

When an object is more than a binding of its features: Evidence for two mechanisms of visual feature integration

Bernhard Hommel and Lorenza S. Colzato

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

People spontaneously integrate the features of visual events into episodic structures that are reactivated if their ingredients match aspects of the current input. Feature integration has been attributed to either the detection of feature conjunctions or the ad hoc binding of feature codes (e.g., by neural synchronization). We report evidence suggesting that both kinds of integration mechanisms coexist. Replicating earlier findings, repeating one visual feature facilitated performance but only if other visual features were also repeated. However, this effect was more pronounced with real objects as compared to arbitrary combinations of shapes and colours. Moreover, the real-object effect was restricted to visual feature integration but did not extend to visuomotor integration, suggesting that the underlying mechanism subserves perception only. We suggest a dual-process feature-integration model that distinguishes between ad hoc binding, which operates on any possible combination of features alike, and conjunction detection, which selectively operates on familiar feature combinations.

Keywords: Object perception; Event file; Feature integration.

Primate brains represent many aspects of the objects and events they perceive and produce in a distributed, feature-based fashion. The human visual cortex, for instance, consists of various neural maps that code for different visual features of perceived objects, such as orientation, shape, or motion (see DeYoe & van Essen, 1988), and the frontal cortex houses maps coding for the direction, distance, and force of intentional actions (see Hommel & Elsner, in press). These observations have been taken to imply various integration or binding problems (e.g., Treisman, 1996), as they raise

Please address all correspondence to Bernhard Hommel, Leiden University, Department of Psychology, Cognitive Psychology Unit, Postbus 9555, 2300 RB Leiden, The Netherlands. E-mail: hommel@fsw.leidenuniv.nl

© 2008 Psychology Press, an imprint of the Taylor & Francis Group, an Informa business
<http://www.psypress.com/viscog> DOI: 10.1080/13506280802349787

the question how the brain knows which of the feature codes activated at a given time are related to the same event.

Authors differ with respect to how serious they consider these problems to be and how easily they think these problems can be solved. For instance, some authors have claimed that serially operating attentional mechanisms are required to bind related information together (e.g., Treisman & Gelade, 1980), whereas others assume that the retinotopic organization of visual maps provides sufficient information for properly integrating at least visual features (van der Heijden, 1995). Most authors have focused on one of two neural principles that may mediate feature integration and help solving binding problems. One principle is that of *convergence*: Lower level neurons may code for simple features, such as <round> or <green>, and project onto higher level neurons that code for feature conjunctions (e.g., <round> AND <green>). This may lead up to even higher level representations of whole objects (Barlow, 1972). Given the considerable variability of objects in terms of their instances and retinal projections, as well as the numerous ways in which features can be potentially combined, the exclusive reliance on convergence mechanisms would lead to a combinatorial explosion and is therefore not particularly plausible. Accordingly, a second mechanism has been suggested in which integration does not (necessarily) rely on conjunction detectors. The idea is that integration comprises of *synchronizing* the firing patterns of feature-coding neurons, in such a way that the neurons coding for features of the same object act as a unit (Engel & Singer, 2001; Raffone & Wolters, 2001; von der Malsburg, 1999).

Even though these two mechanisms are commonly treated as mutually exclusive alternatives, the benefits and costs they imply suggest that they both play a role in dealing with binding problems (Colzato, Raffone & Hommel, 2006; VanRullen, 2009 this issue). Synchronization-based integration has the advantage of being particularly flexible and parsimonious in terms of long-term memory structures but the disadvantage that even features that are very likely to cooccur would need to be bound anew every time they are encountered. Hence, this integration method would be economical in terms of cognitive structure but wasteful in terms of processing time. In contrast, convergence-based integration has the advantage of allowing for the fast and effortless registration of feature combinations of practically unlimited complexity but the disadvantage that (apart from possible genetically hardwired conjunction detectors) this registration presupposes extensive learning and some degree of separability of conjunctions (as conjunction with too much feature overlap would lead to the activation of too many conjunction detectors). Hence, this integration method would be economical in terms of processing time but wasteful in terms of structure. Given that our environment calls for both the recognition of highly reliable feature combinations (as with natural objects) and the

processing of highly arbitrary combinations (as commonly used in psychological experiments), it makes sense to assume that both convergence-based and synchronization-based mechanisms are at work in generating human perception.

The present study was conceived of with this distinction in mind. As we will argue, there is behavioural evidence suggesting that at least two different mechanisms are at work in human visual feature integration, one that operates on any feature conjunction that is encountered (presumably based on synchronization) and another that selectively operates on frequent conjunctions that have been stored (presumably based on convergence). Before addressing the motivation and rationale of our study in more detail we will provide a brief overview of the theoretical background relevant for the study and the experimental paradigm we used.

OBJECT FILES

Most studies addressing feature binding focused on the visual modality. Some studies investigated whether there actually are binding problems in visual feature integration. For instance, Treisman and Schmidt (1982), and many others since then, demonstrated that creating attentionally demanding conditions results in an increasing numbers of incorrect bindings or “illusory conjunctions”—suggesting that feature integration is not a trivial task and raising doubt whether it is solely handled by convergence mechanisms (which have a hard time predicting such observations). Other studies have looked into whether people actually *do* bind features, even under circumstances that do not seem to require any binding. Particularly appropriate for that purpose turned out to be the preview paradigm developed by Kahneman, Treisman, and Gibbs (1992) and used by many others since then.

The simplest, stripped-down form of this paradigm is illustrated in Figure 1 (please ignore the R1 cue and R1 for the moment). The sequence of trial events comprises a nominally task-irrelevant prime or preview display (S1) followed by a probe display (S2). Let's assume in this example that S1 and S2 can either be a circle or triangle, can be red or green, and can be presented in the top or bottom position. The observer's task is simply to identify S2's shape as quickly as possible. Colour and position information are completely irrelevant to the task and can be safely ignored. Because S1 requires no response, it too can be ignored. Critically, S1 may consist of the same or a different shape as S2 and may appear in the same or a different colour and at the same or a different location. Analysing performance on S2 (i.e., R2) as a function of the repetition or alternation of shape, colour, and location (from S1 to S2) may or may not produce main effects, such as better performance if a particular feature repeats. These kinds of effects should

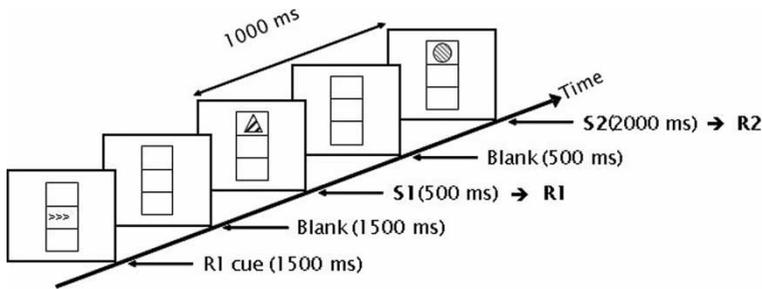


Figure 1. Sequence of events in the present experiment. Participants prepared a left- or right-hand response as indicated by a cue (R1 cue) and carried it out (R1) when the next stimulus (S1) appeared. Then they waited for the next stimulus (S2) and carried out a left- or right-hand response (R2) to its shape. S1 and S2 varied in shape, colour, and location, so that all three features could repeat or alternate. As R1 did not depend on S1 (but on the R1 cue), the response could repeat or alternate irrespective of the stimulus sequence.

reflect the priming of feature codes (e.g., leftover activation in the corresponding codes) and are thus unlikely to reflect binding processes (Kahneman et al., 1992).

