8 Perceptual Organization and Visual Target Selection

Cees van Leeuwen
UNIVERSITY OF LEUVEN, BELGIUM;
UNIVERSITY OF KAIERSLAUTERN, GERMANY

Tina Weis
UNIVERSITY OF KAIERSLAUTERN, GERMANY

Thomas Lachmann
UNIVERSITY OF KAIERSLAUTERN, GERMANY;
UNIVERSITY OF LEUVEN, BELGIUM

Abstract

Visual information processing is linked with feature selection and integration (also known as perceptual organization). The focus of the current chapter will be on the interaction of these processes in different stages of visual information processing, giving rise to various interference effects, such as crowding, Stroop and Garner interference, as well as specific interference characteristic of letters and numbers. The stagewise organization of visual perception is reviewed based on experimental studies using behavioral responses or measurements of brain signals, including event-related potentials (ERP) and the electroencephalogram (EEG). Even though perceptual organization can be fast, it interacts with early (crowding) and late (Stroop-congruence) visuospatial target selection. In letter processing, the association with phonemes makes that letters are preferably processed analytically, giving rise to negative congruence effects; for numbers we observe congruence effects in a generic magnitude space instead of the visual space. Furthermore, depending on the stimulus, perceptual organization can also be slow in the presence of distractors.

8.1 Introduction

The visual system allows us to orient ourselves in our environment, to locate and identify static or moving structures, textures, and shapes, to recognize people, animate and inanimate objects and scenes, and to read signs and symbols. All these functions fall under the canopy of “visual information processing”. Information is visual to the extent that it is influenced by the patterns
of light that enter the visual system via the retinal mosaic. The passage through the retinal mosaic makes visual information a point process; any non-local information is initially lost and must be reconstructed (Koenderink, 2012). Like the furniture from IKEA, the signal must be unpacked and assembled before use. In visual information processing, “unpacking” is known as feature selection, or filtering, “assembling” as feature integration, or perceptual organization. Our focus will be on the stagewise organization of selection and organization processes, with particular emphasis on attentional selection and the role of perceptual bias in perceptual organization.

Assembling furniture could be done without reading the instructions, but sometimes it is unclear what goes with what. Relying on our gut feelings, we may regret the outcome later. In perceptual organization, perceptual bias (Gepshtein & Kubovy, 2005) plays the role of “gut feelings”; for “instructions”, we have attentional selection. We will consider in particular visuospatial attention (C. W. Eriksen & St. James, 1986) advising us where to put things together and where not. Feature or object-based attention will not receive the same consideration. We suspect, however, that perceptual bias might cover some of the associated phenomena.

Sometimes, ignoring the instructions can speed up the process; sometimes, doing so gets into the way of getting things right. If perceptual organization competes with attention, we will speak of interference. Interference can take many forms. We will discuss crowding and Stroop and Garner interference in visual shapes, as well as specific interference characteristic of letters and numbers.

Our perspective will be on perceptual organization and visuospatial attention as neurocognitive processes. Perceptual organization is often identified with what we experience when we look at a display. The precise relationship of conscious experience and the corresponding processes is the hard problem of cognitive neuroscience (Chalmers, 1995). Our views on this matter are presented elsewhere (Fekete, van Leeuwen, & Edelman, 2016; Raffone, Srinivasan, & van Leeuwen, 2014a, 2014b; van Leeuwen, 2007, 2015). These views license us to use experiences as correlates of neurocognitive processes, even when the exact relationship with these processes is unknown. In experimental studies, behavioral responses (such as reaction times (RTs) or detection scores) as well as brain signal measurement (e.g., event-related potentials (ERP) or event-related field (ERF) in the electro- or magnetoencephalogram (EEG or MEG, respectively); see Luck, 2014, for an overview) have been used as measures to analyze perceptual organization processes. The “easy” problem, how exactly behavioral responses and brain signals relate to these processes, is equally unresolved. For where we stand on the latter, see Alexander, Trengove, and van Leeuwen (2015) or Alexander et al. (2013). Here we will follow the common practice of taking these measures as correlates of brain processes.
8.2 Perceptual Organization, Fast and Slow

Perceptual organization plays a central role in all visual functions, many of which take place in dynamic environments. At least some perceptual organization processes, therefore, need to be fast. Non-local properties such as convexity and parallelism are generally identified more rapidly than their constituent parts (Kimchi, 2015). This phenomenon is called configural superiority. To illustrate the power of configural superiority: Pomerantz (1983) asked participants to distinguish between “)” and “(”. Doing so turned out much easier an extra parenthesis is added: “()” versus “()”). The resulting two parentheses configurations gave rise to convexity and parallelism. Notably, the added parenthesis is the same in both stimuli.

