Chapter 1

Introduction
INTRODUCTION

Our visual environment is rich. When we scan the world around us we are bombarded with an incredible amount of visual information. It has been estimated that our retina is exposed to a mind blowing $10^{10}$ bits per second (C. H. Anderson, Van Essen, & Olshausen, 2005; Nørretranders, 1998). Realizing this, it seems remarkable that we still function as well as we do and that we still make sense of the world around us. The key to our success is that we are very well equipped to filter the information we receive. In split seconds, we disregard most of the information and magnify what remains. The mechanisms that allow for the selection of information are generally known as mechanisms of attention. Attention is aimed to select only those signals that intrinsically yield information potentially relevant to us, signals that have proven to be relevant in the past, or signals that correspond to our current goals. Attention helps us cope with the physical capacity limitations of our brain (James, 1890; Marois & Ivanoff, 2005). Even if we were capable of processing all the information that reaches our senses, we would still need attentional selection to enable behavior that is oriented toward our needs rather than behavior that is simply random.

One of the most explicit means to select visual information is by overt selection. Overt selection is the act of prioritizing an object or location over others by moving the eyes to that location in space (Posner, 1980). The opposite, covert selection, is the allocation of one’s focus of attention without an accessory shift of gaze (Helmholtz, 1867; Posner, 1980; Wright & Ward, 2008). It has been suggested that the two involve very similar, or even overlapping, neural mechanisms (e.g., Corbetta & Shulman, 2002; Rizzolatti, Riggio, Dascola, & Umilta, 1987). However, what makes overt selection distinct and prominent is that the direction of gaze determines where information is projected on the retina. When information falls at the center of the retina, it is captured by a region called the fovea. This region yields a relatively high density of photoreceptors (approximately 35,000 in an area of 0.3 mm in diameter (Dowling, 2007)), facilitating an enriched detail with which information can be processed. Hence, overt selection allows for optimal detail in the processing of visual information. Another distinct characteristic of overt selection is that it involves an ‘all or nothing’ process. At any given time the eyes can only be directed to one particular location in space. Therefore, in order to reallocate gaze the visual system is compelled to prioritize one
particular piece of information over others. Movements of the eyes thus explicitly reveal what kind of information is prioritized by observers.

This thesis discusses the main factors that drive overt selection. Intuitively, we might think that we have full control over what we choose to select, but in reality there are many factors that influence selection behavior in an involuntary manner. William James (1890) already recognized this and addressed a distinction between passive and active forms of selection. Nowadays many cognitive scientists tend to categorize the factors that drive selection in terms of the bottom-up versus top-down dichotomy. Bottom-up are those factors that determine selection exogenously, in a stimulus-driven manner. Visual information intrinsically consists of characteristics that can be unique or distinct in a visual scene, i.e., features that are less common elsewhere in the visual field. This typically determines whether information is salient or not. If one region or object in a visual scene is more salient than another, it is more likely to be selected, covertly or overtly, regardless of the intentions of the observer. Top-down factors, on the contrary, affect selection endogenously, in a goal-driven manner. Information that is similar or related to the goals of the observer is prioritized and more likely to draw attention and gaze. In recent days, some scientists came to realize that not all factors that influence selection can be classified as either bottom-up or top-down. Therefore a third category has been identified: selection history (Awh, Belopolsky, & Theeuwes, 2012; Belopolsky, 2015). This class of factors relates to past experience and past selection behavior of the observer. For instance, when a specific visual property has recently been selected, this by itself facilitates selection of that property the next time it is encountered; a phenomenon known as feature-based priming (Kristjansson & Campana, 2010; Maljkovic & Nakayama, 1994). A similar facilitation to select specific information has been observed when particular visual properties turn out to be predictive for obtaining rewards (Chelazzi, Perlato, Santandrea, & Della Libera, 2013). In this thesis, I will elaborate on selection history, as well as stimulus- and goal-driven influences on overt selection and introduce a theory in which these are explained and united in a single framework.
1.2. Basic mechanisms behind overt selection

1.2.1. The oculomotor system

The system that enables movements of the eyes, and thus enables overt selection, is known as the oculomotor system. At the lowest level, shifts of gaze are caused by six muscles at the outside of the eye balls (extraocular muscles). They work in pairs pulling the eyes either horizontally, vertically, or obliquely. Together these muscles enable several types of eye movements. The eyes for instance can follow moving visual objects in smooth pursuit or the position of the eyes can adjust for movements of the body or head. However, with regard to the aims of the current thesis, the most relevant type of eye movement is the saccadic eye movement. Saccadic eye movements are rapid and abrupt shifts of gaze. They typically occur approximately three times a second and are designed to quickly reallocate the fovea from one location to the next. There are two components of saccadic eye movements that the oculomotor system needs to control: amplitude and direction. Generally, the amplitude of a saccade is determined by the firing rates of motor neurons, whereas direction is determined by which of the muscles are activated (Sparks, 2002; Van Gisbergen, Robinson, & Gielen, 1981). The most basic form of control over eye movement behavior is provided by the brainstem. The brainstem, or more specifically the subregions known as the pons and the midbrain directly connect with the three cranial nerves that control the extraocular muscles. A major component of the midbrain is a (paired) structure known as the superior colliculus (SC). The SC contains a retinotopic representation of the visual field (Marino, Rodgers, Levy, & Munoz, 2008; Robinson, 1972) and is assumed to be responsible for inducing a saccade when it receives sufficient stimulation at a particular location (Gandhi & Katnani, 2011; Robinson, 1972). It presumably does so by exciting the pontine reticular formation (PPRF) and the medial longitudinal fasciculus (riMLF), structures which respectively encode for horizontal and vertical components of eye movements. However, to actually trigger a saccade, the SC also needs to inhibit the omnipause neurons (OPN). The OPN continuously down regulate activity in PPRF and riMLF, so only when the OPN themselves are inhibited, the trigger is pulled to execute the actual saccadic movement (Ramat, Leigh, Zee, & Optican, 2007; Sparks, 2002). Importantly, the SC receives its input from a variety of sources as it connects with striate, extrastriate and frontal brain areas (Munoz
CHAPTER 1

& Schall, 2004; Schlag-Rey, Schlag, & Dassonville, 1992; Sommer & Wurtz, 2000). This implies that both exogenous as well as endogenous signals may have access to the SC. The SC has therefore been taken as a core structure in a number of models of oculomotor control (Arai, Keller, & Edelman, 1994; Godijn & Theeuwes, 2002; Meeter, Van der Stigchel, & Theeuwes, 2010; Trappenberg, Dorris, Munoz, & Klein, 2001). Typically these models incorporate two important assumptions about the SC. First, the SC consists of a retinotopic map of the visual field, often referred to as the ‘saccade map’, in which potential saccadic target locations are reflected in terms of enhanced activity levels. Second, the SC is the structure where exogenous and endogenous signals converge. An enhanced peak of activity could thus be elicited exogenously by the presence of visual information in the visual field, but also endogenously by the intentions of the observer. However, exactly how exogenous and endogenous influences arise, evolve, integrate, and shape oculomotor behavior is still largely unknown.