Theoretically more interesting are *interactions* between repetition effects, because they indicate that the impact of repeating one feature depends on the identity of the other. Such interactions have been frequently observed: Repeating one feature produces better performance if other features are also repeated (e.g., Kahneman et al., 1992) but worse performance if these other features alternate (e.g., Hommel, 1998). To take our example, responding to a red circle (as S2) is easier following the presentation of another red circle than following a green circle (as S1), but more difficult following a red than a green triangle. Likewise, responding to a top circle is easier following a top than a bottom circle, but more difficult following a top than a bottom triangle. These observations suggest that registering the cooccurrence of two or more given features is sufficient to create some sort of binding between them, so that reencountering one of them tends to reactivate the whole binding in a pattern-completion-like process (Hommel, 2004). As a consequence, partial repetitions induce the retrieval of no longer valid feature codes, which disturbs current coding processes and induces code conflict. There is thus evidence that people do bind features—no matter how plausible one finds the available arguments for or against the logical necessity of feature binding. And they do so even under the most unlikely circumstances, that is, even if neither the bound features nor the object they refer to need to be attended or reported, even in the absence of any attentionally challenging display or task, and even when no more than 250–500 ms are available for creating such bindings (Hommel & Colzato, 2004).

Direct evidence for the retrieval of object files has been obtained in an fMRI study by Keizer et al. (2008). They presented subjects with preview (S1) and probe displays (S2) that both consisted of two blended pictures showing a face and a house. Either the face or the house moved in one of two possible directions, and subjects were to respond to the direction of S2 irrespective of which object moved. Most interesting were the conditions in which S1 showed a moving house and S2 a moving face: If the direction of motion in these two displays was the same more activation of house-related information in the parahippocampal place area was observed than if the motion differed. This suggests that the direction of motion was integrated with the object that moved, so that repeating the motion reactivated the representation of the object that had just accompanied this motion. Hence, binding features creates episodic cognitive structures that tend to be reactivated as a whole as soon as one of their ingredients matches the current input. This is indeed what underlies the original idea underlying Kahneman et al.'s (1992) *object file* concept: Bindings are functional in establishing object constancy by maintaining information about an object even in the absence of current sensory input and by relating this information to later reoccurrences of this object, even though these reoccurrences may only match part of the maintained information. This is why we can track objects over longer periods of occlusion and across changes in a number of visual features.

Interactions between repetition effects indicative of feature binding have been obtained for various features. Shape, colour, and location features have been shown to interact (Hommel, 1998, 2007; Hommel & Colzato, 2004) just as well as face, house, and motion information (Keizer, Colzato & Hommel, 2008)—suggesting that binding can span ventral and dorsal processing streams. An interesting observation in all these studies is that location does not seem to play a particularly dominant role. Some authors have claimed that object files can only be reactivated or reassessed if the current object matches the respective object file in terms of location (Kahneman et al., 1992; Mitroff & Alvarez, 2007). This would suggest that nonspatial matches are insufficient by themselves to retrieve a previous object file, which again implies that nonspatial features can only interact if this interaction is mediated by a location match. Even though it is clear that spatial location and spatial matches are important in multielement displays, simply because location is commonly crucial to track the identity of an object, spatially unmediated interactions between nonspatial features are possible (e.g., Colzato, Raffone, & Hommel, 2006; Hommel, 1998, 2007), which disconfirms approaches that rely on spatial correspondence as a retrieval cue. On the other hand, spatial location clearly plays a central role in the *encoding* of object files. Most studies on feature integration confound the sharing of spatial location with belongingness to the same perceptual object. Van Dam

and Hommel (2008) disentangled these factors by testing whether two given features appearing in the same location would be still integrated even if they obviously belonged to two different objects. Indeed, orientation and colour features were bound (i.e., orientation- and colour-repetition effects interacted) irrespective of whether they appeared as part of the same object or of different objects (e.g., one stationary and the other moving continuously, or a banana in a particular orientation overlaying an apple of a particular colour). In contrast, integration was markedly reduced when the two objects were separated in space (cf. Xu, 2006). Thus, spatial location is important for the encoding but not the retrieval or reactivation of object files.

MULTIPLE INTEGRATION MECHANISMS

The available evidence suggests that cooccurring visual features are more or less automatically bound into object files, that is, temporary links between, or pointers to the codes representing, the features of a perceived visual event. To take our example, and following Hommel (2004) and Colzato et al. (2006), this process can be captured by the cartoon model sketched in Figure 2a. Registering a red circle appearing in a top position would lead to the activation of corresponding codes in shape, colour, and location maps, and these codes are cross referenced by creating a temporary object file (symbolized by the folder).

Note that Figure 2 considers a further impact from the current attentional set, which is assumed to prime task-relevant feature dimensions (shape in our example). The reason to include such a top-down mechanism derives from a number of observations. On the one hand, the fact that bindings are created under the most unlikely conditions seems to suggest that feature integration is a highly automatic process. Indeed, systematic manipulations of the amount of attention directed to the to-be-integrated features and available for integration failed to modulate feature integration effects (Hommel, 2005, 2007), suggesting that the encoding of event files is a spontaneous process (cf. Logan, 1988). On the other hand, the retrieval of event files turned out to be rather highly controllable. One indication for that is that features varying on task-relevant feature dimensions are more likely to be involved in interactions with other features. For instance, having participants respond to the shape of S2 yields particularly strong binding-related effects involving shape repetition while having them respond to colour yields particularly strong effects involving colour repetition (e.g., Hommel, 1998). Even trial-to-trial shifts between shape- and colour-relevant versions of the task induce stronger binding-related effects for the currently task-relevant feature dimension, suggesting that attentional set has an immediate impact (Hommel, Memelink, Zmigrod, & Colzato, 2008). As it

