Feature Integration across Multimodal Perception and Action: A Review

Sharon Zmigrod [1,*] **and Bernhard Hommel** [2]

[1] Leiden University Institute for Psychological Research, Cognitive Psychology Unit, 2300 RB Leiden, The Netherlands

[2] Leiden Institute for Brain and Cognition, Leiden, The Netherlands

Abstract

The human brain is facing a continuous stream of stimulus information delivered by multiple modalities and sensory channels and processed in distinct cortical regions. We discuss recent empirical and theoretical developments in addressing the question of how this distributed information is integrated into coherent representations (the so-called *binding problem*) with an emphasis on the principles and constraints underlying the integration of multiple (rather than redundant) features across different sensory modalities and across perception and action planning.

1. Introduction

One of the most remarkable aspects of multimodal perception is its coherence, that is, our ability to seamlessly integrate the most diverse kinds of information that our different sensory modalities generate. Our conscious perception is unified at any given moment, although we acquire information from diverse channels operating with distinct transduction mechanisms and time characteristics, and process it in separate cortical areas, often at a different time and pace. For instance, a simple event such as eating a sandwich requires integration of the visual attributes such as the shapes and the colors of the ingredients; tactile attributes such as the sandwich's texture and the degree of hotness, not

---

* To whom correspondence should be addressed. E-mail: szmigrod@fsw.leidenuniv.nl

forgetting the chemical attributes such as the smell and the taste; along with the action of chewing that might produce a sound. Apparently, our brain is able to relate all these different kinds of information processed in different cortical locations and systems to each other, and to bind them into a coherent event representation.

The study of the binding problem in the cognitive and neurocognitive sciences tried to tackle the dialectic of modularity and coherence in the brain. Each sensory modality processes its sensory information independently of others in specialized cortical areas: visual information in various dedicated feature maps in the occipital lobe and higher-order areas in infero-temporal cortex (e.g., Zeki, 1993), auditory information in the temporal lobe (Lee and Winer, 2005), and so forth. In a similar vein, action plans consist of various cortically distributed codes representing the different features of an action (e.g., Jeannerod, 1997; Keele *et al.*, 1990), suggesting that planning an action requires some sort of feature binding similar to representing a perceptual event (Hommel *et al.*, 2001; Singer, 1994). Thus, given the primate brain's preference for representing perception and action events in a distributed fashion, there must be some sort of binding mechanism that integrates the information belonging to a specific event. At some point, the brain should construct some form of integrated representation for control and coherent perception, namely it needs to solve the *binding problem* (Hommel, 2004; Treisman, 1996).

1.1. Feature Integration in Vision and Audition

One of the first and most influential theories of feature integration in (or for) visual perception was the Feature Integration Theory (FIT) of attention developed by Treisman and Gelade (1980). The theory posits that visual features (such as color, orientation, brightness, etc.) of an object are processed in parallel in separate feature maps and are later integrated through spatial attention and/or top down processes. Supporting evidence comes from the *object reviewing paradigm* (Kahneman *et al.*, 1992), a visual task in which participants respond to letters or other objects that are preceded by task-irrelevant prime displays. Better performance was observed when the same letter appeared in the prime and the probe display than if prime and probe letter appeared in the same location. This *object-specific preview benefit* was taken to imply identity–location binding: if in the prime display a letter appears in a particular location, the representation of its identity is assumed to be bound to the representation of that location (creating what Kahneman and colleagues call an *object file*: see Hommel, 1998, 2004), so that repeating the exact combination of identity and location allows for the re-use of the previous binding, which again facilitates performance (see Fig. 1A).

Recent neuroimaging studies support this interpretation. For instance, Keizer *et al.* (2008) observed that repeating a particular stimulus feature does
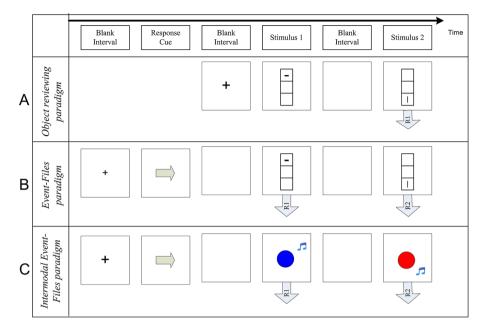
Figure 1. Schematic illustration of the three basic paradigms. (A) *Object-reviewing paradigm* (upper raw) — The first stimulus S1 is a combination of two features: identity and location. The second stimulus S2 is either a complete repetition of S1 with regard to identity and location, complete alternation of S1, or partial repetition of either identity or the location. S2 signaled the response, a speeded left or right response according to the task (either to the identity or to the location). (B) *Event-files paradigm*. A visual response cue signals a left or right response (R1) that should be delayed until presentation of the first stimulus S1 which is again combination of identity and location features (S1 is used as a detection signal for R1). The second stimulus S2, also combination of identity and location features (complete repetition or alternation or partial repetition of the S1's features), appears after responding to S1. S2 signals R2, a speeded left or right response according to the task (either to the identity or to the location). As R1 is independent of the features of S1, this design allows varying response repetition (R1 being the same as, or different from R2) independently of stimulus-feature repetition. (C) *Intermodal event-files paradigm*. Similar to B, but the stimuli are combinations of auditory and visual features (such as pitch and color) that can be presented either in synchronous or asynchronous manner. This figure is published in colour in the online version.