1.2.2. Oculomotor selection and the global effect

Since the oculomotor system can only perform saccades toward one location at a time, saccade programming is commonly assumed to be a process of competition (Arai et al., 1994; Godijn & Theeuwes, 2002; Meeter et al., 2010; Trappenberg et al., 2001). If there is more than one possible target location, more than one enhanced peak of activity in the saccade map of the SC, a competition emerges in the oculomotor system. In order to resolve the competition the activity levels reflecting the objects are presumably suppressed. Eventually, the competition is settled when a threshold activity level is reached at a particular location. This is when the oculomotor system has selected a single location and generates a saccadic movement; a process also known as oculomotor selection.

An important consequence of the mutual suppressive nature of the competitive interactions of activity peaks in the SC is that the threshold activity level may be harder to reach depending on how fierce the competition is. This idea has been supported by findings of Walker, Deubel, Schneider, and Findlay (1997). In their search task, participants were asked to saccade toward a predefined and simple target object. In some of the trials this target was
accompanying a distractor object, either in the same direction of the target or in the opposite direction. The results showed that the presence of the distractor generally led to a slower oculomotor response, i.e., relative to a response generated toward a target presented in isolation, or when the distractor was presented in the opposite direction of the target. Walker et al. (1997) took these results as support for the inhibitory processes in the SC, operating across large distances in the visual field. In the same study, Walker et al. (1997) demonstrated that when the target and distractor were presented in the same direction and relatively close together, the presence of the distractor can facilitate rather than delay a response. An interpretation of this finding is that visual objects in close proximity induce activity peaks in the SC that overlap. Presumably where an overlap occurs, the peaks of activity complement each other rather than inhibit. As a result, the threshold activity level to trigger an oculomotor response is reached sooner at the location of the overlap. This generally leads to a faster oculomotor response that is aimed at a location in between the objects. Walker et al. (1997) indeed found significant deviations in the landing positions of eye movements whenever the distractor was presented relatively near the target (within a 20˚ radius). Eye movements were directed at locations intermediate to the target and distractor object. This phenomenon is known as the global effect (Coren & Hoenig, 1972; Findlay, 1982; Van der Stigchel & Nijboer, 2011).

Oculomotor selection and the global effect are phenomena that can be used to make inferences about the factors that drive overt selection. Eye movements can be seen as the end product of a process in which stimulus- and goal-driven factors play a role. Stimulus qualities, for instance, may have a determining influence on the activity levels in the SC. If one stimulus is more salient than another, it is likely to induce a more pronounced peak of activity in SC, thereby granting itself a competitive advantage and an enhanced probability to be selected by the oculomotor system. Even when stimuli are presented in close proximity, and thereby elicit a global effect, the exact direction of a saccade may still contain a bias toward one stimulus in particular. Such a bias in saccade direction provides information about the role of stimulus characteristics involved in overt selection. Although salient stimuli might have an initial competitive advantage over stimuli that are not salient, as time progresses, the activity peaks in the SC are presumably more and more dictated by the goals of the observer. Therefore, the latency of the oculomotor
response provides vital information as well. Even for the global effect, which typically involves responses that are relatively fast, one should never take for granted the stimulus-driven nature of the response, as we will see in this thesis. Nevertheless, oculomotor selection and the global effect phenomenon in this thesis will serve as the key tools to gain more insight into the mechanisms behind overt selection.

1.3. Control mechanisms of overt selection

1.3.1. Stimulus-driven selection

When a stimulus is distinct from its surroundings, such as a colorful painting on a blank wall, it tends to attract gaze. Selection behavior can be driven by stimulus salience, irrespective of the goals and intentions of an observer (Caputo & Guerra, 1998; Eimer, Kiss, Press, & Sauter, 2009; Kim & Cave, 1999; Theeuwes, 1991, 1992, 1994; Theeuwes, Kramer, Hahn, & Irwin, 1998; Theeuwes, Kramer, Hahn, Irwin, & Zelinsky, 1999; Todd & Kramer, 1994; Yantis, 1993, 1996). Major evidence for the idea that attentional selection can be stimulus-driven was provided by Theeuwes (1992). Theeuwes employed a search paradigm in which participants were presented with an array of search items presented at an equal distance to a central fixation point. Participants were asked to search for a diamond among circles, and manually report the orientation of a line segment presented within the diamond. The array of shapes was occasionally accompanied by a salient distractor item that had a distinct but task-irrelevant color. Despite the fact that the distractor was physically very unlike the target, the presence of the distractor evoked a delay in participant response times. This result suggests that attention was drawn to the distractor before attending the target. In other words, a distractor potentially leads to an involuntary capture of attention when it is relatively salient. Selection behavior can be purely driven by salience, by the physical characteristics of stimuli, rather than the goals of the observer.

The processes that underlie salience-driven effects on selection are often conceptualized in terms of activity levels in 'salience' maps (Itti & Koch, 2000, 2001; Itti, Koch, & Niebur, 1998; Koch & Ullman, 1985). These salience maps are typically considered spatiotopic. A spatiotopic map yields a topographic
representation of the information from the visual environment as it is configured in the outside world. This representation is only marginally, if at all, influenced by the direction of gaze. On the contrary, a retinotopic map takes the spatial configuration of the visual scene exactly how it is received by the retina. This implies that the topographic configuration of the visual environment is critically dependent on the direction of gaze. Movements of the eyes elicit a drastic change in the manner in which information in the map is organized, whereas representations in spatiotopic space hardly change as a consequence of eye movements. Salience map models are not directly associated with the SC. Instead, the salience map is often linked with human cortical brain areas, such as the posterior parietal (Gottlieb, Kusunoki, & Goldberg, 1998), occipital lobe (Li, 2002), or frontal eye fields (Thompson & Bichot, 2005).

One of the most influential salience-map models was proposed by Itti et al. (1998; Itti & Koch, 2000; 2001; Koch & Ullman, 1985). The model has largely been inspired by the Feature Integration Theory (FIT) (Quinlan, 2003; A. M. Treisman & Gelade, 1980). FIT describes that low-level visual information, i.e. basic stimulus features, can be processed pre-attentively. Accordingly, there is an unlimited capacity to obtain information from all locations in the visual field simultaneously, which is accomplished by encoding low-level information in separate (spatiotopic) maps for each feature dimension. This concept of feature maps is the core assumption of the model of Itti et al. (1998). However, instead of a single map for each feature dimension as described by FIT, Itti et al. (1998) assumes there are multiple maps for each dimension which together form a hierarchical structure. At the lower levels, the processing units have smaller receptive fields than at the higher levels. Moreover, to account for a plausible underlying neural mechanism, the model incorporates center-surround suppression (Koch & Ullman, 1985). Center-surround suppression implies that the presence of a particular feature might locally enhance activity in a feature map, but depending on the presence of that same feature in its surroundings, the induced activity levels can be down regulated just as much. In the model of Itti et al. (1998), center-surround suppression together with the model’s hierarchical structure enables salient features to be detected relatively independent of the size of the involved object or region. Eventually this leads to a single or a small number of regions in the visual field that are considered salient based on feature contrasts within the feature dimensions. In
This information is combined into a single master map in which the activity levels reflect the overall salience levels, i.e., the salience levels across all feature dimensions. So if a location has a high feature contrast to neighboring locations in the visual field, it evokes a higher activity in the master map, whereas a less distinct location generates less activity. When there are multiple salient locations in the visual field, the model selects the location generating the highest activity first. Itti et al. (1998) assumes that this is the location in the visual field where the first saccade will be aimed at. Once that location is selected, its activity is suppressed, an operation inspired by the mechanism known as ‘Inhibition of Return’ (Itti & Koch, 2001; Klein, 2000; Posner & Cohen, 1984). This results in the selection of the second most salient location, followed by the third, and so forth. Importantly, the locations that are selected by the model of Itti et al. (1998) largely correspond to the locations that human observers tend to select (Niebur & Koch, 1996; Parkhurst, Law, & Niebur, 2002). This was found in an eye tracking study of Parkhurst et al. (2002) in which participants were asked to freely observe artificial scenes (fractals), as well as complex natural scenes, including natural landscapes, home interiors, and pictures of buildings and cities. In the results, there was a significant correlation between the fixation locations of participants and the salience of those locations in the scene. This correlation was present for all types of scenes and was greatest for gaze responses that immediately followed the onset of the scene presentation. The results therefore support that the salience-based model of Itti et al. (1998) accounts for some of the variance in overt selection behavior.