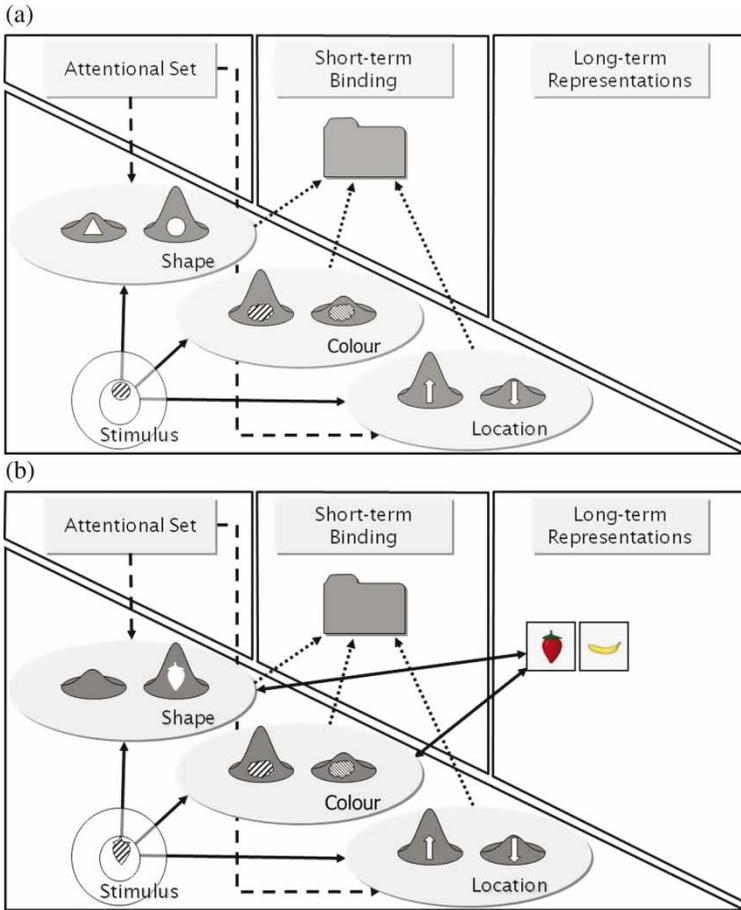


Figure 2. Cartoon model of feature binding and binding retrieval. (a) The coding of arbitrary feature conjunctions (adapted from Colzato et al., 2006). Colours are indicated by fill patterns. Shape is directly task relevant (by having it signalling responses) and location is indirectly task relevant (by defining the responses in terms of spatial locations), so that the shape and location dimensions are primed by the attentional set. Features on primed dimensions are assumed to be more likely integrated with other features and/or retrieved by stimuli that feature-overlap with the respective binding. (b) The coding of highly familiar feature conjunctions or real objects represented in long-term memory. Stimuli that match long-term representations to a sufficient degree activate these representations, which again induce top-down priming of the stimulus features that are coded on object-defining dimensions (shape and colour in the example). Top-down priming may work by increasing the gain of the respective feature dimension, which multiplies the stimulus-induced activation of this dimension's feature codes. Note that short-term bindings and long-term representations relate differently to individual feature codes: Whereas short-term bindings are considered to “point” to, and thus remain linked with the individual codes, long-term representations only respond to the presence of features or feature combinations. Likewise, short-term bindings can reactivate individual feature codes upon retrieval of a code that has been bound with them, whereas the top-down effect of long-term representations is restricted to priming feature dimensions but not individual codes.

does not matter whether the set is established before or after the hypothetical binding process (i.e., before after S1 presentation), it makes sense to assume that the impact of the attentional set selectively targets object file *retrieval* but not encoding (Hommel et al., 2008).

In Figure 2a we have considered two types of impact of the current attentional set. One is rather obvious: Having participants to respond to the shape of stimuli makes the shape dimension task relevant, so that it should receive top-down priming. However, stimuli are not the only events in reaction time tasks; participants also need to prepare and carry out particular actions. Given that actions can be assumed to be cognitively represented in terms of their perceptual effects (reflecting the ideomotor principle; see Hommel, Müsseler, Aschersleben, & Prinz, 2001; James, 1890; Lotze, 1852), preparing and controlling a particular set of actions entails attending the perceptual dimensions on which the actions are defined (Adam, Hommel, & Umiltà, 2003; Fagioli, Hommel, & Schubotz, 2007). If, as in our example, responses are defined by their spatial location (left vs. right keypress), this introduces task relevance of location in general. If we assume that the task relevance of a feature dimension leads to the priming of all feature values defined on it (Fagioli et al., 2007; Found & Müller, 1996; Hommel, 2007), making left and right *response* locations relevant should lead to the priming of any location information (i.e., whether it refers to a stimulus event or a response)—just as indicated in Figure 2a. Accordingly, the impact of (repetitions of) *stimulus* location increases as a consequence of choosing spatial responses. Indeed, Hommel (2007) could provide evidence that repetitions of stimulus location strongly affect binding-related effects with spatially defined response sets but not with nonspatial responses.

The model shown in Figure 2a suffices to account for the basic findings from most studies using the design introduced by Kahneman et al. (1992), and the binding process it implies reflects more characteristics of synchronization-based binding than of convergence-based integration. For one, it is difficult to see how convergence detectors might produce partial-repetition costs. Clearly, activating the same detector twice, as with the complete repetition of a particular shape-colour conjunction, say, should speed up performance and it commonly does. But consider a partial repetition (of either the shape or the colour) and a nonrepetition. Partial repetitions may be thought to also activate the same detector twice, though to a lower degree, or to fail activating the same detector. In the first case one would expect that performance for partial repetitions falls between complete repetitions and nonrepetitions; whereas in the second case one would expect that performance on partial repetitions and nonrepetitions is equally worse than on complete repetitions. As already mentioned, however, the standard finding is that performance on complete repetitions and nonrepetitions is equally good and better than on partial repetitions (e.g., Hommel, 1998). Closer

