

Integrating faces, houses, motion, and action: Spontaneous binding across ventral and dorsal processing streams

André W. Keizer^{*}, Lorenza S. Colzato, Bernhard Hommel

Leiden University Institute for Psychological Research & Leiden Institute for Brain and Cognition Leiden, The Netherlands

Received 12 January 2007; received in revised form 16 April 2007; accepted 21 April 2007

Available online 6 June 2007

Abstract

Perceiving an event requires the integration of its features across numerous brain maps and modules. Visual object perception is thought to be mediated by a ventral processing stream running from occipital to inferotemporal cortex, whereas most spatial processing and action control is attributed to the dorsal stream connecting occipital, parietal, and frontal cortex. Here we show that integration operates not only on ventral features and objects, such as faces and houses, but also across ventral and dorsal pathways, binding faces and houses to motion and manual action. Furthermore, these bindings seem to persist over time, as they influenced performance on future task-relevant visual stimuli. This is reflected by longer reaction times for repeating one, but alternating other features in a sequence, compared to complete repetition or alternation of features. Our findings are inconsistent with the notion that the dorsal stream is operating exclusively online and has no access to memory.

© 2007 Elsevier B.V. All rights reserved.

PsycINFO classification: 2340

Keywords: Feature integration; Binding problem; Object file; Ventral; Dorsal

1. Introduction

Processing a visual object in the human brain involves numerous functionally and spatially distinct cortical areas. For instance, the shape and color of an object are coded in dedicated feature maps in V1-4, the features of a face in motion are registered in the fusiform face area (FFA) (Kanwisher, McDermott, & Chun, 1997) and the motion-sensitive area MT/MST (Tootell et al., 1995; Zeki et al., 1991) while a house or landscape will be coded in the Parahippocampal Place Area (PPA) (Epstein & Kanwisher, 1998). This form of distributed processing creates multiple

binding problems (Treisman & Gelade, 1980; Treisman & Schmidt, 1982) which call for some kind of integration.

A well established method to indicate what kind of information is integrated under what circumstances is the analysis of interactions between sequential effects. The logic is straightforward: if the codes of two given features or objects have been bound together they should from then on act as a pair. If so, reactivating one of the codes (through repeating the corresponding stimulus) should reactivate the other code as well, even if the two coded features are uncorrelated and co-occurred only once. An implication of this mechanism would also be that performance is impaired if one member of the pair is repeated but the other is not. Indeed, repeating the shape of an object but changing its color or location produces slower reaction times (RTs) and more errors than repeating both features or repeating none (Hommel, 1998; Hommel, Proctor, & Vu, 2004; Kahneman, Treisman, & Gibbs, 1992) suggesting that processing an object leads to the

^{*} Corresponding author. Address: Leiden University, Department of Psychology, Cognitive Psychology Unit, Postbus 9555, 2300 RB Leiden, The Netherlands. Tel.: +31 0 71 5273655.

E-mail address: akeizer@fsw.leidenuniv.nl (A.W. Keizer).

spontaneous binding of the neural codes of its features. Interestingly, this logic also seems to apply to perception-action associations: repeating an object feature but changing the action it accompanies produces worse performance than repeating both or neither (Hommel, 1998, 2004) suggesting that stimulus features get bound to the actions they “afford”.

The object features investigated so far in research on binding phenomena, such as shape, color, or allocentric location, can all be considered to be processed by ventral pathways in the human brain (Goodale & Milner, 1992; Milner & Goodale, 1995; Neggers, Van der Lubbe, Ramsey, & Postma, 2006). Ventral pathways are commonly distinguished from dorsal pathways in terms of the information they process. Whereas earlier approaches associated visual ventral and dorsal pathways with the processing of nonspatial (what) and spatial (where) information, respectively (Ungerleider & Mishkin, 1982) more recent accounts assume that ventral pathways process information necessary for object perception, whereas dorsal pathways process action-relevant information (Creem & Proffitt, 2001; Goodale & Milner, 1992; Milner & Goodale, 1995). Importantly, dorsal pathways are assumed to operate exclusively online and, thus, to have no memory (beyond a few milliseconds, see Milner & Goodale, 1995). For instance, Cant, Westwood, Valyeara, and Goodale (2005) found that visually guided actions were not influenced by previewing the goal object, while memory guided actions were. They argue that visually guided actions are entirely fed by dorsal pathways, which because of their nonexistent short-term memory capacity cannot maintain information necessary to produce priming effects. In contrast, memory guided actions involve ventral pathways that do have sufficient short-term memory capacity.