Zhaoping Li (2002) proposed an alternative model to account for salience-driven selection. Her model is based on the idea that the primary visual cortex (V1) could be an eligible region of the brain to represent salience levels. She therefore proposes that some main properties of V1 should be better taken into account when designing a salience-based model. For instance, the manner in which the neurons in V1 are structurally ordered already inherently carries

1 The activity levels in the master salience map are assumed to predict oculomotor behavior although the configuration of the map is spatiotopic. This does not match the retinotopic organization of the SC and the hypothetical saccade map. Exactly how information is translated from spatio- to retinotopic space is left unexplained by Itti et al. (1998). Accordingly, FIT, the theory on which the model of Itti et al. (1998) is based, does not account for eye movement behavior.
information about the spatial structure of a visual scene when a scene is processed. Moreover, although the individual receptive fields of the neurons in V1 are small, the neurons share horizontal connections with one another (Tso & Gilbert, 1988), potentially allowing for representations that span the entire visual field. In particular those neurons are connected that share a sensitivity for similar low-level stimulus features. V1 might therefore signal salience without the need to categorize processed information into different feature maps. Accordingly, the model of Zhaoping Li (2002) assumes that salience is reflected in V1 depending on firing rates at particular locations rather than on the identities of the neurons that fire. Visual input in a neuron’s receptive field leads the neuron to fire, and due to the intracortical inhibitory connections among the neurons, firing rates are modulated by the firing rates of the other V1 neurons. As a result, salience is computed by ‘bottom-up’ mechanisms alone, i.e., without the need to integrate the information from separated feature maps. A number of studies have found evidence for a salience map in V1 (Hupe et al., 1998; Jingling & Zhaoping, 2008; Koene & Zhaoping, 2007; Yantis, 2005; Zhaoping, 2008). For instance Zhaoping (2008) conducted a series of experiments in which distinct visual information was presented separately to each of the eyes of the participants. These dichoptic presentations were designed to examine whether salience is represented by neurons that receive monocular information or neurons that receive binocular information. Crucially, monocular neurons are abundant in V1 and scarce in other cortical areas. In one of the tasks of Zhaoping (2008), participants searched for an orientation singleton embedded in a field of background line segments in a brief (200 ms) presentation. There were three dichoptic presentation conditions: (i) the target element was separated from the rest of the search array by presenting it to one of the eyes, (ii) a single background element was separately presented to one of the eyes, or (iii) all elements were presented together to just one of the eyes. Importantly, the combination of the information that was received by the two eyes was equal in all conditions. Accordingly, participants reported that they could not consciously notice a difference between the conditions. The results demonstrated that search performance was higher when the target was presented separately from the rest of the search array, as compared to the other two conditions. Presumably, participants perceived the target as more distinct, more salient, when it was presented to one of the eyes separately, which would imply that monocular
neurons are involved in signaling the salience of information. This interpretation of the results fits the idea that a salience-based map is constructed in V1, i.e., if one assumes that a cortical area is involved.

Even though the Itti et al. (1998) model and the Zhaoping Li (2002) model both do well in accounting for some of the variance in overt selection behavior, they certainly do not provide a full explanation. Crucially, the models fall short of explaining selection behavior that is not controlled by salience. Itti et al. (1998) acknowledged the importance of top-down influences and the necessity to incorporate these into the model to provide a full understanding of selection control. One approach would be to assume that certain feature maps can be weighted more heavily than the other maps. If a search target is red, for example, the feature maps that process the feature dimension of color could be made more heavily weighted. Particular locations in the master salience map can also be differently weighted based on the goals and motivational state of the observer (Navalpakkam & Itti, 2005; Wolfe, 1994). However, the incorporation of these influences into a salience model and in particular how these top-down influences interact with salience-driven influences is still in its infancy (Borji & Itti, 2013; Itti, 2006; Peters & Itti, 2007).

1.3.3. Goal-driven selection

The models discussed above assume that goal-driven selection is contingent upon stimulus-driven selection (Hochstein & Ahissar, 2002; Hoffman, 1979; Itti & Koch, 2000, 2001; Itti et al., 1998; Julesz & Bergen, 1983; Koch & Ullman, 1985; Lamme & Roelfsema, 2000; Theeuwes, 2010; Theeuwes & Burger, 1998; A. M. Treisman & Gelade, 1980; VanRullen & Koch, 2003). In other words, information always needs to be selected based on low-level characteristics before goal-driven mechanism can gain control. However, opponents assume that (overt) selection behavior is foremost governed by the intentions of the observer (Bacon & Egeth, 1994; Chen & Zelinsky, 2006; Folk, Remington, & Johnston, 1992; Folk, Remington, & Wright, 1994; Gibson & Jiang, 1998; Gibson & Kelsey, 1998; Posner, 1980; Posner, Snyder, & Davidson, 1980). According to this view, stimulus-driven selection is subordinate to goal-driven mechanisms. Evidence for the idea that attentional selection is primarily goal-driven was provided by Folk and colleagues (Bacon & Egeth, 1994; Folk et al., 1992; Folk et
INTRODUCTION

al., 1994; Gibson & Jiang, 1998; Gibson & Kelsey, 1998). One example of an experiment that appears to suggest that goal-driven mechanisms can dictate visual selection is provided by Folk et al. (1992). Participants were asked to report the identity of a target that abruptly appeared in one of four empty boxes that were already in place at fixed positions on the screen. For one group of participants, the target simply appeared alone, without any distractors, leaving three boxes empty in the search display. Hence, in this version of the task, participants searched for an 'onset target'. For another group of participants, the target simultaneously appeared with three distractors, filling all four boxes of the search display. The distractor identities were determined by randomly selecting one of the two potential target identities (either 'X' or '='), individually for each distractor. In order to dissociate the target from the distractors, the target was colored red whereas the distractors were white. Hence, in this version of the task, participants searched for a 'color target'. For both groups of participants, there were four cueing conditions that were separated in blocks of trials: the no cue, center cue, valid cue and invalid cue conditions. In the most interesting of these conditions, the valid cue and invalid cue conditions, a spatial cue was applied prior to the search display. One of the four empty boxes was briefly surrounded by bright dots, eliciting a flash percept at one of the potential target positions. The authors hypothesized that the validity of this abrupt-onset cue should only affect the performance of the participants that searched for an onset target. The results indeed showed that in the invalid cue condition, reaction times were impaired when the target was an onset singleton, but not when the target was a color singleton. As a continuation, Folk et al. (1992) conducted a second experiment in which the valid and invalid cues highlighted all four potential target locations, but one location with a red color and the others with a white color. This subtle change of the cue led to an opposite pattern of results. The validity of the cue now influenced the performance of the participants that searched for a color target. Reaction times were impaired when the cue was invalid and the target was a color singleton, but not when the target was an onset singleton. Together, the results of Folk et al. (1992) suggest that only the cues that share the same features as the search target, only the distractors that match the features that observers search for, are capable of affecting attention. Folk et al. (1992) concluded that selection is dominated by the goals of the observers. Selection
is determined by the ‘attentional set’ and all features that do not match this attentional set are simply ignored.