consideration of the possible noise and competition between alternative conjunction detectors helps a bit. For instance, one may assume that encountering one combination of two possible colours and two possible shapes (a red circle, say) leads to the activation of at least three conjunction detectors: A strong activation of the target detector ($\langle \text{red} \rangle + \langle \text{circle} \rangle$) and some milder activation of feature-sharing detectors (say, $\langle \text{red} \rangle + \langle \text{square} \rangle$ and $\langle \text{green} \rangle + \langle \text{circle} \rangle$). If the target repeats, all three detectors would be reactivated; this would lead to a particularly strong activation of the target detector, which now dominates the feature-sharing detectors even more. More concretely, if we count three activation units per target-induced activation and one unit per activation through feature overlap, this would mean that the target detector as an activation level of 6 as compared to an activation level of 2 for each of the two main competitors. It is easy to see that this approach correctly predicts worse performance with partial repetitions: If a red circle is followed by a green circle, say, this would first activate the same three detectors ($\langle \text{red} \rangle + \langle \text{circle} \rangle = 3$, $\langle \text{red} \rangle + \langle \text{square} \rangle = 1$, and $\langle \text{green} \rangle + \langle \text{circle} \rangle = 1$) as in the previous example and then the target detector ($\langle \text{green} \rangle + \langle \text{circle} \rangle = 3$) and the feature-overlapping detectors ($\langle \text{green} \rangle + \langle \text{square} \rangle = 1$ and $\langle \text{red} \rangle + \langle \text{circle} \rangle = 1$). Now the activation level of the target detector would be 4, just as much as its main competitor ($\langle \text{red} \rangle + \langle \text{circle} \rangle$), a situation that should result in impaired performance. The approach can also account for the observation that nonrepetitions produce better performance than partial repetitions. If a red circle is followed by a green square, this would activate ($\langle \text{red} \rangle + \langle \text{circle} \rangle = 3$, $\langle \text{red} \rangle + \langle \text{square} \rangle = 1$, and $\langle \text{green} \rangle + \langle \text{circle} \rangle = 1$) followed by ($\langle \text{green} \rangle + \langle \text{square} \rangle = 3$) and ($\langle \text{green} \rangle + \langle \text{circle} \rangle = 1$ and $\langle \text{red} \rangle + \langle \text{square} \rangle = 1$). The target activation would now add up to 3 and face activation levels of 2 in each of the two main competitors. Even though the target is now exposed to stronger overall competition than with partial repetitions (3:4 as compared to 4:4), the target would be in the position to outcompete each of the two competitors separately. Hence, choosing the right parameters, one may end up with a model that can account for better performance under nonrepetition than partial repetition. But again, it is difficult to see how it can account for *equal* performance under complete repetitions and nonrepetitions.

These difficulties suggest looking for alternative or at least additional mechanisms that are able to integrate features. Particularly promising with regard to the available findings seems the assumption that features are bound by an ad hoc binding mechanism, such as synchronizing the firing patterns of feature codes. If such a mechanism has just bound, say, the feature codes $\langle \text{red} \rangle$ and $\langle \text{circle} \rangle$, and if this binding has been maintained or stored, repeating one but not the other feature (as with a red square) could retrieve this binding (and thus reactivate the codes $\langle \text{red} \rangle$ and

<circle>), which would result in coding conflict between <circle> and <square> codes (Hommel, 2004). Coding conflict would only occur with partial repetitions but not with complete repetitions (where there is no conflict) or with alternations (where there is no retrieval). Another argument for a role of synchronization derives from the observation that manipulations targeting the muscarinic-cholinergic transmitter system affect both visual binding and synchronization in the visual cortex. Muscarinic-cholinergic agonists and antagonists have been demonstrated to respectively facilitate and impair neural synchrony in the gamma band (~30–70 Hz) in the visual cortex of the cat (Rodriguez-Bermudez, Kallenbach, Singer, & Munk, 2004) and muscarinic-cholinergic antagonists were found to interfere with feature binding in the rat (Botly & de Rosa, 2007). Consistent with that, binding-type interactions between repetitions of visual features in humans are boosted by caffeine (a muscarinic-cholinergic agonist) and reduced by alcohol (a muscarinic-cholinergic antagonist), but unaffected by nicotine (a nicotinic-cholinergic agonist that does not affect muscarinic pathways; Colzato, Erasmus, & Hommel, 2004; Colzato, Fagioli, Erasmus, & Hommel, 2005). These parallels are consistent with the assumption that visual feature binding is mediated by neural synchronization processes that are driven by muscarinic-cholinergic neurotransmitters (Colzato et al., 2005).

Recent observations however suggest that the model sketched in Figure 2a is incomplete in important ways. In a series of experiments, Colzato et al. (2006) investigated the relationship between binding and longer term learning. The basic idea was that learning particular feature conjunctions may change the way the features they entail are bound, by either increasing the strength of the binding (because it would find increasing support by long-term associations) or by eliminating binding effects (because online binding would be no longer necessary). Surprisingly, however, even though more frequent feature combinations facilitated performance as such, binding-related effects were not at all affected by learning. This was true for highly frequent arbitrary conjunctions of features, such as orientation and colour, and for real objects, like red strawberries and yellow bananas. Even though performance was better if strawberries appeared in red and bananas in yellow, there was no indication that, say, strawberries are more strongly (or weakly) bound to red than they are to yellow, pink, or purple. These findings suggest that binding and learning are less intimately related than one may think. Interestingly, however, a comparison between the different experiments of Colzato et al. suggested that real objects created larger partial-repetition costs (i.e., binding-related effects) than arbitrary combinations of simple features did, which may point to a contribution from long-term memory.

To account for this pattern of results, Colzato et al. (2006) suggested that feature integration may proceed via two routes, the ad hoc binding of

cooccurring features (presumably mediated by synchronization processes) and the registration of previously acquired conjunctions by conjunction detectors stored in long-term memory (presumably using convergence mechanisms). This brings into play long-term memory representations the way we sketched in Figure 2b. Whereas ad hoc integration takes place as described in Figure 2a, overlearning feature conjunctions are thought to establish a conjunction detector in long-term memory. These detectors may be of any complexity and thus function as object representations or cardinal cells in the sense of Barlow (1972). However, establishing a new detector makes sense only, so we suggest at least, under two conditions. First, the conjunction it can detect needs to be *significant* in the sense of reliably indicating a particular stimulus event and, second, the conjunction needs to be *diagnostic* in the sense that it should be functional in discriminating the given stimulus event from other events. The rationale of this reasoning is that devoting (presumably limited) cognitive structure to a task that in principle could also be solved by ad hoc binding presupposes some surplus functionality, which would not be given if a new conjunction detector would be unable to reliably detect the stimulus it stands for or discriminate it from alternative stimuli.

We further assume that, whenever a particular stimulus activates such a conjunction detector, the detector will provide top-down support by facilitating the processing of all the features belonging to the stimulus (Colzato et al., 2006)—which may be achieved by priming the respective stimulus dimensions and thus multiplying any stimulus-induced activation of codes falling on them. This assumption is grounded in evidence coming from several lines of research showing that it is easier to attend multiple features of the same object (e.g., Baylis & Driver, 1993; Duncan, 1984) and more difficult to ignore distractors if they are part of the same object (e.g., Baylis & Driver, 1992; Hommel, 1995; Kahneman & Henik, 1981). This implies that processing one feature of an object automatically opens the attentional gate to other feature dimensions of this object, whether this is useful or not. In the example shown in Figure 2b, this kind of top-down priming would facilitate the processing of shape and colour information, and of any other visual feature belonging to a dimension that defines the stimulus object. Given that stimulus location is not an object-defining feature, location information would not benefit from this top-down priming, however.