Considering that binding stimulus (and/or response) features can only affect later performance if the binding is maintained, the apparently different memory characteristics of ventral and dorsal pathways raise the question whether binding takes place across dorsal and ventral pathways at all and/or whether such bindings can be maintained long enough to affect performance a second or more (the typical interval between prime and probe in binding studies) later. We investigated this issue by testing whether binding effects can be demonstrated between visual object features (or even whole objects) that are presumably processed in different pathways. In particular, we tested whether motion (a dorsal feature) can be bound to faces and houses (ventral features), and to manual responses. We carried out four experiments using the standard paradigm introduced by Hommel (1998). Given that our crucial experiments, 3 and 4, used faces and houses as “ventral” stimuli, and given that these stimuli were never used in sequential studies before, we first ran two more experiments (1–2) to make sure that the previous demonstrations of spontaneous binding between shape, color, and location extend to these more complex stimuli.

2. Experiments 1 and 2

We used two modified versions of the S1–S2 paradigm introduced by Hommel (1998; for an overview, see Hommel, 2004). In the task employed in Experiment 1, subjects are confronted with two objects, separated in time by a short interval, and they respond to one feature of the second object (S2) while ignoring the first (S1). As discussed before, such setups create (typically binary) interactions that are indicative of feature integration processes: repeating one of two features but not the other yields worse performance than repeating both or none (Hommel, 1998). In Experiment 1, we presented blended face-house compounds as S1 and S2, and S2 could repeat or alternate the picture of the face and the picture of the house to create an orthogonal 2×2 design.

As already discussed, integration can also include the response, leading to interactions between stimulus (feature) repetition and response repetition (i.e., better performance if stimulus and response are both repeated or both alternated). To investigate whether this pattern extends to faces and houses, participants in Experiment 2 were to respond to S1 by means of a precued manual reaction (R1; see Hommel, 1998 and Fig. 1). This design creates temporal overlap between S1 and R1 (which is known to be a sufficient condition for integration: Hommel, 2004) without making R1 contingent on S1, which allows for the orthogonal manipulation of stimulus and response repetition.

2.1. Methods

2.1.1. Participants

22 and 20 healthy, young undergraduates participated in Experiment 1 and Experiment 2, respectively. All subjects participated in exchange for course credit or money.

2.1.2. Stimuli and task

Following O’Craven, Downing, and Kanwisher (1999) each stimulus was composed by transparently superimposing one of eight grayscale front-view photographs of male (4) and female (4) faces on one of eight grayscale photographs of houses. The images were cropped to fit a square size (10° by 10°). All images were adjusted to assure the same average luminance. The house-face combinations of the 128 trials of Experiment 1 were constructed by randomly drawing from the eight possible houses and faces, except that the stimuli were chosen to result in equal proportions (32 trials) in the four cells of the 2×2 analytical design (house repetition vs. alternation \times face repetition vs. alternation). The trials of Experiment 2 were composed the same way, except that adding the response-repetition manipulation increased the design cells to eight (house repetition vs. alternation \times face repetition vs. alternation \times response repetition vs. alternation) and the number of trials to 256.

In Experiment 1 (see Fig. 1a), subjects were presented with a picture of a face transparently superimposed on a

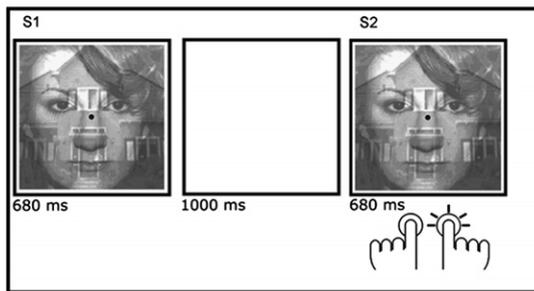
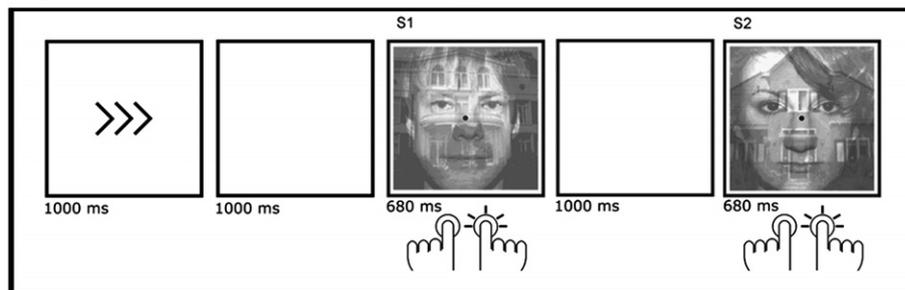
Experiment 1**Experiment 2**