Not all perceptual organization processes are fast. Consider, for instance, the time it takes to detect the hidden structure in Archimboldo’s and some of Dali’s paintings. When we observe an ambiguous figure like the Necker cube, it can take several seconds before we experience a switch in its orientation. Switches appear sudden but, as far as the corresponding brain activity is concerned, are complex processes that take about one second to complete (Nakatani, Orlandi, & van Leeuwen, 2011; Nakatani & van Leeuwen, 2006, 2013).

These effects illustrate that even without changes to the stimulus, perceptual organization does not always provide definitive “output” but rather belongs to an ongoing cycle (Neisser, 1976) in which the percept is continually being revised and updated. In an ongoing cycle, there is no unambiguous ordering of stages. This basic observation is routinely ignored when experimenters focus on the processes that follow the onset of a stimulus and study how these are ordered in time.

In this restricted framework, we may think of configural superiority as reflecting an early stage of perceptual organization. Further subdivision might seem possible. In occluded figures, a mosaic interpretation was believed to precede an occlusion interpretation (Sekuler & Palmer, 1992). Apparently, before we see a square partially occluded by a circle, we represent the stimulus as a mosaic of a square and a circle segment. This result suggests that at least initially, the fast processes are insensitive to the 3-D structure of the world. This study, however, used behavioral responses (RTs), which makes it difficult to determine the precise time course of processing. A finer temporal resolution can be obtained by measuring evoked brain activity. ERF studies revealed mosaic and completion to arise in the same time periods. Moreover, these alternative organizations activated the same visual areas, in particular, the right fusiform gyrus (Plomp, Liu, van Leeuwen, & Ioannides, 2006). These results suggest that fast configural processing supports alternative structures, in particular also including ones that give rise to occlusion interpretations.
Plomp and colleagues showed further that either interpretation could be precipitated, depending on preceding context and expectation (Plomp et al., 2006; Plomp & van Leeuwen, 2006). This is commonly understood as follows: The fast processes initiate alternative candidate organizations, resulting in competition that is resolved through attentional processes that bias the competition (Desimone & Duncan, 1995).

Evoked brain activity apparently reflects this stagewise ordering of processing. The earliest ERP component, C1, peaks at 60–90 ms after stimulus onset is selective to configurations arise within 60 ms (e.g., Nikolaev & van Leeuwen, 2004), before the onset of attention. Attention kicks in 80–100 ms from stimulus onset, although the transition is gradual (Clark, Fan, & Hillyard, 1994; Di Russo, Martinez, & Hillyard, 2003; Martinez et al., 1999). Best, therefore, to consider the stages as cascaded. The next components, P1, typically 80–150 ms after stimulus onset and N1, typically 150–200 ms after stimulus onset, reflect perceptual organization (Han, Jiang, Mao, Humphreys, & Qin, 2005; Khoe, Freeman, Woldorff, & Mangun, 2006). These components, however, also show effects of visuospatial attention (reviewed by Hillyard, Vogel, & Luck, 1998), plausibly to selectively enhance the target configuration or suppress neighboring distractors (Desimone & Duncan, 1995; Mounts & Tomaselli, 2005), seemingly in accordance with the biased competition view.

We should be mindful, however, that this understanding was reached in the restricted framework of stimulus-evoked brain activity. These methods fail to consider that perception is ongoing and that processes in anticipation of the stimulus may also play a role. Spontaneous, pre-stimulus EEG or MEG activity recording might be informative in this respect (Nikolaev, Gepshtein, & van Leeuwen, 2016). In stimuli with ambiguous orientations, the authors showed that these stimuli were preferably organized according to the principle of proximity. However, there were episodes lasting several trials, in which the stimuli were predominantly oriented vertically, irrespective of proximity. Nikolaev et al. (2016) explained this result as an effect of intrinsic perceptual bias.

In contrast to decision bias, which is a systematic preference for a certain response category (P. R. Jones, Moore, Shub, & Amitay, 2015; Wenger & Rasche, 2006), perceptual biases involve a systematic preference for a certain stimulus attribute (Albright, 2012; Allport, 1967; Chopin & Mamassian, 2011; Gepshtein & Kubovy, 2005; Harrison & Backus, 2010; van Dam & Ernst, 2010; Weiss, Simoncelli, & Adelson, 2002). Many animals, for instance, prefer symmetric over asymmetric features in mates (Moller & Thornhill, 1998); human symmetry bias emerges early in infancy, around four months of age (Bornstein, Ferdinandsen, & Gross, 1981; see Lachmann, 2002 or Lachmann & Geyer, 2004, for a review).

Perceptual bias plays a role in perceptual organization when the stimulus is ambiguous (Gepshtein & Kubovy, 2005), or when stimulus information is uncertain, for instance when it is partially occluded. Note that, in this
understanding, perceptual biases are to be distinguished from attention. In Nikolaev et al. (2016), bias was more pronounced in trials where allocation of attention was lacking, as reflected by a reduction in the N1 component. As a result, the stimulus effectively became more ambiguous.

