 810 ms) than when only one of them repeated and the other alternated (861 and 834 ms). A response × identity interaction [F (1, 10) = 10.64, MSE = 19,912, p < 0.005] reflected the binding of facial identity with response (see Fig. 4b). Responses were faster when both identity and response features repeated or alternated (810 and 830 ms) than when only one of them repeated and the other alternated (841 and 841 ms). In addition to these visuo–motor bindings that indicated the presence of "event files", a significant identity × emotion interaction [F (1, 10) = 9.10, MSE = 17,321, p < 0.05], indicated the binding of identity and emotion (see Fig. 4c) and, therefore, the emergence of an "object file". Faster responses were recorded when both identity and emotion repeated or alternated together (824 and 817 ms) than when only one of them repeated and the other alternated (853 and 828 ms). Two-tailed t tests verified that the benefits and costs associated with all these pairwise bindings were significantly different from zero (all ps < 0.05). The response × identity × emotion interaction was significant [F (1, 10) = 5.28, MSE = 18,503, p < 0.05].

Error analysis A similar four-way ANOVA was performed on error rates. The analysis revealed a main effect of emotion [F (1, 10) = 8.84, MSE = 0.016, p < 0.05], indicating that more errors (5.0 %) were committed when emotion was repeated than when alternated (3.1 %). No other significant effects have been found on the error analyses.
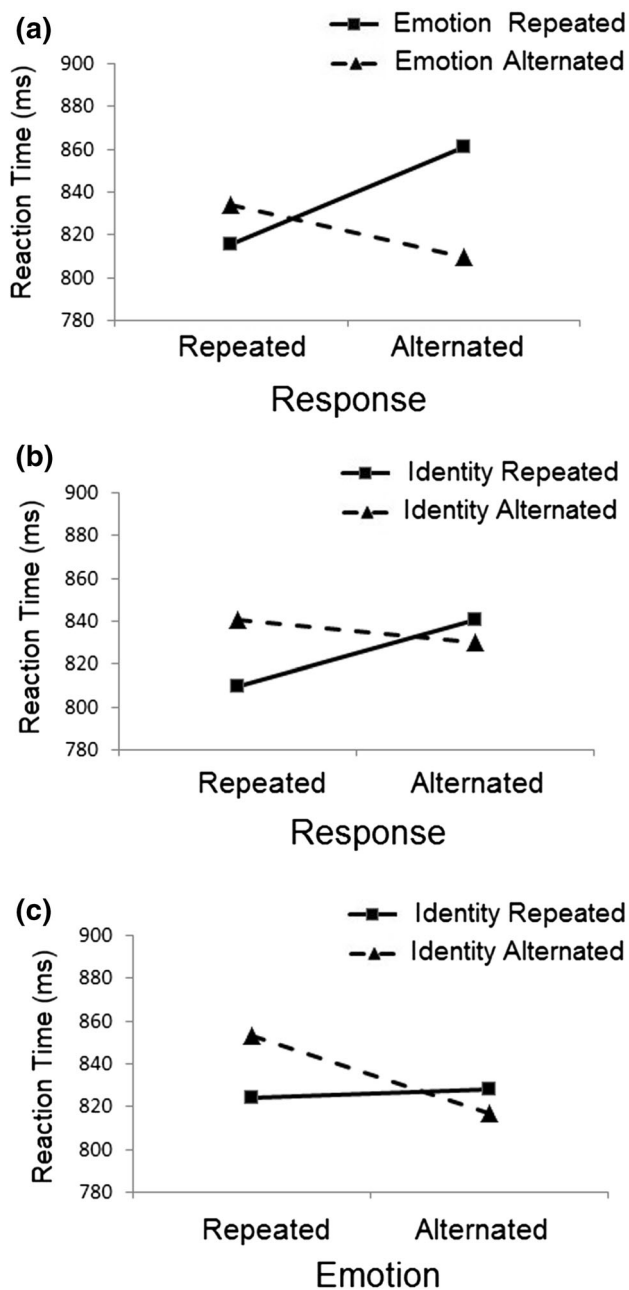
Fig. 4 Experiment 2. Partial-repetition costs indicating the binding of **a** response and facial emotion, **b** response and facial identity, and **c** facial emotion and identity

Discussion

In Experiment 2, facial emotion served as the task-relevant dimension. The results provided further evidence for the formation of "face files". The existence of these episodic structures proposes that visual and motor attributes are abstracted, extracted, and integrated into temporary constructs in visual short-term memory. Both visuo–visuo and visuo–motor bindings obtained with facial emotion and identity. The task-relevant feature of emotion was

combined with response, as did facial identity. Emotion and identity were also bound together. This outcome supports the hypothesis that facial identity is an essential attribute in the formation of face files, as it was automatically integrated into a "face file", even though it was not relevant for response.

Another conclusion that can be made is that spatial location has not played a significant role in the binding process. The almost non-existent involvement of spatial location in binding effects in the last two experiments is quite surprising. The previous studies with geometric colored shapes have shown that spatial location is often integrated with visual and motor features (Hommel, 1998; Kahneman et al., 1992; van Dam, & Hommel, 2010). Why were bindings with spatial location missing here? It should be noted that an auxiliary experiment has been conducted using the same displays and methods adopted in Experiments 1–2. In this experiment, I have replicated Hommel's (1998) Experiment 1, including the presence of the exact patterns of partial-repetition costs with spatial location. Thus, the absence of bindings with spatial location in the last two experiments is not due to differences in methods. There are two possibilities that can account for the lack of such binding effects with faces. One is that spatial location is not operative in the binding of facial features. An alternative account is that spatial location may be active only when it becomes relevant for the processing of the task-relevant dimension. The testing of these two hypotheses becomes possible in the next set of two experiments by introducing the dimension of eye-gaze direction.