A well-known model in which it is also assumed that the goals of the observer can directly tune the manner in which low level features are processed is Wolfe’s Guided Search model (Cave & Wolfe, 1990; Wolfe, 1994, 2007; Wolfe, Cave, & Franzel, 1989; Wolfe & Gancarz, 1996). The model is also largely based on the Treisman’s FIT. Guided Search consists of two stages: an initial parallel stage and a subsequent serial stage. However, a major difference with FIT is that Guided Search assumes that goal-driven influences can already operate at the first stage. The stage that FIT considers fully pre-attentive can be altered in accordance to the goals of the observer in Guided Search. As a result, the first stage can guide the attentive stage of visual processing. For instance when observers search for a red target, the pre-attentive stage will be sensitized for red, and in turn, the attentive stage will be more prone to select red objects. Importantly, more recent versions of the Guided Search model also take into account that features can be salient regardless of the search goals (e.g., Wolfe, 2007). Stimulus-driven activations can still derive from local contrasts within particular feature dimensions, relatively similar to the salience models.

To account for overt selection behavior, Zelinsky and colleagues (Rao, Zelinsky, Hayhoe, & Ballard, 1997; Zelinsky, 2008; Zelinsky, Rao, Hayhoe, & Ballard, 1997) designed a computational model inspired on the theoretical framework of Guided Search, emphasizing the role of goal-driven influences. The model predicts landing positions of oculomotor responses in real world scenes, assuming that the features of a search target determine how visual information is processed already at a pre-attentive parallel stage. Simultaneously for all regions in the visual field, the model computes the extent to which the features in the scene match the features of the search target. However, although the model computes correlations between the target and all positions of the scene in parallel, the model cannot process in great detail the information that is located in the periphery. Therefore, when there are multiple potential targets located in the periphery of the visual field, uncertainty remains regarding the exact position of the target. To solve this issue, the model initiates a saccadic movement to bring the potential target locations closer to the center of the visual field, the region with the highest visual acuity. Commonly, the model aims these saccades towards a location in
between the potential targets because such an intermediate location allows for optimal information gain in resolving the competition between potential targets. Depending on the complexity of the search display, multiple saccadic movements are executed before the location of the target is determined. Many of these intermittent movements are directed at locations in between the potential targets. This is an interesting aspect of the model because the simulated behavior resembles the global effect phenomenon in human observers. However, whereas Zelinsky et al. (1997) explains the intermediate landing positions as a consequence of goal-driven mechanisms, the global effect is believed to be caused by stimulus-driven factors. Moreover, Zelinsky et al. (1997) finds these effects not only for the first saccadic movement. A second and a third response can land at a location in between objects just as well. This is incompatible with the time course considered typical for the global effect. The global effect is known to be most prominent for fast initial responses (Coeffe & Oregan, 1987; Findlay, 1982; Ottes, Vangisbergen, & Eggermont, 1985; Van der Stigchel & Theeuwes, 2005). If an oculomotor response is initiated relatively late in time, the eyes tend to land closer to individual objects. In Chapter 4 of this thesis, I will discuss in greater depth how the findings of Zelinsky et al. (1997) relate to the global effect. New empirical findings will be introduced to examine to what extent the global effect can be caused by uncertainty regarding the spatial positions of objects in a scene, or by uncertainty regarding the identity of objects in a scene.

1.4. Time course of overt selection

1.4.1. Transience in the representation of salience

Since both stimulus- and goal-driven influences turn out to play an important role in visual selection, one question that needs to be answered is which of the two effects predominates. One fruitful approach is to investigate the time course in which one or the other type of influence prevails (N. C. Anderson, Ort, Kruijne, Meeter, & Donk, 2015; Donk & van Zoest, 2008, 2011; Kim & Cave, 1999; Lamy & Egeth, 2003; Ludwig & Gilchrist, 2002; Nakayama & Mackeben, 1989; van Zoest & Donk, 2004, 2005, 2006, 2008, 2010; van Zoest, Donk, & Theeuwes, 2004; van Zoest, Donk, & Van der Stigchel, 2012). One example of a study that has systematically investigated this time course is van Zoest et al.
(2004). In a series of experiments, participants were asked to make a speeded saccade towards a singleton target line in a grid of homogenous background line segments. The target consisted of a distinct orientation relative to a large number of background lines. The search display also contained a distractor line that was also distinctly oriented relative to the background lines, but in the opposite orientation of the target. The orientation contrast between a singleton (target or distractor) and the background lines determined the relative salience of the singletons. In one of the experiments, in any given trial, either the target or the distractor line was more salient. The results showed that when the distractor was more salient, performance was severely impaired relative to performance in trials in which the target was more salient. That is, observers often made an incorrect eye movement towards the distractor rather than towards the target. Crucially, this was particularly the case for the short-latency responses. This finding suggests that the fastest oculomotor responses are mainly driven by the relative salience of the singletons whereas this was not the case for the slower responses. In fact, as saccadic latencies increased, the proportions of saccades aimed at the target gradually increased. Together these results suggest that salience-driven effects prevail immediately after the presentation of a search display whereas goal-driven influences dictate oculomotor selection behavior later on.

One reason why the influence of salience-driven control on oculomotor selection behavior decreases with increasing saccadic latency may be that salience-driven influences are ultimately overruled by goal-driven processes. Alternatively, it may be that the influence of salience is inherently short-lived, i.e., salience might only transiently affect selection, irrespective of the presence of opposing goal-driven processes. This latter interpretation has been supported by findings of Donk and van Zoest (2008). In one experiment, participants performed a speeded saccade that had to be aimed toward the most salient orientation singleton (among two) that was presented in a field of homogenous line segments. The absolute orientations of the two singletons did not vary across trials, but because the orientation of the background line segments did vary (randomly altered between horizontal and vertical), the absolute orientation of the singletons could not be used to identify the salience of the singletons. In other words, the identity of the singletons was dissociated from the salience of the singletons. Remarkably, the results demonstrated that participants more accurately directed the eyes to the most
salient singleton when they moved their eyes quickly. When saccadic latencies increased, oculomotor selection performance gradually decreased even to chance level. In another experiment, participants were asked to report which of two simultaneously presented singletons was most salient in a search display similar to the displays used in the first experiment. Importantly, the duration of the presentation of the display was varied and participants had to report their answer with a button-press. The results were remarkable again. Participants performed better when displays were briefly presented, compared to when displays were presented for a more prolonged duration. Together these results suggest that representations of salience are transient. It appears that the influence of salience is inherently short-lived. Importantly, this conclusion appears to reject one of the main assumptions of the salience map models that were earlier discussed: the assumption that salience is a continuous factor in overt selection behavior. However, so far the role of salience has been investigated merely for the initial oculomotor response. It is still an open question whether the influence of salience may resurrect after each response. The role of salience might still be considered long lasting, i.e., if one examines overt selection behavior more generally, beyond the initial eye movement.

