Levels of visual information processing: perception of dynamic properties and events

Dissertation
der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
M.Sc. Alisa Brockhoff
aus Iserlohn

Tübingen
2016

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	13.03.2017
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Jun. Prof. Dr. Markus Huff
2. Berichterstatter:	Prof. Dr. Stephan Schwan

Declaration of Authorship

I, Alisa BROCKHOFF, declare that this thesis titled, "Levels of Visual Information Processing: Perception of Dynamic Properties and Events" and the work presented in it are my own.

Unfortunately, I have no financial interest to declare.

I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date: 20.12.2016

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Abstract

Levels of Visual Information Processing: Perception of Dynamic Properties and Events

Objective: The here presented studies explore automatic and controlled perceptual processes in two dynamic paradigms and support a rationale of a multi-level approach to dynamic visual perception.

Method: I investigate different perceptual levels of dynamic scenes, including factors within the perceiver, within the objects, and within the environment. Automatic processes are explored with a simple 3-D tracking task and an event perception recognition task; controlled processes are observed in a modified tracking task with specific object properties and an identification task.

Results: Through analysis of tracking and report errors measured in the two paradigms, I observed similarities in the automatic processing of artificial 3-D tracking environments (Study 1: the scene-based relations are more important than positions of individual objects) and real-life video clips (Study 2: core aspects are preferred over fine details). Despite the assumption that tracking is a cognitive-impenetrable mechanism, results of the modified tracking task (Study 3) point towards the ability of participants to strategically weigh visual information based on task-demands.

Conclusion: The results of this dissertation illustrate that the identification of influential internal and external factors is important to enhance our understanding of the multidimensional nature of perception – an understanding that will eventually and hopefully bring research to move beyond questions of how resources are limited, and start to focus on fundamental issues like how we can use mental resources to our benefit.

Keywords: Multiple Object Tracking, event perception, cognition, scene-based, object based.

Übersetzung

Ziel: Die hier präsentierten Studien untersuchen automatische und gesteuerte Prozesse mit zwei dynamischen Paradigmen und unterstützen die Grundüberlegung bezüglich eines Mehr-Ebenen-Ansatzes zur dynamischen Wahrnehmung.

Methode: Ich untersuche verschiedene perzeptuellen Ebenen von dynamischen Szenen, darunter Faktoren innerhalb des Wahrnehmendens, innerhalb der Objekte und innerhalb der Umgebung. Automatische Prozesse werden hier mit einer simplen 3-D Tracking-Aufgabe sowie mit einer Wiedererkennungsaufgabe bei der Geschehenswahrnehmung untersucht; gesteuerte Prozesse werden in einer modifizierten Trackingaufgabe mit spezifischen Objekteigenschaften und einer Identifikationsaufgabe beobachtet.

Ergebnisse: Durch eine Analyse der Trackingfehler und der Fehler in der Wiedererkennungsaufgabe in den beiden Paradigmen konnte ich gewisse Gleichartigkeiten in der automatischen Verarbeitung von artifiziellen 3-D Tracking-Umgebungen (Studie 1: die szenen-basierten Beziehung waren wichtiger als die Positionen der einzelnen Objekte) und von realitätsnahen Videosequenzen (Studie 2: Kernaspekte wurden feinen Details vorgezogen) feststellen. Ungeachtet der Annahme, dass Tracking ein kognitiv-unzugänglicher Mechanismus sein könnte, weisen die Ergebnisse von Studie 3 signifikant darauf hin, dass Teilnehmer dazu in der Lage sind, die visuellen Informationen basierend auf den Aufgabenanforderungen strategisch abzuwägen.

Fazit: Die Ergebnisse dieser Dissertation zeigen deutlich, dass die Ermittlung von einflussreichen internalen und externalen Faktoren wichtig ist, um unser Verständnis für die Vielschichtigkeit der visuellen Wahrnehmung zu verbessern – ein Verständnis das hoffentlich eines Tages dazu führt, dass die Forschung über Fragen nach der Art der begrenzten mentalen Kapazität hinausgeht und sich stattdessen fundamentaleren Fragen widmet, zum Beispiel wie wir unsere mentale Kapazität bestmöglich nutzen können.

Schlüsselbegriffe: Multiple Object Tracking, Geschehenswahrnehmung, Kognition, szenenbasiert, objektbasiert.

Acknowledgements

I would like to express my gratitude to my advisor Jun. Prof. Dr. Markus Huff for the access to the laboratory and research facilities, and the scientific guidance at some critical points along the way. His strategic support forced me to work independently and guided my judgment in terms of distinguishing helpful feedback from biased opinions. I am also grateful to the second examiner Prof. Dr. Stephan Schwan who has been a calm and understanding consultant to me, scientifically and personally.

Disagreement and ceaseless critics helped me to find my own path. I have experienced the hard way that one should only criticize to drive improvement – not to justify rejection. For giving me the freedom to make my own mistakes and for truly changing my perception, I thank Dr. F. Papenmeier and Dr. H. Meyerhoff. I look forward to future work that falsifies my hypotheses.

I sincerely thank Annika Maurer (the encapsulated colleague and companion) for all the laughter, the talks, the winter sports, the dogs, the cocktails, the horse riding, the Vampys, the coffees ... I still don't know *where* the castle is.

My deepest gratitude goes to Steffen Hildebrandt (the lobster). I love you for your patience and cleverness. For tolerating my tantrums. For spontaneously programming everything, especially the google-scholar-laziness-transformer. For the volleyball and for all the joy you have brought back to my life. And for the rum. Don't forget the rum.

Papa, thank you for telling me what I'm capable of. For saying that money is no object. For believing that I have the talent to reach my goals. And for sharing your sense of humor, however bleak the situation. Mama, thank you for listening to my nagging. For your candid opinions and for recognizing when I need them to hear. For being a mother to me, even when you clearly didn't have the energy to do so. And for all the books. I owe you both.

I still have not mentioned everyone who brought me here:
Dominik Frommeyer and Lia Heubner (the assistants), Rain (the horse), Ben (the dog), Corinne McCafferty (the intern), Laura and Jochen (the physio therapists), Hanne Ermann (the Wendy pen pal), Rob den Otter (the tiger-and-gun guy), Anna Lipinski (the reason), Linda Kersten (the soap star), Phoebe, Monica, Rachel, Chandler, Joey, and Ross (the English teachers), Rosi (the secretary), Karl-Heinz (the parking card ruler) & technically, Aygo (the car).

Contents

Declaration of Authorship iii

Abstract v

Acknowledgements vii

Contents ix

List of Figures xi

List of Abbreviations xiii

Preface 1

Practical information 3
- General organization . . . 3
- Statistical Analyses . . . 4
- Dynamic Paradigms . . . 8

1 Introduction 13

I AUTOMATIC PERCEPTUAL PROCESSES 17

2 Study 1:
The development of dynamic representations 19
- 2.1 Perceiving static and dynamic objects . . . 20
- 2.2 Study 1:
  Scene over objects . . . 22

3 Study 2:
Cognitive-perceptual processes and events 33
- 3.1 Cognition and perception . . . 35
- 3.2 Event perception . . . 36
- 3.3 Study 2:
  Internal structures and conceptual representations of events . . 38
- 3.4 Explaining the results with the event horizon model . . . 54

II CONTROLLED PERCEPTUAL PROCESSES **57**

4 Study 3:
Attention, intention, and perception **59**
4.1 Visual Search and Features 62
4.2 Attentional Capture 63
4.3 Study 3:
Feature-based tasks in multiple object tracking 64

III Review **93**

5 General discussion and conclusion **95**
5.1 Summary of results 95
5.2 A multi-process approach to visual perception? 97
5.2.1 Limited resources 98
Limited object tracking 99
Limited event perception 100
5.2.2 Weight 100
Weighted object tracking 100
Weighted event perception 101
5.2.3 Processing 102
5.2.4 Competition 104
5.2.5 Bottom-up and top-down levels 106
What and where? 106
Unidirectional or iterative? 108
5.3 Theories of Attentional Selection 110
5.3.1 Cause Theories 110
5.3.2 Effect Theories 111
5.3.3 Attention in MOT: mixed cause-effect models? 113
Spatial, temporal, and shape: limits of tracking 113
5.3.4 Attention in Figure 7: a multi-layered process? 115
5.4 Some philosophical final remarks 116
5.4.1 Utilitarian perception 117
5.4.2 The human need for simplexity 119

Bibliography **121**

List of Figures

1 A typical MOT experiment. 9
2 Event sequence with visible ball contact. 10
3 Event sequence with ball contact cut out. 10
4 Second part of the clip: causal sequence. 11
5 Second part of the clip: non-causal sequence. 11

6 Conscious Awareness and Attention: 2 Models 61

7 Diagram of visual perception as a dual process 97
8 Attention as a cause and as an effect: an overview. 112

List of Abbreviations

ACS	**A**ttentional **C**ontrol **S**etting
AIC	**A**ikaike's **I**nformation **C**riterion
ANOVA	**AN**nalysis **O**f **VA**riances
DFG	**D**eutsche **F**orschungs-**G**emeinschaft
DMC	**D**ual **M**echanism of **C**ontrol
EST	**E**vent **S**egmentation **T**heory
FIFA	**F**édération **I**nternationale de **F**ootball **A**ssociation
FIT	**F**eature **I**ntegration **T**heory
MIT	**M**ultiple **I**dentitiy **T**racking
MOT	**M**ultiple **O**bject **T**racking
MOMIT	**MO**del of **M**ultiple **I**dentitiy **T**racking
NHST	**N**ull **H**ypothesis **S**ignificance **T**esting
RM	**R**epresentational **M**omentum

Dedicated to my left knee. That bastard.

