Psychol Res (1993) 56: 35–43

Psychological Research
Psychologische Forschung

Space-based visual attention models and object selection: Constraints, problems, and possible solutions

W. X. Schneider

Ludwig-Maximilians University, Munich, Germany

Summary. One of the functions of visual attention is the selection of object information. This seems to be in line with an influential group of attentional models that assume that attentional selection is space based. These models assume that the selection of an object in vision is realized by selection of the location of that object. Whether this relatively simple idea of space-based attention and the corresponding, more elaborated space-based models are sufficient to handle selected constraints and problems of object selection is the main issue of this article. The first step toward an answer is to describe the common computational structure of space-based attentional models. Two model classes will be distinguished: capacity-limited models (e.g., Treisman, 1988; LaBerge & Brown, 1989) and models that do not assume a capacity limitation (e.g., Van der Heijden, 1992). Next, three kinds of task and data on object selection are introduced that are especially challenging for space-based models. The first type of data refers to experiments that require selection between overlapping objects. The second type of data concerns the influence of early perceptual grouping – a strong object-defining factor – on late response competition, and the third type consists of a selection task in which a high-level (semantic) attribute defines an object and controls selection. In all three cases, problems of space-based models are analyzed and possible solutions are sketched. Finally, a brief evaluative summary is given.

Introduction

What is selective attention? Frequently, an answer is that attentional processes can be regarded as brain processes that "select only part of the information momentarily available to the organism" (e.g., LaBerge, 1990, p. 358). Any model or theory about these selection operations requires – as a first step – a more specific description of the functions of selection – function in the sense of Marr (1982). At least three different functions can be found in the literature.

A first and widely accepted function of attentional processes is *selection-for-object-recognition* (e.g., Neisser, 1967).[1] This function is motivated by the assumption that object recognition, especially shape-based object recognition, is a computationally expensive task (e.g., Marr, 1982; Ullman, 1989; Poggio & Edelman, 1990; Hummel & Biederman, 1992). One of the main difficulties is due to the fact that even a simple object (e.g., a coffee cup) can generate very different retinal projections when the object varies in distance, locus, and orientation within the visual field. Nevertheless, it can easily be recognized. To make these complex computational processes tractable it is assumed that the brain has invented a capacity limitation – that is, only one object (or just a few objects) at a time can be recognized. If the visual field contains several objects, attentional selection is needed. It determines which of the many input patterns (of possible external objects) is delivered to the recognition system (e.g., Koch & Ullman, 1985; LaBerge & Brown, 1989; Olshausen, Anderson & Van Essen, 1992; Goebel, 1993).

The second functional ascription of attention is in a sense similar to the first one and may be called *selection-for-feature-integration.* Attention is assumed to solve the "binding problem" (e.g., Treisman & Gelade, 1980; Treisman, 1988). The brain computes different elementary visual features of external objects (e.g., colors, local oriented contrasts, movement directions, etc.) in different areas. To make sure that features that belong to one external object are correctly integrated into an internal "object-file" (e.g., Kahneman & Treisman, 1984; Treisman, 1988), attentional selection comes into play.[2] This mechanism permits the integration of features from one selected location in space that is assumed to correspond to an external object.

The third function of attentional processes is in a way more basic than the previous ones and may be termed

[1] This term follows Allport's (1987) "selection-for-action" phrase.

[2] Other integration mechanisms are also assumed, such as top-down knowledge about the featural structure of objects.

Correspondence to: W. X. Schneider, MPI für psychologische Forschung, Leopoldstrasse 24, D-80802 München, Germany

selection-for-action (Allport, 1987; Neumann, 1987). It implies as a selection problem the "specification of parameters" for action control (Neumann, 1987).[3] For instance, grasping a certain object (e.g., a glass of beer) presupposes that spatial coordinates of this object are used to compute the movement to the target object. If the visual field contains several objects suitable for the action (e.g., several glasses of beer) then only the spatial coordinates of the intended object (e.g., one's own glass) should be temporarily coupled to the motor control structure that is used for computing the grasping trajectory.

As has already been stated, all three postulated functions of attentional selection assume that selection is object based. No matter whether selection for action, for object recognition, or for feature integration is required, the content of the selection process is always information from an object of the external world. This requirement for object selection seems to be in line with a very prominent, large subgroup of current attentional models. These models that refer to the visual domain assume that selection in vision is *space based* (e.g., Treisman & Gelade, 1980; Crick, 1984; Koch & Ullman, 1985; LaBerge & Brown, 1989; Treisman, 1988; Olshausen et al., 1992; Van der Heijden, 1992; Goebel, 1993).[4] A suggestive abbreviation for the idea of space-based selection is caught by the metaphor that attention acts like a spotlight or zoom-lens in the visual field, leading to priorized processing inside the spot (e.g., Posner, 1980; Eriksen & St James, 1986). The popularity and the high level of acceptance of these visual attentional models is probably due to the large amount of experimental data from psychology (e.g., Eriksen & Hoffman, 1973; Posner, 1980; Hoffman & Nelson, 1981; Tsal & Lavie, 1988; cf. above all van der Heijden, 1992, this issue, for a convincing and focused argument on this issue) and from neurobiology (e.g., Bushnell, Goldberg & Robinson, 1981; Moran & Desimone, 1984; for overviews, see Robinson & Petersen, 1986; Posner & Petersen, 1990) which strongly support the claim of space-based selection.