1.4.2. Salience beyond the first eye movement

As was put forward in the previous section, the effects of salience on initial oculomotor behavior are transient rather than continuous. Several studies reported a clear reduction in the strength of the effect of salience with increasing saccadic latencies (Donk & van Zoest, 2008; Franconeri, Simons, & Junge, 2004; Kim & Cave, 1999; Lamy & Egeth, 2003; Ludwig & Gilchrist, 2002; Nakayama & Mackeben, 1989; van Zoest & Donk, 2004, 2005, 2006, 2008, 2010; van Zoest et al., 2004), and the results of Donk and van Zoest (2008) even suggested that the brain only briefly represents salience. However, so far it has remained unclear how salience affects subsequent oculomotor behavior. The first indications were provided by free-viewing tasks in which stimulus salience correlated with the fixation locations of observers (Niebur & Koch, 1996; Parkhurst et al., 2002). That is, salient locations were more likely to be fixated than non-salient ones. Importantly, this correlation was demonstrated
to be strongest for the first selected fixation location and declined as a function of fixation number. Nevertheless, even later in the trial, eye movements were still influenced by salience, suggesting that salience effects do not completely fade away (Parkhurst et al., 2002). One explanation for the presence of this correlation is that salience continues to affect oculomotor selection behavior because the salience representation is reinstated each time an eye movement is made. Thus, even though the salience representation may dissipate during each fixation, it may be reinstated each time an eye movement is made. This interpretation paralleled with a well-known neural mechanism called neural adaptation (Hartline, 1940; Kuffler, 1952). The visual system might lose sensitivity to exogenous salience-driven signals because these signals are continuous and constant during fixation, but as soon as gaze direction is shifted, the signals change and the effects of salience are re-instated. Alternatively, salient regions simply have an enhanced chance to be selected due to the correlation between salience and the presence of interesting information (Elazary & Itti, 2008; Masciocchi, Mihalas, Parkhurst, & Niebur, 2009). Salience might only directly affect fast initial responses, but exerts no control in later eye movements at all. In fact, it is possible that observers preferably select salient locations because they have a top-down preference to fixate interesting objects rather than non-interesting backgrounds (Einhauser, Spain, & Perona, 2008). A final explanation is that people exhibit a central fixation bias (Tatler, 2007; Tatler, Hayhoe, Land, & Ballard, 2011), a tendency to look at the center rather than the peripheral regions of a picture. Since the center of a picture commonly contains an interesting object or some other salient region (since this is simply the way humans take pictures of natural scenes), the correlation between salience and fixation is an inevitable outcome of the central bias, irrespective of the true occurrence of salience-driven effects.

To examine oculomotor selection behavior beyond a first eye movement with stimuli that are better controlled than natural scenes, Siebold, van Zoest, and Donk (2011) presented observers with three orientation singletons concealed in a background of homogenously oriented line segments. The singletons were equidistant from a central fixation point, and from each other, and the singletons varied in salience. One singleton was highly salient, the second was medium salient, and the third was of a low salience. Observers had the task to search for a small dot (present in one third of the trials), which was
superimposed on one of the three singletons. The results showed that the proportion of initial eye movements aimed at the individual singletons depended on salience. The singleton that was more salient was more likely to be selected. This tendency decreased with increased latencies, but subsequent (second) eye movements turned out unaffected by the salience of the singletons completely. Salience did not affect oculomotor selection behavior beyond the first eye movement. The influence of salience was not reinstated when an eye movement was performed. Continuing on these results, Siebold and Donk (2014) attempted to establish whether salience can have an influence beyond a first eye movement as a result of a dynamic change in the visual field. Observers were presented with a display containing two orientation singletons, one target and one distractor. The task instruction was to make a speeded saccade to the target singleton, either directly from fixation, or preceded by an eye movement to an additional fixation dot. The presence of rapid luminance flicker (Experiment 2) or the presence of a higher luminance contrast (Experiment 3) determined whether the target or the distractor was more salient. Finally, the luminance contrast or flicker was either implemented from the start of the search trial, or during the first eye movement. The results showed that direct eye movements were profoundly affected by the relative salience difference between the singletons, irrespective of the specific salience manipulation, but more importantly, second eye movements were not. Salience had no influence on oculomotor selection beyond the first eye movement\(^2\). However, when the salience difference between the singletons was implemented later in time, i.e., during an initial eye movement aimed at an additional fixation dot, a new salience effect was observed for the second oculomotor response. Together these results show that, without any change, salience does not affect oculomotor selection behavior beyond the first eye movement, but its effects can be reinstated if salience values change, even when these changes occur contingent upon a first eye movement. If new highly salient information is introduced in a scene, the oculomotor system can be guided by salience once again.

\(^2\) Changes during an eye movement cannot be perceived directly. The inability to perceive changes during saccadic movements is presumably caused by a mechanism known as saccadic suppression (e.g., Ross, Burr & Morrone, 1996).
Importantly, this finding is incongruent with observations of prior studies that also applied saccade-contingent changes to visual scenes (Brockmole & Henderson, 2005, 2008; Matsukura, Brockmole, Boot, & Henderson, 2011; Matsukura, Brockmole, & Henderson, 2009). For instance, Brockmole and Henderson (2005) performed a study in which participants were asked to free view natural scenes in which stimuli occasionally abruptly appeared. These abrupt onsets could either occur during saccades or during an eye fixation. In the results, the timing of the onset turned out to be a crucial factor. When an onset was applied during fixation, the onset quickly captured gaze, whereas an onset during a saccadic movement affected primarily long-latency and later saccades. Furthermore, the onsets had more impact on oculomotor behavior when introduced relatively late after the start of the presentation of the scene. Brockmole and Henderson (2005) therefore concluded that an onset presented during an eye movement created an identity mismatch between the visual scene and internal (memory) representations of the observer. This provoked mismatch is what drives subsequent eye movements towards the onsets, not salience-driven influences. Whether the findings of Siebold and Donk (2014) are indeed also the result of an identity mismatch, rather than an effect of salience, needs to be further examined. In Chapter 2, I will discuss and provide new observations to solve this unsettled issue.

1.4.3. Towards a model of overt selection on and beyond initial responses

To better understand what drives overt selection behavior beyond the initial response, the essential function of salience must be illuminated. Most researchers agree that an abrupt change in salience, e.g., a luminance increase, may have a prominent influence on selection behavior because it commonly signals the appearance of a new object (Egeth & Yantis, 1997; Franconeri & Simons, 2003; Hollingworth, Simons, & Franconeri, 2010; Rauschenberger, 2003; Yantis, 2005). From an evolutionary perceptive, newly detected objects could be a potential threat or reward, so the system should be adapted to immediately select regions that suddenly contain enhanced salience levels. Moreover, regions in visual scenes that signal relatively high salience commonly correspond with regions that participants judge as relatively interesting (Einhauser et al., 2008; Elazary & Itti, 2008; Masciocchi et al., 2009).
Based on these characteristics of salience, there are two alternative interpretations of the essential function of salience in overt selection. On the one hand, salience may be processed as signaling the relevance of particular locations in space. This role of salience could be enabled by a spatial representation of the visual scene, such as a salience map, in which relative variations in salience directly affect the selection priority of particular locations, possibly with or without the involvement of goal-driven mechanisms. Accordingly, one might expect salience to be a property that influences selection continuously. On the other hand, salience could be the means to segregate objects from the background. In line with this idea, salient objects presumably are detected faster than objects that are less salient. Rather than being represented continuously, salience might be coded in time. A salient object does not cause the corresponding location in a salience map to be more active. A salient object simply causes the corresponding location to be active at an earlier point in time. In this sense salience is processed as an emergent property, rather than a continuous one, a property that is determined by the speed with which individual objects are detected by the system (Donk & Soesman, 2010, p. 299; see also Donk & Soesman, 2011; Donk & van Zoest, 2008) A selection bias for selecting salient over less salient objects may arise initially, but eventually a state of equivalence will emerge between all locations that stand out from the background.