In the next set of experiments, the facial dimension of eye-gaze direction was varied in a newly created set of face stimuli. Gaze direction is a facial attribute of considerable social import. It is extremely useful in reading other people's attention, intentions, and actions. In recent years, this facial attribute has been under extensive scrutiny (Calder, Beaver, Winston, Dolan, Jenkins, Eger, & Henson, 2007; Engell, & Haxby, 2007; Friesen, & Kingstone, 1998; Frischen, Bayliss, & Tipper, 2007). One of the most intriguing discoveries is that gaze direction induces spatial attentional shifts by acting as a visual cue for location (Friesen, & Kingstone, 1998). When perceivers see a leftward looking face, their responses are faster to targets located on the left, while the reverse also holds true. In Experiment 3, gaze direction will serve as the target dimension. Because gaze direction is coded in a spatial location code, it seems likely that it will interact with spatial location and response features.

Experiment 3

The goal of Experiment 3 has been to further test the binding process with facial and non-facial attributes. The focus of this experiment was the possibility that the spatial

Fig. 5 Four face images used in Experiments 3 and 4. The faces were created by crossing two levels of facial emotion (anger and fear) with two levels of gaze direction (*leftward* and *rightward*)

Table 3 Means reaction times (RT) in ms and percentage of error (PE) for R2 in Experiment 4 for conditions of repetition and alteration in S1 and S2 and in R1 and R2

Face dimension repeated	Response					
	Repeated		Alternated		Single	
	RT	PE	RT	PE	RT	PE
Neither	743	2.0	683	4.2	705	1.8
Emotion (E)	744	2.5	721	4.9	730	2.6
Gaze (G)	771	3.6	736	3.6	751	3.0
Location (L)	740	3.2	730	3.1	724	1.9
EG	745	2.0	733	4.4	725	2.6
EL	739	3.2	709	3.6	737	3.8
GL	758	0.0	735	3.5	751	1.8
EGL	693	4.8	736	1.6	720	1.1

codes of location and response are integrated with gaze direction. A new set of faces was constructed in which the eye gaze of the face could be directed either to the left or to the right (see Fig. 5). Participants indicated the direction of the face's eye gaze while ignoring variations in facial emotion, location, and response. Based on the spatial qualities of gaze direction (Friesen, & Kingstone, 1998), it was predicted that spatial location would come to play a vital role in the formation of "face files", such that it would be integrated with gaze direction as well as with other facial and non-facial features, including spatial location.

Method

Participants

A new sample of seventeen young volunteers from Ariel University took part in Experiment 3. These were male and female undergraduate students who participated in partial fulfillment of course credit. All reported normal or corrected-to-normal vision, normal hearing, and unencumbered use of their two hands.

Apparatus and stimuli

A new set of stimuli was created for this experiment. The face images were downloaded with permission from the Karolinska directed emotional face (KDEF) database (Lundqvis et al., 1998). The images were modified using

the free GIMP software to create the four images that are presented in Fig. 5. Two levels of gaze direction (left, right) were crossed with two levels of facial emotion (anger, fear). The same female identity from Experiment 1 (person B) was used. The change of the emotion values from those used in the previous experiments was done to extend the conceptual replicability of the stimuli. The images were equated on size, brightness, and overall shape.

Procedure and design

The procedure and design were identical to those reported in Experiment 1. The task-relevant dimension for response was gaze-direction. Participants pressed a right-hand key ("m") if the gaze was averted to the right. They pressed a left-hand key ("z") if the gaze was averted to the left. Participants were asked to ignore all the other irrelevant dimensions.