Fig. 1. Overview of the display and timing of events in Experiments 1 and 2.

house, twice within a single trial and they were instructed to make a discriminative response (R2) to the gender of the second stimulus (S2). Half of the participants responded to the male and the female face by pressing the left and right key of a computer keyboard, respectively, while the other half received the opposite mapping. S1 appeared for 680 ms, followed by a blank interval of 1000 ms. S2 appeared and stayed until the response was given or 2000 ms had passed. S2 was followed by a fixation circle (diameter: 0.5°), which stayed for a randomly chosen duration of between 1000 and 2500 ms (varied in 100-ms steps). If the response was incorrect, auditory feedback was presented.

The procedure of Experiment 2 was the same, with the following exceptions. Participants carried out two responses per trial. R1 was a simple reaction with the left or right key, as indicated by a 1000-ms response cue (three arrows pointing either leftward or rightward) appearing 2000 ms before S1. R1 was to be carried out as soon as S1 appeared, disregarding S1's attributes. As in Experiment 1, R2 was a binary-choice reaction to the gender of S2.

2.2. Results and discussion

RTs and error rates were analyzed by means of repeated-measures ANOVAs with the factors face repetition (vs. alternation) and house repetition in Experiment 1, and with face repetition, house repetition, and response repetition in Experiment 2. The RTs revealed a main effect of face repetition in Experiment 2, $F(1,19) = 73.918$, $p < .001$. More importantly, there were significant interactions between face repetition and house repetition in

Experiment 1 (Fig. 2a), $F(1,21) = 13.373$, $p < .01$ and in Experiment 2 (Fig. 2b), $F(1,19) = 6.831$, $p < .05$, indicating significantly faster RTs when both features were repeated or alternated, as compared to when only one was repeated but the other alternated. Moreover, Experiment 2 provides evidence for binding between faces and responses, as indicated by the significant interaction between face repetition and response repetition, $F(1,19) = 30.184$, $p < .001$.

Error rates showed comparable results: a main effect of face repetition was obtained in Experiment 1, $F(1,19) = 20.958$, $p < .001$, and in Experiment 2, $F(1,21) = 11.059$, $p < .01$, a response repetition effect in Experiment 2, $F(1,19) = 5.208$, $p < .05$, and a significant interaction between face and response repetition in Experiment 2, $F(1,19) = 47.805$, $p < .001$.

Experiments 1 and 2 provided evidence for the spontaneous integration of blended faces and houses: repeating a face was beneficial if the house was also repeated, but turned into a cost if the house was also changed. Hence, the mere co-occurrence of a face and house was sufficient to create a binding between their representations. Furthermore, Experiment 2 provides evidence for a binding between the task-relevant stimulus feature (face) and the response, even though the latter was not determined, but only triggered by the former. This extends previous findings of stimulus-response integration obtained with simpler stimuli, but it also shows that face-house compounds were not treated as a single stimulus. If they were, the hint to the integration of faces and houses would be of less theoretical interest—even though this would fail to explain why “complete alternations” were not associated with the worst