Evidence for this view stems from work of Donk and Soesman (2011). In a temporal order judgment task, two orientation singletons briefly changed color and observers had to indicate which of the singletons changed color first. These singletons were embedded in a grid of background lines. In the first experiment, both lines had a distinct tilt relative to the background elements, and since this tilt was larger for one singleton than the other, the singletons differed in salience. A second manipulation involved the interval prior to the color change. Participants had to wait during a brief (58 ms) or long interval (800 ms) for the color change to occur. The results showed that the temporal order judgments were affected by salience after a brief interval, while the judgments were unaffected by salience after a long interval: participants tended to report that the more salient target line had changed first. However, this bias was only present when the change occurred after a brief interval and not after a long interval. This finding suggests that salience is only briefly represented. A second experiment was similar to the first except that one of
the singletons was substituted by a background line. The results now showed that the temporal order judgments were affected by salience both after a brief and a long interval. The authors concluded that there must be a difference between object salience and object presence. Whereas the salience of a singleton is only represented transiently, its presence (as opposed to the background) is continuously represented.

The conception of a relatively robust representation of object positions in a scene is congruent with a theoretical framework proposed by Henderson, Malcolm, and Schandl (2009) which is known as the cognitive relevance framework. Henderson et al. (2009) explain that low-level features of a visual scene form the basis for figure-ground parsing. Similar to the salience map models, low-level stimulus properties are processed already early in time and subserve the generation of a topographical representation of the scene. However, in contrast to the salience map models, Henderson et al. (2009) assumes the topographical representation to consist of regions that are not yet ranked in priority. The map is basically a flat landscape in which objects are merely segregated from the background. Then, as soon as the spatial figure-ground representation has been built, potential saccade targets are ranked by their cognitive relevance. Cognitive relevance is based on prior knowledge of the task, general semantic knowledge, and episodic knowledge (Henderson & Ferreira, 2004), which becomes available from the overall structure of the scene (the gist) and from detailed information that is obtained by fixating. In essence, overt selection behavior is thus determined by knowledge structures, not by visual salience. Visual salience merely segregates objects from the background, after which cognitive relevance determines which regions in the visual scene are prioritized and targeted. In Henderson et al. (2009), the cognitive relevance theory was put to the test with two empirical studies. In both, participants were asked to search for a target in natural scenes, a target with which they were either familiar (Experiment 1) or not familiar (Experiment 2). Targets were always presented at the semantically appropriate location in the scene. Crucially, the target could either be salient or not. This implied that many of the search scenes contained regions that were more salient than the target and thus potentially attracted gaze away from the target. Nevertheless, the results showed that search performance was fast and efficient. This was regardless of whether participants were familiar with the visual form of the search targets or not. Furthermore, in the majority of the trials, high-salience
regions were not fixated. In fact, participants were more likely to allocate gaze at the targets than at alternative regions in the scene that were more salient, suggesting that the semantic appropriateness of a location in relation to the target was a more prominent factor in search than the visual salience of a location.

A potential shortcoming of the theoretical framework of Henderson et al. (2009) is that it disregards salience-driven selection altogether. The cognitive relevance framework therefore does not reconcile with many of the empirical observations discussed in the previous sections, including the initial transient influence of salience on first oculomotor responses (e.g., van Zoest et al., 2004), and the reinstatement of salience-driven capture that results from newly introduced salient information (Siebold & Donk, 2014). Moreover, it neglects that salience signals the presence of information that is potentially vital; salience might signal the presence of new objects (Egeth & Yantis, 1997; Franconeri & Simons, 2003; Hollingworth et al., 2010; Rauschenberger, 2003; Yantis, 2005) or objects that are commonly judged as relatively interesting (Einhauser et al., 2008; Elazary & Itti, 2008; Masciocchi et al., 2009). This thesis therefore introduces a new model, which is largely analogous to the ideas of Henderson et al. (2009), but emphasizes on a more direct and prominent influence of salience early in time. Salience indeed subserves the generation of a spatial figure-ground representation of the visual field. However, instead of a simultaneous detection of all salient regions in the visual field, regions that are more salient are detected earlier in time than regions that less salient (see Figure 1.1). Accordingly, an initial selection bias might arise for highly salient over less salient objects. However, as the system rapidly develops a more elaborate representation of the spatial positions of objects in the scene, a state of equivalence will emerge between the regions that stand out from the background, a state similar to the flat landscape described by Henderson et al. (2009). In other words, obtaining knowledge concerning the presence of objects in a scene removes the selection bias that might emerge early in time due to relative differences in salience.
Figure 1.1 An illustration of the generation of a topographical representation of a visual scene. The visual input is a grid of line segments containing two orientation singletons. There is a salience difference between the singletons which leads to a differentiation in the time required to detect the presence of the singletons, explaining why initially the system might be more prone to select one singleton over the other, a selection bias that gradually fades away.
INTRODUCTION

Taken together, the model proposed here leads to a clear hypothesis: once the locations of the objects in a visual scene are represented in a (hypothetical) spatial object-ground map of the visual field, a relative difference in salience that may exist between the objects, or even a difference in salience that arises later in time, should not influence overt selection to a similar extent as would be the case when the locations of objects are not yet detected. Chapter 3 will elaborate on this prediction and will examine it with new empirical observations.

1.5. Beyond stimulus-driven and goal-driven factors

1.5.1. Selection history

According to the model outlined above, visual salience creates the proto-objects for goal-driven selection. While building this representation, selection behavior is biased towards the most salient locations since these are the first to be recognized as objects. This initial influence of salience on selection is considered to be stimulus-driven and involuntary. In turn, as time progresses, selection becomes more and more aligned with the task goals and selection becomes goal-driven. In this manner, stimulus- and goal-driven processes undoubtedly play a lead role in overt selection behavior. However, these factors certainly do not explain everything. As was mentioned earlier, there are also factors that determine where we attend and look, that do not fit this dichotomy. These factors do not include (physical) salience nor the conscious intentions of observers. An example of such a factor is known as feature-based priming. A classic demonstration of this phenomenon can be found in Maljkovic and Nakayama (1994). In a series of experiments, participants were asked to search for a color oddball among a number of other objects. In the standard task, the target and distractors were colored diamonds, each with a cut-off either on the left or on the right side. Participants were required to manually report the side of the cut-off of the oddball target as fast as possible. As the colors of the objects varied trial-by-trial, the color of the target could either repeat or change in relation to the previous trial. The results showed that a repetition of the color of the target decreased response times in comparison with response times in trials in which the color of the target changed. Selecting the target color in a previous trial thus accelerated selection
of the same target color in the next trial. Maljkovic and Nakayama (1994) observed that a selected color could even influence behavior for 5 to 8 trials. Moreover, the facilitation to select a previously attended color was unintentional. Participants could not even control the effects when they were informed about the stimuli in the upcoming trial. The observation that feature-based priming cannot be controlled by the observer puts priming outside of the category of goal-driven factors. However, the colors used by Maljkovic and Nakayama (1994) were close to equiluminant and counter-balanced across trials. In other words, the priming effects were also not elicited by the intrinsic stimulus properties that made one object physically more salient than another. Yet, selection behavior was clearly affected. It is for this reason that Awh et al. (2012) defined a new type of factor distinct from stimulus- and goal-driven influences: selection history. Selection history encompasses more than just feature-based priming alone. For instance, there are experiments that show that the configuration of targets and distractors in a search scene can facilitate responses in a similar fashion as does feature-based priming (Umemoto, Scolari, Vogel, & Awh, 2010). Another example is that the likelihood of objects to be a target can impact selection behavior without the conscious awareness of the observer (Chun & Jiang, 1998). Generally, selection history seems to involve implicit learning mechanisms. Once a selection bias is installed, it tends to linger on in a passive manner which may even bridge a time interval of several days, and up to a whole week (Chun & Jiang, 2003; Kruijne & Meeter, 2015; Leber, Kawahara, & Gabari, 2009).