Results

Trials in which RTs for R1 and R2 were incorrect, longer than 1900 ms or shorter than 150 ms were removed from the analysis. These amounted to 6.9 % of the total number of trials. Mean RTs and mean proportion of errors were calculated as a function of the four possible relationships between the stimuli and the responses of the two subtasks (R1 and R2). That is according to whether the emotion, gaze direction, or location of S1 and S2 was repeated or alternated, and whether R2 was preceded by a same, different, or no response. Table 3 presents the mean RTs and error rates in the different conditions. A four-way analysis of variance (ANOVA) with response (repeated, alternated), emotion (repeated, alternated), gaze direction (repeated, alternated), and location (repeated, alternated) as factors
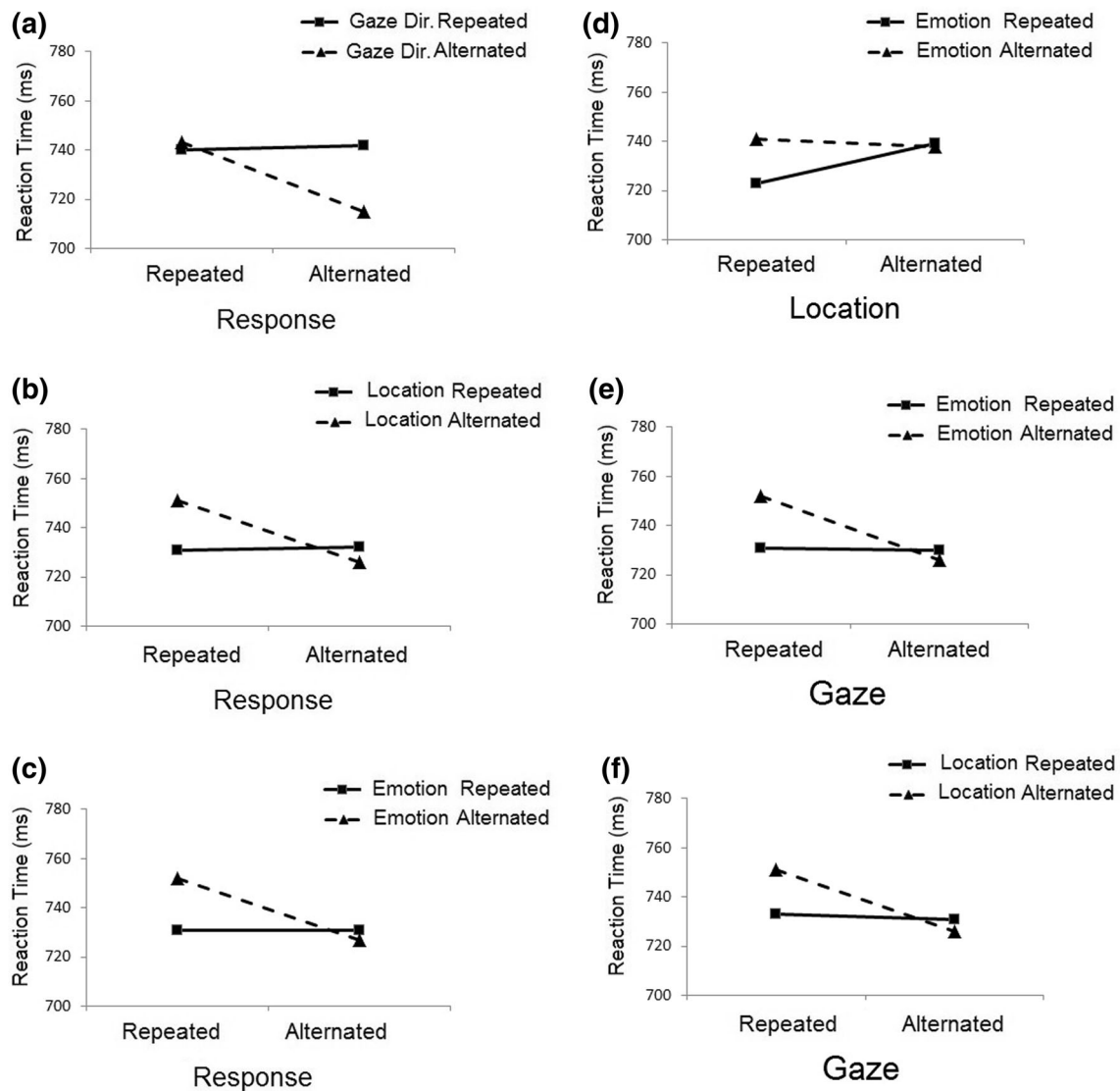
Fig. 6 Experiment 3. Partial-repetition costs indicating the binding of **a** gaze direction and response, **b** spatial location and response, **c** emotion and response, **d** emotion and location, **e** gaze direction and emotion, and **f** gaze direction and location

was performed. The full list of ANOVA effects is presented in Table 7 at the Appendix.

Reaction time A main effect of emotion [F (1, 16) = 5.38, MSE = 4954, $p < 0.05$] revealed that responses were faster when emotion repeated (731 ms) than when emotion alternated (739 ms). Most importantly, various two-way interactions signaled the obtainment of visuo–motor and visuo–visuo bindings (see Fig. 6a–f). First, a significant response × gaze interaction [F (1, 16) = 7.08, MSE = 16,394, $p < 0.05$] indicated that response and the task-relevant feature of gaze direction were bound together. RTs were faster when these features both repeated or alternated together (740 and 715 ms) than when one repeated and the other alternated (743 and 742 ms). A response × location interaction [F (1,

16) = 13.17, MSE = 11,800, $p < 0.005$] revealed the binding of response with location. RTs were faster when both response and location repeated or alternated (731 and 726 ms) than when one repeated, but the other alternated (732 and 751 ms). Response also integrated with emotion [F (1, 16) = 7.61, MSE = 10,312, $p < 0.05$], as indicated by faster RTs when the two features repeated or alternated together (731 and 727 ms) than when one of them repeated and the other alternated (731 and 752 ms).

Evidence for visuo–visuo bindings was indicated by the significant gaze × emotion interaction [F (1, 16) = 9.923, MSE = 10,451, $p < 0.01$]. Faster RTs were recorded when both features repeated or alternated (731 and 726 ms) than when only one of them repeated and the other alternated (730 and 752 ms). A marginally significant interaction of

location $\times$ emotion [F (1, 16) = 4.45, MSE = 6449, p = 0.05] was found (723 and 738 ms in mutual repetitions or alternations vs 739 and 741 ms in the contrasting case); and a marginally significant interaction of location and gaze [F (1, 16) = 4.31, MSE = 8464, p = 0.054] was recorded (733 and 726 ms in mutual repetitions or alternations vs 731 and 751 ms in the contrasting case). These two interactions pointed to the binding of location with emotion and location with gaze. T tests revealed that most of the partial-repetition effects were significant.

Error analysis Similar four-way ANOVA was performed on the error rates. No significant effects have been observed.

Discussion

In Experiment 3, gaze direction served as the relevant dimension for response. Participants were instructed to ignore variations on facial emotion, response, and spatial location. A number of pairwise bindings were detected across motor and visual face attributes. In addition to the expected binding of the task-relevant dimension of gaze direction with the motor response feature, gaze was integrated with spatial location and emotion. It seems that including the dimension of gaze direction in the stimuli set, and rendering it the relevant face attribute, activated the spatial location code. This in turn, led to the binding of spatial location with other features. Such an outcome supports the hypothesis that the absence of binding effect with spatial location in Experiments 1 and 2 is not due to some unique characteristic of faces, but rather due to the facial features used. The findings of genuine bindings with spatial location in Experiment 3 are commensurate with our initial hypothesis that gaze direction is coded in some sort of a spatial code that is shared with the spatial location code. These results support Hommel's (1998) conjecture that binding is more likely across features that share common codes. The binding of emotion with gaze direction is consistent with recent studies showing interactions between the two dimensions (Adams, Gordon, Baird, Ambady, & Kleck, 2003; Adams, & Kleck, 2003, 2005).