Two parallel mechanisms of feature integration could account for the observations of Colzato et al. (2006) in the following way. With arbitrary, not highly overlearned feature combinations that do not signify a unique object (as geometric shapes are commonly not related to or correlated with particular colours), the situation would be as depicted in Figure 2a: Shape coding would be primed, due to the task relevance of shape, but colour coding would not (stimulus location was not varied in that study).

Accordingly, even if shape–colour conjunctions would be automatically integrated, colour-induced retrieval would be weak at best and the corresponding effects would be modest and fragile. Indeed, Colzato et al. found only small effects reflecting shape–colour binding and even these effect tended to disappear with increasing practice (presumably due to increased focusing on the relevant shape information). Real objects with which participants are familiar would be more likely to have led to the creation of reliable and discriminative conjunction detectors or object representations in long-term memory. As depicted in Figure 2b, this would lead to a match between stimuli and the long-term representation and thus induce top-down facilitation of the object features, including both shape and colour. Accordingly, it would matter less that colour is actually not relevant for the task, implying stronger and more stable shape–colour binding effects—exactly what Colzato et al. observed.

AIM OF STUDY

The present experiment was set up to test the post hoc considerations of Colzato et al. (2006) in a more systematic fashion. We used a similar task (as sketched in Figure 1) but compared real objects and arbitrary feature conjunctions that varied on three dimensions (shape, colour, and stimulus location) in a within-subjects design. Shape was directly relevant for the task as participants were to discriminate and respond to the shape of S2. Even though stimulus location varied randomly and could safely be ignored, using spatially defined *responses* (left vs. right keypress) made location indirectly task relevant. Colour was entirely irrelevant. We expected the standard interactions between feature-repetition effects indicative of feature binding but were particularly interested in testing three more specific hypotheses.

First, we expected that interactions between shape and colour repetition (indicating shape–colour binding) would be more pronounced, and perhaps even restricted to, real objects. As explained already, real objects are likely to match representations stored in long-term memory, which should induce top-down priming of all object-related features. As shape is primed by task relevance anyway, it would be colour coding that benefits from this priming process, so that colour codes would interact more strongly with shape codes in the real-object condition.

Second, we expected that the difference between arbitrary feature combinations and real objects would not affect stimulus-location coding and, thus, not mediate location-related interactions. As explained earlier, long-term representations are unlikely to contain information about the location of a given object in space, as location is not an object-defining

attribute. Accordingly, location coding would not receive or benefit from top-down priming.

Third, we expected that the hypothetical real-object effect would be restricted to *stimulus-related* feature integration (and/or retrieval). Previous studies have shown that feature binding as such is not restricted to stimulus processing but operates across perception and action. Hommel (1998) has extended the classical preview design to include response repetitions by having participants to respond to the first stimulus (S1) with a previously cued and already prepared response (R1; see Figure 1). As in this design R1 does not correlate with the features of S1, stimulus features and responses can vary independently, so that stimulus–feature repetition and response repetition can be orthogonally manipulated (so to avoid the acquisition of stimulus–response associations). If this is done, the same type of crossover interaction as with stimulus–feature repetitions can be observed: Repeating a stimulus feature facilitates performance if the response also repeats but impairs performance if the response alternates (Hommel, 1998). Again, it seems that the mere single cooccurrence of a stimulus attribute and a response is sufficient to create a binding between their codes, so that repeating either the attribute or the response is sufficient to reactivate both or all members of this binding. Indeed, repeating some of the stimulus attributes induces a tendency to repeat the response as well in a free-choice reaction task (Hommel, 2007). However, visuomotor integration clearly differs from the process responsible for the integration of visual features. Not only do the two integration processes operate at a different point in time (visual binding seems to be stimulus locked, whereas visuomotor binding seems to be response locked; Hommel, 2005) but they are also driven by different neurotransmitter systems: Whereas muscarinic-cholinergic manipulations affect visual but not visuomotor binding (Colzato et al., 2004, 2005), manipulations targeting dopaminergic pathways affect visuomotor but not visual binding (Colzato & Hommel, 2008; Colzato, Kool, & Hommel, 2008; Colzato, van Wouwe & Hommel, 2007a, 2007b). Moreover, if it is true that establishing detectors for highly frequent, unique, and reliable feature conjunctions serves the purpose of facilitating object perception, it makes sense to assume that the impact of such detectors are restricted to perceptual processes—visual integration that is. Accordingly, we expected the standard interactions between visual-feature repetition and action repetition (Hommel, 1998) but no mediation of these effects by the arbitrary-conjunction vs. object manipulation.

METHOD

Thirty students of the Leiden University served as subjects for partial fulfilment of course credit or a financial reward. All reported having normal or corrected-to-normal vision, and were not familiar with the purpose of the experiment.

The experiment was controlled the Experimental Run-Time System (ERTS™) running on a PC attached to a 17-inch monitor. Participants faced three grey square outlines, vertically arranged, as illustrated in Figure 1. From a viewing distance of about 60 cm, each of these frames measured $2.6^\circ \times 3.1^\circ$. A banana ($0.3^\circ \times 0.6^\circ$), a strawberry ($0.5^\circ \times 0.6^\circ$), a triangle ($0.3^\circ \times 0.6^\circ$), and a circle ($0.5^\circ \times 0.6^\circ$) served as S1 and S2 alternatives, which were presented in blue or (close-to-magenta, purplish) pink (to avoid any preexperimental object–colour associations)¹ in the top or bottom frame. The stimuli were taken from Experiment 4 of Colzato et al. (2006; see p. 711 for bitmaps) and presented in the same colours (parameters were red = 0, green = 0, blue = 255, hue = 160, saturation = 240, and luminance = 120, for blue and red = 255, green = 0, blue = 255, hue = 200, saturation = 240, and luminance = 120, for pink). Response cues were also presented in the middle frame (see Figure 1), with rows of three left- or right-pointing arrows indicating a left and right keypress, respectively. Responses to S1 and to S2 were made by pressing the left or right shift-key of the computer keyboard with the corresponding index finger.