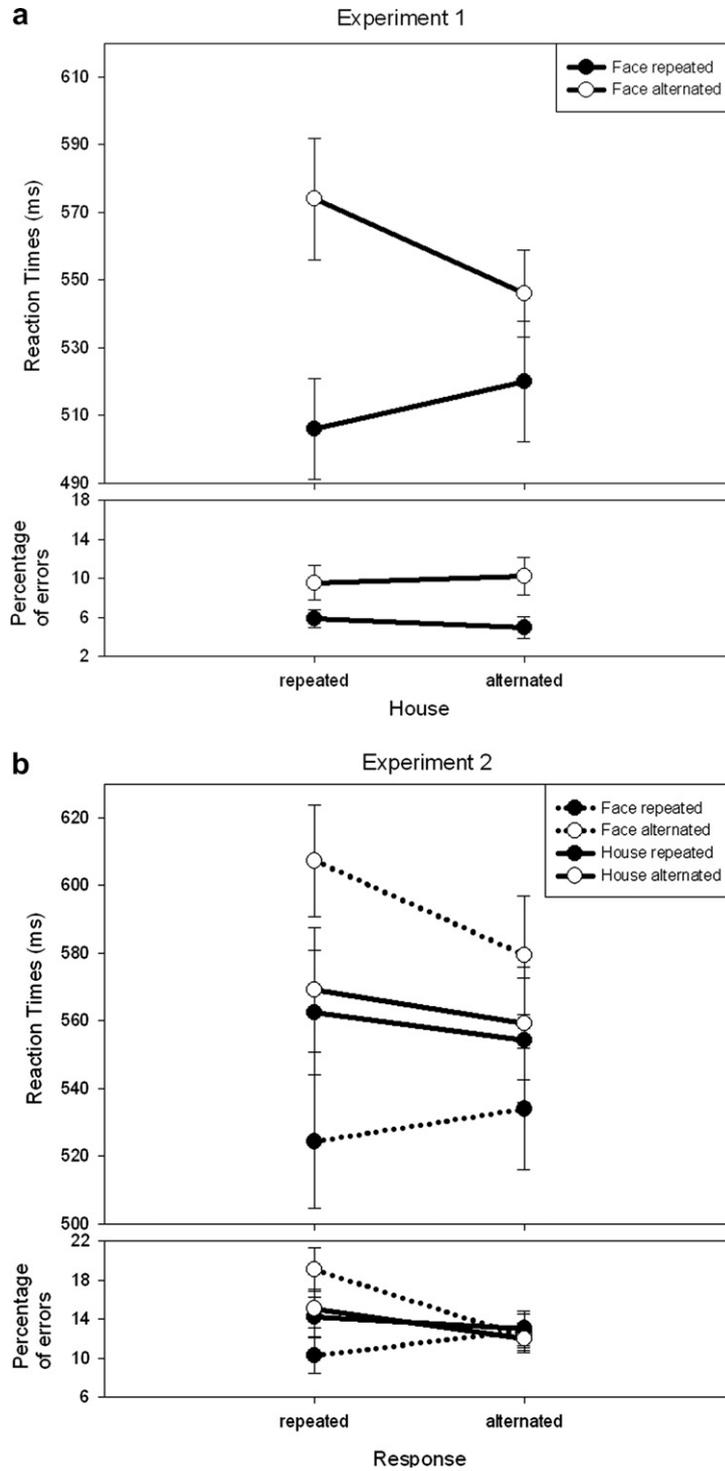


Fig. 2. Error bars represent standard errors in all graphs. (a) and (b) Mean reaction times and error percentages for Experiments 1–2, as a function of repetition vs. alternation of stimulus face and stimulus house (Experiment 1), or of stimulus face, stimulus house, and response (Experiment 2). (c) and (d) Mean reaction times and error percentages for Experiments 3–4, as a function of repetition vs. alternation of stimulus motion and the moving object (face or house; Experiment 3), or of stimulus motion, moving object, and response (Experiment 4).

performance. Also in line with previous findings (Hommel, 1998), sensorimotor integration was restricted to the task-relevant stimulus information (faces), suggesting that the creation and/or the retrieval of bindings is under attentional control (Hommel, 2007).

3. Experiments 3 and 4

Experiments 3 and 4 studied whether bindings can link information processed in ventral pathways with motion, which is processed in the dorsal system MT/MST (Tootell