It is interesting to note the involvement of spatial location in binding with gaze direction, as well as with emotion and response features. A plausible explanation for this might be the relevancy of gaze direction for task completion in this experiment. As mentioned earlier, gaze direction is known to induce reflexive shifts of spatial attention (Friesen, & Kingstone, 1998). Gaze direction is also instrumental in providing the observer with cues for emotion (Adams et al., 2003; Adams, & Kleck, 2003, 2005). This might account for the observed bindings of location with gaze direction and facial emotion. If this

explanation is correct, one would expect to find a reduction or even a complete abolishment of the involvement of spatial location in binding when gaze direction stops serving as a relevant dimension for response. This prediction has been tested in Experiment 4 by turning the feature of gaze direction into an irrelevant (though existent) dimension in the stimuli set.

Experiment 4

Experiment 4 was identical to Experiment 3 in terms of procedure, design, and stimuli. The only difference was that facial emotion was made the relevant dimension for response instead of gaze direction.

Method

Participants

A new sample of seventeen young volunteers from Ariel University took part in Experiment 4. Participants were young male and female undergraduate students who participated in partial fulfillment of course credit. None of them participated in the previous experiments. All of them reported normal or corrected-to-normal vision and unencumbered use of their two hands.

Apparatus and stimuli

Apparatus and stimuli were identical to those reported in Experiment 3.

Procedure and design

Procedure and design were identical to those reported in Experiment 3. The only difference was that facial emotion served as the relevant dimension for response in this experiment. Response assignment was counterbalanced across participants.

Results

Trials in which RTs for R1 and R2 were incorrect, longer than 1900 ms or shorter than 150 ms were removed from the analysis. This amounted to 9.8 % of the total number of trials. Table 4 gives the mean RTs and error rates in the various conditions. These means were entered into a four-way analysis of variance (ANOVA) with response (repeated, alternated), emotion (repeated, alternated), gaze direction (repeated, alternated), and location (repeated, alternated) as factors. The full list of ANOVA effects

Table 4 Means reaction times (RT) in ms and percentage of error (PE) for R2 in Experiment 3 for conditions of repetition and alteration in S1 and S2 and in R1 and R2

Face dimension repeated	Response					
	Repeated		Alternated		Single	
	RT	PE	RT	PE	RT	PE
Neither	838	1.3	768	8.9	789	1.1
Emotion (E)	838	4.4	824	2.1	831	4.2
Gaze (G)	816	0.0	787	8.5	784	2.7
Location (L)	821	1.7	752	7.0	831	2.6
EG	778	4.7	851	5.6	815	3.1
EL	806	2.8	845	3.3	851	4.6
GL	791	1.3	774	7.5	825	2.2
EGL	755	7.3	827	3.9	807	4.2

appears in Table 8 at the Appendix. A main effect of emotion repetition [F (1, 16) = 10.05, MSE = 65,877, $p < 0.005$] was recorded, suggesting that repeating emotion led to slower responses (843 ms) than alternating it (811 ms). Most importantly, evidence for the existence of "face files" was indicated by a highly significant response × emotion interaction [F (1, 16) = 33.76, MSE = 140,273, $p < 0.000001$] and a significant response × gaze-direction interaction [F (1, 16) = 6.50, MSE = 54,862, $p < 0.05$] (see Fig. 7a, b). The response × emotion interaction indicated partial-repetition costs due to bindings of the response feature with the task-relevant facial feature of emotion. Responses were faster when both the features of emotion and response repeated or alternated (825 and 783 ms) than when only one of them repeated and the other alternated (840 and 860 ms). Paired comparisons confirmed that the costs and benefits associated with response and emotion were significantly greater than zero (all ps < 0.05).

The response × gaze-direction interaction reflected partial-repetition costs due to binding of the response feature (left, right) with the task-irrelevant facial feature of gaze direction (left, right). Responses were faster when both gaze direction and response features repeated or alternated (814 and 811 ms) than when only one of them repeated and the other alternated (832 and 850 ms). Paired comparisons confirmed that the costs and benefits associated with the response and emotion features were significantly greater than zero (all ps < 0.05). This result suggests that although it was not relevant for the task at hand, gaze direction was integrated into a "face file". No other effects have reached significance level.

Error analysis Similar four-way ANOVA was performed on error rates. The analysis revealed a main effect of response [F (1, 16) = 4.73, MSE = 0.047, $p < 0.05$],
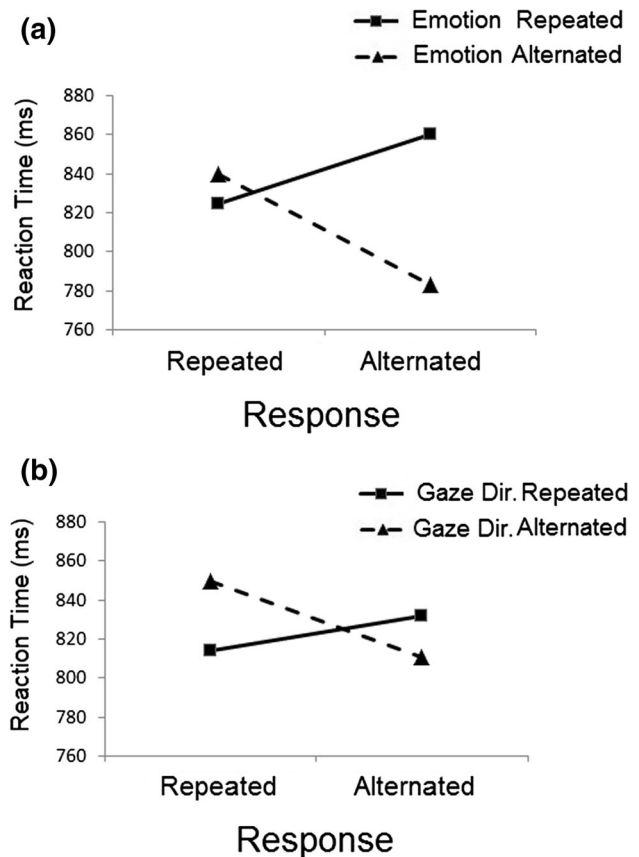