Apart from priming mechanisms, another phenomenon that could be categorized within the selection history class is the experience of reward. In behavioral experiments, rewards have been used often as a means to boost the motivational engagement of participants, since reward stimulates performance relatively directly (Bucker & Theeuwes, 2014; Pessoa, 2009; Pessoa & Engelmann, 2010; Sarter, Gehring, & Kozak, 2006), but a large body of recent work demonstrates that reward can influence behavior without necessarily involving the willful intentions of the observer (B. A. Anderson, Laurent, & Yantis, 2011a, 2011b; B. A. Anderson & Yantis, 2013; Bucker, Belopolsky, & Theeuwes, 2014; Bucker, Silvis, Donk, & Theeuwes, 2015; Della Libera & Chelazzi, 2006, 2009; Failing & Theeuwes, 2014; Theeuwes & Belopolsky, 2012), and seemingly at a very fundamental level (Engelmann, Damaraju, Padmala, & Pessoa, 2009; Engelmann & Pessoa, 2007;
INTRODUCTION

Navalpakkam, Koch, Rangel, & Perona, 2010; Pessoa & Engelmann, 2010; Serences, 2008; Small et al., 2005). A recent example of an oculomotor study demonstrating this role of reward is the work of Bucker et al. (2015). Using a global effect paradigm, participants were asked to perform a speeded saccade toward two differently colored disks that abruptly appeared on the screen. There was no specific target object. The experiment consisted of three phases. Only in the second phase, the training phase, rewards were emitted depending on the color of the disk that the participant tended to aim at. There were three reward values consistently linked with the three object colors used in the experiment. The results showed that in the training phase the differently valued colors led to modulations in the direction of the global effect. Participants were biased to move the eyes more closely towards the highly rewarded color. In the post-training phase of the experiment, rewards were not delivered anymore, but the bias in the global effect, as they were induced by the incentive rewards in the training phase, remained. Importantly, whereas the oculomotor responses were relatively slow in the training phase, the response times in the post-training phase were relatively fast, as fast as they were in the first phase of the experiment, the pre-training phase. This finding suggests that participants may have strategically adapted their responses in the training phase in order to optimize the rewards they received, but once rewards were no longer delivered, participants changed back to a more stimulus-driven approach. Nevertheless, oculomotor selection continued to be affected by the learned stimulus associations. In conclusion, although researchers have long associated reward merely with an enhanced motivational engagement, recent work demonstrates that reward can install an overt selection bias that does not necessarily involve the willful intentions of the observer. This puts reward, together with effects of priming, in the category of selection history, a category representing a broad class of factors which could have a determining relevance to our daily life selection behavior. Selection history can be considered a key addition to the bottom-up top-down framework.
1.5.3. Working memory

Another research area that is related to selection by history is working memory. The influence of working memory content on overt selection does not fit the description of either stimulus- or goal-driven selection. Working memory is a concept that is typically considered higher-order and controlled by the intentions of the observer. For instance in the component model of Baddeley and Hitch (1974), which is one of the most widely accepted models of working memory, one of the fundamental components is the ‘executive function’, the homunculus in charge of all our intentional behavior and cognitions. It may therefore come as a surprise that working memory can influence visual selection in an involuntary manner. A reason why this could be the case is the strong overlap in structure and function between visual working memory and visual attention (Awh & Jonides, 2001; Awh, Vogel, & Oh, 2006; Desimone, 1996; Desimone & Duncan, 1995; Duncan, 1998; Olivers, 2008; Theeuwes, Belopolsky, & Olivers, 2009). Desimone (1996) even suggested that attentional selection is derived from memory mechanisms, at least in part. Accordingly, there is a body of literature that describes evidence for automatic effects of visual working content on attentional (Downing, 2000; Olivers, Meijer, & Theeuwes, 2006; van Moorselaar, Battistoni, Theeuwes, & Olivers, 2015; van Moorselaar, Theeuwes, & Olivers, 2014) and oculomotor selection (Hollingworth, Matsukura, & Luck, 2013; Silvis & Van der Stigchel, 2014; Soto, Heinke, Humphreys, & Blanco, 2005; Soto, Humphreys, & Heinke, 2006). For instance Olivers et al. (2006) adapted the attentional capture paradigm of Theeuwes (1992) to examine whether the impact of a distractor in a visual search task could be modulated by a match between the color of the distractor and a color maintained in memory. Participants were asked to search for a diamond among a number of squares and manually report whether the diamond contained a vertical or horizontal bar. Occasionally, a colored disk, the distractor, occurred with the onset of the search display. Prior to this search task, participants were shown a memory color that would be probed as soon as the search task was finished. The distractor that occasionally appeared in the search task could either match or mismatch the memorized color. The results showed that the presence of the distractor generally impaired search performance, but predominantly so when the distractor color matched the color that participants were memorizing.
More evidence comes from the work of Hollingworth et al. (2013). Rather than a response time task, Hollingworth et al. (2013) embedded an oculomotor task in a memory task. Participants memorized a particular shade of a color while directing the eyes as quickly and accurately as possible towards a predefined target object. The objects varied in color on a trial-by-trial basis, which was irrelevant to the oculomotor task objectives. The results revealed significantly faster and more precise oculomotor responses when the target color happened to match the memorized color. Importantly, even taking the fastest saccades, with a latency below 150 ms, eye movements still turned out to be more precise when there was a color match. The authors concluded that working memory content is capable of affecting the fastest possible responses, responses that are typically assumed to be stimulus-driven. Hence, working memory content modulates selection processes from very early on. Similar observations were obtained in an experiment conducted by Silvis and Van der Stigchel (2014). Their oculomotor task was designed to elicit a global effect response pattern, and they did not instruct participants to select one object over the other. Still, participants showed a preference to aim the eyes to a position closer to the objects that consisted of the same color as the color held in memory (an exact match or a different shade), despite the possible damage caused to the memory representation by selecting the similar object. This finding demonstrates that the influence of working memory content on oculomotor behavior is not only early, but also automatic and involuntary. Both Hollingworth et al. (2013) and Silvis and Van der Stigchel (2014) additionally examined whether the effects that appeared to be driven by working memory content could also be caused by feature-based priming. In an altered version of the task, participants were asked to observe a certain color without holding it in active memory before performing the oculomotor task. Hollingworth et al. (2013) as well as Silvis and Van der Stigchel (2014) demonstrated that the match between the previewed color and the color of a target object did not influence either saccadic latencies or the precision of landing positions of the eye movements. The effects of active memory thus appear to be distinct from effects of feature-based priming. However, one concern with the method of ruling out the involvement of feature-based priming is that the passive observation of a color could be insufficient to allow for priming. Several studies point out that feature-based priming critically depends on prior active visual selection (Asgeirsson, Kristjansson, & Bundesen,
The mere presentation of a feature may not suffice for priming effects to occur. This knowledge puts doubt to the claim that match effects in oculomotor selection should be attributed solely to active working memory content. More research is needed to elucidate whether working memory is indeed a driving factor of oculomotor behavior, independently from priming influences. Chapter 5 will describe empirical work that provides new insights into the distinction between feature-based priming and the effects of working memory content.