These results show that an understanding, in which fast processing is biased by spatial attention is incomplete at best and at worst confounding. At the least, the role of bias needs to be reconsidered. Perceptual bias is not a component of attention. In Nikolaev et al. (2016), bias favored vertical over horizontal orientations. Horizontally oriented contours are often discriminated better than vertical ones, and these better than oblique ones (Appelle, 1972). However, the opposite occurs in more natural stimulation conditions, with broadband spatial frequency characteristics (Essock, DeFord, Hansen, & Sinai, 2003). These authors suggested that orientation bias is deployed actively by the visual system, to suppress the more common (and therefore less informative) aspects of natural stimulation. The active role of bias was consistent with the observation in Nikolaev et al. (2016) that early during the experiment, when participants are still learning the task, biased responses are preceded by reduced pre-stimulus alpha power in the visual areas. This effect is generally understood as top-down influence on perception (Bonnefond & Jensen, 2012; Handel, Haarmeier, & Jensen, 2011).

The results of Nikolaev et al. (2016) showed bias to influence perceptual organization independently of attention. We must reconsider the question, whether attention is involved in perceptual organization. According to Roelfsema (2006), we must distinguish configurations that are standing out from ones that arise in a context of a set of distracting elements. The former configurations are detected fast, without involvement of attention (Mitroff & Scholl, 2005). According to Roelfsema, this happens within a single, feedforward sweep of activity through the visual hierarchy (Roelfsema, 2006; Roelfsema, Lamme, & Spekreijse, 1998; Schmidt & Schmidt, 2009).

Since Hubel and Wiesel (1959) feedforward, hierarchical organization has been our default understanding of visual information processing. The visual input signal proceeds from the retina through the lateral geniculate nuclei to the primary visual cortex. Hubel and Wiesel (1974) found the primary visual cortex to encode basic features of oriented contour segments, systematically laid out in an orientation preference map. Likewise, for other features such as color, form, location, and spatial frequency maps are found in the early visual cortex. One step higher in the hierarchy, more complex feature maps are found, responsible for curvature and closure (visual area V4); still higher up the hierarchy (e.g., the right fusiform gyrus) increasingly complex patterns are encoded in an increasingly sparse and location-independent fashion.

Detection happens, according to this view, where and when the distinctive feature of the configuration becomes available. Closure, for instance, becomes available as soon as closure detectors, probably in V4, are
activated through converging signals entering from V1 and V2. Mosaic and occlusion information will become available only tens of milliseconds later, when the feedforward sweep reaches the right fusiform gyrus.

But the notion that a configuration is detected in a single, feedforward sweep ignores the puzzle that configural superiority poses to the hierarchical account. Simple features are activated in the system before configurations. Why, then, are the latter detected first? More generally, activation of feature detectors in some brain region does not equal detection, which is a response at system level. How is the local activation communicated to the areas of the brain that initiate the detection response (de Wit, Alexander, Ekroll, & Wagemans, 2016).

Let us not forget, moreover, that intrinsic perceptual bias can also be a factor in the rapid detection of a configuration. If biases inform our decision making, at least it could be understood how local activity in a certain area (V4 or the right fusiform gyrus) could give rise to a rapid response. We might consider the possibility that biases are imposed in a top-down manner to gate feedforward activity to the areas responsible for initiating a response (Jensen & Mazaheri, 2010). This might also explain why familiarity can facilitate perceptual organization before figure-ground segregation (Peterson, 1994) or even before edges are assigned to objects (Kahan & Enns, 2014).

To obtain a configuration in the presence of distractors, according to Roelfsema (2006), several rounds of updating involving interactions with attentional processes are needed. Attentional involvement is needed in the presence of distractors because of a phenomenon known as crowding (Bouma, 1970). Crowding is the effect that target identification is reduced in the presence of nearby distractors. Chicherov, Plomp, and Herzog (2014) showed that evoked activity shows a crowding effect around 175/200 ms from stimulus onset. The timing of this effect is consistent with its dependence on visuospatial selection. Crowding is stronger in peripheral vision because spatial resolution is lower there, but also occurs in the fovea (e.g., Lev, Yehezkel, & Polat, 2014). Crowding effects are based on perceptual grouping of target and nontarget features (Herzog & Fahle, 2002), rather than on their proximity per se. Grouping strength is influenced by proximity, but also by similarity and uniform connectedness (Rosen, 2013), target complexity and familiarity (Chanceaux, Mathot, & Grainger, 2014). We may therefore consider perceptual grouping as belonging to the fast perceptual organization processes involved in configural superiority. Crowding shows that these processes can be a liability; the configuration can lead to reduced discrimination of its components (Mathis & Kahan, 2014; Poljac, de Wit, & Wagemans, 2012). This effect may interfere with visuospatial target selection.