Fig. 7 Experiment 4. **a** Partial-repetition costs indicating the binding of response feature and facial emotion. **b** Partial-repetition costs indicating the binding of response feature and gaze direction

indicating that fewer errors were committed (2.9 %) when response was repeated than when response was alternated (5.5 %). A two-way interaction of response × emotion [F (1, 16) = 30.32, MSE = 0.085, $p < 0.0000$], signaled higher error rates (4.6, 7.4 %) when response and emotion repeated than when one of them repeated and the other alternated (3.7 and 1.2 %). This result looks like a mirror image of the RT result, and thus might suggest a speed-accuracy tradeoff strategy.

Discussion

In Experiment 4, a single facial identity varied on gaze direction, facial emotion, and spatial location. Facial emotion served as the target dimension. The results attested once more to the primacy of the task-relevant dimension (i.e., emotion) in binding with the response feature. The irrelevant dimension of gaze direction has also been integrated with response. Commensurate with our initial hypothesis, spatial location has not been involved in any of the bindings. It seems that once gaze direction becomes an irrelevant dimension for response, its close associate and

code-sharing dimension—spatial location—turns into a dormant feature. Such an account can be attributed to a possible reduction in the amount of attention allocated to gaze direction, and consequently to the weaker activation levels that might spread out to the spatial location codes.

General discussion

A series of four experiments provided substantial evidence for the formation and retrieval of transient memory structures with face attributes. In Experiment 1, participants responded to the identity of two unfamiliar faces varying on facial emotion and spatial location. Partial-repetition costs (Hommel, 1998) indicated the bindings of identity with response features, location with response, and identity with emotion. In Experiment 2, facial emotion was made the relevant dimension for response. Similar pairwise bindings were recorded. In Experiment 3, a single identity varied on gaze direction, emotion, and spatial location, with gaze-direction serving as the relevant feature for response. Several visuo–motor and visuo–visuo bindings were documented, including pairwise conjunctions with spatial location. In Experiment 4, facial emotion served as the target feature. The task-relevant dimension of emotion and the task-irrelevant dimension of gaze direction were both integrated with the response feature.

Taken collectively, the results from the four experiments converged on the conclusion that social and physical attributes of faces are integrated into temporary memory structures of pairwise visuo–visuo and visuo–motor bindings. The current empirical patterns extend the previous findings with the low-level features of color, shape, or location (Hommel, 1998). Here, it has been shown that binding can take place with attributes of higher representational complexity than the routine colored shapes (Hommel, 1998, 2004). At the neuronal level, the integration of facial attributes with each other, as well as with other motor features, requires the engagement of a broader network of neuronal substrates (Haxby et al., 2002; Sugase et al., 1999). This is implied by the logic that correct and efficient binding of facial attributes necessitates the activation of long-term representations as well as prior knowledge of social categories (Freeman, & Ambady, 2011).

The present study is not the first to incorporate face stimuli in a binding task. Mitroff et al. (2007) presented evidence for the integration of faces with objects, showing that response to facial identity is enhanced when the face reappears within the boundaries of the same object. Keizer et al. (2008) demonstrated spontaneous integration between blended images of faces and houses. The present study extends these results in several important ways. First, the binding unit in the Mitroff et al. (2007) and the Keizer et al.

(2008) studies is the whole face itself, whereas in the present study, it is the face attribute (e.g., identity and emotion). The current study demonstrates that binding can occur at the mid-level of a spectrum expanding—on the one end—the whole face as the unit of integration, and—on the other end—the face attribute (e.g., identity) as a unit of integration. Second, the classification of face images is more difficult than that of house images. Faces share many overlapping low-level features and thus belong to the same basic-level category (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Consequently, within-faces discriminations should rely more heavily on the activation of long-term nodes.

A closer look at the contents of "faces files" in the current study reveals that they all consisted of pairwise bindings of task-relevant and response features. This type of integration was evident in each and every one of the experiments. Other visuo–motor bindings with task-irrelevant features were also prevalent, depending on whether those features shared a representational code (e.g., a spatial code in the case of gaze direction and response). These findings are commensurate with Hommel's claims concerning the likelihood of a feature to be integrated into an "event file" given its role in a particular task (Hommel, 2004). According to this idea, task-relevant feature dimensions are "intentionally weighted" (Hommel, Memelink, Zmigrod, & Colzato, 2014), and thus have better chances of being integrated. This can also explain why spatial location has been activated when gaze direction became the task-relevant dimension, but remained dormant when gaze direction stopped been relevant for task completion. The current results are also consistent with an embodied cognition view of face recognition. According to this approach, faces are embodied entities (Spivey, & Dale, 2004) that, in addition to conveying perceptual information, also activate a rich network of action programs in the viewers, depending on context.

One issue that deserves a comment concerns the possibility that the binding effects observed here capture some type of configural learning rather than the genuine integration of complex facial attributes. According to this argument, the small number of face stimuli used may have encouraged participants to respond to a learned configuration of the low-level features rather than to the criterial abstracted facial attribute. Several lines of evidence militate against such a possibility. First had there been any configuration effects, the outcome pattern should have been different than that observed. In the case that people respond to the configuration, the $2 \times 2 \times 2$ stimulus-feature design should have partitioned into two conditions; in one condition, S2 is the exact replica of S1, and in the second condition, it is not. This should have resulted in a four-way interaction, in which only exact repetitions speed up

response repetitions and slow down response alterations, while all seven other conditions do the opposite. Since the current findings do not seem to reflect such a case, they provide positive evidence against configurational processing. Another source of evidence against the possibility of configurational processing comes from studies that demonstrated that learning does not interact with binding effect (Colzato, Raffone, Hommel, 2006; Hommel, & Colzato, 2009). Colzato et al. (2006) examined how color-shape binding is affected by conjunction probabilities and learning. They found that the effects of binding and learning were independent of each other. There is good reason to believe that this is also the case here.