not only reactivate the neural representation of that feature in perceptual cortices but also reactivates the neural codes of the feature that accompanied the repeated feature in the previous trial. In particular, encountering the presentation of an image of a face moving in a specific direction after having seen an image of a house moving in the same direction increases activation not only in the fusiform face area (FFA — a brain area coding for face information) but also in the right parahippocampal place area (PPA — an area coding for house information). In other words, repeating parts of a previous feature combination leads to the retrieval of all components of that combination. This is beneficial

if the present combination is identical to the previous one but creates feature confusion if the present combination is different — the *partial-repetition cost* (see Note 1).

Further research in this domain revealed that object files integrate information with respect to spatiotemporal criteria (Mitroff and Alvarez, 2007) and seem to persist for at least 8 s (Noles *et al.*, 2005). Additional support for FIT comes from a study in the auditory domain by Hall *et al.* (2000), who presented conjunctions of pitch and timbre in different lateralized positions. The results demonstrated more frequent illusory conjunctions when pitch and timbre features were separately presented, suggesting that the auditory system binds its features with reference to their location, just as claimed by FIT for the visual domain. Other studies have provided further support for the binding of auditory (Mondor *et al.*, 2003; Zmigrod and Hommel, 2009) and tactile features (Zmigrod *et al.*, 2009).

1.2. Feature Integration across Perception and Action

Traditional approaches in the cognitive sciences consider perception and action as two separate entities; however, there is ample evidence suggesting a rather intimate relationship between these domains (for reviews see Hommel *et al.*, 2001; Noë, 2004). For example, looking at an apple requires the action of moving one's head and eyes (saccades), even without considering eating it. Thus, perceiving an event is almost always accompanied and in a certain sense even produced by action. Hommel (1998, 2004) claimed that object files in the sense of Kahneman *et al.* may contain action-related information, i.e., representations of the features of an action that is associated with a given object. To test this hypothesis, he designed a task that is based on the object reviewing paradigm including a prime display and a probe display (Kahneman *et al.*, 1992) but that requires participants to carry out a response to both displays. To be more specific, each trial (see Fig. 1B) starts with the presentation of a response cue for the first response (R1), which is to be carried out after the presentation of the first stimulus (S1). The features of this first stimulus can vary but they are picked randomly and not relevant to the task. The second stimulus (S2) is composed of the same perceptual features as S1, partly the same features, or entirely different features, thus creating conditions with complete feature repetition, partial repetition, and alternation, respectively. The participant then responds (R2) to one of the values of S2s perceptual features, such as color (e.g., green indicating one response key and red indicating another). As R1 is independent of the features of S1, this design allows varying response repetition (R1 being the same as, or different from R2) independently of stimulus-feature repetition. The general findings from such tasks indicate partial-repetition costs (in terms of reaction time and accuracy), that is, worse performance is a stimulus feature that is repeated while the response alter-

nates, or *vice versa*. Obviously, this is the same pattern as indicated by the object-specific preview benefit.

Further evidence for comparable integration mechanisms in perception and across perception and action was provided by a recent fMRI study. Kühn *et al*. (2011) showed that repeating a stimulus feature leads to the reactivation of the response that in the preceding prime display accompanied that feature, and that repeating a response leads to the reactivation of the neural codes representing the feature that previously accompanied that response. Accordingly, it makes sense to assume that the mere co-occurrence of multiple features (of a stimulus or response) is sufficient to create a binding between the neural codes representing these features, so that reactivating one of these codes spreads activation to the other members of the binding — which impairs performance if the reactivated representations no longer match the present stimulus or response (Hommel, 1998, 2004).