But how does the requirement for object selection fit in to the concept of space-based selection? A first illustrative and intuitively plausible answer can be given by the spotlight metaphor. According to this metaphor, the selection of an object and its information for action control, object recognition, or feature integration is realized by the directing of the spotlight to the location in space that is occupied by the object. The main issue of this article will be whether this relatively simple idea of space-based attention and the corresponding, more elaborated models are sufficient to handle selected constraints and problems of object selection (see also Duncan, 1984). The first step toward an answer will be to describe the basic common computational structure of space-based attentional models. This will include the distinction between two classes of model, namely capacity-limited (e.g., Treisman, 1988; Koch & Ullman, 1985; LaBerge & Brown, 1989; Goebel, 1993) and non-capacity-limited models (e.g., van der Heijden, 1992), which give different computational accounts for space-based attentional selection. The next step will be to introduce experimental tasks and data on object selection that are especially challenging for space-based attentional models. The first type of experimental task requires the selection of information from overlapping objects (Duncan, 1984), while the second refers to a late response-competition experiment (Eriksen & Eriksen, 1974) in which the effects of early perceptual grouping – a strong object-defining factor – are investigated (Kramer & Jacobsen, 1991; Baylis & Driver, 1992). A third type of selection task relates to target objects that are defined and should be selected by a high-level attribute (e.g., the meaning of a word). The analysis of all three types of task and corresponding data will include not only the discussion of problems and constraints for space-based models, but also possible solutions and modifications.

Models of space-based visual attention: Common computational assumptions

Although different versions of space-based attention models to exist, there are four central computational assumptions that can be regarded as defining characteristics of this class of model.[5] These assumptions are meant to set the framework for the discussion.

First, space-based models postulate that the visual-selection operation is carried out at a representational level of the brain where information is still spatially structured in a topographical manner (e.g. Treisman, 1988; LaBerge & Brown, 1989; van der Heijden, 1992; Goebel, 1993).[6] This means that basic visual-feature information (e.g., local oriented contrast, color, etc.) is coded in terms of a two-dimensional retinotopic map. Thus, the representation of external objects in this map does not contain high-level information (e.g., recognized shapes), but consists of representations that make explicit information about elementary visual attributes of objects. Neurobiological data (e.g., Felleman & Van Essen, 1991) suggest that only early visual areas (e.g., V1 to V4) have corresponding topographically (retinotopically) structured neuron populations and thus are suitable sites for this spatial-attentional operation.

[3] Another selection-for-action problem concerns the output side and has been termed the problem of "effector recruitment" (Neumann, 1987), indicating a selection process "that regulates which skills are allowed to recruit which effector at a given moment in time" (Neumann, 1987, p. 376).

[4] Two recent, elaborated non-space-based attentional models in vision have been proposed by Duncan and Humphreys (1989) and Bundesen (1990) – for older models see, for instance, Broadbent (1958) and Posner (1978). Both models address the issue of object-based selection, but my discussion is limited to space-based models and their capabilities. Problems of these non-space-based models in explaining experimental data which indicate the location is special nature of attentional selection are discussed forcefully by Van der Heijden (this issue).

[5] The four computational assumptions are to a varying degree explicitly contained in different versions of space-based models, see, e.g., footnote 8. Furthermore, a few recently proposed models deliver specific and elaborated suggestions for neurocomputational implementation of these assumptions (e.g., LaBerge, 1990; Olshausen et al., 1992; Goebel, 1993).

[6] Treisman's feature integration theory (e.g., Treisman, 1988) does also assume that the early featural information is – at least implicitly – topographically structured. Otherwise, features could not be accessed selectively from the "master map of location."

Higher cortical areas are nontopographical (see Felleman & Van Essen, 1991, p. 6) – for instance, temporal-lobe areas, which are probably involved in object recognition (e.g., Mishkin, Ungerleider & Macko, 1983; Desimone, 1991). The early visual areas with topographical organization, in which the selection operations are carried out, should be called an *attentional working map*. The term *working map* should be neutral with respect to the question of how many submaps of features (e.g., local oriented contrast, color, etc.) are assumed.

Second, all of these space-based models share the assumption of an additional map that controls the selection operation. This attentional-control map contains "location information" (e.g. Treisman, 1988; LaBerge & Brown, 1989; van der Heijden, 1992) only, and is often called the "salience map" (e.g., Koch & Ullman, 1985; Goebel, 1993). It is a structure for the coordination of selection control and is sometimes equated with the subcortical thalamic nucleus pulvinar (e.g., LaBerge & Brown, 1989; but see Desimone et. al., 1990). PET studies (see LaBerge, 1990) have shown increased activation of this nucleus during the selection process. The salience map could be considered as a sort of blackboard for selection operations on which different brain areas can "write down their vote" for a location to be selected (corresponding to an object).