Preface

Visual perception is the complex organization and interpretation of sensory information. Perception is not only a passive processing of signals, but it is influenced by attention, memory, expectation, and learning – thus, it operates on different processing levels. What we accept as our reality is shaped (and sometimes distorted) by multiple factors including the perceiver (e.g. motives or experience), the situation (e.g. dual-task constraints or setting), or the target (e.g. proximity or similarity). Limiting the enormous amount of available sensory input, for example with perceptual grouping or goal-directed attention, is necessary to function in a complex and constantly moving world. The human brain applies numerous ingenious tricks on such a regular basis that we do not even notice the constant lack of information.

Research presented in the following dissertation covers different perceptual levels of dynamic scenes. However, this work cannot even come near to cover the hybrid and multiform nature of perception in its entire complexity and will focus on three factors that influence the perception of dynamic scenes: how perception is shaped by factors within the perceiver (experience), within the environment (task setting), and within the targets (frameworks and features). Overall, this work provides further aspects on dynamic perception to understand how dynamic visual stimuli are processed and interpreted.

PART I of this thesis will be concerned with the interplay of perception and visual memory. It will be discussed how a global (scene-based) organization and experienced schemata influence the perception of *interrupted* dynamic stimuli on a basic developmental level (tracking study) and on a more complex level involving cognitive-perceptual expertise (event perception).

PART II will look at the more complex connection of perception and attention. It will be discussed how changes in the attentional level could lead to a more local (object-based) perceptual processing. Theoretical ideas will be supported by a comprehensive study that shows how top-down control (task setting) influences the allocation of attention in a dynamic environment (tracking study).

PART III reviews the results, presents them together with discussed literature in a diagram, and puts the individual components into context with existing approaches to visual perception. This work tried to shed light on *perceptual* dynamic processes, however, especially tracking is often described as an *attentive* process. Perception and attention are closely intertwined and hard to differentiate which is why attention is sometimes mentioned as an equivalent term without further explanation throughout the text. To reflect at least a little bit on the complexity of attentional processes and their immense research literature, I devote a section to attentional processes (mainly connected to tracking tasks) and describe theoretical accounts of attention in more detail.

Practical information

Here, I provide practical information on the organization of the thesis, my contribution to the three articles published, statistical analyses, and paradigms. A few words on the statistics are needed because next to "standard" analyses, the underlying data sets and hypotheses of some of the experiments presented here required less well-known approaches like generalized models and contrasts. I introduce the basic paradigms (multiple object tracking and event completion). The specific modifications made will be described in the corresponding articles/chapters.

General organization

Each chapter starts with a short summary of the most important points given in an info text box (as done above). The thesis is divided into three parts: Part 1 is concerned with perceptual processes based on global structuring and schemata, Part 2 addresses perceptual processes enhanced by top-down attention allocation, and Part 3 provides an overall summary of the results and discusses them with regard to their novelty and value for existing research and theories.

I published three articles as first author during the doctoral phase of 3 years:

1. Viewpoint Matters: Exploring the Involvement of Reference Frames in Multiple Object Tracking from a Developmental Perspective in *Cognitive Development*,

2. Seeing the Unseen? Illusory Causal Fillings in FIFA Referees, Players, and Novices in *Cognitive Research: Principles and Implications*, and

3. All Eyes On Relevance: Strategic Allocation Of Attention As A Result Of Feature- Based Task Demands In Multiple Object Tracking in *Attention, Perception, and Psychophysics*.

The three articles are included in their original published versions. Rights to use and reproduce them in this thesis were approved by each journal.

Concerning my contribution to each paper, for Paper (1), I was given a data set that was collected 8 years ago in a 3-D tracking task designed by Frank Papanmeier. Back then, Markus Huff, Frank Papenmeier, Kerstin Wolf, and Till Pfeiffer had the opportunity to collect tracking data from 120 children, however, their research question was not elaborated. I researched the tracking and developmental literature, configured the hypothesis, analyzed the data with intensive help from Markus Huff and Frank Papenmeier, and compiled the manuscript. For Paper (2), Frank Papenmeier, Annika Maurer, and Markus Huff programmed the event completion task and tested FIFA referees in Switzerland. I joined the project afterwards and suggested further groups of experts and non-experts to complete the experimental design. I was responsible for literature research, hypotheses, and writing. Markus Huff provided the result section, however, the reviewers of the journal requested a generalized design to meet the requirements of the proportional data, so I replaced his main analysis with a contrast analysis. I was solely responsible for Paper (3) including research idea, research design, analysis of results, and publication, however, many good ideas were generated in discussions with Markus Huff. The stimuli of the experiment were suggested in a DFG project but were supposed to measure effects of gaze cues indicating motion direction in tracking environments. After two experiments that did not show any effects of gaze cues on motion perception, the applicants of the DFG project aborted the experimental series. I kept working on my own ideas for the stimuli and developed the design further so that it met the needs of my research ideas (attentional control and feature-based tracking).

Statistical Analyses

In this section, I shortly explain the inferential statistics we applied to the different data sets. In each experiment, we measured task performance or object preferences (or both) as proportions, for example, participants may have tracked 65% of the target objects correctly. Such a limited dependent variable is a problem for ordinary linear regression because (a) a linear regression model can predict impossible values below 0 or above 1, and (b) the relationship between the dependent and independent variables is not linear but sigmoidal (comparable to a flattened S-shape). A linear approach is thus only justified when the data fall between the middle (linear) section of the curve (i.e., between .2 and .8). In that case, there is a linear relationship that a regression can catch nicely. While this was the case in some of our experiments, we had to use different approaches in others to account for the distribution of the underlying data.

Null Hypothesis Significance Testing

The traditional approach to make inferences has been to frame the scientific question in terms of two contrasting hypotheses. The null hypothesis (H_0) represents no difference between the population parameters of interest, the alternative hypothesis (H_a) represents either a unidirectional (one-sided) or bidirectional (two-sided) alternative. A test statistic is computed from the sample data and compared to the hypothesized null distribution ($\mu_1 = \mu_2$). If the test statistic differs, one assumes that the sample data is not consistent with the null hypothesis. The generally used but arbitrary cutoff level α is .05, defining the probability to make a type I error, that is, to reject the null hypothesis when it is actually true. Or stated differently, how often does one have to conduct the experiment to get a sample that shows the observed effect due to pure coincidence. If the answer is: not more than once in 20 samples (5%) we believe that there is a significant effect in the population, thus we reject the null hypothesis.

As mentioned above, the null hypothesis follows a specific distribution. If one wishes to compare two sets of data and determine whether their two means are equal, one assumes that the test statistics follows a Student's *t*-distribution under the null hypothesis. In the present work, these tests appear as independent t-tests, that is, comparing two means of two independent samples, and as paired t-tests, that is, comparing two means measured on the same statistical unit (e.g. tracking performance before and after a given manipulation).