1.3. Feature Integration across Sensory Modalities

Even though research on feature integration has mainly focused on single modalities such as vision or audition, everyday events are commonly multimodal in nature, which raises the question of how multimodal features are integrated. Feature integration across modalities is more difficult and complex than within a single modality, due to the differences in the physical attributes being coded (such as properties of light, sound, and touch), the respective transduction mechanisms (which use different criteria and operation characteristics), processing times/rates and temporal relations (which emerge from different processing speeds and neural travel times to the brain), and cortical areas in which the different types of information are processed. Evidence that there is a crosstalk between different sensory modalities comes from studies demonstrating perceptual illusions such as the McGurk effect where visual information about a syllable along with discrepant auditory syllable produces a novel combination of the two (McGurk and MacDonald, 1976). Along the same lines is the 'ventriloquism effect' where participants tend to perceive the locations of discrepant visual and auditory stimulus sources somewhere in between the two actual locations (e.g., Bertelson *et al*., 2000; Vroomen *et al*., 2001). In addition, Shams *et al*. (2000) demonstrated the 'double flash' effect in which a single visual flash is perceived as multiple flashes when accompanied by sequences of auditory beeps. This illusion was also found for combinations of other modalities such as auditory–tactile (Hötting and Röder, 2004) and visual–tactile (Kunde and Kiesel, 2006), providing evidence for online *interactions* between different sensory modalities.

Evidence that information from different modalities is not just processed in parallel but really integrated into a coherent representation comes from studies extending the object-reviewing paradigm (Kahneman *et al*., 1992) to multi-

modal stimuli. These studies have demonstrated binding effects for intermodal combinations of visual and auditory (Evans and Treisman, 2010; Jordan *et al.*, 2010; Zmigrod *et al.*, 2009), auditory and tactile features (Zmigrod *et al.*, 2009), and visual and tactile (Cinel *et al.*, 2002). For instance, in an auditory discrimination task using an event-file paradigm (see Fig. 1C), repeating an auditory feature (i.e., pitch) of an audio–visual stimulus facilitates performance if the visual feature (i.e., colour) is also repeated, but impairs performance if the visual feature changes. Thus, the performance depends on the combination of visual and auditory features, providing evidence for automatic integration between different modalities regardless of the task's relevant features (Zmigrod *et al.*, 2009). Further evidence comes from recent studies regarding the temporal dynamics of unimodal (visual or auditory) and multimodal (audio–visual) bindings (Hommel and Colzato, 2004; Zmigrod and Hommel, 2010). The decay rate functions of partial repetition costs (see Note 1) across various response–stimulus intervals (RSI) showed similar patterns in unimodal and multimodal stimuli (see Fig. 2A) and across different modalities and responses (see Fig. 2B), suggesting comparable integration effects within and across perception and action. The similarity of the binding effects reported for single modality and across modalities and domains suggests perceptual features from various sensory modalities and action planning are being integrated into a coherent representation by means of either the same mechanisms or at least mechanisms with comparable operation characteristics and criteria.

The event file paradigm (Fig. 1) was used in many studies to explore further principles and constraints regarding the creation, maintenance (updating) and revision of such episodic representations (Hommel, 2005, 2007, 2009; Hommel and Colzato, 2004; Zmigrod and Hommel, 2009, 2010, 2011), providing essential information for understanding the underlying mechanisms of binding.

2. Principles Underlying Feature Integration across Multimodal Perception and Action

Binding multimodal codes in a distributed brain calls for a rather flexible mechanism that can integrate over learned feature conjunctions and novel feature combinations alike; it must allow for the recombination of features and it must distinguish between features belonging to different events even if these overlap in time. Even though an exhaustive theory of feature integration is still lacking, increasing evidence suggests a number of principles and criteria that feature integration processes seem to handle in order to tackle these challenges. In the following, we discuss some of them.
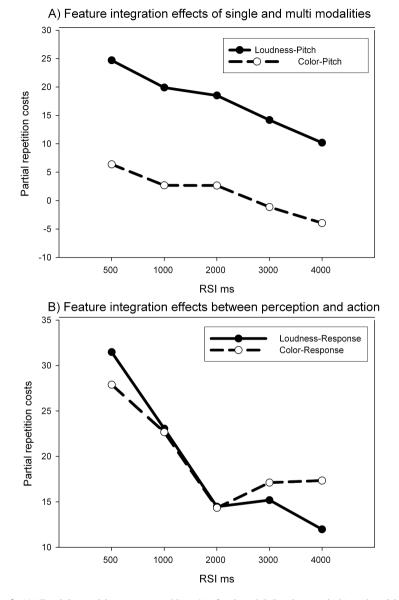
A) Feature integration effects of single and multi modalities

B) Feature integration effects between perception and action

Figure 2. (A) Partial repetition costs (see Note 1) of unimodal (loudness–pitch) and multimodal (color–pitch) feature integration as a function of response–stimulus interval (RSI). (B) Partial repetition costs of feature integration across perception and action (loudness-response and color-response) (Zmigrod and Hommel, 2010).