Third, space-based models assume that the selection of a certain object among other objects implies the activation[7] of the corresponding location representation in the salience map (e.g., LaBerge & Brown, 1989; Van der Heijden, 1992; Goebel, 1993). This activation is generated by other brain areas in which the *selection attribute* is computed.[8] Selection attribute refers to the information used for the intentional selection[9] of an object (see Van der Heijden, 1992). For instance, if a task requires the naming of red items, the color red is the selection attribute. So the corresponding cortical map for computing selection attributes may be called a *selection-attribute map* and has to have access to locations in the salience map.

Fourth, space-based models postulate that the attentional-selection operation leads to *position-based priorized processing* (for action control, object recognition, or feature integration). What this statement means can be explained by a description of the overall process of space-based attentional selection. First of all, the selection attribute is computed within one brain area (selection-attribute map). The corresponding locationally indexed activation is propagated to the salience map and subsequently to the working map. Within the working map only featural information of the activated location is processed in a priorized way to fulfill the suggested selection functions (for action control, object recognition, or feature integration). Thus, the selection operation itself can be equated with position-based priorized processing. What priorized means differs for capacity-limited and nonlimited selection models and will be explained in the next section.

However, before this is done the space-based selection concept should be illustrated by means of a simple selection task. For instance, subjects are instructed to name a red letter in a row of green letters. When the display contains a red H within green Xs, the selection attribute, color, is computed for all objects of the visual field in one brain area (e.g., V4). At the next step, only activation of the color representation of the intended object (selection attribute red) is propagated to the corresponding representation (of the location of the red letter) in the salience map – this is the instruction-mediated part of the selection process (e.g., Van der Heijden, 1992). The activated location in the salience map is then propagated to the working map where it causes the priorized processing of featural representations at that location. In higher response-related brain areas – which receive their activation from the working map – this leads – preceded by the computation of the correct name representation (letter H) – to the generation of the correct motor response (saying H).

Two classes of space-based attention models: Capacity-limited processing and unlimited processing

The assumption of a topographically organized salience and working map is shared by all space-based models mentioned so far. However, the models differ with respect to the question what the term *location-based priorized processing* means. Two classes of model can be distinguished: capacity-limited processing models and models with no presupposed capacity limitation (see also Van der Heijden, this issue).

The *capacity-limited models* (e.g., Koch & Ullman, 1985; Treisman, 1988; Olshausen et al., 1992; Goebel, 1993) postulate that only priorized information is further processed for the assumed functions (such as object recognition).[10] In the framework of the spotlight metaphor this further processing would relate to information within the spotlight. This makes sense if it is assumed that the postattentive processes of object recognition are capacity limited. To handle this limitation the brain uses an attentional mechanism that selects the spatially structured configuration of features that should be further analyzed.

The second class of space-based models (e.g., Van der Heijden, 1992) assumes that there are *no capacity limitations* at postattentive processing levels. Object recognition and other higher-level processes can work – within the acuity limits – for all objects of the visual field in parallel (Van der Heijden, 1992). Selection in the sense of position-based priorized processing means here that activation of selected (priorized) representation in an attentional working map is higher than the activation of non-priorized fea-

[7] The term activation should be understood in the usual connectionist sense (e.g., Rumelhart & McClelland, 1986; Phaf, Van der Heijden & Hudson, 1990). It might be viewed as the mean firing rate of neurons.

[8] Treisman's (1988) model leaves open the question how the spotlight-controlling "master map of location" itself is controlled (see also Van der Heijden's, 1992, p. 251, discussion of Treisman's model). Her additional "feature inhibition" mechanism (Treisman, 1988, p. 226) does not address this function.

[9] The case of involuntary, bottom-up attentional control by salient features (e.g., onset- or popout-stimuli) is not discussed here (see, e.g., Treisman, 1988; LaBerge & Brown, 1989, or Van der Heijden, 1992).

[10] The LaBerge and Brown (1989) model also considers the possibility of shape-based object recognition without attentional involvement.

ture information. This implies that both priorized and non-priorized information are available at postattentive levels, but with different activation levels (see Van der Heijden, 1981, 1992; Allport, 1987). In other words, this selection concept implies that no information at or beyond the level of the attentional working map is excluded – as in the capacity-limited models. All information is processed to the highest level (e. g., the motor level), but with a different degree of activation. The plausibility of this assumption of unlimited processing capacity will be discussed below.