The need of attentional involvement to exclude distractor features means that perceptual organization will become incremental. Attention and integration operate in tandem, whereby attention guides the search for stimulus
features to extend the configuration. Yet again, how would attention select the candidates for extending the configuration? Again there is a potential role for perceptual biases. Rather than integration being driven by a free-floating attention, perceptual configurations may accumulate through an interaction, in which the current integration, gated by the bias, direct attentional selectivity to the candidate location(s) to be integrated at the next iteration.

In order to understand bias, we must appreciate the role of top-down activity in perception. In Hochstein and Ahissar (2002) reverse hierarchy theory, the visual system is actively anticipating visual stimulation. High-level representations are pre-activated, selectively enhancing (Lamme, 1995) or suppressing (Corthout & Supèr, 2004) lower-level activity. Successful top-down projections, in turn, yield sustained activity feeding into the higher-level representations. Top-down feedback, all the way down to primary visual cortex has been observed in animal studies. Macaque neurons in visual cortex responded more strongly to texture elements belonging to a figure than to elements belonging to a background (Lamme, Super, & Spekreijse, 1998; Roelfsema et al., 1998; Zipser, Lamme, & Schiller, 1996).

Specifically, bias could perhaps be understood as priors in predictive coding theory: At each level of the hierarchy, the prediction error between them is minimized by modifying the neural activity, such that only the error signal is propagated upward to the next-higher level (Rao & Ballard, 1999), where it can be used to adjust the prediction. Seeing a configuration (e.g., convexity), thus equates the degree to which convexity detectors are effectively silenced.

Opposite to what prediction coding would expect, however, Kok, Jehee, and de Lange (2012) found that reduction in lower levels corresponded to silencing of detectors that are inconsistent with the current expectations. Such active sparsification may be a way to establish selectivity dynamically (e.g., Lorincz, Szirtes, Takacs, Biederman, & Vogels, 2002). Then again, multi-unit recording studies in ferrets and rats have provided evidence against active sparsification in visual cortex (Berkes, White, & Fiser, 2009). Since both selectively reinforcing (Lamme, 1995) and silencing (Corthout & Supèr, 2004) of lower-level activity have been observed, both process may play a role when integration proceeds incrementally (Roelfsema, 2006; Roelfsema et al., 1998). However, their precise role in perceptual organization remains unclear.

Intrinsic holism is the view that biases reflect the coherent representations arising spontaneously at each level of the visual hierarchy. For these representations, lateral connections are to be held responsible. Feature maps in primary visual cortex are linked by long-range intrinsic connections (see Alexander & van Leeuwen, 2010, for a review). These connections are excitatory, and link subareas with similar selectivity (Lund, Yoshioka, & Levitt, 1993; Malach, Amir, Harel, & Grinvald, 1993). Their function
may be pooling local orientation preferences for detection of curves and texture flows (Ben-Shahar, Huggins, Izo, & Zucker, 2003). Texture flows can be used to identify regions. This may illustrate that information about textures, regions, and occlusion between surfaces is already available early in the perceptual hierarchy. This information may be modulated by top-down activity from higher areas. Unlike top-down activation, lateral processes are fast enough to play a role in configural superiority, provided that population-level selectivity could effectively be read out from the early visual areas to bias the incoming stimulation (Rentzeperis, Nikolaev, Kiper, & van Leeuwen, 2014).

8.3 Stroop and Garner Interference

Competition of perceptual organization and attentional selectivity has also been demonstrated in Stroop- and Garner-interference effects (Garner, 1974, 1988; Stroop, 1935). Both Stroop- and Garner-interference show perceptual organization by observing how responding to categorical properties of the target is made more difficult by a nontarget feature (distractor). Stroop interference is generally defined as a deterioration of a behavioral response to a target category, resulting from one or more irrelevant other features, presented at the same trial, that call for an opposite response. The classical example involves naming the ink color, “red”, of the word BLUE as opposed to the word RED (Stroop, 1935). The label “Stroop” has been applied to a broad variety of tasks (for an extensive review of the early literature see MacLeod, 1991), including auditory versions (Hamers & Lambert, 1972), the Eriksen flanker paradigm (B. A. Eriksen & Eriksen, 1974, mentioned in MacLeod, 1991 as “analog of the Stroop task”), tasks using individual faces and names (Egner & Hirsch, 2005), numerical values and physical sizes (Algom, Dekel, & Pansky, 1996) or names of countries and their capitals (Dishon-Berkovits & Algom, 2000).