The ANOVA (Analysis Of VAriance) generalizes the *t*-test to more than two groups, analyzing the differences among groups based on the variation among and between them. That is, H_0: $\mu_1 = \mu_2 = \mu_3 = ... = \mu_k$, where *k* is the number of groups. The underlying distribution of an ANOVA is the *F*-distribution. ANOVA is an omnibus test statistic, meaning that all means are tested against each other. A significant result can thus only tell us that at least two group means differ, not which ones. A possible solution are *post hoc* comparisons that look for patterns between subgroups, unspecified before the experiments. When specific hypotheses, for example about the strength and direction that a manipulation has on performance, are formulated, not all pairwise comparisons as given by an ANOVA are of interest (all the facts above were learned from: Moore, McCabe, & Craig, 2012). An approach that reduces the risk of Type I errors (i.e. rejecting the null hypothesis when it is true) is to predefine *a priori*/planned contrasts. To further control the Type I error rate at the level of the set of contrasts, it is possible to correct the α level for each contrast. In this work, the most pessimistic estimation has been chosen: the Bonferroni inequality. Next to the statistical advantages, a priori contrasts provide straightforward interpretations and allow for the accommodation of complex comparisons (e.g. testing the performance of group A against performance of group B, C, and D). More details on the advantages of a priori contrasts are discussed in Ruxton and Beauchamp (2008).

Mixed Effect models

Instead of testing the mean of different groups, it is possible to express the relationships in the data in terms of a function, for example, modeling tracking performance as a function of age: Proportion correct $\sim$ age + ε. Age would be a fixed effect while ε represents the non-specific error term, that is, the deviations from the model predictions we cannot control with the experiment's design. For example, personal experience of the participant that may increase tracking performance in the lab. In a mixed effect model, the error term that is normally an unsystematic part of a model, can be structure by adding a random effect, here, for "participant". The model can then account for idiosyncratic variation that arises from individual differences. In statistical terms, we assign different intercepts to each participant, telling the model that there are multiple responses, that is, individual variation, per subject. These error specifications can be done with different variables, leading to a close-to-fully specified error term. Results of the model are usually interpreted as the likelihood or the probability of receiving the collected data given the model, thus, the Likelihood Ratio Test is usually utilized to attain *p*-values. The (log-)likelihood of a model given the data is estimated as the Akaike's information criterion (AIC; Akaike, 1981). To obtain a *p*-value, one must compare two models with an ANOVA approach, one with the factor of interest and one without. For example, model 1: *Proportion correct* $\sim$ *age* + *training* compared to model 2: *Proportion correct* $\sim$ *age*. If the difference between the likelihood of the models is significant, we conclude that the fixed effect "training" is significant (see Bolker et al., 2009; Pinheiro, Bates, DebRoy, Sarkar, et al., 2007, for in-depth statistical details).

A generalized linear mixed model is an extension of the linear mixed model that allows the response variable to be of a different distribution, such as binary responses and proportions. The basic ideas are the same, however, in a generalized linear model, a link function (here: logistic) is specified that converts the expected value μ_i of the outcome variable, which is assumed to fall within the exponential family of distributions, to the linear predictor. In Experiment 1 we tested two generalized linear mixed effect models against each other with an ANOVA because the response variable was a proportion based on a binary variable. A similar approach has been applied in Experiment 2 and 3 in which such a generalized, binomial model was the basis for an analysis of a priori contrasts. More details and in-depth statistical understanding of generalized models can be found in McCulloch and Neuhaus (2001).

Bayes

The null hypothesis significance testing (NHST) described above relies on the calculation of a certain probability of a phenomenon occurring under specific conditions (i.e. the *p*-value). The whole idea behind NHST is to find enough evidence to favor the alternative hypothesis. But especially when an experiment is under-powered (e.g. due to a small sampe size as it was the case in Experiment 2), the *p*-value may not be meaningful evidence. Bayesian statisticians argue that NHST only allows the researcher to say: since my null hypothesis is wrong, then my null hypothesis is wrong. NHST does not allow for a confident statement that the effect found is due to the alternative hypothesis; a significant *p*-value only allows for a rejection of the null hypothesis. It might be the case that either the null nor the alternative are supported by the data. With Bayesian statistics it is possible to compare two models at the same time: Bayesian statistics can give an idea of how well the null *and* the alternative hypothesis explain a phenomenon in certain conditions. This is done with the Bayes Factor (BF), the a ratio of how likely one model will occur over the other model.

In Experiment 2, we reported null-findings for a study design with a small sample size that may have been under-powered and, by that, may have led to a type II error in the NHST statitics. We may have erroneously denied that there was an effect due to different levels of expertise between the groups. Using Bayesian statistics, we were able to show how much the null (no influence of experience on causal gap filling) and how much the alternative hypothesis (experiences changes perceptual causal gap filling) contributed to the study. We backed up the null hypothesis by a Bayes factor of 4.99, which is classified as a substantial evidence in favor of the null model (only main effects) against the alternative model (including an interaction of expertise and condition).

Dynamic Paradigms

Multiple Object Tracking

The Multiple Object Tracking Paradigm (MOT; Pylyshyn & Storm, 1988) is commonly used to investigate the architecture of visual cognition and dynamic information processing under sustained attention conditions. In an unmodified version of the paradigm, the distinctive pattern of successes and failures in tracking can be attributed to limits of specialized cognitive structures or attentional resources. By modifying the paradigm with occlusions or viewpoint changes, MOT can also be used to measure processes in memory and allows for insights into the mental representation of dynamic objects. In a typical MOT task, the participant is presented with a given number of identical objects, typically white circles on a black screen. A subset of the objects is marked as targets in the beginning of each trial. Participants are asked to keep track of these target objects so that they can identify them in the end of the trial. The identical objects move for a finite time on screen. When they stop, the participant identifies the tracked targets – or guesses them. Please refer to Figure 1 for a graphical presentation of the different sequences in a typical MOT experiment[1].

Event Perception

Implicit causal inferences distort the perception of, or memory for, events only seconds after viewing. Strickland and Keil (2011) used a simple paradigm to study how the human brain fills gaps in a dynamic event.

[1]In classical MOT studies participants are asked to fixate on a central location throughout the trial to avoid body or eye movements. This however was not practiced in our studies due to two reasons: (1) eye movements are believed to aid tracking performance via extrapolation (Luu & Howe, 2015) – however, this result is outperformed by literature indicating that extrapolation does not occur during tracking (e.g,. Atsma, Koning, & van Lier, 2012; Keane & Pylyshyn, 2006), and (2) unconstrained eye movements enhance the ecological validity of MOT performance, that is, the performance measured in the laboratory with unconstrained eye movements is a better predictor for human tracking performance in the real world.

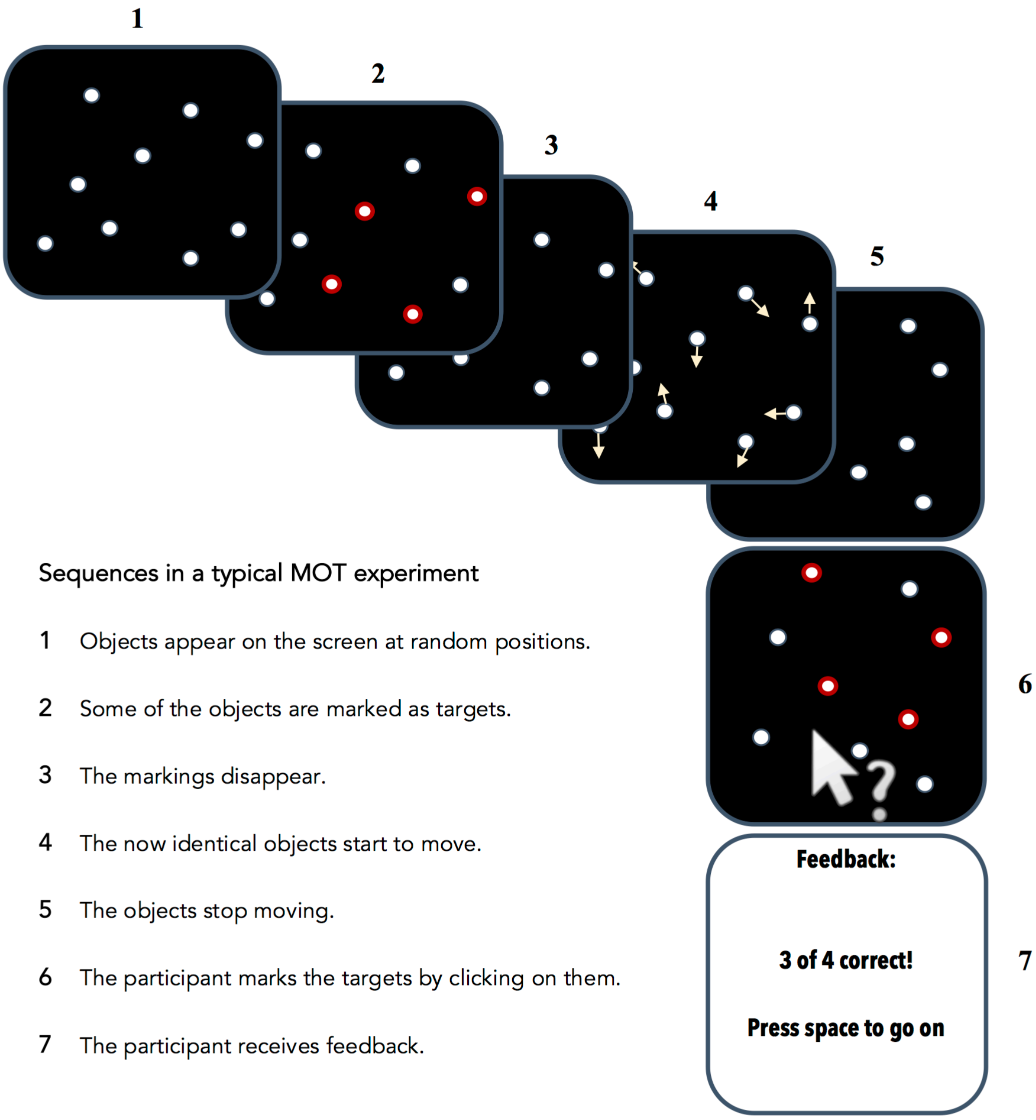

FIGURE 1: A typical MOT experiment.