The ultimate selection in the sense of the excluding of information – certainly required for selection-for-action (e.g., Allport, 1987) – is achieved at the "late late" output side of the brain (Phaf et al., 1990; Van der Heijden, 1992). It relies on competitive coupling between motor programs that represent the action alternatives and that cannot be initiated unless the activation level of one program is higher than the level of the others. If all motor programs receive similar sensory-based bottom-up activation, none of them can win the competition and therefore none can be elicited. But space-based attentional priorizing of the intended information in the working map increases the activation flow to the corresponding motor program. At the motor-program level, this increased activation leads to a state in which the intended program wins the competition and generates and open motor action. A connectionist implementation of this "postcategorical filtering" idea of response selection (van der Heijden, 1981) can be found in Phaf at al.'s (1990) SLAM model.[11]

To summarize how both classes of space-based attention models work, a reversion to the selection task described above should be made. Remember that the task was to name a red letter among green letters. The first step is equal for both types of model. The position representation of the intended object in the salience map is activated by a brain area in which the selection attribute (red) is computed. As a consequence, the early visual-featural information at the corresponding position in the working map is processed in a privileged way. Models with a postattentive capacity limitation (at the stage of object recognition or feature integration) may simply assume that only the red letter (H) is identified (or identified first) and then coupled to action-control structures of the motor system (pronouncing the H). Models that do not assume a capacity limitation postulate simultaneous processing of all letters up to the level of letter recognition and motor-program activation. Owing to the low-level attentional priorizing in the working map (at the location of the red letter) the corresponding motor program (pronouncing H) receives more activation than its competitors (motor programs for the other letters S) and elicits an open motor response.

Tasks and data on object selection: I. Overlapping objects

How are space-based attentional models – capacity-limited and unlimited versions – able to deal with object-selection tasks and data? A simple selection situation was described in the example above (pronouncing a red letter), but what about more intricate object attentional tasks? For instance, imagine a task that requires subjects to select between overlapping objects.

Duncan (1984) conducted experiments of this kind (see, e.g., also Rock & Gutman, 1981; Tipper, 1985). For overlapping objects he investigated whether the selective report of several object attributes is easier when both attributes belong to a common object than when they belong to different objects. Subjects were presented with two overlapping objects, such as a line inside a box. The first object, the box, had a certain size (small or large) and a gap at a certain location (on the left- or right-hand side of the box). The second object, the line, had a certain orientation (tilted to the left or to the right) and texture (dashed or dotted). The task was to report two attributes of the objects while the stimuli were shown for a short time and were pattern-masked. In the condition *same object* both attributes to be reported belonged to one object (e. g., size of the box and location of the gap) and in the condition *different attributes* one attribute of each object had to be reported (e. g., size of box and texture of line). The data showed that reporting performance is better for the condition of same objects than for that of different objects. In other words, it is easier to report attributes that belong to one object than those belonging to two objects, despite the fact that the spatial distance between the attributes is the same.

Three comments should be made on Duncan's task and data. The first two comments elaborated below refer to two possible extensions of space-based models in handling the selection between overlapping objects. The third comment is concerned with the question of whether and how these extensions can explain Duncan's (1984) specific data pattern.

Those versions of space-based attentional models that assume that there is one spotlight-like attentional-selection mechanism (and no further mechanisms) have a problem in explaining selection between overlapping objects. *Spotlight-like* means that selection in the sense of priorization is restricted to a spatially coherent region in the working map – for example, in the form of a circle (e.g., Posner, 1980; Eriksen & St-James, 1986). Two overlapping objects cover the same coherent region, so the spotlight covers both overlapping objects at the same time, and it is not possible to have one object within the spot while the other one is left outside.[12]

Two nonexclusive and maybe complementary attempts to solve the problem of selecting between overlapping objects on the basis of the models described earlier should be sketched. The first attempt simply assumes that there is a *second high-level non-space-based selection mechanism*

[11] SLAM (SeLective-Attention-Model) uses a non-space-based attentional mechanism where color, form, and position are equally potent selection attributes. Thus, SLAM cannot be considered a connectionist implementation of Van der Heijden's (1991) space-based model.

[12] Even if several spatial scales are allowed, this would only be of help to a certain subpopulation of overlapping objects that differ in spatial frequency – see, e. g., Humphrey and Bruce's (1989) explanation of Duncan's data. However, overlapping objects like line drawings in the negative-priming experiments (e. g., Allport, Tipper & Chmiel, 1985; Tipper, 1985) do not differ in spatial frequency.

(e.g., Yantis & Johnson, 1990; LaBerge, 1990; Goebel, 1993) which may select between representations of objects already recognized. Unfortunately, up to now it has not yet been specified in detail – at least to my knowledge – what such an additional mechanism might look like. This would, for instance, require the specification of the representational level of high-level object information and their corresponding selection mechanism – see Goebel (1993) for first steps in this direction.

The second attempt goes back to suggestions claiming that visual attention selects between perceptual objects (e.g., Neisser, 1967; Kahneman, 1973; Kahneman & Henik, 1981; Prinzmetal, 1981; Kahneman & Treisman, 1984). The term *perceptual object* can be explained by reference to representations of external objects in the working map that consist of topographically structured representations of elementary visual features (e.g., in areas V1 to V4). These representations might be called *perceptual chunks*. Several of the space-based models discussed above have incorporated this idea of "perceptual-chunk-based selection" (e.g., Treisman, 1988; Goebel, 1993; Van der Heijden, this issue). The selection between overlapping objects would be made by the priorized processing of one of the two perceptual chunks. One possible realization of this idea is to assume that information within the salience map corresponds exactly to perceptual chunks (e.g., Van der Heijden, this issue).