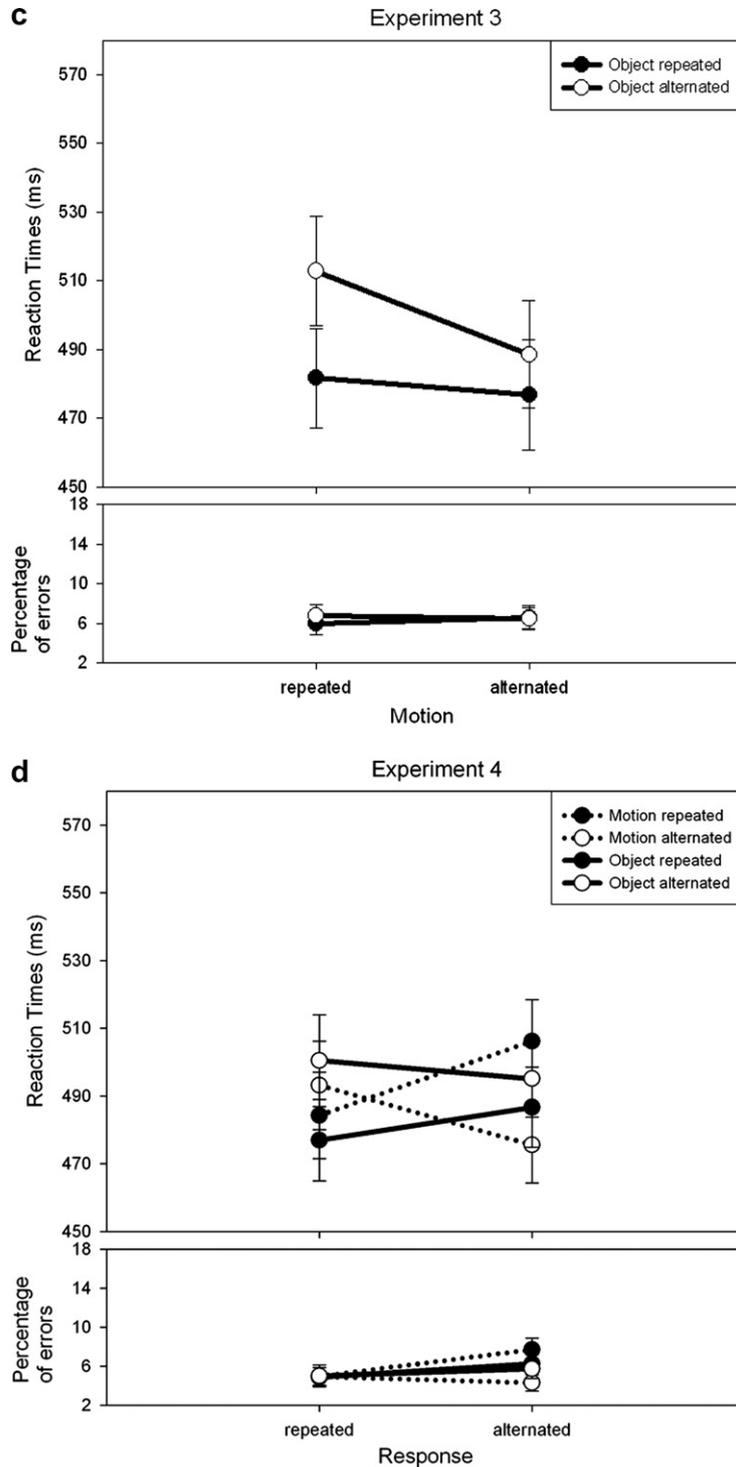


Fig. 2 (continued)

et al., 1995). We still presented face-house compounds, but now faces and houses were always identical in S1 and S2 and either one or the other was continuously oscillating on a diagonal path. In Experiment 3, participants responded to the motion direction of S2 but were to ignore S1 altogether. In S1 and S2 the moving object could be the face or the house, and it could move on one or the other diagonal, so that the moving object and the direction of

the motion could repeat or alternate. If encountering S1 would lead to the spontaneous integration of object and motion, repeating the object but not the motion, or repeating the motion but not the object, should lead to worse performance than complete repetitions or alternations. Experiment 4 added a precued response (R1) to the onset of S1, analogous to the design of Experiment 2. Here we expected the integration of the task-relevant stimulus

feature (motion) and the response, as indicated by an interaction between motion repetition and response repetition.

3.1. Methods

3.1.1. Participants

19 and 20 young, healthy undergraduates participated in Experiments 3 and 4, respectively.

3.1.2. Stimuli and task

The procedure of Experiment 3 was as in Experiment 1, with the following exceptions. Faces and houses were always the same for S1 and S2. Either the face or the house oscillated in a straight path on one of two possible non-cardinal directions (left-up/right-down vs. right-up/left-down), while the total size of the combined images remained the same (10° by 10°). The maximal displacement caused by the motion was less than 10% of the size of the image. The moving image oscillated 2.5 cycles with a constant speed of 9° per second. Subjects performed left-right key presses (R2) to the direction of the motion of S2, disregarding the moving object of S2 and the object and motion of S1. After every seven trials, a fixation circle was presented for 10 s. This rest period was included to allow for a later transfer of exactly the same task to a planned fMRI-study (Keizer, Colzato, Teeuwisse, Rombouts, & Hommel, submitted for publication) where such rest periods are needed to prevent non-linearity effects of the BOLD-signal. The procedure of Experiment 4 was the same, except that they performed a pre-cued, simple response to the onset of S1, just like in Experiment 2. Experiment 3 comprised 182 trials, which were randomly drawn from all combinations of the eight possible houses and faces (the particular combination was identical for S1 and S2), the two possible motions for S1 and S2, and the two possible objects that could move (either face or house); however, the stimuli were chosen to result in roughly equal proportions (averages ranging between 45 and 46) in the four cells of the 2×2 analytical design (repetition vs. alternation of motion \times repetition vs. alternation of moving object). Experiment 4 comprised 378 trials, randomly drawn from all combinations used in Experiment 3 plus the repetition vs. alternation of the response. The stimuli were chosen to result in roughly equal proportions (averages ranging between 45 and 49) in the eight cells of the $2 \times 2 \times 2$ analytical design.