Binding of facial features and "person construal"

The results of the present investigation accord well with recent social cognition studies. In particular, the results fit nicely with research on "person construal". This burgeoning area of study investigates the low-level perceptual mechanisms that generate social phenomena (Fiske, & Neuberg, 1990; Freeman, & Ambady, 2011; Macrae et al., 1995). Studies in this domain have shown that when people categorize faces on a predefined category (e.g., gender), other social categories (e.g., age and race) are activated (Cloutier, Freeman, & Ambady, 2014; Freeman, 2014). To account for these findings, a dynamic model has been recently proposed by Freeman and Ambady, (2011). The model postulates interactive and simultaneous influences of bottom–up face processing of all possible category representations (e.g., male, female, White, Black), and top–down information sources (e.g., attentional states due to task demands). This mode of processing implies that all perceptual, cognitive, and motor attributes associated with the face are processed in parallel (Freeman et al., 2010, 2011). The model also predicts that the most important facial categories, those that are utmost relevant, or those that were recently active, will be activated more strongly.

A certain limitation of Freeman and Ambady (2011) model is that it is not clear regarding whether—beyond the interactive influences postulated—facial features are bound together in short-term memory, and if so, how. A recent study by Martin, Swainson, Slessor, Hutchison, Marosi, and Cunningham (2015) has yielded results that can speak directly to this point. When participants categorized faces on sex (i.e., man vs woman), repetitions and alternation of the previous trial's irrelevant face category (i.e., race and age) affected performance. Responses were faster when the relevant feature and the irrelevant feature repeated or alternated together than when one of the features alternated and other repeated. Martin et al. (2015) intention was not to study integration across social face attributes, although the

partial-repetition costs they have documented seem to support the notion of binding. These researchers have not couched their results in terms of feature binding, and their paradigm does not permit a clear dissociation between visual and motor components of bindings. This has become possible using the "event file" paradigm (Hommel, 1998) deployed in the current study. Future work should seek to reveal the mechanisms that govern the binding of action and perception of facial attributes.

Binding of facial features and the dual-route model

Haxby and his colleagues (Haxby et al., 2000, 2002; see also Bruce, & Young, 1986) have proposed a dual-route model of face recognition. In this model, the representations of facial dimensions (e.g., sex, expression, identity, and gaze direction) are distributed along two separate routes; one route is dedicated to the processing of invariant facial dimensions (e.g., facial identity), while the other route is responsible for the processing of variant facial dimensions (e.g., emotion and eye gaze). The model predicts the emergence of perceptual interactions between any two invariant (e.g., identity and gender) or between any two variant (e.g., gaze and emotion) facial attributes. In contrast, the dual-route model predicts independence between a variant and an invariant feature (e.g., identity and emotion). A vast literature has been dedicated to testing predictions from the dual-route model (Bartlett, Searcy, & Abdi, 2003; Calder et al., 2007; Le Gal, & Bruce, 2002; Fitousi, & Wenger, 2013; Soto, Vucovich, Musgrave, & Ashby, 2014). The accumulated bulk of evidence to date has been generally in agreement with the dual-route model (Haxby et al., 2000, 2002). However, recently, several studies have yielded results that are inconsistent with the dual-route model. In particular, violations of independence between facial identity and emotion have been reported (Fitousi, & Wenger, 2013; Soto et al., 2014; Yankouskaya, Booth, & Humphreys, 2012).

The results of the current investigation cannot be fully accommodated by the dual-route model; some of them are in agreement with the model, but some are much less so (see also Calder, & Young, 2005). As predicted by the dual-route model, variant and variant features, such as gaze direction and facial emotion, did interact in the binding process (see also Adams et al., 2003; Adams, & Kleck, 2003, 2005; Graham, & LaBar, 2007; Hietanen, & Leppänen, 2003), but that held true only when gaze direction was the task-relevant dimension. In contrast to the prediction of the dual-route model, it was found here that the variant and invariant facial attributes of emotion and identity do interact in the binding process. These results suggest that binding can take place within- and across the

two allegedly independent routes and that the binding process is not symmetric. These patterns cannot be fully accommodated by the dual-route model. Another limitation of the dual-route model (Haxby et al., 2000, 2002) is that it does not address the possible interactions between facial features and action codes. The current "face file" approach can guide future theorizing on the dual-route model in this respect. Future work should further study in more detail the formation and function of "face files".

Appendix

See Tables 5, 6, 7 and 8.

Table 5 ANOVA results of Experiment 1

Effect	df	MSE	F	p
Response (R)	1, 19	5400	1.64	
Identity (I)	1, 19	31	0	
Emotion (E)	1, 19	24,055	12.47	***
Location (L)	1, 19	535	0.05	
R × I	1, 19	184,767	43.43	***
R × E	1, 19	14.6	0	
R × L	1, 19	10,586	5.41	*
I × E	1, 19	11,712	4.60	*
I × L	1, 19	11,112	3.32	
E × L	1, 19	10,080	2.80	
I × E × L	1, 19	133	0.09	
R × I × E	1, 19	724	0.12	
R × I × L	1, 19	3853	1.84	
R × E × L	1, 19	2707	0.96	
R × I × E × L	1, 19	832	0.60	