Their participants saw a person running towards a ball. In the *complete* condition, participants saw how the ball was kicked and how it flew. In the *incomplete* condition, the moment of ball contact was omitted, either followed by a logical, causal scene (e.g. the ball flying) or an illogical, non-causal scene (e.g. an injured player or cheering fans). The authors asked the participants whether they have seen the contact picture or not. Results showed higher false-alarm rates for the incomplete-causal compared to the incomplete-non-causal conditions. Please refer to Study 2 in section 3.3, page 38, for further details of the replicated experiments. See Figures 2,3, 4, and 5 for exemplary options. The design allows manipulations of the ball contact moment (visible or cut out) and of the second sequence (causal or non-causal).

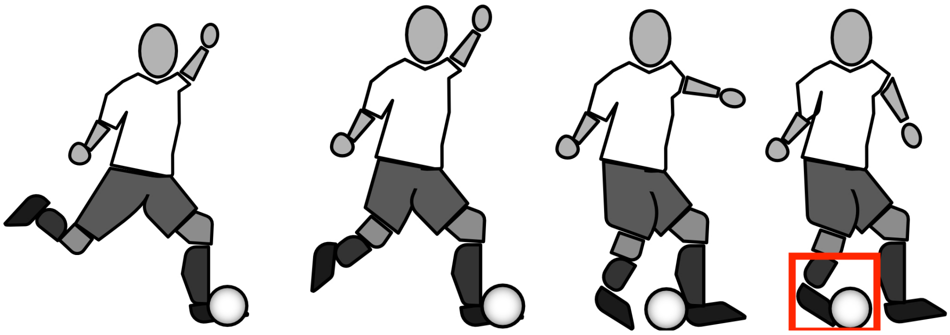

FIGURE 2: Event completion sequence with visible ball contact.

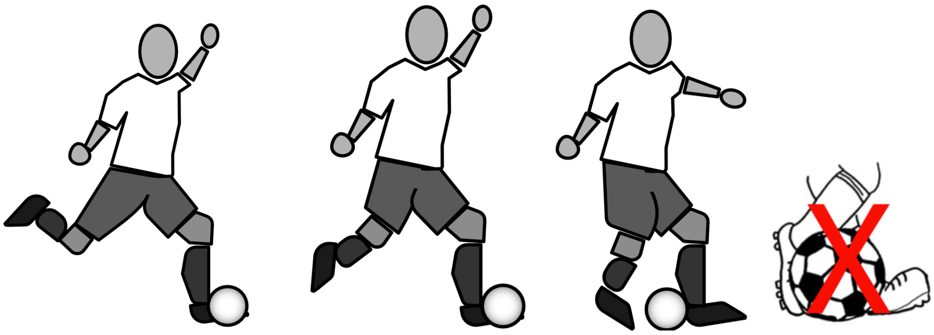

FIGURE 3: Event completion sequence with ball contact cut out.

Figures 2 and 3 show a complete and an incomplete first part of a presented sequence. Figure 4 and Figure 5 on the next page present the two possible second parts of a sequence: a logical continuation of the first part (e.g. ball bounces down the field, reaches teammate) or a non-causal continuation (e.g. cheering fans).

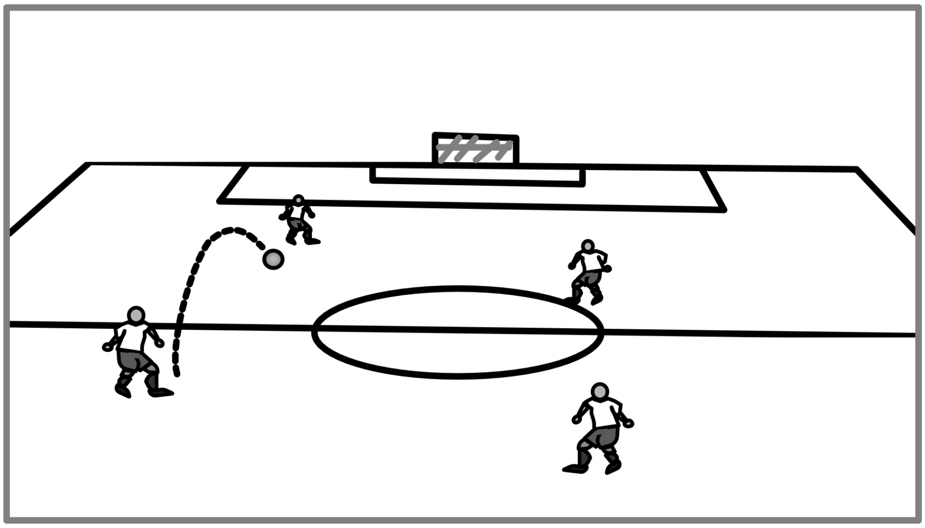

FIGURE 4: Second part of the clip: causal sequence.

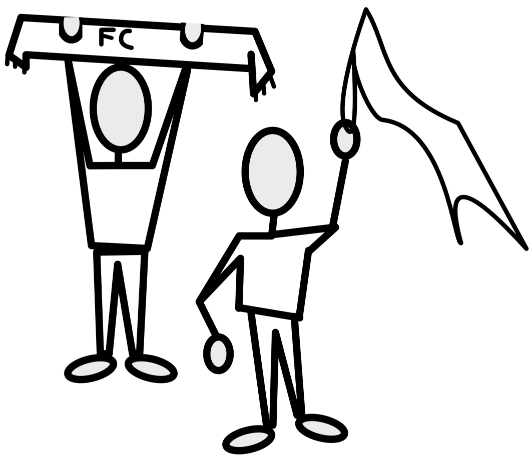

FIGURE 5: Second part of the clip: non-causal sequence.

Chapter 1

Introduction

> The study of perception has inspired research for more than a century but still has not lost its relevance or attractiveness to cognitive psychologists, neurologists, philosophers, and many others (see Hochberg, 1988, for a review). The purpose of this introductory chapter is to discuss the role of perceptual processes in the chapters that follow, trying to place them along earlier and current research knowledge with a focus on mental representations, memory, attention, and top-down influences.

Perception is defined as the awareness of environmental elements through physical sensations (e.g. color perception) but also as physical sensations that are interpreted in the light of experience (from Merriam Webster online dictionary: Perception, n.d.).

The first part of the definition describes the ecological approach of Gibson and Gibson (1955) who proposed that sensation *is* perception. Information is directly processed from the senses, that is, from the bottom end of the visual system. One famous example for such an automatic bottom-up perceptual processing comes from Navon (1977). He showed that observers have a global-over-local precedence, meaning that the perceptual system always starts to process a visual scene as a global structure before it eventually zooms in to integrate details for a clearer picture (discussed in more detail in 2.1). Gibson and Gibson (1955) strongly believed in a pure bottom-up perception and argued against the notion that past experience can influence present experience. According to them, such an enhanced processing would allow the observer to perceive more information about the environment than can actually be transmitted through the receptor system. Facilitated perception in a bottom-up fashion, however, may be possible when the observer organizes multiple stimuli into a (global) *group* (e.g. Navon, 1977; Wertheimer, 1938; Yantis, 1992).