The experiment consisted of two sessions of 35 min, one with real objects (banana and strawberry) and one with arbitrary feature conjunctions (triangle and circle). In both sessions participants carried out two responses per trial, a previously cued simple response (R1) and a binary-choice

¹ More specifically, our idea was to get the hypothesized object representations in long-term memory involved—which required the use of stimuli that were likely to have memory representations—without letting them do the integration job on their own (i.e., without the need for ad hoc feature binding)—which required the use of feature combinations that were unlikely to be covered by these representations. Following Colzato et al. (2006), we thus used shapes of real objects (which should suffice to activate the memory representations) but presented them in colours that were unlikely to be part of the memory representation—using two colours that according to the findings of Colzato et al. are not associated with either of the two object shapes. Theoretically speaking, we expected that this manipulation would activate object representations and the corresponding conjunction detectors but still require ad hoc binding of the uncommon shape–colour conjunction. The former was considered to provide top-down priming of the latter with real objects but not with arbitrary feature conjunctions. As an example, facing a banana should activate a banana-related conjunction detector, which would lead to top-down priming of *all* identity-relevant features belonging to the *present* banana (i.e., to both the familiar shape and the in this case unfamiliar colour). This would prime the colour and increase the likelihood that it is being integrated. As the geometric shapes were not considered to have conjunction detectors linking them to particular colours, no top-down priming should occur for these shapes and present their colours.

response (R2) to the shape of the second of two target stimuli (S1 and S2; see Figure 1). They first prepared a left- or right-hand response as indicated by a cue (R1 cue) and carried it out (R1) upon presentation of the next stimulus (S1). S1 thus merely triggered the already prepared response; its features were entirely irrelevant and uncorrelated with the response. Then participants awaited the next stimulus (S2) and carried out a left- or right-hand response (R2) to its shape. Participants were informed that there would be no systematic relationship between S1 and R1, or between S1 and S2, and they were encouraged to respond to the onset of S1 only, disregarding the stimulus' attributes. The mappings of stimuli to response keys (S2 → R2) and the order of sessions were balanced across participants. The sequence of events in each trial is shown in Figure 1. The experiment was composed of 512 trials resulting from a factorial combination of the two possible shapes, colours, and locations of S2, the stimulus-type (real objects vs. arbitrary feature conjunctions) and the repetition vs. alternation of shape, colour, stimulus location, and the response, and three replications per condition.

RESULTS AND DISCUSSION

After excluding trials with missing (> 1500 ms) or anticipatory responses (< 200 ms), mean reaction times (RTs) and proportions of errors for R2 were analysed (see Table 1 for means and Table 2 for ANOVA terms). ANOVAs were run with stimulus type (real objects vs. arbitrary feature conjunctions), the repetition versus alternation of stimulus shape, colour, and location (S1 → S2), and of the response (R1 → R2) as within-participant factors.

In RTs, the main effect of stimulus type indicated that subjects reacted faster to the arbitrary feature conjunctions than to real objects and the stimulus-location repetition costs in RTs and errors reflect “inhibition of return”—the common observation that attending to an irrelevant stimulus impairs later responses to relevant stimuli appearing in the same location (Posner & Cohen, 1984). More interesting for present purposes were the interactions. First, shape repetition interacted with stimulus location (in RTs) and with colour (in RTs and errors)—repeating one object feature but not the other feature impaired performance as compared to complete repetitions and alternations. Importantly, only the shape-by-colour interaction was modified by stimulus type, thus producing a three-way interaction. As suggested by Figure 3, the shape-by-colour interaction was considerably more pronounced with real objects, $F(1, 29) = 28.14$, $p = .0001$, than with arbitrary feature conjunctions, where the interaction was not reliable, $F(1, 29) = 1.39$, $p = .247$. We checked whether the stronger interaction with real objects might be due to the higher RT level in this condition. However, the

TABLE 1
 Means of mean reaction times and standard deviations (*SD*) for responses to Stimulus 2 (RT, in ms) and percentages of errors on R2 (PE), as a function of stimulus type (real objects vs. arbitrary feature conjunctions), the match between Response 1 and Response 2, and the feature match between Stimulus 1 and Stimulus 2

Match	Response							
	Arbitrary feature conjunctions				Real objects			
	Repeated		Alternated		Repeated		Alternated	
	RT (<i>SD</i>)	PE (<i>SD</i>)	RT (<i>SD</i>)	PE (<i>SD</i>)	RT (<i>SD</i>)	PE (<i>SD</i>)	RT (<i>SD</i>)	PE (<i>SD</i>)
Neither	482 (19)	10.6 (2.4)	474 (21)	2.3 (1.9)	511 (20)	10.6 (1.9)	482 (22)	2.1 (1.0)
Shape (S)	499 (25)	7.9 (1.7)	492 (21)	10.2 (2.0)	510 (20)	7.7 (1.2)	515 (21)	7.9 (1.7)
Location (L)	522 (27)	9.6 (2.0)	495 (21)	2.7 (0.1)	532 (23)	7.7 (1.9)	509 (17)	5.8 (1.3)
Colour (C)	497 (26)	14.2 (2.9)	463 (21)	2.7 (0.7)	512 (24)	10.8 (2.2)	483 (19)	4.3 (1.2)
S × L	475 (19)	3.9 (1.3)	499 (17)	13.7 (1.9)	497 (20)	4.8 (0.8)	523 (16)	14.6 (2.2)
C × L	512 (24)	10.4 (2.1)	498 (23)	6.9 (1.1)	549 (27)	10.2 (2.5)	528 (22)	7.1 (2.0)
S × C	494 (27)	5.4 (1.4)	478 (22)	8.1 (1.6)	495 (22)	5.2 (0.8)	510 (20)	5.0 (1.1)
S × L × C	462 (20)	3.7 (1.1)	497 (20)	14.6 (3.0)	477 (19)	2.5 (0.9)	499 (20)	12.7 (2.6)

outcome of two analyses speaks against this possibility. First, we median-split participants by their mean RT in the real-object condition and reran the ANOVA with level as additional between-participant variable. Whereas the three-way interaction was still reliable, $F(1, 28) = 4.80, p = .04$, there was no hint for any mediation by level, $F(1, 28) < 1$. Second, we computed, for each participant, the increase in shape-colour effect size and the increase in RT from the arbitrary-feature-conjunction condition to the real-object condition.² Given that the two measures were uncorrelated, $r = .06, p > .7$, it seems safe to conclude that a higher RT level as such does not increase the shape-by-colour interaction. The observation that the interaction is mediated by the type of stimulus fully supports our expectation that real objects provide top-down priming for colour coding that compensates for the lack of task relevance. Accordingly, colour did not interact with any other feature dimensions in the case of arbitrary conjunctions but it did interact with shape, the other object-specific feature dimension, in the case of real objects.