2.1. Convergence vs. Synchronization Mechanism

What is the basic operational characteristic of feature integration? An early attempt to explain how feature combinations can be detected and processed was

based on the idea of neural convergence (Barlow, 1972). It was motivated by the discovery of neurons that operate as conjunction detectors, that is, neurons that fire not in the presence of a particular feature, such as a particular color orientation, but rather in response to specific combinations of feature values, such as a particular orientation associated with a particular color. One problem with this idea is the so-called 'combinatorial explosion' problem: considering the enormous number of possible combinations between features, one may wonder whether the conjunctive coding of all these combinations exhausts the neural capacity of the human brain. Moreover, if all perception relies on complex hierarchies of conjunction detectors, higher-order concepts would need to be represented by single neurons, so that the loss of this neuron would imply the elimination of this concept in the given individual — a rather implausible implication.

Another proposed mechanism through which binding might be achieved is temporal synchrony (see: Engel and Singer, 2001; Raffone and Wolters, 2001; von der Malsburg, 1999). The basic idea is that individual features would be represented by the neurons that fire when facing these features, while the belongingness to the same event would be represented by neural synchronicity. In other words, neurons representing features of the same event fire in the same rhythm. This mechanism has the advantage of not suffering from the capacity problem that convergence mechanisms are facing and it can deal better with novel information and feature combinations, which makes it rather attractive from a binding perspective.

There is a growing body of empirical evidence, from both human and animal studies, that supports the idea that temporal synchrony is related to feature integration. For instance, integration-related synchronization in the gamma range ($\sim$30–100 Hz) has been observed in visual areas (Engel *et al.*, 1991), auditory areas (deCharms and Merzenich, 1996; Joliot *et al.*, 1994), and somatosensory areas (Murthy and Fetz, 1992; Nicolelis *et al.*, 1995). Also, evidence for neural synchronization across different sensory modalities in the beta range ($\sim$12–20 Hz) has been observed between visual and auditory areas (von Stein *et al.*, 1999), visual and motor areas (Roelfsema *et al.*, 1997), and motor and somatosensory areas (Murthy and Fetz, 1992, 1996). Even though these findings do not show how integration is actually achieved, they do suggest a close connection between local synchronized activity and feature integration/object representation (Tallon-Baudry and Bertrand, 1999).

However, recent studies suggest that synchronization and conjunction detectors may not represent mutually exclusive alternatives but, rather, two mechanisms actually working in parallel (Hommel and Colzato, 2009; VanRullen, 2009). One motivation for this conclusion comes from studies on the relationship between the learning of feature combinations and the after-effects of spontaneous feature bindings (Colzato *et al.*, 2006a; Hommel and Colzato,

2009). Even though these studies could show that binding and learning affect performance in a preview task (see above), the two types of effects did not interact, not even in the processing of highly overlearned feature conjunctions such as the object of a banana and the color yellow. As the authors suggest, repeatedly facing specific combinations of features, like in the banana case, might lead to the implementation of conjunction detectors that speed up object recognition by means of top-down priming. At the same time, however, the perceptual system must be able to process any possible combination of features, irrespective of the probability of the particular combination, so that synchronization-related integration may proceed concurrently with the activation of conjunction detectors. Hence, more than one neural mechanism may be responsible for feature integration.

2.2. *Automaticity* vs. *Attention*

Findings from the preview task (Kahneman *et al.*, 1992) and its various versions suggests that features are integrated in a rather automatic fashion: even though the preview display is entirely irrelevant and can safely be ignored, people apparently bind the features of objects presented therein. Likewise, the demonstration that observers can apparently not help but averaging across different modalities, as indicated by the McGurk effect, suggests that integration is automatic. Indeed, partial-repetition costs are unaffected by the amount of attention available when facing visual feature combinations (Hommel, 2005) and feature integration in the auditory modality can occur outside the focus of attention (Takegata *et al.*, 2005).