Grouping may be simple in static environments but the level of difficulty increases when the objects are moving. A comparable concept to global and local processing in dynamic environments is processing moving objects based on their retinocentric or allocentric coordinates. Such reference frames can be constructed based on objects (object-based), the constellation of objects in the environment (scene-based), or the viewer itself (egocentric, fixation-based). A fundamental question to understand the nature of basic representational processes in a dynamic environment with *more than one* object concerns the type of information used that contributes to a successful reallocation after an occlusion or, similarly, a sudden viewpoint change. Study 1 presents the results of an experiment that explored the development of two possible representation frameworks used in object tracking: an object-based or scene-based perceptual organization.

The second part of the definition of perception considers the constructivist approach: because the sensory (bottom-up) information is incomplete, humans need to fill in gaps through interpretation. They do so by "predicting" the scene and compare those hypotheses against the incoming sensory information (Gregory, 1980). Thus, how we perceive the world around and the way we organize and interpret sensory input is a creation of reality in order to give meaning to our environment: after all, the perception of an object without recalling and linking it to stored information would be meaningless. And in fact, later research showed that neural processes responsible for visual perception and visual working memory are intertwined (Harrison & Tong, 2009; Serences, Ester, Vogel, & Awh, 2009) and interacting (e.g. Agam & Sekuler, 2007).

In order to fill in gaps produced by the limited perceptual system, the observer needs to draw on previous experiences. These are stored in memory, a general term for the recollection of what was earlier experienced or learned as well as a term for a mental information processing system that receives, modifies, stores, and retrieves informational stimuli (from Farlex Partner online dictionary: Memory, n.d.).

For dynamic and more complex scenes (that involve e.g. relations between stimuli or goal-related behavior), it has been proposed that individuals organize the continuous visual input into discrete actions (Zacks, Speer, Swallow, Braver, & Reynolds, 2007). For each observed object or event, a series of bottom-up "hypotheses" and top-down "queries" are answered by mapping stimuli to a memorized schema. This procedural system allows the observer to make a sufficient interpretation with just a few structural levels (schemata).

Prediction mechanisms rely on experience and may involve post-dictive reconstructions that are not always efficient. That is, the observer may compare the current event to a previous event and weigh its plausibility in order to fill in perceptual gaps with causal content that may not be correct in the specific situation. For example, in the German Bundesliga, the "phantom goal" became famous as an error in event perception. The side net of the goal had a hole through which the ball entered the goal without actually crossing the line. Because a hole in the net is such a rare event, especially in such a high league, fans, referees, and players were convinced that they had seen the ball crossing the goal line in a typical and correct way. Only the camera behind the goal was able to catch the truth.

Given that the mind fills in perceptual gaps with pictures based on knowledge and context: How much is perception facilitated by, or impaired with, shortcuts in more complex environments (such as a soccer match) that require semantic knowledge of actions and events, as well as considering the behavioral intentions of others? Study 2 explores how expertise influences the perception of dynamic events in soccer. Similar to the viewpoint changes in Study 1 with simple stimuli, the more complex visual stimuli presented to the participants in Study 2 were discontinuous. By deleting only seconds of a continuous action (kicking or throwing a ball), we measured the influence of experience on perceptual gap filling processes.

Another cognitive process closely connected to perception is attention. It is defined as the act or state of applying the mind to something and as a condition of readiness for selective focusing of consciousness and receptivity (from Merriam Webster online dictionary: Attention, n.d.). In Reisberg (2013), attention is described as the factor within the observer that decreases perceptual performance in a visual tracking task when the number of objects increases. Attention is the allocation of a limited resource that enables various operations, but it is also associated with intensity and clarity of perception, selection, and consciousness (see Hatfield, 1998).

Attention can be shaped by bottom-up and top-down mechanisms. Stimulus-driven (exogenous) attention is influenced by the object's properties and attracts attention in a non-volitional, preconscious manner. Goal-driven (endogenous) attention allows the observer to direct attention, for example, in accordance with a current task or based on experience. For the Navon (1977) example: perceiving a scene in a global manner is an automatic bottom-up process, while the perception of local properties (the zooming-in) requires additional perceptual resources that are enabled via attentional processes.

What happens to perception when attention acts as a control mechanism? Based on the signal enhancement approach, attention strengthens the representation of relevant stimuli (e.g. Carrasco, Ling, & Read, 2004; Carrasco, Williams, & Yeshurun, 2002). Another possibility is that attention reduces external noise to decrease the impact of distracting stimuli outside the focus (Lu & Dosher, 1998). However, as explored and addressed in Study 3, both approaches may not be mutually exclusive. Of special interest in Study 3 is the allocation of resources in a dynamic environment that requires sustained attention, such as tracking multiple moving objects simultaneously. I present research that shows how goal-driven attentional foci can affect sustained spatial attention, resulting in the conscious perception of features and cues during a tracking task. From a theoretical point of view, I discuss attention in tracking as a cause and as an effect: next to a resource that *causes* spatiotemporal limitations, the observed competition between location and feature representations matches the idea of attention as an effect, that is, a byproduct of intentional perceptual processing.

Part I

AUTOMATIC PERCEPTUAL PROCESSES

The first part of this thesis is concerned with automatic processes in dynamic perception. Chapter 2 sheds light on the developmental process of scene-based (global) processing in dynamic environments and demonstrates that the mind's short-cut to perceive multiple objects as a group can facilitate human perception effortlessly from early age on. Chapter 3 presents a study on automatic causal gap filling – a mechanism that may have evolved to accelerate visual perception but has its downsides in the observation of those kinds of events that do not comply with experienced schemata and causality.

Chapter 2

Study 1: The development of dynamic representations

> In this chapter I review literature that provides further background studies that were not mentioned in the original introduction (Paper (1)) of Study 1. We found that children from the age of 6 years on apply the same perceptual organization strategy in an interrupted tracking environment as adults: they use a global, scene-based approach. The global-over-local precedence has been observed in children as young as 3 to 4 months for static objects, but has never been tested with dynamic objects.

Human perception and their thought processes were researched early by Wilhelm Wundt. By opening the first institute for psychology in 1879 and by analyzing the human mind in a structured, experimental way, he marked the beginning of modern psychology. Wundt defined heuristics of perceptual organization, including the sensation of object connectivity when they are similar or close to each other. Identifying a global pattern by grouping objects is a fundamental part of organizing the overwhelming visual input the brain has to deal with. Study 1 explores whether such a scene-based organization is also applied for moving objects – and if so, does this grouping ability during object tracking depend on age?

2.1 Perceiving static and dynamic objects

In static environments, Navon (1977) presented convincing evidence that observers can adopt a local (perceiving objects as single objects, probably one by one) or global (perceiving objects as a group with a focus on the global pattern) perceptual organization strategy. Navon used letters that were composed of smaller letters (a big E composed of small Es in the congruent condition; a big E composed of small Ss in the incongruent condition) and asked participants to identify the big or the small letter as fast as possible. In the incongruent condition, participants were slower when they had to name the small letters, that is, global identity interfered with local identity - but this was not observed the other way around. Results were interpreted as showing that perception of coarse-grain global properties of an object or scene in an automatic manner comes first. He proposed that directing attention to the scene reveals fine-grain local properties but requires additional resources. Stoffer (1993) backed up the theoretical ideas of Navon (1977), proving that *attention* has to change between local and global representational levels – which is time-consuming. He compared reaction times of participants who were cued to different areas of the screen after having adopted a local or global focus. Results suggested that zooming to the local level took longer than perceiving the global level, again indicating that local processing, compared to global processing, takes on additional resources. In the same vein, Shiffrin and Czerwinski (1988) proposed that attention changes in spatial extent, and thus changes its processing state of global or local, depending on experimental conditions.

Findings of a similar study with 3 to 4 month-olds provided evidence that infants show the same order of perceptual precedence: global before local (Ghim & Eimas, 1988). Others wondered whether the configuration of objects into meaningful entities would last even if the objects appeared in a novel region. The general findings: (1) infants are sensitive to common region; as adults, they are able to use extrinsic factors (here: regions) to perceptually organize, and (2) these processing units can further be of an abstract nature, a result that points towards perceptual organization that follows context-based grouping principles (see Bhatt, Hayden, & Quinn, 2007; Bhatt & Quinn, 2011; Goldstone, 2003; Hayden, Bhatt, & Quinn, 2008).

Is such a perceptual organization similar for dynamic objects? Yantis (1992) used an object tracking task and found that adult participants who were presented with targets in a canonical formation (thus providing the possibility to form a virtual polygon) tracked the moving objects better than participants who tracked randomly configured target objects. The grouping advantage of the virtual polygon condition over the random condition was only observed in the initial stages of the experiment, indicating that participants in the random condition developed individual grouping strategies over time.