For perceptual organization, it is relevant that Stroop effects occur in tasks involving shape configurations (Boenke, Ohl, Nikolaev, Lachmann, & van Leeuwen, 2009; Patching & Quinlan, 2002; van Leeuwen & Bakker, 1995). In Boenke et al. (2009), a closed shape consisted of two connected parts, which could either be triangular or rectangular in shape. One of these was the target, the other the distractor. Perception of the target was facilitated when both were of the same shape (congruent) as opposed to when both were different in shape (incongruent). Congruence effects (for a review: Marks, 2004; Patching & Quinlan, 2002) are specific to the domain of visual configurations, which include a target and distractor component. Both are presented together, large enough to be clearly visible, in an environment free of clutter, limiting the influence of crowding effects.
Whereas according to Pomerantz (1983), congruence-based Stroop effects are tied to the experience of non-local properties such as convexity or closure, van Leeuwen and Bakker (1995) showed that this is not necessary. The same mechanism covers interactions between closed figures (van Leeuwen & Bakker, 1995) or between remote stimulus components. For instance, a congruence-based Stroop effect occurs in classifying as “vertical” or “horizontal” segments that are either horizontally or vertically aligned. Differences between congruent (same orientation for the segments and their alignment) and incongruent (different orientations) conditions are not limited to presentations of the stimuli closely together the fovea, but also occur when the stimuli are presented far apart in peripheral vision on opposite sides of the fovea (Bennett & Banks, 1991; Berardi & Fiorentini, 1991; Fahle, 1986; McGraw & Whitaker, 1999; Rentzeperis, Alexander, Kiper, & van Leeuwen, 2015; Rovamo, Virsu, Laurinen, & Hyvarinen, 1982; Sasaki et al., 2006; Scobey & van Kan, 1991; Temme, Malcus, & Noell, 1985; Westheimer, 2003, 2005). This effect depends on the color of the stimuli (Rentzeperis et al., 2015), possibly as a result of organization of the color and orientation mappings in primary visual cortex (Alexander & van Leeuwen, 2010).

Dropping the requirement of a perceptually dominant configuration allows us to see the similarity between congruence-based and other Stroop effects, in which a range of cognitive processes, including non-visual ones, may result in a multifaceted representation, where the presence of one component may conflict with naming another. The similarity justifies that all these tasks go by the same name of Stroop effect. This is not to deny the differences: Resource allocation (Collin, Large, & McMullen, 2003; Egner, Delano, & Hirsch, 2007) and stimulus material (Lachmann & van Leeuwen, 2004; Pomerantz & Pristach, 1989; van Leeuwen & Bakker, 1995; van Leeuwen & Lachmann, 2004) may all affect the Stroop conflict. We may distinguish tasks in which the conflict arises between attributes of the same feature domain (i.e., flanker task, Navon task or shape classification task) or different domains (such as word meaning and ink-color—i.e., the classical Stroop task). In the latter, the conflict may arise only after the target and nontarget have been processed up to some point in independent streams (Aine & Harter, 1984a, 1984b; Cohen & Shoup, 2000; Treisman, 1991). The family of Stroop tasks will therefore involve of partially overlapping, and partially distinct mechanisms (see, for instance, a recent meta-analysis in Nee, Wager, & Jonides, 2007).

Garner-interference (Garner, 1974, 1988) is the result of irrelevant variation: It consists in the difference in responding to a feature of a configuration where another, irrelevant feature is varied and where it stays the same. In Boenke et al. (2009), figures of two components were used in two conditions: One in which both target and distractor both were varied independently between rectangular and triangular shapes; another in which the target varied while the nontarget shape was held constant (triangular or rectangular in different blocks).
Why would a variable nontarget part interfere more with a target than a fixed one? One possibility is that the fixed part is easier for selective attention to suppress. This would imply that the mechanism of Garner interference is similar to that of Stroop interference. Another possibility, more in the spirit of Garner, is that task gives rise to a visual working memory representation of all target-nontarget configurations (Berti, Geissler, Lachmann, & Mecklinger, 2000; Carmo et al., 2017; Lachmann & Geissler, 2002; Lachmann & van Leeuwen, 2005). Performing the task then consists of matching the current item with one of the configurations in memory. This process involves memory search, which will take longer and produce more errors the larger the number of items stored. The variable nontarget condition requires a greater number of configurations to be stored than the fixed nontarget condition. In this explanation, Garner interference still implies that the stimuli are processed as configurations. However, the mechanism differs from Stroop interference.

Stroop and Garner effects can be studied as independent factors (Pomerantz, 1983) in a design in which congruent or incongruent stimuli are presented in one condition in separate blocks and randomly intermingled in another. In RT studies Marks (2004) and Pomerantz and Pristach (1989) observed an interaction: Congruence-based Stroop effects were found only in the condition where the irrelevant feature varied, not within the blocks in which it was kept constant. A Garner effect, in contrast, was found in both congruent and incongruent conditions. Thus, the study provides an instance of a Garner effect without a corresponding Stroop effect. Van Leeuwen and Bakker (1995) and Patching and Quinlan (2002) found evidence of Stroop without a corresponding Garner effect. Combining these observations provides evidence for a double dissociation of Stroop and Garner effects. Despite arguments to the contrary (Orden, Pennington, & Stone, 2001), double dissociations in effects are often taken as evidence for process dissociations.