How can both ideas for a solution account for Duncan's (1984) data pattern – that is, the superior performance in reporting two attributes from one object compared to reporting one attribute from each of the overlapping objects? The first idea seems not yet specific enough to me, but the second one, which assumes spatially chunk-based selection, can do the job. According to this idea, the superior performance in the report of two attributes from one object should be due to the fact that only one perceptual chunk at a time (which corresponds to an object) can be selected (e.g., activated within the salience map; see Van der Heijden, this issue). Thus, reporting two attributes from one object can be done by the priorizing of information from one chunk in the working map. Reporting two attributes from two objects requires the serial selection of two chunks, which, under conditions of data limitation (pattern masking), leads to worse performance.

Tasks and data on object selection: II. Response competition and perceptual grouping

The second set of data that is challenging for space-based attentional conceptions is concerned with the question of how *early* perceptual grouping – a prominent factor in defining objects – influences *late*-located response competition (e.g., Kramer & Jacobsen, 1991; Baylis & Driver, 1992; see also Banks & Prinzmetal, 1976). Perceptual-grouping principles, sometimes called Gestalt laws (Wertheimer, 1923), specify the factors (such as similarity, proximity, common fate, etc.) that determine what information is grouped into a common perceptual object. Response competition refers to an experimental effect in the task of Eriksen and Eriksen (1974). Their task is a two-choice reaction-time task requiring a subject to react to the identity of a central target object in a linear row of other distractor objects, for instance, to press the left button when the target is an H, the right button when it is an S. The typical result is that distractors that would require the opposite reaction (incompatible distractors, e.g., H S H as a typical configuration) slow down the reaction time to the target considerably as compared to neutral distractors (no response association, e.g., D S D) or compatible distractors (e.g., S S S). This reaction-time difference between incompatible and neutral or compatible condition is sometimes called the Eriksen interference and is regarded as a result of *response competition.*

The existence of *late* response-competition effects is especially challenging for capacity-limited space-based models because they assume an *early* site of attentional selection. Response-competition effects presuppose the computation of target and distractor-letter identities that are located at a processing level beyond space-based selection processes. Furthermore, there is evidence that distractors are processed up to a *late late* output side level, – that is, lead to a "preliminary response activation" at the motor level (e.g., Gratton, et al., 1988). Imperfect early space-based selection has therefore to be assumed. Perfect selection would have prevented the computation of distractor identity.[13] In other words, the postattentive locus of response competition implies that target as well as distractor information were selected by the capacity-limited space-based attentional system for further analysis. Thus, the resolution of interference, the response competition, has to rely on a second *later* high-level selection mechanism whose existence was already claimed in connection with Duncan's data.

Non-capacity-limited models of selection can easily explain the response-competition effects. Van der Heijden's earlier (1981) concept of "postcategorical filtering" is especially suited for this task. It assumes unlimited parallel processing of target and distractors from the retina up to the level of motor-program activation and competition for response initiation. For the condition incompatible distractors this implies competition between conflicting motor programs that are related to the corresponding target and distractor identities. Neutral or compatible distractors do not activate the wrong motor program and thus no response competition can follow. Space-based attentional selection-for-action within the condition of incompatible distractors is realized at the level of the early working map. It causes a higher information flow from the target location compared to that from the distractor locations so that the corresponding target-motor program wins the competition in the end. To achieve this state, the target-motor program has to show higher activation than the distractor-motor programs for a certain amount of time (see SLAM, Phaf et al., 1990). This principle involved in winning the competition can be re-

[13] For a insightful discussion of selection and response competition, see Eriksen, Pan, and Botella (this issue), and of the connected *early-late-selection* controversy, see Allport (1989).

garded as a second selection mechanism of models with unlimited processing capacity.

Now, the more challenging aspect of the second object-selection task can be discussed – that is, the influence of early grouping factors – which define perceptual objects – on late response competition. Studies of Kramer and Jacobson (1991) and of Baylis and Driver (1992) will be reported briefly.

Kramer and Jacobson (1991) manipulated the degree of *perceptual grouping* between target and distractors and measured the effect on response competition. The experimental condition, same object, consisted of a target and distractors that had the same color and were integrated into one shape. The condition of different objects consisted of a target and distractors of diffeent colors, and there was no common shape. The distance between target and distractors was constant in both conditions so that a simple space-based model would not predict any differences in selection efficiency between them. Yet, the results show that the perceptual grouping of target and distractors into a joint object (by color and shape) within the condition of same object reduced selection efficiency and increased the response competition as compared to the condition of different objects (which still generates interference). Again, this suggests that selection performance depends decisively on perceptual grouping, despite equal physical distance between target and distractors.

Baylis and Driver (1992) extended these results by showing that further Gestalt factors – such as good continuation in the form of a common row of target and distractors – also influence the degree of response competition and, hence, selection performance.