3.2. Results and discussion

Main effects on RTs were obtained for motion repetition in Experiment 3, $F(1, 18) = 9.709$, $p < .01$, and Experiment 4, $F(1, 19) = 10.956$, $p < .01$, and for (moving-) object repetition in Experiment 3, $F(1, 18) = 49.901$, $p < .001$, and Experiment 4, $F(1, 19) = 52.122$, $p < .001$. More importantly, reliable interactions between motion repetition and object repetition provided evidence for visual integration in Experiment 3, $F(1, 18) = 5.752$, $p < .05$, and in

Experiment 4, $F(1, 19) = 12.779$, $p < .01$. Separate analyses showed that it did not matter whether a face or a house was integrated with motion on S1, as indicated by an absence of a three-way interaction between the object that moved on S1 (face or house), repetition or alteration of the object that moved on S2 and repetition or alteration of the direction of motion on S2 in Experiment 3, $F(1, 18) < 1$, and in Experiment 4, $F(1, 19) < 1$. Experiment 4 points to the binding of motion and response, as indicated by the interaction between motion repetition and response repetition, $F(1, 19) = 34.637$, $p < .001$. Even though less pronounced, the interaction between object repetition and response repetition was also significant, $F(1, 19) = 6.553$, $p < .05$. Error rates of Experiment 3 did not yield reliable results and the errors of Experiment 4 showed a significant interaction between motion and response $F(1, 19) = 9.844$, $p < .01$.

The results show significant binding between motion and the object that moved. This demonstrates that bindings between ventral and dorsal features can be created in principle and, what is more, that such bindings actually are spontaneously created even if integration is not required by the task. Experiment 4 included a response to the first stimulus, following the same logic as Experiment 2. Apart from replicating the face-motion and house-motion interactions, we found evidence for bindings between motion and response and between the moving object (be it face or house) and the response.

Interestingly, the results suggest that faces were integrated with motion in the same way as houses were. Considering that both houses and faces were not task-relevant, this outcome pattern is in line with the findings of O'Craven et al. (1999) and their claim that visual attention spreads from task-relevant features of an attended object (motion in our case) to the task-irrelevant features of that object. Apparently, then, this object-specific attentional spreading does not only affect online processing, as studied by O'Craven et al. (1999) but also affect the creation and maintenance of feature bindings. The observation that faces and houses were comparable in this respect is particularly relevant in view of claims that face information may be processed differently than house information. Even though it is clear that cortical face- and house-related areas (FFA and PPA) are both located in the ventral stream (Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999) it has been argued that especially faces may be processed more holistically than places or objects are (Farah, 1996). This raises the question whether faces are integrated with other features just like house features are—a question to which our observations provide an affirmative answer.

4. General discussion

The first experiment extended previous demonstrations of bindings between simple features, such as shape, color, or relative location, to complex stimuli, such as faces and houses. These findings bear significance with regard to the scope of the concept of event files (Hommel, 1998,

2004) in particular, but also for the related Theory of Event Coding (TEC, Hommel, Müsseler, Aschersleben, & Prinz, 2001). The observations that motivated and supported the event-file concept were commonly related to simple features, such as line orientations and color patches, but the present findings show that the same logic applies to more complex stimulus configurations, such as faces and houses. One may ask whether stimuli like faces and houses can be still described as features, since these stimuli are composites of numerous simple features and may therefore be more accurately described as event files themselves. If so, we can conclude that event file logic seems to apply to several levels of stimulus representation, ranging from individual features to composites. Hence, event files can apparently enter new ‘higher order’ bindings with other event files. This possibility is also suggested by the findings of Waszak, Hommel, and Allport (2003). They found that when subjects were presented with pictures and overlapping words, they found it more difficult to switch from one task to another when the concrete stimulus had already appeared in the alternative task. It seems that stimuli and stimulus compounds can be bound to a specific task context, which is reactivated automatically when the stimulus material is repeated. Future research may determine if it is possible to distinguish between different hierarchies of bindings or even binding mechanisms.

Experiment 2 confirmed that complex stimuli also enter sensorimotor bindings, and our findings showed consistently that feature binding seems to cross border between ventral and dorsal processing pathways. Experiment 3 provided evidence that motion is automatically integrated into enduring object representations and, as confirmed by Experiment 4, into sensorimotor event representations.