A second cluster of interactions in both RTs and error rates involved response repetition. It interacted with shape repetition and with stimulus-location repetition, and was involved in a three-way interaction with shape

² Effect sizes were computed by subtracting the mean RT for complete repetitions and alternations from the mean RT for partial repetitions (i.e., shape repetition and colour alternation or shape alternation and colour repetition). Note that this amounts to the interaction term corrected for possible main effects.

TABLE 2
 Results of analysis of variance on mean reaction time of correct responses (RT) and percentage of errors (PE)

Effect	RT_{R2}		PE_{R2}	
	MSE	F	MSE	F
Stimulus type (T)	16069.44	5.04*	52.98	0.07
Shape (S)	6254.84	2.39	112.02	0.84
Colour (C)	6909.71	0.58	61.11	0.43
Location (L)	3292.18	8.92**	38.78	6.06*
Response (R)	55221.34	1.32	223.37	0.45
T × S	1362.66	0.50	45.02	0.01
T × C	1991.67	0.10	62.18	0.09
T × L	1902.42	0.05	48.32	0.05
S × C	1277.82	13.62***	47.95	16.16**
T × S × C	916.60	4.97*	31.46	0.25
C × L	1297.17	0.02	81.96	0.87
T × C × L	1523.76	0.13	61.09	1.41
S × L	8461.69	10.01**	65.27	1.56
T × S × L	1800.15	0.36	36.28	0.54
S × C × L	4844.92	0.28	53.57	0.30
T × C × S × L	2747.68	2.06	53.76	1.74
T × R	2574.38	0.08	80.22	0.24
S × R	2005.57	36.98***	169.77	44.73***
T × S × R	1665.66	1.64	54.28	2.74
C × R	1942.08	0.02	47.21	0.22
T × C × R	1374.84	0.37	30.53	0.01
S × C × R	1672.61	0.13	45.61	0.60
T × S × C × R	1450.68	0.05	35.13	0.12
L × R	1281.34	12.20**	132.17	26.25***
T × L × R	1038.80	2.21	49.11	0.48
C × L × R	1715.44	1.36	59.88	1.69
T × C × L × R	3157.48	1.61	48.81	1.76
S × L × R	1086.53	7.92**	44.76	10.21**
T × S × L × R	1759.39	2.56	44.42	0.29
C × S × L × R	2297.17	0.37	55.09	0.29
T × C × S × L × R	862.35	0.01	50.56	1.85

* $p < .05$, ** $p < .01$, *** $p < .001$; $df = 1, 29$ for all effects.

and stimulus location. The latter indicated that location only interacted with the response if shape was repeated, $F(1, 29) = 31.92$, $p < .001$ (RTs), and $F(1, 29) = 24.89$, $p < .001$ (errors), but not if shape alternated, $ps > .4$. The reliable interactions all followed the standard form with better performance for shape repetitions and for stimulus-location repetitions if the response repeated but worse performance for shape repetitions and stimulus-location repetitions if the response alternated. Most interesting for present purposes, stimulus type and response repetition were not involved in any reliable

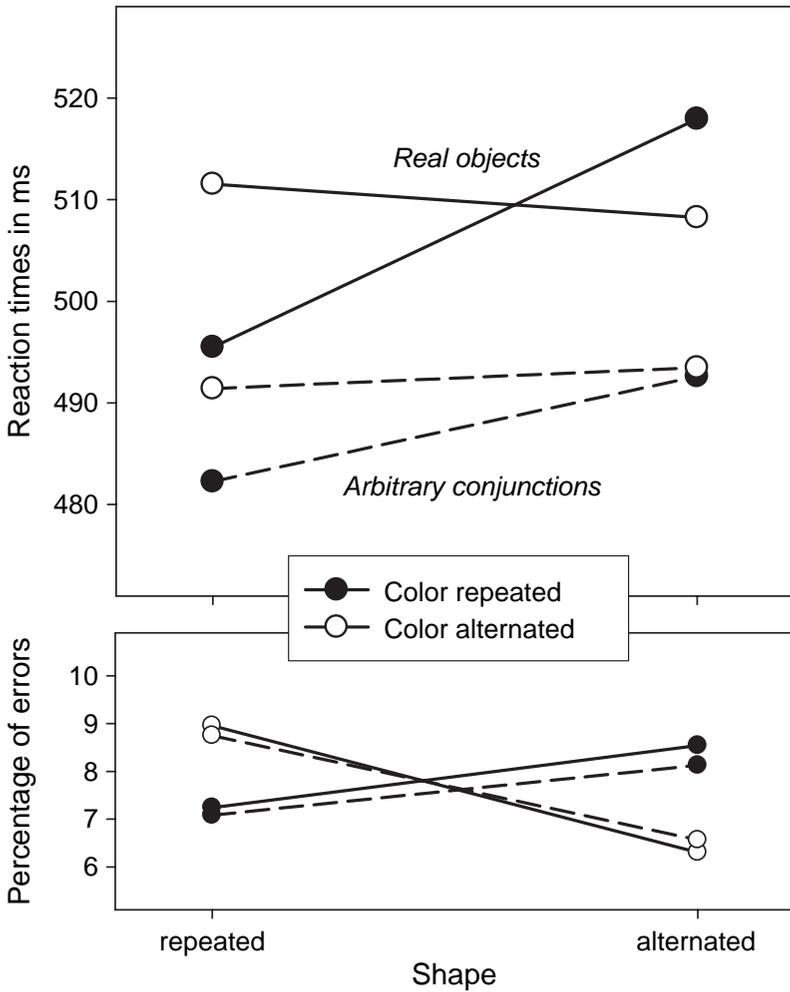


Figure 3. Mean reaction times and error percentages for R2 as a function of stimulus type and of repetition versus alternation of stimulus shape and colour.

interaction. This suggests that visual integration is influenced by stimulus type but visuomotor integration is not.

CONCLUSIONS

Taken altogether, our findings support the assumption of two different feature-integration mechanisms in visual perception (see also VanRullen, 2009 this issue). One mechanism seems to be agnostic about the familiarity

or possibility of particular combinations of features and integrates any feature that falls into a given temporal integration window (Akyürek, Toffanin, & Hommel, 2008). We speculate that this mechanism is mediated by, or relies on neural synchronization processes, as is suggested by the observation that both visual integration and synchronization in the visual cortex seem to be driven by the same neurotransmitter system. The other mechanism is sensitive to the familiarity with the stimulus and it seems to provide attentional top-down support for real, familiar objects. Recognizing such objects presupposes a stored detector of the underlying feature combinations, and we speculate that such detectors are created for frequent combinations only—even though we cannot rule out the possibility that the naturalness of the stimulus also plays a role. In any case, there are reasons to assume that feature integration can take place in more than one way and that the principles of integration-through-convergence and integration-through-synchronization do not exclude but complement each other.