In Boenke et al. (2009), Stroop and Garner effects were studied in combination with evoked brain activity measurement. The evoked signals confirm the dissociation of Stroop and Garner effects. Stroop effect emerges earlier than Garner effect. Two correlates of the Stroop effect occurred in the brain signal, starting at 172 and 268 ms after stimulus onset. The first overlaps in time with crowding (Chicherov et al., 2014), the second occurred considerably later. The first correlate of the Garner effect started even later, appearing at 328 ms. The reason why Garner interference appears later in the ERP than Stroop interference is plausibly related to visual working memory search; the timing of the Garner effect in ERP corresponds to that of attention switching within working memory (Berti, 2015). This mechanism differs from the one proposed for Stroop interference, but nonetheless implies that the stimuli were represented as configurations.
8.4 Letters Versus Shapes

Unlike crowding, Stroop and Garner interference depend on categorical distinctions. In van Leeuwen and Lachmann (2004) an interesting difference in form congruence effects appeared between two categories of material: letters and non-letters. Non-letters were either simple, discrete geometrical shapes, pseudoletters, or letters upside-down (for illustration see examples in Figure 8.1). Non-letters invariably showed a congruence-based Stroop effect. They were classified faster if the target and its surrounding were form-congruent as compared to when they were form-incongruent. For letter targets, however, the opposite result was found: Letters were categorized faster when surrounded by an incongruent nontarget than when the nontarget was congruent—i.e., a negative congruence effect.

Negative congruence effects sometimes occur also in non-letter shapes (Bavelier, Deruelle, & Proksch, 2000; Briand, 1994; van Leeuwen & Bakker, 1995). These tasks have in common that the stimuli are complex and the target hard to detect among the distracting stimulus features. Bavelier et al. (2000) therefore suggested that, akin to crowding, attentional suppression of the surrounding nontarget information was needed to successfully perform the task. Congruence would then make it more difficult to suppress the nontarget information. Unlike in normal adult readers, negative congruence effects occur for both letters and non-letters in illiterates without letters being faster (Lachmann, Khera, Srinivasan, &

![Figure 8.1 Examples for the stimuli used in a series of experimental studies: Letter “A” (upper part) and corresponding pseudo-letter (lower part) surrounded by a form-congruent (left side) or a form-incongruent nontarget surrounding (right side) (adapted from Lachmann & van Leeuwen, 2004)
van Leeuwen, 2012). Given their general unfamiliarity with task and material, they probably found both conditions hard, as expressed by very long reaction times and standard deviations across conditions when compared with controls (see Figure 8.2).

This observation is, however, implausible for the negative congruence effects observed for letters in adult literates (see Figure 8.2, right side). Letters are processed faster and with greater accuracy than non-letters across a variety of tasks (Burgund, Guo, & Aurbach, 2009; Burgund, Schlaggar, & Petersen, 2006; Lachmann & van Leeuwen, 2007; Poirel, Pineau, & Mellet, 2008; van Leeuwen & Lachmann, 2004). A possible interpretation is that negative congruence effects emerge with letters because skilled readers prefer processing them in isolation from their surroundings (Lachmann & van Leeuwen, 2008). This is harder when the surrounding context is congruent, giving rise to spurious target-nontarget grouping by similarity. Processing letters in isolation guarantees a rapid grapheme-phoneme mapping, given the serial structure of acoustic/phonological working memory (D. M. Jones, Macken, & Nicholls, 2004) undisturbed by irrelevant visual information of the surrounding (Lachmann

![Figure 8.2](image)

**Figure 8.2** Mean RT (ms) with error bars (5% confidence interval) for a group of Indian illiterates (left side) and a control group of Indish skilled readers (right side) for letter and shape targets surrounded by either a or a form-congruent or a form-incongruent, task-irrelevant shape (as illustrated by two item samples on the left side: congruent letter and incongruent letter condition). Please note the different scaling of y-axes (data from Lachmann et al., 2012).
van Leeuwen, 2014). Early reading instruction (Ehri, 1998; Frith, 1985) requires the knowledge and use of individual graphemes and phonemes and their correspondences. Fast processing on the basis of these correspondences must be trained, which takes several years (Froyen, Bonte, van Atteveldt, & Blomert, 2009; Lachmann & van Leeuwen, 2008), after which children have a fully developed cross-model representation of letters and, accordingly, grapheme-phoneme associations, which are applied automatically in reading situations (Froyen et al., 2009). Although letter-by-letter reading is no longer dominant in skilled reading, it is still required for reading new or unfamiliar words (Lachmann & van Leeuwen, 2014).