And yet, attentional manipulations are not entirely ineffective. For instance, features that are relevant to the task or that are particularly salient for other reasons are commonly more likely to produce reliable partial-repetition costs (Hommel, 1998). For instance, Zmigrod and Hommel (2009) observed more reliable binding effects related to stimulus location in tasks with spatial responses as compared to tasks with non-spatial responses. Likewise, more salient features were found to be more likely to produce reliable binding effects than less salient features (Dutzi and Hommel, 2009; Zmigrod *et al.*, 2009). This suggests that feature dimensions are weighted according to their salience and relevance (Found and Müller, 1996; Hommel, 2004; McGinnis and Keil, 2011; Memelink and Hommel (in press); Zmigrod and Hommel, 2009), with higher weights producing stronger activation of the respective feature codes and, thus, more impact on behaviour. However, this leaves open the question of which process is actually affected by dimensional weighting and task relevance. For one, it could be the integration process proper, that is, higher weights would increase the activation of a feature code during integration and thus modulate the likelihood that it becomes part of the object file. For another, it could be the retrieval process. Hence, encountering a feature

repetition would lead to the retrieval of object files containing that feature but only to the degree that it falls on a task relevant dimension.

Even though direct evidence revealing whether dimensional weighting affects the creation or the retrieval of bindings is still lacking, there are numerous indications that are pointing towards retrieval. For instance, partial-repetition costs are less — rather than more — pronounced in individuals with high fluid intelligence (Colzato *et al.*, 2006b), a population that is unlikely to have weak binding processes. Given the link between fluid intelligence and executive-control functions (Duncan *et al.*, 2000), individuals with higher intelligence might be suspected to be more efficient in managing episodic bindings, be it by inhibiting bindings from previous encounters whenever necessary and/or by updating these bindings to make them fit the present feature combination (Colzato *et al.*, 2006b). This scenario fits with observations that partial-integration costs are increased in children and older adults as compared to young adults (Hommel *et al.*, 2011); in children with Autistic Spectrum Disorder, which is assumed to implicate impaired executive control functions (Corbett *et al.*, 2009; Hill, 2004), as compared to a healthy control group (Zmigrod *et al.* (in press)); in cannabis users (Colzato and Hommel, 2008); and as a consequence of positive affect (Colzato *et al.*, 2007), which has been shown to counteract control processes (van Steenbergen *et al.*, 2009). It is also consistent with the observation that increasing neural gamma-band activity over the frontal cortex by means of neurofeedback improves explicit memory retrieval and reduces partial-repetition costs (Keizer *et al.*, 2010).

2.3. Integration Criteria

We have pointed out that one challenge for feature integration consists in the problem that the time needed to process different features is likely to differ, especially if they are coded in different sensory modalities. This raises the question of which criteria feature integration processes are being used to determine which features are going with which, that is, which of the activated feature codes are belonging to the same event (see also Ernst and Bülthoff, 2004). Research on feature integration has revealed that at least two criteria play a role. First, as suggested by the original FIT (Treisman and Gelade, 1980) already, features coming from the same location are likely to be integrated into the same object file. This is obvious from a recent study of van Dam and Hommel (2010), in which participants were presented with visual prime and probe displays. Findings demonstrate that participants spontaneously created bindings between features belonging to different objects (e.g., the shape of one object with the color of another), as long as these objects overlapped in space. Spatial separation, in contrast, prevented any binding. Moreover, similar suggestions also came from findings in the auditory domain where Hall *et al.* (2000) showed more illusory conjunctions when pitch and timbre fea-

tures were separately presented. Hence, spatial proximity seems to be a key criterion for integrating perceptual features.

Indeed, spatial processing was found to be prominent in multimodal integration under conditions in which conflict information is reached by more than one modality, as in the McGurk effect (McGurk and MacDonald, 1976) or the ventriloquist effect (Vroomen *et al.*, 2001). Yet, even though in intermodal feature-integration experiments the features are often presented at different locations (e.g., visual features on a screen and auditory features through headphones), participants can easily integrate the signals belonging to the same event (Zmigrod and Hommel, 2010, 2011). Thus, spatial proximity is not the sole criterion for feature integration.

Another important criterion is time. Even though processing times and rates are likely to differ for different features, particularly in different modalities and domains, the activation of the corresponding neural codes is likely to overlap in time. If we assume that activated codes are prone to decay, this suggests that the relative timing of stimulus features (or at least their hypothetical cortical arrival times) should determine whether two different features are integrated. Indeed, Zmigrod and Hommel (2009, 2010) have manipulated factors that determine the time of onset and the decay rate of feature code and were able to predict the likelihood of feature integration based on the assumed temporal overlap between the corresponding neural codes.