More generally speaking, our findings provide support for the assumption that the retrieval of object files is cocontrolled by two types of top-down priming processes. Offline priming, as one may call the impact of the current attentional set, precedes the stimulus and reflects the task relevance of feature dimensions for selecting the stimulus and the response (see Figure 2a and b). This priming is offline in the sense that it can be established any time before a given stimulus or response event occurs. Online priming, as it may be characterized, can be induced by stimuli that have entries in long-term memory, such as familiar real objects. Sensory information coming from these stimuli is likely to access corresponding memory entries in a first fast forward sweep, followed by a recurrent top-down process refining and contextualizing the input (Lamme & Roelfsema, 2000). Whereas the first, bottom-up part of this scenario is likely to be rather nonselective, the recurrent wave will be shaped by the current attentional settings. Given that this wave follows the first contact between the sensory information and the memory content, the outcome of this contact will contribute as well. Accordingly, the eventual representation of the present stimulus and the degree to which this representation is permitted to reactivate available object files will thus be codetermined by the task set and the stimulus-induced memory activation—provided that the stimulus matched some memory content it could activate.

REFERENCES

- Adam, J., Hommel, B., & Umiltà, C. (2003). Preparing for perception and action I: The role of grouping in the response-cuing paradigm. *Cognitive Psychology*, *46*, 302–358.
- Akyürek, E. G., Toffanin, P., & Hommel, B. (2008). Adaptive control of event integration. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 569–577.

- Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology. *Perception, 1*, 371–394.
- Baylis, G. C., & Driver, J. (1992). Visual parsing and response competition: The effect of grouping factors. *Perception and Psychophysics, 51*, 145–162.
- Baylis, G. C., & Driver, J. (1993). Visual attention and objects: Evidence for hierarchical coding of location. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 451–470.
- Botly, L. C. P., & de Rosa, E. (2007). Cholinergic influences on feature binding. *Behavioral Neuroscience, 121*, 264–276.
- Colzato, L. S., Erasmus, V., & Hommel, B. (2004). Moderate alcohol consumption in humans impairs feature binding in visual perception. *Neuroscience Letters, 360*, 103–105.
- Colzato, L. S., Fagioli, S., Erasmus, V., & Hommel, B. (2005). Caffeine, but not nicotine enhances visual feature binding. *European Journal of Neuroscience, 21*, 591–595.
- Colzato, L. S., & 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., Kool, W., & Hommel, B. (2008). Stress modulation of visuomotor binding. *Neuropsychologia, 46*, 1542–1548.
- 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.
- Colzato, L. S., van Wouwe, N. C., & Hommel, B. (2007a). Feature binding and affect: Emotional modulation of visuo-motor integration. *Neuropsychologia, 45*, 440–446.
- Colzato, L. S., van Wouwe, N. C., & Hommel, B. (2007b). Spontaneous eyeblink rate predicts the strength of visuomotor binding. *Neuropsychologia, 45*, 2387–2392.
- DeYoe, E. A., & van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neuroscience, 11*, 219–226.
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113*, 501–517.
- Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Science, 5*, 16–25.
- Fagioli, S., Hommel, B., & Schubotz, R. I. (2007). Intentional control of attention: Action planning primes action-related stimulus dimensions. *Psychological Research, 71*, 22–29.
- Found, A., & Müller, H. J. (1996). Searching for unknown feature targets on more than one dimension: Investigating a “dimension weighting” account. *Perception and Psychophysics, 58*, 88–101.
- Hommel, B. (1995). Attentional scanning in the selection of central targets from multi-symbol strings. *Visual Cognition, 2*, 119–144.
- 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. (2005). How much attention does an event file need? *Journal of Experimental Psychology: Human Perception and Performance, 31*, 1067–1082.
- Hommel, B. (2007). Feature integration across perception and action: Event files affect response choice. *Psychological Research, 71*, 42–63.
- Hommel, B., & Colzato, L. S. (2004). Visual attention and the temporal dynamics of feature integration. *Visual Cognition, 11*, 483–521.
- Hommel, B., & Elsner, B. (in press). Acquisition, representation, and control of action. In E. Morsella, J. A. Bargh, & P. M. Gollwitzer (Eds.), *The psychology of action, Vol. 2*. Oxford, UK: Oxford University Press.

- Hommel, B., Memelink, J., Zmigrod, S., & Colzato, L. S. (2008). How information of relevant dimension control the creation and retrieval of feature–response binding. *Manuscript submitted for publication*.
- 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*, 849–878.
- James, W. (1890). *The principles of psychology*. New York: Dover Publications.
- Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 181–211). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175–219.
- 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.
- Keizer, A. W., Colzato, L. S., Theeuwisse, W., Nieuwenhuis, S., Rombouts, S. A. R. B., & Hommel, B. (2008). When moving faces activate the house area: An fMRI study of object file retrieval. *Manuscript submitted for publication*.
- Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neuroscience*, *23*, 571–579.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.
- Lotze, R. H. (1852). *Medicinische Psychologie oder die Physiologie der Seele*. Leipzig, Germany: Weidmann'sche Buchhandlung.
- Mitroff, S. R., & Alvarez, G. A. (2007). Space and time, not surface features, underlie object persistence. *Psychonomic Bulletin and Review*, *14*, 1199–1204.
- Posner, M. I., & Cohen, Y. A. (1984). Components of visual orienting. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X* (pp. 531–554). Hove, UK: Lawrence Erlbaum Associates Ltd.
- Raffone, A., & Wolters, G. (2001). A cortical mechanism for binding in visual working memory. *Journal of Cognitive Neuroscience*, *13*, 766–785.
- Rodriguez-Bermudez, R., Kallenbach, U., Singer, W., & Munk, M. H. (2004). Short- and long-term effects of cholinergic modulation on gamma oscillations and response synchronization in the visual cortex. *Journal of Neuroscience*, *24*, 10369–10378.
- Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*, 171–178.
- Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*, 107–141.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.
- Van Dam, W. O., & Hommel, B. (2008). How object-specific are object files? Evidence for integration by location. *Manuscript submitted for publication*.
- Van der Heijden, A. H. C. (1995). Modularity and attention. *Visual Cognition*, *2*, 269–302.
- VanRullen, R. (2009). Binding hardwired vs. on-demand feature conjunctions. *Visual Cognition*, *17*, 103–119.
- Von der Malsburg, C. (1999). The what and why of binding: The modeler's perspective. *Neuron*, *24*, 95–104.
- Xu, Y. (2006). Encoding objects in visual short-term memory: The roles of feature proximity and connectedness. *Perception and Psychophysics*, *68*, 815–828.