One may argue that at least some of our findings (Experiment 2) may not necessarily reflect binding across ventral and dorsal pathways but integration at earlier stages of visual processing (before the ventral–dorsal split), e.g., involving the thalamic nuclei and/or V1/V2. However, there are several reasons to discount this possibility. First, the results from Experiment 2 seem to suggest that face-house compounds were treated as consisting of two distinct objects, as faces selectively formed a persistent binding with action while houses did not. Second, a recent fMRI study of ours (Keizer et al., submitted for publication) showed that encountering a face moving into a particular direction after having seen a house moving into the same direction leads to an increase in activation of the PPA—the area coding house information. This suggests that processing a particular motion direction automatically retrieved the stimulus that just moved in the same direction, which again implies that a binding between this motion and that previous stimulus has been created. Reactivating this binding reactivates PPA, but not earlier visual areas, which strongly suggests that the binding includes information from both dorsal and ventral pathways.

Taken altogether, our findings thus suggest that stimulus information coded in the ventral stream is automati-

cally integrated with information coded in the dorsal stream, and both types of information can be integrated with temporarily overlapping actions. The integration process creates memory structures that survive at least one second (the time between the presentations of the two stimuli in our experiments), and there are reasons to believe that this is a conservative estimate (Hommel & Colzato, 2004). Primate studies have shown that the dorsal area MT/MST projects to the ventral area V4 (Maunsell & Van Essen, 1983; Ungerleider & Desimone, 1986) and it has been suggested that this projection allows for the recognition of the semantic characteristics of biological motion (Oram & Perrett, 1994; Perret, Harries, Benson, Chitty, & Mistlin, 1990) or form defined by motion (Sary, Vogels, & Orban, 1993). Our findings suggest a far more extensive and reciprocal connectivity between dorsal and ventral processing, connectivity that apparently allows for the fast and automatic integration of information about ventral and dorsal aspects of perception and action. Thus, even though physiological findings suggest that visual information processing is distributed across two anatomically separable streams, our present observations show that this separation by no means implies poor communication between them.

Our observations also question the characterization of the dorsal stream as exclusively online and as lacking memory beyond a few milliseconds (Cant et al., 2005; Goodale & Milner, 1992; Milner & Goodale, 1995). This does not necessarily contradict the claim that the dorsal stream is particularly well-suited to inform ongoing action (Hommel et al., 2001), but it does show that dorsally coded information is involved in off-line processing and in the integration of perception and action. Our findings are in accordance with studies showing priming effects of visual motion (Campana, Cowey, & Walsh, 2002; Pinkus & Pantle, 1997) i.e., of a feature processed in the dorsal stream; (Tootell et al., 1995) suggesting that dorsally coded information can be retained for a nontrivial period of time. In addition, Chun and Jiang (1999) studied the effect of predictable, but irrelevant motion patterns of items in a search display (one target item among distractor items). They found that subjects were apparently able to use these consistencies, as target localization reaction times were faster when all items moved in a predictable manner versus an unpredictable manner. It seems that the subjects formed long-term associations between particular motion patterns and the items in the search display, which would require integration of form and motion. Our results show that these findings can be extended to online, single-trial integration of complex forms (faces and houses) and motion. A phenomenon called the ‘McGurk aftereffect’ can also be explained in a similar way. When subjects are presented with a sound and an incongruent mouth movement, the perception of the sound is modulated by this mouth movement to produce the well-known McGurk effect (McGurk & MacDonal, 1976). Bertelson, Vroomen, and de Gelder (2003) showed that the perception of a subsequent presentation

of that same sound in isolation is still modulated by the mouth movement that accompanied the sound in the initial presentation; the McGurk aftereffect. Apparently, mouth movement and sound can form an enduring association, which results in retrieval of the mouth movement when its associated sound is presented in isolation.

Soto-Faraco, Spence, and Kingstone (2005) showed that the integration between sound and motion occurs automatically, which suggests that the McGurk aftereffect found by Bertelson et al. (2003) is not due to top-down influences (see also Vatakis & Spence, *in press*, for a related discussion).

This raises the question of why Cant et al. (2005) observed priming effects for memory guided, but not for visually guided actions. As the authors themselves acknowledge, the conclusions of Cant et al. (2005) are based on a null effect, which makes it difficult to exclude the possibility that memory guided actions are only more sensitive to priming effects than visually guided actions are—which, given the fact that continuous visual input can easily overwrite the contents of the visual short-term memory buffer, is not implausible. Also, it is theoretically possible that visual guided actions are processed via the dorsal stream, but that they are functionally distinct from the dorsal features that were used in the current study (motion and responses). This may be so for the motion-sensitive area MT, because of its previously discussed connections with ventral area V4. Moreover, the responses used in our study may be inherently different than the visual guided actions used by Cant et al. (2005) as the former may be based on a relatively more semantic judgment. If this is indeed the case and visual guided actions cannot be bound to ventral, or other dorsal features like motion and the actions used in our study, the conclusions of Cant et al. (2005) would need to be moderated accordingly.