$* \ p < 0.05$, $** \ p < 0.01$, $*** \ p < 0.005$

Table 6 ANOVA results of Experiment 2

Effect	df	MSE	F	p
Response (R)	1, 10	45,520.33		
Identity (I)	1, 10	3939	2.49	
Emotion (E)	1, 10	11,399	3.22	
Location (L)	1, 10	11,641	4.37	(*)
R × I	1, 10	19,912	10.64	**
R × E	1, 10	50,889	18.12	**
R × L	1, 10	3112	1.17	
I × E	1, 10	17,321	9.10	*
I × L	1, 10	6962	1.46	
E × L	1, 10	2489	1.26	
I × E × L	1, 10	149	0.09	
R × I × E	1, 10	18,503	5.28	*
R × I × L	1, 10	1947	0.34	
R × E × L	1, 10	240	0.08	
R × I × E × L	1, 10	104	0.80	

(*) $p < 0.06$, $* \ p < 0.05$, $** \ p < 0.01$, $*** \ p < 0.005$

Table 7 ANOVA results in Experiment 3

Effect	df	MSE	F	p
Response (R)	1, 16	10,177	0.93	
Emotion (E)	1, 16	4954	5.38	*
Gaze (G)	1, 16	12,157	3.84	
Location (L)	1, 16	3157	0.53	
R × E	1, 16	10,312	7.61	*
R × G	1, 16	16,394	7.08	*
R × L	1, 16	11,800	13.17	***
E × G	1, 16	10,451	9.92	***
E × L	1, 16	6449	4.45	(*)
G × L	1, 16	8464	4.31	(*)
E × G × L	1, 16	1012	0.83	
R × E × G	1, 16	5947	5.44	*
R × E × L	1, 16	0.4	0.01	
R × G × L	1, 16	840	0.71	
R × Ex G × L	1, 16	8907	2.90	

(*) $p < 0.06$, $* \ p < 0.05$, $** \ p < 0.01$, $*** \ p < 0.005$

Table 8 ANOVA results in Experiment 4

Effect	df	MSE	F	p
Response (R)	1, 16	78,870.25		
Emotion (E)	1, 16	65,877	10.05	**
Gaze (G)	1, 16	4031	1.14	
Location (L)	1, 16	19,621	3.02	
R × E	1, 16	140,273	33.76	***
R × G	1, 16	54,862	6.50	*

Table 8 continued

Effect	df	MSE	F	p
R × L	1, 16	9913	1.02	
E × G	1, 16	1374	0.44	
E × L	1, 16	519	0.14	
G × L	1, 16	6530	0.93	
E × G × L	1, 16	1428	0.38	
R × E × G	1, 16	2434	0.48	
R × E × L	1, 16	121	0.01	
R × G × L	1, 16	454	0.10	
R × Ex G × L	1, 16	284	0.02	

(*) $p < 0.06$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$

References

Adams, R. B., Gordon, H. L., Baird, A. A., Ambady, N., & Kleck, R. E. (2003). Effects of gaze on amygdala sensitivity to anger and fear faces. *Science, 300*, 1536.

Adams, R. B., & Kleck, R. E. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychological Science, 14*, 644–647.

Adams, R. B., & Kleck, R. E. (2005). Effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion, 5*, 3–11.

Bartlett, J. C., Searcy, J. H., & Abdi, H. (2003). What are the routes to face recognition. *Perception of Faces, Objects and Scenes: Analytic and Holistic Processes*, 21–52.

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., … Rosen, B. R. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron, 17*(5), 875–887.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*(3), 305–327.

Burton, A. M., Bruce, V., & Johnston, R. A. (1990). Understanding face recognition with an interactiveactivation model. *British Journal of Psychology, 81*(3), 361–380.

Burton, A. M., Kelly, S. W., & Bruce, V. (1998). Cross-domain repetition priming in person recognition. *The Quarterly Journal of Experimental Psychology: Section A, 51*(3), 515–529.

Calder, A. J., Beaver, J. D., Winston, J. S., Dolan, R. J., Jenkins, R., Eger, E., & Henson, R. N. (2007). Separate coding of different gaze directions in the superior temporal sulcus and inferior parietal lobule. *Current Biology, 17*(1), 20–25.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience, 6*(8), 641–651.

Cloutier, J., Freeman, J. B., & Ambady, N. (2014). Investigating the early stages of person perception: The asymmetry of social categorization by sex vs. age. *PLoS One, 9*, e84677.

Colzato, L. S., Raffone, A., & Hommel, B. (2006). What do we learn from binding features? Evidence for multilevel feature integration. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 705–716.

Ellis, A. W., Young, A. W., Flude, B. M., & Hay, D. C. (1987). Repetition priming of face recognition. *The Quarterly Journal of Experimental Psychology, 39*(2), 193–210.

Engell, A. D., & Haxby, J. V. (2007). Facial expression and gaze-direction in human superior temporal sulcus. *Neuropsychologia, 45*(14), 3234–3241.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is ''special'' about face perception? *Psychological Review, 105*(3), 482–498.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*(1), 1–47.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology, 23*, 1–74.

Fitousi, D. (2013). Mutual information, perceptual independence, and holistic face perception. *Attention Perception and Psychophysics, 75*, 983–1000.

Fitousi, D. (2015). Composite faces are not processed holistically: Evidence from the Garner and redundant target paradigms. *Attention Perception and Psychophysics, 77*, 2037–2060.

Fitousi, D. (2016). Comparing the role of selective and divided attention in the composite face effect: Insights from Attention Operating Characteristic (AOC) plots and cross-contingency correlations. *Cognition, 148*, 34–46.

Fitousi, D., & Wenger, M. J. (2013). Variants of independence in the perception of facial identity and expression. *Journal of Experimental Psychology: Human Perception, and Performance, 39*, 133–155.

Freeman, J. B. (2014). Abrupt category shifts during real-time person perception. *Psychonomic Bulletin and Review, 21*(1), 85–92.

Freeman, J. B., & Ambady, N. (2009). Motions of the hand expose the partial and parallel activation of stereotypes. *Psychological Science, 20*(10), 1183–1188.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review, 118*(2), 247–279.