Furthermore, studies on multimodal perception showed the existence of an 'intersensory temporal synchrony window' (Lewkowicz, 1996) where features that presented within a temporal window of up to about 100 ms (Lewald *et al.*, 2001) or even 200 ms (van Wassenhove *et al.*, 2007) are still perceived as belonging to the same event. This likelihood decreases as the temporal interval between the features increases (Zmigrod and Hommel, 2011). However, the unified perception was found to be more sensitive to the temporal window than the binding effects: feature integration effects were observed with asynchronous audiovisual stimuli of up to 350 ms between the features regardless of the subjective unified experience (Zmigrod and Hommel, 2011). Thus, it seems that both spatial and temporal factors play a significant role in establishing a coherent event representation, and the combination of spatiotemporal continuity can contribute to the persistence of this representation over time (Goa and Scholl, 2010; Spapé and Hommel, 2010).

3. Conclusions

The aim of the present review was to sketch our current understanding of when, how, and according to which principles distributed feature codes (from same and different domains) are integrated into coherent event representation. The empirical evidence presented suggests that feature integration occurs

rather spontaneously, within and across perceptual and action dimensions, and by administering general criteria, such as spatial and temporal overlap. Even though more research is necessary, the available evidence is consistent with the assumption that the retrieval of bindings is more tightly controlled than the integration process itself.

Note

1. Partial repetition costs for a given interaction between factors X and Y are commonly expressed in terms of the size of the interaction term (see Fig. 1B). More formerly, the partial repetition cost PRC is calculated as the difference between the RTs for partial repetitions (feature X repeated and feature Y alternated, or *vice versa*) and the RTs for complete repetitions and 'complete' alternations. For example, the partial repetition costs for the Identity X location interaction at a given RSI would be $PRC_{identity X location} = (RT_{identity repeated/location alternated} + RT_{identity alternated/location repeated})/2 - (RT_{identity repeated/location repeated} + RT_{identity alternated/location alternated})/2$.

References

Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* **1**, 371–394.

Bertelson, P., Vroomen, J., de Gelde, B. and Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention, *Percept. Psychophys.* **62**, 321–332.

Cinel, C., Humphreys, G. W. and Poli, R. (2002). Cross-modal illusory conjunctions between vision and touch, *J. Exper. Psychol.: Human Percept. Perform.* **28**, 1243–1266.

Colzato, L. S. and Hommel, B. (2008). Cannabis, cocaine, and visuomotor integration: evidence for a role of dopamine D1 receptors in binding perception and action, *Neuropsychologia* **46**, 1570–1575.

Colzato, L. S., Raffone, A. and Hommel, B. (2006a). What do we learn from binding features? Evidence for multilevel feature integration, *J. Exper. Psychol.: Human Percept. Perform.* **32**, 705–716.

Colzato, L. S., van Wouwe, N. C., Lavender, T. J. and Hommel, B. (2006b). Intelligence and cognitive flexibility: fluid intelligence correlates with feature 'unbinding' across perception and action, *Psychonom. Bull. Rev.* **13**, 1043–1048.

Colzato, L. S., van Wouwe, N. C. and Hommel, B. (2007). Feature binding and affect: emotional modulation of visuo-motor integration, *Neuropsychologia* **45**, 440–446.

Corbett, B. A., Constantine, L. J., Hendren, R., Rocke, D. and Ozonoff, S. (2009). Examining executive functioning in children with autism spectrum disorder, attention deficit hyperactivity disorder and typical development, *Psychiatry Res.* **166**, 210–222.

deCharms, R. C. and Merzenich, M. M. (1996). Primary cortical representation of sounds by the coordination of action-potential timing, *Nature* **381**, 610–613.

Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., Newell, F. N. and Emslie, H. (2000). A neural basis for general intelligence, *Science* **289**, 457–460.

Dutzi, I. B. and Hommel, B. (2009). The microgenesis of action-effect binding, *Psycholog. Res.* **73**, 425–435.

Engel, A. K. and Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness, *Trends Cognit. Sci.* **5**, 16–25.

Engel, A. K., Konig, P. and Singer, W. (1991). Direct physiological evidence for scene segmentation by temporal coding, *Proc. Nat. Acad. Sci. USA* **88**, 9136–9140.

Ernst, M. O. and Bülthoff, H. H. (2004). Merging the senses into a robust percept, *Trends Cognit. Sci.* **8**, 162–169.

Evans, K. K. and Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features, *J. Vision* **10**, 1–12.

Found, A. and Müller, H. J. (1996). Searching for unknown feature targets on more than one dimension: investigating a 'dimension weighting' account, *Percept. Psychophys.* **58**, 88–101.

Gao, T. and Scholl, B. J. (2010). Are objects required for object files? Roles of segmentation and spatiotemporal continuity in computing object persistence, *Visual Cognition* **18**, 82–109.