In identifying letters, component features rather than configural features are important (consider O vs. Q or P vs R or b vs. d). In one version of their experiment, van Leeuwen and Lachmann (2004) varied the task in the following way: One condition had response alternatives based on feature content (Category 1 = “A” or “circle” versus Category 2 = “C” or “triangle”) the other had response alternatives based on global shape similarity (Category 1 = “A” or “triangle” vs. Category 2 = “C” or “circle”). As a result, the negative congruence effect for letters was maintained in the former but eliminated in the latter condition, where letters now showed the same congruence effects as non-letters. The result shows that the negative congruence effect depends on the prominence of component features in letter identification (see also Lachmann & van Leeuwen, 2004).

It is therefore arguable that letter reading specifically recruits an analytic perceptual strategy. Letters are preferably encoded in an analytic fashion whereas non-letters are preferably given a holistic perceptual representation (Lachmann & van Leeuwen, 2014). Since this distinction depends on our ability to read, an interesting question is what about children who have difficulties in learning to read (developmental dyslexia)?

Lachmann and van Leeuwen (2008) and Fernandes, Vale, Martins, Morais, and Kolinsky (2014) found that normally reading children of 8–10 years old show positive congruence effects for both letters and non-letters. A possible reason is that skills involved in phoneme processing may not have sufficiently developed for children to benefit from the analytic strategy in offering rapid grapheme-phoneme conversion (Froyen et al., 2009; Wetzel, Widmann, & Schroger, 2011).

Compared to their normal reading peers, children who have difficulties in learning to read were either anomalously slow on the task or showed anomalous congruence effects. The first subgroup shows in diagnostic reading tests to have impairments in reading frequent words. This subgroup shows congruence effects for letters as well as non-letters, just like normally reading children, but are much slower overall. We consider them to have generic difficulties with activation speed, in accordance with their generically low psychophysiological responses to both phonological and non-phonological stimulation (Lachmann, Berti, Kujala, & Schroger, 2005).
The other subgroup showed reading impairments on non-words. This subgroup exhibit negative congruence effects for letters, pointing to an analytic strategy. Unlike adults, this is probably because, like the illiterates, they experience the task as hard.

8.5 Numbers

Like letters, digits are visual objects with a specific symbolic value. For the latter, this is a numerical value. Visual properties of digits such as length, size, and luminance also are readily expressed on a numerical scale. This could give rise to Stroop interference effects. One such effect is known as the size-congruity effect: It is easier to respond to a congruent number (a small number printed in small size or to a large number printed in large size), than to an incongruent one (a small number printed in large size or vice versa).

The size congruity effect was first investigated by Banks and Flora (1977) and Paivio (1975); several studies since then have further explored the effect (Ansari, Fugelsang, Dhital, & Venkatraman, 2006; Besner & Coltheart, 1979; Henik & Tzelgov, 1982; Kaufmann et al., 2005; Pinel, Piazza, Le Bihan, & Dehaene, 2004; Ren, Nicholls, Ma, & Chen, 2011; Santens & Verguts, 2011; Schwarz & Heinze, 1998). Whereas Besner and Coltheart (1979) presented pairs of numbers and asked participants to choose the numerically larger (or smaller) stimulus, Henik and Tzelgov (1982) asked participants to choose the physically larger stimulus. Both tasks showed the same results: Congruent conditions are preferred over incongruent ones. The same effects occur when judging whether a single number is larger or smaller than a given standard (Schwarz & Heinze, 1998; Schwarz & Ischebeck, 2003).

ERP studies show an effect of the congruity on N2 potential over the frontal and central midline regions, suggesting that early attentional processing is involved in the effect (Schwarz & Heinze, 1998). Neuroimaging studies using fMRI observed that incongruent trials lead to higher neural activity in higher regions, such as the prefrontal as well as the anterior cingulate cortex (Ansari et al., 2006; Kaufmann et al., 2005).

Santens and Verguts (2011) modeled their behavioral data based on two alternative hypotheses: A shared representation account in which the congruence effect is based on a task-invariant representation of overlapping size and number dimensions (Kornblum, Hasbroucq, & Osman, 1990) and a shared decision account involving a common domain-general decision system. An example of such an account is the theory of magnitude (ATOM) by Walsh (2003), in which pace, time, and quantity are mapped into a generalized space for deciding their magnitude. Consistently with the fMRI data by Ansari et al. (2006) and Kaufmann et al. (2005) and the ERP data by Schwarz and Heinze (1998), Santens and
Verguts (2011) concluded that the latter account best explained their data.

The size-congruity effect as a Stroop interference in numbers is to be contrasted with effects that can arise because of stimulus-response compatibility (SRC). In choice-reaction tasks, responses to stimuli on the left-hand side with the left and to stimuli on the right-hand side with the right hand are generally faster and more accurate than responses to a stimulus on the left with the right hand and vice versa (SRC, Fitts & Deininger, 1954; Fitts & Seeger, 1953). When the hands are crossed, however, it is still easier when stimulus and the corresponding response button are positioned on the same side than when they are on opposite sides (e.g., Fitts & Deininger, 1954; Fitts & Seeger, 1953; Kornblum et al., 1990; Proctor & Cho, 2006). SRC effects are extremely widespread and occur in several modalities (e.g., auditory or visual: Nuerk, Wood, & Willmes, 2005, and independent of task: Hommel & Prinz, 1996; Kornblum & Lee, 1995, Simon, 1990).