To explain these perceptual-grouping effects on response competition, the models of nonlimited capacity can rely on a modification already described in the analysis of Duncan's (1984) data, which postulated perceptual chunks within the salience map (see also Van der Heijden, this issue). Gestalt manipulation of common color, shape, or common row between target and distractors should generate a corresponding chunk in the salience map that reflects the perceptual-grouping operation. If target and distractor share a common color and form (e. g., Kramer & Jacobson, 1991; Baylis & Driver, 1992) their representations should be grouped into one common chunk. But the selection of the target requires restriction of the priorized processing to the subpart of the chunk that corresponds to the target only. This presupposes that the common chunk of target and distractors has to be segregated into two chunks representing both elements. After this segregation operation the target chunk can win the competition in the salience map against distractor chunks and then control the activation flow from the working map.

For the results of the perceptual-grouping manipulation on response competition, the following explanation can be sketched. The condition of same object in Kramer and Jacobsen's (1991) experiment should lead to the computation of a common chunk in the salience map that includes target and distractor information. Thus, selection of the target requires the additional step of chunk segregation. This step is not needed for the condition of different objects, in which grouping is not expected. Therefore, the reaction time for the condition of same object is increased as compared to the reaction time for the condition of different objects.[14]

How can capacity-limited models of space-based attention explain the effects of perceptual grouping on response competition? If differences in response competition reflect only the working of second high-level selection mechanism (see the explanation of response competition), then there is a problem. Perceptual-grouping effects are spatially based and should therefore be located at an elementary level of visual information processing – between V1 and V4 (see Felleman & Van Essen, 1991) which precedes the cortical site of the second high-level working map. This map is probably somewhere in the temporal areas as part of the "what" pathway (e. g., see Mishkin et al., 1983; Desimone & Ungerleider, 1989; Felleman & Van Essen, 1991) where the topographical representation – a prerequisite of spatially mediated perceptual grouping – is already lost (see Felleman & Van Essen, 1991).

Goebel's neural-network model (1993), which locates the grouping effects between the level of the spatially organized first working map and that of the second high-level non-spatially organized one, may offer a more plausible attempt to tackle this problem. According to his model, the output of first space-based selection consists of a representation that is still spatially structured (see also Olshausen et al. 1992). The subsequent object-recognition process and the second attentional mechanism works – by means of interactive processing – on the basis of this grouped representation. To explain the basic data of Kramer and Jacobson (1991) and of Baylis and Driver (1992), it would be necessary to specify how the second attentional mechanism depends on object recognition and its relationship to perceptual grouping, and how it resolves response competition. Goebel's (1993) model which was constructed primarily to explain other tasks and data is not specific enough yet to answer these questions, but it might be a promising starting point.

All in all, we are left with the impression that data concerning Gestalt manipulations in response-competition experiments (e. g., Kramer & Jacobson, 1991; Baylis & Driver, 1992) can be handled by non-capacity-limited models, but are still a challenge for current limited-capacity models. Further elaborations on how object recognition, perceptual grouping, and the second high-level selection mechanism work together are therefore required.

Tasks and data on object selection: III. Attentional control by high-level attributes

The last selection problem to be analyzed consists of a relatively simple selection task in which a *high-level object attribute* (meaning) is used to control the selection of low-

[14] Interference is still present in the experimental condition different objects. In such non-capacity-limited models this is due to the fact that incompatible distractors and their corresponding segregated chunk-based information in the working map are nevertheless processed up to the level of activating motor programs, influencing the competition at this level (see SLAM, Phaf et al., 1990).

level-based motor responses. Imagine that a subject is presented with a list of words printed in different colors and well within the acuity limits. She or he is instructed to name the color of those words that are animal names – the other words are names of nonliving things. The intricate aspect of this task is that high-level information (the category of animal names), which constitutes the target object, is used to control the low-level-based response to that object (color naming).

The accomplishment of this task can be explained within the framework of capacity-limited models of space-based selection by reference to the second high-level categorical-based selection mechanism. This mechanism might work as follows: the first space-based mechanism may begin to select a certain chunk that corresponds to a potential word (e. g., the utmost left one); shape information from word after word is fed by space-based attention into the object recognition system, and temporary object-files are created (Kahneman & Treisman, 1984). If a word is recognized as an animal name, then the corresponding color information is used for specifying the motor program.

How can non-capacity-limited models deal with such a selection task? All words should be recognized in parallel (simultaneously), that is, their high-level category representations should be computed. If an animal name is recognized, the corresponding low-level location (object chunk) of that word in the salience map is activated. This causes, within the working map, priorized processing of the corresponding feature information – inter alia the color information – that occupies the same location. The priorized processing of the target color helps to win the competition of the motor program for naming the color.

These computational steps sound plausible, but they contain an assumption whose realization cannot easily be reconciled with conventional conceptions about the cortical functioning of a primate brain. The assumption is that high-level categorical representations within the "what pathway," which are probably not spatially (topographically) coded (see Desimone & Ungerleider, 1989; Felleman & Van Essen, 1991), have direct access to a purely spatially organized salience map. Even if it is supposed that the salience map is not located within the thalamus, but consists of a high-level area in the "where pathway" (e. g., in the parietal cortex; e. g., van der Heijden, 1992), the question still remains as to how the high-level area for computing animal names within the what pathway contacts the high-level salience map within the where pathway.