Freeman, J. B., Ambady, N., Midgley, K. J., & Holcomb, P. J. (2011). The real-time link between person perception and action: Brain potential evidence for dynamic continuity. *Social Neuroscience, 6*(2), 139–155.

Freeman, J. B., Pauker, K., Apfelbaum, E. P., & Ambady, N. (2010). Continuous dynamics in the real-time perception of race. *Journal of Experimental Social Psychology, 46*(1), 179–185.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin and Review, 5*(3), 490–495.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological bulletin, 133*(4), 694–724.

Gordon, R. D., & Irwin, D. E. (1996). What's in an object file? Evidence from priming studies. *Perception and Psychophysics, 58*(8), 1260–1277.

Graham, R., & LaBar, K. S. (2007). Garner interference reveals dependencies between emotional expression and gaze in face perception. *Emotion, 7*(2), 296–313.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*(6), 223–233.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry, 51*(1), 59–67.

Henderson, J. M. (1994). Two representational systems in dynamic visual identification. *Journal of Experimental Psychology: General, 123*, 410–426.

Hietanen, J. K., & Leppänen, J. M. (2003). Does facial expression affect attention orienting by gaze direction cues? *Journal of Experimental Psychology: Human Perception and Performance, 29*, 1228–1243.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience, 3*(1), 80–84.

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus–response episodes. *Visual Cognition, 5,* 183–216.

Hommel, B. (2000). The prepared reflex: Automaticity and control in stimulus–response translation. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 247–273). Cambridge: MIT Press.

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences, 8,* 494–500.

Hommel, B. (2005). How much attention does an event file need? *Journal of Experimental Psychology: Human Perception and Performance, 31*(5), 1067–1082.

Hommel, B., & Colzato, L. S. (2009). When an object is more than a binding of its features: Evidence for two mechanisms of visual feature integration. *Visual Cognition, 17,* 120–140.

Hommel, B., Memelink, J., Zmigrod, S., & Colzato, L. S. (2014). Attentional control of the creation and retrieval of stimulus–response bindings. *Psychological Research, 78*(4), 520–538.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). Codes and their vicissitudes. *Behavioral and Brain Sciences, 24*(05), 910–538926.

Hubel, D. H., & Wiesel, T. N. (1977). Ferrier lecture: Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London B: Biological Sciences, 198*(1130), 1–59.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 29–61). New York: Academic Press.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 174–219.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience, 17*(11), 4302–4311.

Keizer, A. W., Colzato, L. S., & Hommel, B. (2008). Integrating faces, houses, motion, and action: Spontaneous binding across ventral and dorsal processing streams. *Acta Psychologica, 127,* 177–185.

Le Gal, P. M., & Bruce, V. (2002). Evaluating the independence of sex and expression in judgments of faces. *Perception and Psychophysics, 64*(2), 230–243.

Livingstone, M. S., & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *Journal of Neuroscience, 7*(11), 3416–3468.

Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science, 240*(4853), 740–749.

Lundqvist, D., Flykt, A., & Ohman, A. (1998). The Karolinska Directed Emotional Face—KDEF, CD ROM from Department of Clinical Neuroscience. Psychology section, Karolinska Instituet. ISBN 91-6307164.

Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology, 69*(3), 397–407.

Martin, D., Swainson, R., Slessor, G., Hutchison, J., Marosi, D., & Cunningham, S. J. (2015). The simultaneous extraction of multiple social categories from unfamiliar faces. *Journal of Experimental Social Psychology, 60,* 51–58.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748.

Mitroff, S. R., Scholl, B. J., & Noles, N. S. (2007). Object files can be purely episodic. *Perception, 36,* 1730–1736.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. *Attention and Performance X: Control of Language Processes, 32,* 531–556.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach.* Cambridge: MIT Press.

Rolls, E. T., & Tovee, M. J. (1995). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology, 73*(2), 713–726.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3), 382–439.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience, 18*(1), 555–586.

Soto, F. A., Vucovich, L., Musgrave, R., & Ashby, F. G. (2014). General recognition theory with individual differences: A new method for examining perceptual and decisional interactions with an application to face perception. *Psychonomic Bulletin and Review, 22*(1), 88–111.

Spivey, M. J., & Dale, R. (2004). On the continuity of mind: Toward a dynamical account of cognition. *Psychology of Learning and Motivation, 45,* 87–142.

Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature, 400*(6747), 869–873.

Tipper, S. P., Weaver, B., Jerreat, L. M., & Burak, A. L. (1994). Object-based and environment-based inhibition of return of visual attention. *Journal of Experimental Psychology: Human Perception and Performance, 20*(3), 478–499.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology, 6*(2), 171–178.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14*(1), 107–141.

van Dam, W. O., & Hommel, B. (2010). How object-specific are object files? Evidence for integration by location. *Journal of Experimental Psychology: Human Perception and Performance, 36*(5), 1184–1192.

Von der Malsburg, C. (1999). The what and why of binding: The modeler's perspective. *Neuron, 24*(1), 95–104.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRI study. *Neuron, 30*(3), 829–841.

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research, 37*(1), 25–43.

Yankouskaya, A., Booth, D. A., & Humphreys, G. (2012). Interactions between facial emotion and identity in face processing: Evidence based on redundancy gains. *Attention, Perception, and Psychophysics, 74*(8), 1692–1711.

Young, M. P., & Yamane, S. (1992). Sparse population coding of faces in the inferotemporal cortex. *Science, 256*(5061), 1327–1331.

Zmigrod, S., de Sonneville, L. M., Colzato, L. S., Swaab, H., & Hommel, B. (2013). Cognitive control of feature bindings: Evidence from children with autistic spectrum disorder. *Psychological Research, 77*(2), 147–154.