Local and global processing of static events seem, in a broad sense, comparable to the use of reference frames in dynamic scenes. These can be constructed based on objects (object-based), the constellation of objects in the environment (scene-based), or the viewer itself (egocentric, fixation-based). Liu et al. (2005) and Huff, Jahn, and Schwan (2009) wondered about the reference frame used in tracking environments. Liu et al. (2005) speculated that only allocentric, scene-based coordinates are used. Huff et al. (2009) tested tracking performance under viewpoint changes in a 3D-tracking environment and concluded that the visual system compensates for minor rotations but not for large rotations. Adults are known to use reference frames spontaneously or flexibly according to the situation or task (e.g. Iglói, Zaoui, Berthoz, & Rondi-Reig, 2009), but less is know about children's use of reference frames.

How do infants and children perceptually organize *dynamic* objects? Is the effective (global) scene-based perception of a *dynamic* scene learned or already existent in early childhood? In a comprehensive study by Hespos and Rochat (1997), infants of 4- to 8-months of age were tested in their ability to track and anticipate the orientation of an object after an invisible spatial transformation (from 60 to 150 degrees). Adults were found to rely on the *available* information in circumstances of fewer perceptual cues (Gibson, 1966), compared to younger infants who need rich cues from the environment. In Hespos and Rochat (1997), from 6 months of age, infants were able to track and anticipate the orientation outcome for a variety of situations, hence, approaching the flexibility of adults. These findings show that children can use environmental cues to anticipate the orientation outcome of a rotated object, but nothing is known about their perceptual organization of *dynamic multiple* objects in tracking environments.

2.2 Study 1: Scene over objects

In Study 1, we hypothesized that the global-over-local processing approach and the early-developed ability to track objects through brief occlusions are reflected in a tracking performance of children that is comparable to those of adults. Since a global perceptual organization is an essential building block of the architecture of the visual system, we did not expect that the decreased tracking performance would be explained by an interaction of age and viewpoint changes. Hence, we only hypothesized main effects for age and viewpoint changes on overall tracking performance.

The design by Huff et al. (2009) has been conceptually replicated to be used in Study 1, an experimental study of age differences in tracking objects in a 3D environment with viewpoint changes. The viewpoint changes after a short occlusion of all objects allowed us to test whether the participants adopted a global scene-based or a local object-based perceptual organization strategy. Participants were asked to track 3 targets out of 8 objects. In contrast to a mark-all tracking task in which the participant has to actively indicate all of the tracked objects, here, in the end of each trial, one object flashed. Children were asked to indicate whether the flashing object was among the targets they had tracked, or not. This resulted in a binomial independent variable (Correct/Incorrect) that we analyzed by comparing generalized mixed effect models assuming a logistic distribution, while specifying participants as the random effect. In order to test for the contribution of local object-based cues, we calculated a continuous variable (*object displacement*) that presented the distance on the screen between the location of the target right before and after the rotation. We compared the fit of the models with either both variables (viewpoint changes and object displacement) to models with only one of the variables. The logic: if a model with *both* variables does not differ significantly from a model with only *one* variable, the model with one variable explains tracking performance sufficiently. If they differed significantly, we would conclude that both variables explain tracking performance best. This logic resulted in two models for each variable that we called objects-over-scene and scene-over-objects.

Brockhoff, A., Papenmeier, F., Wolf, K., Pfeiffer, T., Jahn, G., & Huff, M. (2016). Viewpoint matters: exploring the involvement of reference frames in multiple object tracking from a developmental perspective. *Cognitive Development*, *37*, 1–8. doi:10.1016/j.cogdev.2015.10.004.

Cognitive Development 37 (2016) 1–8

Contents lists available at ScienceDirect

Cognitive Development

ELSEVIER

Viewpoint matters: Exploring the involvement of reference frames in multiple object tracking from a developmental perspective

Alisa Brockhoff[a,*], Frank Papenmeier[a], Kerstin Wolf[b], Till Pfeiffer[b], Georg Jahn[c], Markus Huff[a]

[a] *Department of Psychology, University of Tübingen, Germany*
[b] *University of Education Karlsruhe, Germany*
[c] *University of Lübeck, Germany*

A R T I C L E I N F O

Article history:
Received 28 July 2014
Received in revised form 13 October 2015
Accepted 17 October 2015

Keywords:
Multiple object tracking
Grouping
Allocentric
Retinocentric
Global-local
Development

A B S T R A C T

Earlier studies demonstrated that visual tracking of dynamic objects is supported by both scene-based and object-based reference frames, depending on the magnitude of scene displacement (Huff, Jahn, & Schwan, 2009; Liu et al., 2005). The current experiment tests if this pattern also applies to younger participants, i.e. school-age children, by comparing the effects of abrupt scene rotations on tracking performance of multiple dynamic objects in a 3D scene across five age groups (grade 1, 3, 5, 7 and adults). Scene rotations have two consequences: displacement of (1) the whole scene and, (2) individual objects. Tracking accuracy of 123 participants was measured across five age groups (grades 1, 3, 5, 7, and adults). Either 1 or 3 targets moved independently among a total of 8 identical objects for 5 s. The scene remained constant or was rotated by 10° or 20° after 3 s. Tracking performance of all participants was well above chance level (probability of 0.5) and an age-related increase in performance was observed. Contrasting the two factors revealed that scene rotation had a greater impact on performance than object displacement. Further, the effect of abrupt rotations was independent of age. These findings suggest that allocentric reference frames support attentive tracking across abrupt viewpoint changes and that scene-based tracking is already applied early in human development. Findings are discussed in light of new studies that link MOT to grouping processes (local and global). We propose that scene-based or allocentric processing abilities undergo a similar development as, or are connected to, grouping skills.

1. Introduction

The ability to keep track of multiple moving objects within a scene is critical to the successful negotiation of complex visual environments. For instance, crossing the street requires several attentional skills (Dunbar, Hill & Lewis, 2001; Tabibi & Pfeffer, 2007), but mainly to keep track of multiple moving objects. Although research has demonstrated a clear developmental trajectory in children's multiple object tracking with respect to the total number of objects that can be tracked, little is known

* Corresponding author at: University of Tübingen, Department of Psychology, Schleichstr. 4, D-72076 Tübingen, Germany.
E-mail address: alisa.brockhoff@uni-tuebingen.de (A. Brockhoff).

http://dx.doi.org/10.1016/j.cogdev.2015.10.004
0885-2014/

about the mechanisms responsible for these changes. The goal of this paper is to better understand these mechanisms by exploring how children's tracking abilities are affected by reference frames and accordingly, global and local processing of several moving objects.

Attention allocation in complex dynamic environments is experimentally tested using the Multipe object tracking paradigm (MOT; Pylyshyn & Storm, 1988). While watching several identical moving objects, observers are asked to maintain focus on a pre-assigned group of target objects. Developmental studies demonstrated that the number of objects children can track simultaneously increases markedly between 3 years of age and adulthood (Dye & Bavelier, 2010; O'Hearn, Hoffman, & Landau, 2010; Trick, Audet, & Dales, 2003; Trick, Hollinsworth, & Brodeur, 2009; Trick, Jaspers-Fayer, & Sethi, 2005). However, the majority of studies has focused on children over the age of 5, except for O'Hearn et al. (2010) who tested typically developing 3- and 4-year-olds and people with Williams Syndrome on multiple object tracking (MOT) and memory for static spatial location. Less is known about which maturing system is contributing to or is responsible for the observed improvement. O'Hearn et al. (2010) suggest that the developing visuospatial working memory (see also Klingberg, 2006) or attentional resolution (Wolf & Pfeiffer, 2014) play a role, whereas others see the number of tracked objects as reflecting the limited capacity of the maturing attentional system (Alvarez & Franconeri, 2007; Trick et al., 2005). MOT studies involving young individuals with disorders (e.g. Autism Spectrum Disorders (ASD), Williams Syndrome, Fragile X, syndrome, and, Turner's syndrome) who typically showed a lower mean of successfully tracked objects (Farzin, Rivera, & Whitney, 2010; Beaton et al., 2010; O'Hearn, Landau, & Hoffman, 2005; O'Hearn et al., 2010; O'Hearn, Hoffman, & Landau, 2011) suggest that MOT may even be utilized as a screening tool to measure a developmental delay in different developing groups during childhood.