A possible solution to this problem is to postulate that selection is not mediated by the pathway of selection attribute (e. g., animal-name category map) – salience map – working map (e. g., Van der Heijden, 1992), but by a pathway of selection attribute – working map – salience map – working map instead. This might be feasible by the use of those feedback neural connections that correspond to the connections of the feedforward pathway (from the lower-level working map to the higher-level selection attribute *name* map).

This high-level object-selection task makes explicit a rather radical claim of all space-based models without capacity limitations – that is, the assumption of parallel processing of visual information within the acuity limits up to the highest processing level (e. g., the motor level). In other words, these models assume parallel recognition of *all objects* in the functional visual field. For the high-level selection task this would imply that the meanings of all words that are within the acuity limits are computed in parallel. However, this assumption of unlimited processing is at variance with detailed neurocomputational models of visual-based (shape-based) object recognition (e. g., Hummel & Biederman, 1992). Recognition processes in the human brain require a complex exchange of information between highly processed representations – such as part-whole relationships of identified subobjects in different brain areas and pathways. This kind of expensive computation makes a restriction to one (or a few) object(s) at a time very plausible (see Hummel & Biederman, 1992). Furthermore, and most important, there are – at least to my knowledge – no experimental data to support this claim of unlimited spatial and temporal parallel processing of all visual-based information within the acuity limits up to the highest level.

In my view, the assumption of unlimited processing capacity is not central to the important idea of attentional selection in the sense of feedback-based postcategorical filtering (Van der Heijden, 1981) and its later space-based version (Van der Heijden, 1992). This assumption of unlimited capacity could be replaced by the more plausible standard view (e. g., Neisser, 1967; Treisman & Gelade, 1980), which claims that the parallel-processing capabilities are restricted up to a certain level – for instance, the level of simple categorical representations (e. g., categorical-color representation or simple shape primitives). The recognition of real-world objects, which requires the computation of more complex categorical representations, could still be a capacity-limited postattentive process whose input is controlled by a space-based postcategorical (in the sense of simple categories) attentional mechanism.[15]

Conclusion

Finally, a brief evaluative summary is given of our analysis of the capabilities of space-based attentional models to handle the object selection constraints and problems.

First, the problem of selection between overlapping objects and the explanation of Duncan's (1984) data patterns are solved by both classes of model if adequate extensions or modifications are introduced. The first applies to both classes of model and assumes chunk-based visual selection. The second modification is required only for capacity-limited models and implies the assumption of a second high-level selection mechanism.

Second, data on the influence of early perceptual grouping – a low-level object-defining factor – on late response competition can be handled without further modifications

[15] This idea can be viewed as a combination of Van der Heijden's model (1992) and the capacity-limited "guided search model" (Cave & Wolfe, 1990).

by non-capacity-limited models of space-based selection. Current capacity-limited models are not yet specific enough for this purpose, but future attempts may be able handle this problem – that is, they may offer computational assumptions about the relationship between perceptual grouping, object recognition, and the second high-level attentional mechanism.

Third, selection tasks in which high-level attributes constitute an object and control attentional selection can be explained by the capacity-limited models already modified (second selection mechanism), but they may require a revision of the salience-map-access concept for models with no capacity limits. Furthermore, their assumption of totally unlimited capacity will probably have to be revised.

Fourth, further steps in the direction of the handling of object-selection problems could be taken by a type of a space-based model in which elements of both classes of models are mixed. This model type might combine the idea of early attentional selection in order to deal with capacity-limited high-level brain processes (e. g., limited object-recognition capabilities; Treisman, 1988; LaBerge & Brown, 1989; Goebel, 1993) with the idea of feedback-based post-categorical attentional control (Van der Heijden, 1981, 1992). Moreover, the inclusion of important insights from non-space-based models – for instance, Duncan and Humphreys's (1989) similarity principles for visual search – might also be a productive step forward. These or similar modeling attempts that try to satisfy the constraint of object selection, on the one hand, and the constraint that selection in vision is "special-about-position" (Van der Heijden, this issue), on the other hand, clearly sacrifice the ideal of simple psychological theories that do not care about task and data analysis in terms of detailed neurophysiological and computational considerations. However, even if we think that these types of complex model I have discussed here are the royal road, we should also keep in mind that "while we continue the debate within the old framework, we should remain alert to the possibility that it could soon become obsolete" (Kahneman & Treisman, 1984, p. 57).

Acknowledgements. I wish to thank Heiner Deubel, Alan Allport, Thomas Stoffer, and Lex van der Heijden for insightful comments, and Heidi John and the copy editor for helpful advice in improving the language.

References

Allport, D. A. (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 395–419). Hillsdale, NJ: Lawrence Erlbaum.

Allport, D. A. (1989). Visual attention. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 631–682). Cambridge, MA: The MIT Press.

Allport, D. A., Tipper, S. P., & Chmiel, N. R. (1985). Perceptual integration and postcategorical filtering. In M. I. Posner & O. S. Marin (Eds.), *Attention and performance XI* (pp. 107–132). Hillsdale, NJ: Lawrence Erlbaum.