1.6. Summary and preliminary conclusions

Our understanding of overt selection has progressed much over the past decades. We have come to know about three distinct ways in which selection behavior can be influenced: stimulus-driven, goal-driven, and a recently identified category linked to the prior selection history of the observer. First, stimulus-driven influences derive from the intrinsic characteristics of the stimuli and operate in an automatic and involuntary manner. Some have theorized that any type of selection behavior is always contingent upon stimulus-driven selection. Accordingly, the influence of visual salience on selection is a key component in a number of influential computational models that, rather successfully, predict which regions in a visual scene are more likely to attract gaze. A major weakness of these models is that they fail to account for overt selection behavior that is not merely driven by salience. They might explain free viewing behavior, but they do not provide a satisfactory explanation for search behavior.

Besides the issues with models that are either stimulus-driven or goal-driven, there are also some issues regarding the interpretations and implications of recent empirical findings. A number of studies suggest that salience is represented only very briefly, not continuously. Moreover, influences of salience do not reinstate after eye movements, except when novel highly salient information is introduced. To account for these observations, a new model is introduced in this thesis. This model is largely analogous to the cognitive relevance framework of Henderson et al. (2009), but whereas Henderson et al. (2009) merely assumes a subordinate role of salience, this
model presented here addresses a more direct and prominent influence of salience on early overt selection. Salience is utilized to compute object-background segregation, and crucially, while doing so, highly salient objects are detected earlier in time than objects that are less salient. Accordingly, early in time observers are more prone to select highly salient objects, simply because these are already represented in the topographic figure-ground representation of the visual scene. Shortly after, when the system has developed a more elaborate representation of the spatial positions of the objects in the scene, a state of equivalence will emerge: a flat landscape in the topographical map reflecting the locations that stand out from the background.

The descriptive model proposed in this thesis potentially offers a framework in which also a third type of influence on overt selection can be incorporated: selection history. This type of influence originates from the prior selection behavior of the observer, rather than the intrinsic characteristics of the stimuli, and influences oculomotor selection outside the observer’s conscious will. Possibly, selection history facilitates the detection of objects in a background, similar to the function of object salience. In the model of Henderson et al. (2009), there are knowledge frameworks that assist in the generation of a figure-ground representation of the scene, possibly incorporating an effect of selection history. However, according to the model of Henderson et al. (2009), these knowledge frameworks always serve the goals of the observers, which makes the model hard to reconcile with the effects of selection history that are clearly involuntary. Therefore, the model of the current thesis takes selection history as a component in the generation of figure-ground representations. Selection history elicits a bias in the early selection behavior of observers, not because they are in accordance with the search goals, but because selection history allows for an earlier detection of a particular object.

In the next chapters of the thesis, a number of ambiguities in the literature are further explored with novel observations from empirical studies. Together the chapters offer a more solid foundation for the newly introduced model. The first issue concerns the conflicting findings and interpretations between the studies of Siebold and Donk (2014) and the studies of Brockmole and Henderson (2005). On the one hand, oculomotor capture beyond the first response might be due to a mismatch between the visual scene and memory representations of the scene. If correct, one would expect various types of
changes in visual scenes to generally affect oculomotor behavior. On the other hand, capture beyond the first eye movement may be induced by an abruptly increased salience level. If true, one could similarly expect a variety of changes to induce capture but only when these changes involve an increased salience level. Chapter 2 describes a series of experiments that tested these hypotheses. The results turned out in favor of the interpretation of Siebold and Donk (2014). Only when the salience of particular objects in a search display was increased, search performance was affected. Changes that did not involve an increased salience level could not influence the proportion of oculomotor responses correctly aimed at the target. This rules out that general memory mechanisms are responsible for oculomotor capture. The results also oppose the assumption that overt selection is affected by relative salience differences among objects. Rather, the results suggest that salience-driven capture is the result of the early or novel detection of an object in a scene, which supports the view that salience is relevant for object-ground segregation.

Chapter 3 dives deeper into the segregation functionality of salience. It might be incorrect to think of salience as a physical property that leads to a continuous prioritization of particular regions in space. Instead, salience may merely allow for a faster detection of some objects than others. In turn, a faster detection of particular objects could lead to an initial selection bias. This bias is eventually removed by obtaining knowledge concerning the presence of other objects in a scene. The empirical work discussed in Chapter 3 will provide novel observations that support the idea that salience-driven effects disappear as a function of the availability of object location information in a visual scene. In a set of search task experiments, the onset asynchrony of singletons (target and distractor) relative to the background elements was varied. At the moment when the singletons and background elements were both present in the display, the background elements elicited a salience difference between the singletons. The results demonstrated that this salience difference only had a prominent influence on oculomotor selection performance when the onset of the singletons and background elements was simultaneous. This finding confirms the idea that salience effects are caused by the speed with which objects can be detected.

Another conflict between different lines of research involves the global effect phenomenon. Zelinsky et al. (1997) described eye movement behavior beyond
the first response to be directed at intermediate locations, whereas the global effect is traditionally seen as an effect that occurs for fast initial responses. **Chapter 4** describes a series of experiments that attempted to replicate the observations of Zelinsky et al. (1997) with tasks involving simple, more controlled, search objects rather than complex real world search objects. The results showed that eye movements beyond the first saccade do not land in between objects when these object are relatively simple. However, saccades did land in between the objects when the target and distractor identities are initially masked and therefore initially unavailable. The global effect can thus be induced, at least in eye movements beyond the first, due to a lack of knowledge concerning which of the objects is the target, not due to a lack of spatial knowledge about the positions of the objects. In line with the proposed model, oculomotor responses beyond the first, which are responses initiated later in time, do not lack knowledge about the positions of relatively distinct visual objects.

The final issue concerns the effects of working memory content on oculomotor selection. Features that are actively held in memory appear to install an automatic prioritization to select that feature as soon as it is encountered in a visual scene. What has remained unclear is whether the selection bias caused by active memorization is distinct from a selection bias caused by mere prior selection. **Chapter 5** presents empirical work to affirm the unique nature of oculomotor capture induced by active working memory content. In two experiments, participants performed a color memory task and an oculomotor task each trial. In the oculomotor task, the colors of the target and distractor item were randomly varied. They were task irrelevant, but could potentially match the color of the color memory task. The results showed that the memory color biased selection performance in the oculomotor task, and importantly, both when the color had to be memorized (Experiment 1) and when the memory task was already completed (Experiment 2). However, the selection bias was more pronounced when a color was actively maintained in memory, suggesting that working memory content indeed has an influence on overt selection that is unique from influences associated with prior selection.

Together the empirical studies discussed in Chapters 2 to 5 are in support of the model of this thesis. The model’s relatively simple set of assumptions can
account for and integrate the known factors that drive overt selection into a common theoretical framework.