In addition to a developmental trend in tracking ability, tracking may change qualitatively with age and experience, for example referencing objects in relation to other objects and the presented scene. MOT tasks presenting objects in 3-D scenes enable the exploration of visuospatial attention during tracking, with regard to the question of whether reference frames are used during MOT tasks, and if so, which ones. Humans use reference frames to transform scattered visual information input into one stable and detailed representation. When constructing a reference frame, it is possible to use objects, the environment, or the viewer as reference points (Howard, 1982). At present, there is little agreement on the form of reference used during tracking. Liu et al. (2005) have speculated that MOT mechanisms in 3-D scenes only rely on allocentric, scene-based coordinates. Thus, referencing objects in relation to each other would make tracking robust against abrupt viewpoint changes—that is, the displacement of objects by cuts from one camera perspective to another should not influence tracking performance. To test this speculation, Huff et al. (2009) introduced scene rotations of 10°, 20°, and 30° to a MOT task that was adapted to 3-D. The authors hypothesized that allocentric representations are only necessary for a successful relocation of objects in cases of large viewpoint changes. Minor rotations, however, change retinocentric coordinates only minimally. Because tracking performance was significantly decreased in 20° and 30° conditions, but not for 10° rotations, they concluded that the visual system relies on the retinocentric framework and compensates for small displacements when tracking multiple moving objects. The authors attempted to test for the involvement of retinocentric processes by using the screen coordinates of objects and calculated their displacement in conditions with rotation. The extent of object displacement was analyzed for trials with 30° viewpoint changes and two targets, finding no effect between large and small displacement for targets far and close to the center of rotation, respectively. Thus, not the displacement of an object but the rotation of the whole scene determined tracking performance.

Scene-based processing presupposes the ability to integrate local sensory information into one global whole. The ability to reference objects in relation to each other, perceiving them globally as one dynamic structure, overcomes the capacity limitations of selective attention (Yantis, 1992) and makes tracking robust against abrupt viewpoint changes (Jahn, Papenmeier, Meyerhoff, & Huff, 2012). In MOT, this ability was discussed in light of the target grouping approach by Yantis (1992) who argues that tracking benefits from grouping the single targets into one higher-order object, such as three targets into a triangle. Recent studies by Evers et al. (2014) and Van der Hallen et al. (2015) modified a MOT task to explore grouping interference in normally developing children and children with ASD (autism spectrum disorder). Both research teams picked up the approach by Scholl, Pylyshyn and Feldman (2001), namely that target objects in MOT are units of attentional selection. They paired each target with a distractor by displaying a connecting line between them and compared the tracking performance to trials in which objects were left ungrouped. If the performance in the grouped condition was significantly worse than in the ungrouped condition, one can assume that global processing, which means that objects are perceived as connected to each other, interfered with the tracking task. And in fact, global processing in MOT was measured based on a weaker tracking performance in the grouped condition, supporting the idea that grouping may shape sensory processing throughout the whole life span (Carey & Xu, 2001). Another recent study by O'Hearn, Franconeri, Wright, Minshew, & Luna (2013) compared adults, children, and matched participants with autism on a modified MOT task. The multiple objects were grouped in two ways, first by arranging them (i.e. by varying the space between them), to imply a grouped element and second, by letting them move together. This design allowed the authors to compare performance, for example, on target–target and target-distractor trials. They found children aged 9–12 years to show the same influence of motion-based, as well as element-based grouping as adults. Processing of the scene rather than single objects may evolve to enhance tracking performance, for example, when target objects are perceived as connected. Scene-based, global processing has been observed in various studies using dynamic stimuli and different samples of clinical and typically developing children but it has not been explored whether this ability is under development (i.e. whether this ability partially explains the developmental curve of tracking performance in children).

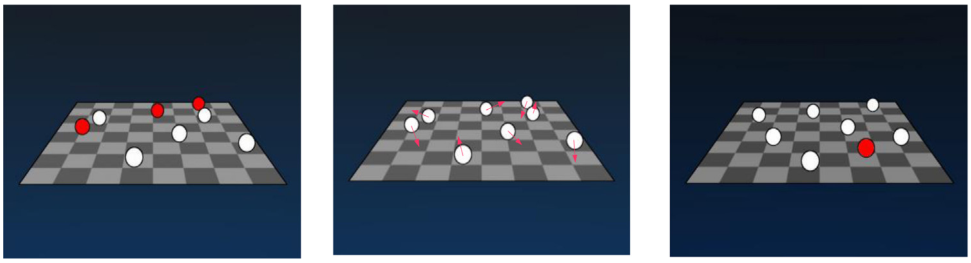

Fig. 1. Target designation, visual tracking and decision/marking phase.

Taken together, the current paper strives to answer the question whether tracking performance in children is determined only by object-based (local) processes or also by scene-based (global) processes. We assume that allocentric, scene-based processing and a global perception of multiple objects as a single grouped element are closely related, if not the same in a task in which the objects are displayed on a floor plane that is abruptly rotated in 3D (Jahn et al., 2012). The abrupt rotation of the floor plane has two consequences: the displacement of the individual objects and the rotation of the whole scene. If observers are tracking multiple objects in an object-based local manner, only the displacement of each individual object should determine tracking performance (lower tracking performance the further an object is displaced). However, if observers also utilize scene-based information such as grouping multiple objects into a higher-order object, the amount of scene change (angle of abrupt scene rotation) should explain tracking performance over and above the displacement of individual objects alone. Based on what we know about the effect of grouping in MOT studies (e.g. Van der Hallen et al., 2015), we expected to find scene-based effects across all age groups tested.

To shed light on how attentionally-demanding visuospatial skills mature with age, a more detailed analysis will focus on the strength of each impact on different age levels. To our knowledge, this is the first study that tested different age groups to see whether scene rotations impair tracking performance less with increasing age. Because adults are more experienced in global processing, an alternative finding would be that adults' tracking performance is even more impacted by scene rotations than children's performance.

2. Method

2.1. Participants

The sample consisted of 123 participants. Twenty-seven children were in grade 1 (age in years: $M = 6.45$, $SD = 0.57$), 31 in grade 3 ($M = 8.71$, $SD = 0.55$), 23 in grade 5 ($M = 11.51$, $SD = 0.49$) and 23 in grade 7 ($M = 13.34$, $SD = 0.50$). In sum: 104 children completed the experiment at the University of Education in Karlsruhe after written consent was obtained from parents. Seventeen adults participated (15 from the University of Education in Karlsruhe and 2 from the University of Tübingen). Three participants were excluded due to technical issues during the experimental session. The children received a small present for their participation and the adults were given monetary compensation.

2.2. Stimuli and procedure

Stimuli were presented using the Blender game engine (www.blender.org) and custom software written in Python. They were 8 small white, black-bordered 3-D spheres moving on a checkerboard floor plane (see Fig. 1). At the beginning of each trial, the 8 spheres were randomly positioned on screen. After 2 s, 1 or 3 spheres flashed red 4 times within 1.6 s and remained red for another 2 s. These spheres were the target objects. The target spheres turned white again and all spheres began to move at a constant speed of 3°/s for 5 s. The spheres moved in random directions and were allowed to touch or to overlap. Reaching the boundaries of the checkerboard, the spheres were reflected in a physically consistent manner (comparable to billiard balls), however, the spheres did not bounce off of each other.

The rotation of the scene was characterized by 3 conditions: the scene either remained constant, or it was rotated by 10°, or 20° (around the vertical axis through the center of the floor rectangle). It appeared abruptly (as if a camera cut in a movie displayed the same scene from another person's view) and did not influence the movement of the spheres. Rotations occurred after 3 s. Half of the rotations were directed to the left, the other half to the right. Fig. 2 illustrates a simplified rotation to the right and the two emerging variables we used for the analysis (see the next section for more details).

Following randomized movement, the spheres came to a stop and one turned red. The observer, then, had to indicate whether the marked object was part of the original target set seen at the beginning of the trial. Demo videos can be found here: https://homepages.uni-tuebingen.de/frank.papenmeier/mot-develop/.

Participants proceeded to the next trial by pressing the spacebar. Each participant performed 6 practice trials (2 levels of target number × 3 levels of scene rotation). The final experiment was comprised of 72 trials (2 target numbers (1 or 3) × 3 levels of scene rotation (0°, 10°, or 20°) × 12 repetitions). The order of conditions was randomized throughout the experiment. The participants had the option to take self-paced breaks between the trials. The within-subjects design allowed

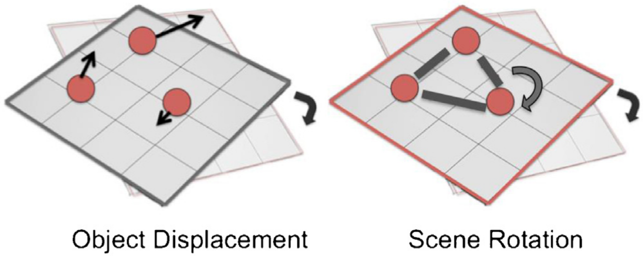

Fig. 2. A simplified visualization of the moment of rotation with different foci: perceiving objects as individually displaced (left) or as a group of objects rotated (right).

for controlling individual differences, reducing the associated error variance. The different grades (1, 3, 5, 7, and adults) served as a between-factor.