Acknowledgements

We thank M. Spapé for programming parts of the experiments and N. Kanwisher and P. E. Downing for providing the parameters necessary for constructing the images.

References

- Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science*, *14*, 592–597.
- Campana, G., Cowey, A., & Walsh, V. (2002). Priming of motion direction and area v5/mt: A test of perceptual memory. *Cerebral Cortex*, *12*, 1047–1051.
- Cant, J. S., Westwood, D. A., Valyeara, K. F., & Goodale, M. A. (2005). No evidence for visuomotor priming in a visually guided action task. *Neuropsychologia*, *43*, 216–226.
- Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*, 360–365.
- Creem, S. H., & Proffitt, D. R. (2001). Defining the cortical visual systems: “What”, “Where”, and “How”. *Acta Psychologica*, *107*(1–3), 43–68.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601.
- Farah, M. J. (1996). Is face recognition ‘special’? Evidence from neuropsychology. *Behavioral Brain Research*, *76*, 181–189.
- Goodale, M. A., & Milner, D. A. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*, 20–25.
- Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus–response episodes. *Visual Cognition*, *5*, 183–216.
- Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences*, *8*, 494–500.
- Hommel, B., & Colzato, L. S. (2004). Visual attention and the temporal dynamics of feature dynamics of feature integration. *Visual Cognition*, *11*, 483–521.
- Hommel, B. (2007). Feature integration across perception and action: Event files affect response choice. *Psychological Research*, *71*, 42–63.
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*, 849–937.
- Hommel, B., Proctor, R. W., & Vu, K. P. L. (2004). A feature-integration account of sequential effects in the Simon task. *Psychological Research*, *68*, 1–17.
- Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 9379–9384.
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175–219.
- 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*, 4302–4311.
- Keizer, A.W., Colzato, L.S., Teeuwisse, W., Rombouts, S.A.R.B. & Hommel, B. (submitted for publication). When moving faces activate the house area: An fMRI study of object file retrieval.
- Maunsell, J. H. R., & Van Essen, D. C. (1983). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *The Journal of Neuroscience*, *3*, 2563–2586.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.
- Milner, D. A., & Goodale, M. A. (1995). *The visual brain in action*. NY: Oxford University Press.
- Neggers, S. F. W., Van der Lubbe, R. H. J., Ramsey, N. F., & Postma, A. (2006). Interactions between ego- and allocentric neuronal representations of space. *NeuroImage*, *31*, 320–331.
- O’Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*, 584–587.
- Oram, M. W., & Perrett, D. I. (1994). Responses of anterior superior temporal polysensory (STPA) neurons to biological motion stimuli. *Journal of Cognitive Neuroscience*, *2*, 99–116.
- Perrett, D. I., Harries, M. H., Benson, P. J., Chitty, A. J., & Mistlin, A. J. (1990). Retrieval of structure from rigid and biological motion: An analysis of the visual responses of neurones in the macaque temporal cortex. In A. Blake & T. Troscianko (Eds.), *AI and the eye* (pp. 181–199). London: Wiley.
- Pinkus, A., & Pantle, A. (1997). Probing visual motion signals with a priming paradigm. *Vision Research*, *37*(5), 541–552.
- Sary, G., Vogels, R., & Orban, G. A. (1993). Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science*, *260*, 995–997.
- Soto-Faraco, S., Spence, C., & Kingstone, A. (2005). Assessing automaticity in the audiovisual integration of motion. *Acta Psychologica*, *118*, 71–92.
- Tootell, R. B. H., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., et al. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *The Journal of Neuroscience*, *15*, 3215–3230.

- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.
- Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*, 107–141.
- Ungerleider, L. G., & Desimone, R. (1986). Cortical connections of visual area MT in the macaque. *Journal of Comparative Neurology*, *248*, 190–222.
- Ungerleider, L., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.
- Vatakis, A., & Spence, C. (2008). Evaluating the influence of the ‘unity assumption’ on the temporal perception of realistic audiovisual stimuli. *Acta Psychologica*, *127*, 12–23.
- Waszak, F., Hommel, B., & Allport, A. (2003). Task-switching and long term priming: Role of episodic stimulus-task bindings in task-shift costs. *Cognitive Psychology*, *46*, 361–413.
- Zeki, S., Watson, J. D. G., Lueck, C. J., Friston, K. J., Kennard, C., & Frackowiak, R. S. J. (1991). A direct demonstration of functional specialization in human visual cortex. *The Journal of Neuroscience*, *11*, 641–649.