Banks, W. P., & Prinzmetal, W. (1976). Configurational effects in visual information processing. *Perception & Psychophysics, 19*, 361–367.

Baylis, G., & Driver, J. (1992). Visual parsing and response competition: The effect of grouping factors. *Perception & Psychophysics, 51*, 145–162.

Broadbent, D. E. (1958). *Perception and communication.* New York: Pergamon Press.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97*, 523–547.

Bushnell, M. C., Goldberg, M. E., & Robinson, D. L. (1981). Behavioral enhancement of visual responses in monkey cerebral cortex. I. Modulation in posterior parietal cortex related to selective visual attention. *Journal of Neurophysiology, 46*, 755–772.

Cave, K. R., & Wolfe, J. M. (1990). Modelling the role of parallel processing in visual search. *Cognitive Psychology, 22*, 225–271.

Crick, F. (1984). Functions of the thalamic reticular complex: The searchlight hypothesis. *Proceedings of the National Academy of Sciences, 81*, 4586–4590.

Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *Journal of Cognitive Neuroscience, 3*, 1–8.

Desimone, R., & Ungerleider, L. G. (1989). Neural mechanisms of visual processing in monkeys. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology (Vol. 2, pp. 267–299). New York, NY: Elsevier Science Publishers (Biomedical Division).*

Desimone, R., Wessinger, M., Thomas, L., & Schneider, W. (1990). Attentional control of visual perception: Cortical and subcortical mechanisms. *Cold Spring Harbor Symposia on Quantitative Biology, 55*, 963–971.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113*, 501–517.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*, 433–458.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*, 143–149.

Eriksen, C. W., & Hoffman, J. E. (1973). The extent of processing of noise elements during selective encoding from visual displays. *Perception & Psychophysics, 14*, 155–160.

Eriksen, C. W., & St James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics, 40*, 225–240.

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

Goebel, R. (1993). Perceiving complex visual scenes: An oscillatory neural network model that integrates location-based attention, perceptual organisation, and invariant recognition. In C. L. Giles, S. J. Hanson, & J. D., Cowan (Eds.), *Advances in neural information processing systems 5.* San Mateo, CA: Morgan Kaufman Publishers.

Gratton, G., Coles, M. G. H., Sirevaag, E. J., Eriksen, C. W., & Donchin, E. (1988). Pre- and poststimulus activation of response channels: A psychophysiological analysis. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 331–344.

Hoffman, J. E., & Nelson, B. (1981). Spatial selectivity in visual search. *Perception & Psychophysics, 30*, 283–290.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99*, 480–517.

Humphreys, G. W., & Bruce, V. (1989). *Visual cognition.* Hillsdale, NJ: Lawrence Erlbaum.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

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.

Kahneman, D., & Treisman, A. M. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 28–61). New York: Academic Press.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4*, 219–227.

Kramer, A. F., & Jacobson, A. (1991). Perceptual organization and focused attention: The role of objects and proximity in visual processing. *Perception & Psychophysics, 50*, 267–284.

LaBerge, D. (1990). Thalamic and cortical mechanisms of attention suggested by recent positron emission tomographic experiments. *Journal of Cognitive Neuroscience, 2*, 358–372.

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review, 96*, 101–124.

Marr, D. (1982). *Vision*. New York: W. H. Freeman.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences, 6*, 414–417.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*, 782–784.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crafts.

Neumann, O. (1987). Beyond capacity: A functional view of attention. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 361–394). Hillsdale, NJ: Lawrence Erlbaum.

Olshausen, B., Anderson, C., & Van Essen, D. (1992). *A neural model of visual attention and invariant pattern recognition* (CNS Memo 18). Pasadena, CA: California Institute of Technology.

Phaf, R. H., Van der Heijden, A. H., & Hudson, P. T. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology, 22*, 273–341.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature, 343*, 263–266.

Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Lawrence Erlbaum.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3–25.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*, 25–42.

Prinzmetal, W. (1981). Principles of feature integration in visual perception. *Perception & Psychophysics, 30*, 330–340.

Robinson, D. L., & Petersen, S. E. (1986). The neurobiology of attention. In J. E. LeDoux & W. Hirst (Eds.), *Mind and brain*. Cambridge. Cambridge University Press.

Rock, I., & Gutman, D. (1981). The effect of inattention on form perception. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 275–285.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing*. Cambridge, MA: MIT Press.

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology, 37*, 571–590.

Treisman, A. M. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology, 40*, 201–237.

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

Tsal, Y., & Lavie, N. (1988). Attending to color and shape: The special role of location in selective visual processing. *Perception & Psychophysics, 44*, 15–21.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition, 32*, 193–254.

Van der Heijden, A. H. C. (1981). *Short-term visual information forgetting*. London: Routledge & Kegan Paul.

Van der Heijden, A. H. C. (1992). *Selective attention in vision*. New York: Routledge & Kegan Paul.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt, II. *Psychologische Forschung, 4*, 301–350.

Yantis, S., & Johnson, J. C. (1990). On the locus of visual selection: Evidence from focused attention tasks. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 135–149.