Hall, M. D., Pastore, R. E., Acker, B. E. and Huang, W. (2000). Evidence for auditory feature integration with spatially distributed items, *Percept. Psychophys.* **6**, 1243–1257.

Hill, E. L. (2004). Evaluating the theory of executive dysfunction in autism, *Developmental Review* **24**, 189–233.

Hommel, B. (1998). Event files: evidences 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 Cognit. Sci.* **8**, 494–500.

Hommel, B. (2005). How much attention does an event file need? *J. Exper. Psychol.: Human Percept. Perform.* **31**, 1067–1082.

Hommel, B. (2007). Feature integration across perception and action: event file affect response choice, *Psycholog. Res.* **7**, 42–63.

Hommel, B. (2009). Action control according to TEC (theory of event coding), *Psycholog. Res.* **73**, 512–526.

Hommel, B. and Colzato, L. S. (2004). Visual attention and the temporal dynamics of feature integration, *Visual Cognition* **11**, 483–521.

Hommel, B. and 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., Müsseler, J., Aschersleben, G. and Prinz, W. (2001). The Theory of Event Coding (TEC): a framework for perception and action planning, *Behav. Brain Sci.* **24**, 849–937.

Hommel, B., Kray, J. and Lindenberger, U. (2011). Feature integration across the lifespan: stickier stimulus–response bindings in children and older adults, *Frontiers Psychol.* **2**, 268.

Hötting, K. and Röder, B. (2004). Hearing cheats touch, but less in congenitally blind than in sighted individuals, *Psycholog. Sci.* **15**, 60–64.

Jeannerod, M. (1997). *The Cognitive Neuroscience of Action*. Blackwell Publishers, Oxford, UK.

Joliot, M., Ribary, U. and Llinás, R. (1994). Human oscillatory brain activity near 40 Hz coexists with cognitive, *Proc. Nat. Acad. Sci. USA* **91**, 11748–11751.

Jordan, K. E., Clark, K. and Mitroff, S. R. (2010). See an object, hear an object file: object correspondence transcends sensory modality, *J. Vision* **18**, 492–503.

Kahneman, D., Treisman, A. and Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information, *Cognit. Psychol.* **24**, 175–219.

Keele, S. W., Cohen, A. and Ivry, R. (1990). Motor programs: concepts and issues, in: *Attention and Performance: Motor Representation and Control*, M. Jeannerod (Ed.), Vol. 13, pp. 77–110. Erlbaum, Hillsdale, NJ, USA.

Keizer, A. W., Nieuwenhuis, S., Colzato, L. S., Theeuwisse, W., Rombouts, S. A. R. B. and Hommel, B. (2008). When moving faces activate the house area: an fMRI study of object file retrieval, *Behav. Brain Functions* **4**, 50.

Keizer, A. W., Verment, R. and Hommel, B. (2010). Enhancing cognitive control through neurofeedback: a role of gamma-band activity in managing episodic retrieval, *Neuroimage* **49**, 3404–3413.

Kühn, S., Keizer, A., Colzato, L. S., Rombouts, S. A. R. B. and Hommel, B. (2011). The neural underpinnings of event-file management: evidence for stimulus-induced activation of, and competition among stimulus-response bindings, *J. Cognit. Neurosci.* **23**, 896–904.

Kunde, W. and Kiesel, A. (2006). See what you've done! Active touch affects the number of perceived visual objects, *Psychonom. Bull. Rev.* **13**, 304–309.

Lee, C. C. and Winer, J. A. (2005). Principles governing auditory cortex connections, *Cerebral Cortex* **15**, 1804–1814.

Lewald, J., Ehrenstein, W. H. and Guski, R. (2001). Spatio-temporal constraints for auditory–visual integration, *Behav. Brain Res.* **121**, 69–79.

Lewkowicz, D. J. (1996). Perception of auditory–visual temporal synchrony in human infants, *J. Exper. Psychol.: Human Percept. Perform.* **22**, 1094–1106.

McGinnis, E. M. and Keil, A. (2011). Selective processing of multiple features in the human brain: effects of feature type and salience, *PLoS ONE* **6**, e16824. DOI:10.1371/journal.pone.0016824.

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

Mitroff, S. R. and Alvarez, G. A. (2007). Space and time, not surface features, guide object persistence, *Psychonom. Bull. Rev.* **14**, 1199–1204.

Memelink, J. and Hommel, B. (in press). Intentional weighting: a basic principle in cognitive control, *Psycholog. Res.*; published online in 2012: DOI: 10.1007/s00426-012-0435-y.

Mondor, T. A., Hurlburt, J. and Thorne, L. (2003). Categorizing sounds by pitch: effects of stimulus similarity and response repetition, *Percept. Psychophys.* **65**, 107–114.