Interestingly, Dehaene, Bossini, and Giraux (1993) showed that smaller numbers (e.g., 1, 2) were responded to faster with the left hand, whereas larger numbers (e.g., 8, 9) were responded to faster with the right hand, a phenomenon that became well known as the Spatial Numerical Association of Response Codes (SNARC) effect. SNARC effects occurred in explicit task conditions, where the numbers were judged as large or small, as well as in implicit task conditions, where numbers were judged for parity (even or uneven). The authors attributed these effects to an internal representation of numbers aligned horizontally and in ascending order from left to right the mental number line (Dehaene, Dupoux, & Mehler, 1990; Restle, 1970).

Whereas an effect of a mental number line would reveal, in our terms, a perceptual bias stemming from long-term memory, another possible explanation invokes working memory (van Dijck, Abrahamse, Acar, Ketels, & Fias, 2014). Compatible alignment of associated linguistic markers (Nuerk, Iversen, & Willmes, 2004), valuation (Casasanto, 2009), or polarity correspondence for stimulus and response dimensions (Proctor & Cho, 2006) reduce working memory load, assuming both stimulus and response dimensions overlap (Kornblum et al., 1990). A third alternative is to consider SNARC effects as a product of a common domain-general space for deciding quantities, as in Walsh’s (2003) ATOM theory. In this approach, spatially misaligned numbers interfere with the decision making.

To decide between these alternatives, let us consider an effect similar to SNARC. This effect is known as the Spatial Pitch (or Musical) Association of Response Codes (SPARC, Lidji, Kolinsky, Lochy, & Morais, 2007; SMARC, (Rusconi, Kwan, Giordano, Umilta, & Butterworth, 2006)). Mapping low pitches to the lower and high pitches to the upper ends of a continuum constitutes a compatible assignment, as opposed to vice versa.
SNARC and SPARC (in)compatibility can be varied as independent factors by singing numbers in a low or high pitch (Fischer, Riello, Giodano, & Rusconi, 2013; Weis et al., 2016). Weis et al. (2016) obtained both SNARC and SPARC effects, as well as a superadditive interaction between the two (Figure 8.3). Perceptual bias could, in principle, explain the SPARC main effect by postulating a mental pitch line, in analogy to the mental number line, but it would be unclear why the two would interact. The joint representations account would have predicted an interaction because a spatial dimension is used to encode both numerical and musical quantities (e.g., Beecham, Reeve, & Wilson, 2009; Krumhansl, 2000). However, this dimensional overlap (Kornblum et al., 1990) means that, in addition to the SNARC and SPARC effects, there will be an advantage if both stimulus dimensions are consistently aligned (either compatible or both incompatible) as opposed to inconsistently aligned (one compatible, the other incompatible). In other words, the interaction predicted by the shared representations account would have to be sub-additive. By contrast, a shared decisions account (e.g., Walsh, 2003) could predict a superadditive interaction if we generalize this account to accommodate pitch. If an item is simultaneously incompatible on several dimensions, mapping it into the decision space may increase the decision load. Thus, as with the size-congruity effect, the SNARC and SPARC interaction point to an explanation in terms of a shared decision space.

8.6 Concluding Observations

Perceptual organization can be fast, leading to the configural superiority effect. But fast integration can be a nuisance for both early (crowding) and late (Stroop-congruence) visuospatial target selection. Letters and numbers show interference profiles that differ from other visual
configurations, due to the specific symbolic load they carry. Their association with phonemes makes that letters are preferably processed analytically, giving rise to negative congruence effects; for numbers we observe congruence effects in a generic magnitude space instead of the visual space. Depending on task demands, letters can also show regular congruence effects. Plausibly, the same will apply to numbers. Depending on the stimulus, besides fast, perceptual organization can also be slow in the presence of distractors, or when a structure is hidden as in an Archimboldo painting. As in our IKEA furniture, we observe in the products of perceptual organization, once assembled, a certain amount of flexibility. For perception at least, this is to be valued. Perhaps, the power of perceptual organization resides, not in its speed, but in its versatility, making it a pervasive attribute of visual processing and experience.

Note

This effect nicely illustrates Hans Geissler’s famous principle of seeming redundancy and how this can enhance processing effectivity (Buffart & Geissler, 1984; Geissler, 1985; Geissler, Klix, & Scheidereiter, 1978; Geissler & Puffe, 1983; Lachmann & Geissler, 2002).

References


Dishon-Berkovits, M., & Algom, D. (2000). The Stroop effect: It is not the robust phenomenon that you have thought it to be. Memory & Cognition, 28(8), 1437–1449. doi: 10.3758/BF03211844