Murthy, V. N. and Fetz, E. E. (1992). Coherent 25- to 35-Hz oscillations in the sensorimotor cortex of awake behaving monkeys, *Proc. Nat. Acad. Sci. USA* **89**, 5670–5674.

Murthy, V. N. and Fetz, E. E. (1996). Oscillatory activity in sensorimotor cortex of awake monkeys: synchronization of local field potentials and relation to behavior, *J. Neurophysiol.* **76**, 3949–3967.

Nicolelis, M. A., Baccala, L. A., Lin, R. C. and Chapin, J. K. (1995). Sensorimotor encoding by synchronous neural ensemble activity at multiple levels of the somatosensory system, *Science* **268**, 1353–1358.

Noë, A. (2004). *Action in Perception*. MIT Press, Cambridge, MA, USA.

Noles, N. S., Scholl, B. J. and Mitroff, S. R. (2005). The persistence of object file representations, *Percept. Psychophys.* **67**, 324–334.

Raffone, A. and Wolters, G. (2001). A cortical mechanism for binding in visual working memory, *J. Cognit. Neurosci.* **13**, 766–785.

Roelfsema, P. R., Engel, A. K., Konig, P. and Singer, W. (1997). Visuomotor integration is associated with zero time-lag synchronization among cortical areas, *Nature* **385**, 157–161.

Shams, L., Kamitani, Y. and Shimojo, S. (2000). What you see is what you hear, *Nature* **408**, 788.

Singer, W. (1994). A new job for the thalamus, *Nature* **369**, 444–445.

Spapé, M. and Hommel, B. (2010). Actions travel with their objects: evidence for dynamic event files, *Psycholog. Res.* **74**, 50–58.

Takegata, R., Brattico, E., Tervaniemi, M., Varyagina, O., Näätänen, R. and Winkler, I. (2005). Preattentive representation of feature conjunctions for concurrent spatially distributed audition objects, *Cognit. Brain Res.* **25**, 169–179.

Tallon-Baudry, C. and Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation, *Trends Cognit. Sci.* **3**, 151–162.

Treisman, A. (1996). The binding problem, *Curr. Opinion Neurobiol.* **6**, 171–178.

Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention, *Cognit. Psychol.* **12**, 97–136.

van Dam, W. O. and Hommel, B. (2010). How object-specific are object files? Evidence for integration by location, *J. Exper. Psychol.: Human Percept. Perform.* **36**, 1184–1192.

VanRullen, R. (2009). Binding hardwired *versus* on-demand feature conjunction, *Visual Cognition* **17**, 103–119.

van Steenbergen, H., Band, G. P. H. and Hommel, B. (2009). Reward counteracts conflict adaptation: evidence for a role of affect in executive control, *Psycholog. Sci.* **20**, 1473–1477.

van Wassenhove, V., Grant, K. W. and Poeppel, D. (2007). Temporal window of integration in bimodal speech, *Neuropsychologia* **45**, 598–607.

Violentyev, A., Shimojo, S. and Shams, L. (2005). Touch-induced visual illusion, *NeuroReport* **16**, 1107–1110.

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

von Stein, A., Rappelsberger, P., Sarnthein, J. and Petsche, H. (1999). Synchronization between temporal and parietal cortex during multimodal object processing in man, *Cerbral Cortex* **9**, 137–150.

Vroomen, J., Bertelson, P. and de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention, *Percept. Psychophys.* **63**, 651–659.

Zeki, S. (1993). *A Vision of the Brain*. Blackwell Scientific, Oxford, UK.

Zmigrod, S. and Hommel, B. (2009). Auditory event files: integrating auditory perception and action planning, *Attention Percept. Psychophys.* **71**, 352–362.

Zmigrod, S. and Hommel, B. (2010). Temporal dynamics of unimodal and multimodal feature binding, *Attention Percept. Psychophys.* **72**, 142–152.

Zmigrod, S. and Hommel, B. (2011). The relationship between feature binding and consciousness: evidence from asynchronous multi-modal stimuli, *Conscious. Cognit.* **20**, 586–593.

Zmigrod, S., Spapé, M. and Hommel, B. (2009). Intermodal event files: integrating features across vision, audition, taction, and action, *Psycholog. Res.* **73**, 674–684.

Zmigrod, S., de Sonneville, L. M., Colzato, L. S., Swaab, H. and Hommel, B. (in press). Cognitive control of feature bindings: evidence from children with autistic spectrum disorder, *Psycholog. Res.*; published online in 2011: DOI: 10.1007/s00426-011-0399-3.