2.3. Data analyses

The application of a mixed-factor ANOVA on the proportion of correct answers provided a first impression of the data. In a further analysis, we fit logistic generalized mixed-models (*glmer*) due to the non-linear response variable that was expressed as a categorical variable with two levels (Yes/No). The aim was to quantify age-related and inter-individual differences that might influence the factors *scene rotation* and *object displacement* that, in turn, were thought to determine the variability in the number of correct responses. *Object displacement* was calculated as the distance on the screen between the location of the target probe right before and right after the rotation. We constructed *object displacement* as a continuous factor. Because a 0° scene rotation would automatically result in a displacement of 0 degrees of visual angle, only 10° and 20° trials were analyzed within *the glmer* analysis.

The *lme4* package for R (Bates, Maechler, Bolker & Walker, 2014) was used to perform the binomial logistic analysis. In a first step, using likelihood-ratio tests, the fit of the model with only object displacement as a fixed effect was compared to the fit of a model including both scene rotation and object displacement as fixed effects, in order to investigate whether scene rotation has a beneficial contribution. We called this the "Scene over Objects" logic: a significant result would lead to the acceptance of the model with both effects. Thus, both scene rotation and object displacement would contribute to successful tracking. A non-significant result would lead to the rejection of the model with both effects, indicating that object displacement explains the variance sufficiently. In a second step, we compared the fit of the model with only scene rotation to the fit of the model including both. We tested whether object displacement has an additional explanatory benefit. This "Objects over Scene" logic is similar to the "Scene over Objects" logic with the order of including the fixed effects into the models interchanged. Participants, specified as a random effect, allowed a separation of between-subjects (inter-individual) and within-subjects (responses to the variable of interest depending on individual differences) variance in the data.

3. Results

3.1. Repeated-measures analysis of variance

The mixed factor ANOVA has been executed with the between-subject factor *age* (grades 1, 3, 5, 7 and adults) and the within-subjects factors *number of targets* (1 or 3) and *level of scene rotation* (0°, 10°, and 20°) on the mean proportion of correctly identified targets. As predicted from previous research, statistically significant main effects of age, $F(4116) = 16.35$, $p < .001$, scene rotation $F(4116) = 23.87$, $p < .001$, and number of targets $F(4116) = 138.94$, $p < .001$ on mean proportion correct were observed. The effect of level of scene rotation was the same for all age groups, $F(8232) = 0.55$, $p = 0.82$, whereas age and number of targets as well as scene rotation and number of targets appeared to interact, $F(4116) = 4.58$, $p < .001$; $F(2232) = 4.79$, $p < .001$, respectively (see Fig. 2).

Based on established findings in the literature it is not surprising that an increased tracking load decreased performance in young participants. Further, the influence of the number of targets was higher in conditions with larger scene rotations. Finally, the interaction of age, scene rotation, and number of targets was not significant, $F(8232) = 0.73$, $p = 0.67$.

3.2. Generalized mixed-effects models for object displacement and scene rotation

The repeated measures ANOVA provided a first impression of the data, suggesting an exponential, developmental nature of tracking skills, with scene rotation having the same effect for all age groups. In a further analysis, we explored which reference frame (allocentric or retinocentric) participants used across the age groups. Therefore, we ran a separate analysis on all target probe trials with scene rotations of 10° and 20° and calculated the object displacement of the target probe. The direction of the data (see Fig. 3) mirrors the predicted, developmental trajectory nicely, but also points to a lower performance in trials with 20° rotation independent of object displacement. To better understand the different effects of the

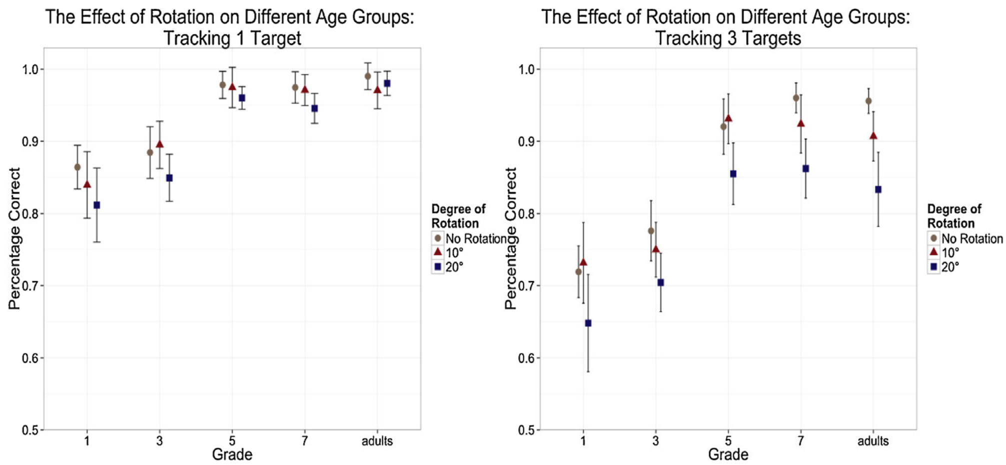

Fig. 3. Influence of age and number of targets on proportion correct, separated by scene rotation. Error bars indicate the 95% confidence interval of the mean.

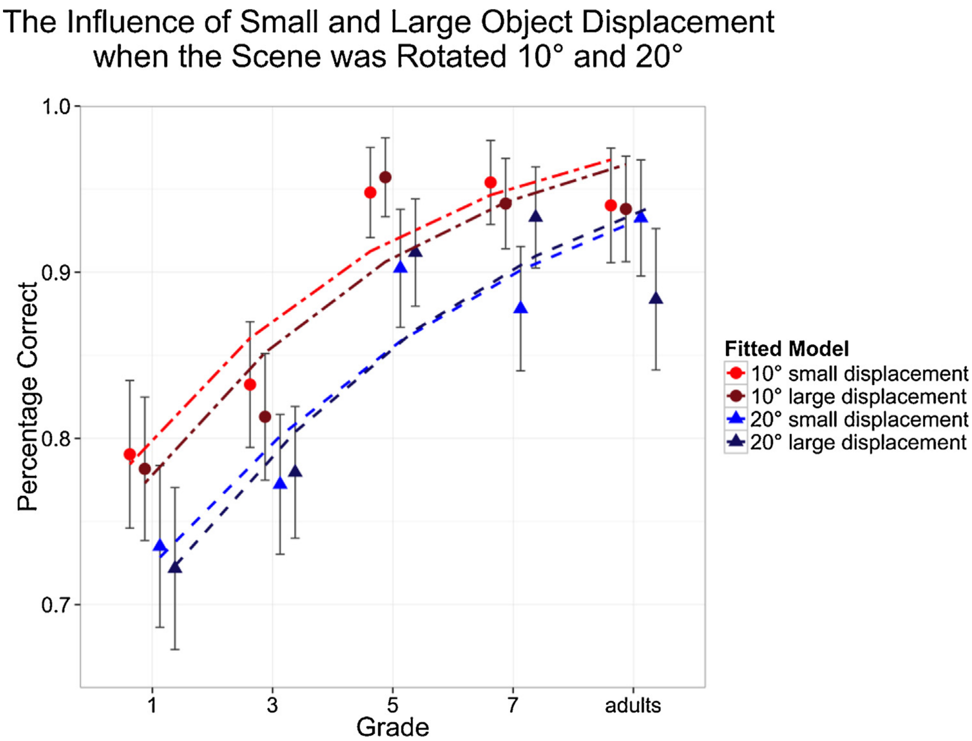

Fig. 4. Influence of object displacement and rotation on tracking performance over all target-trials. The red line and blue lines depict 10° and 20° rotation, respectively.

two predictors (scene rotation and object displacement), tracking performance (correctness of response) was subjected to a binomial *glmer* with the factors scene rotation (10° or 20°) and object displacement ($M = 24.53$ pixels, range [0.00; 125.40] pixels).

For better visualization only, we utilized the median as a cut-off score and displayed *object displacement* with two levels (small/large). Note that the factor was continuous in the analysis. Error bars represent 95% confidence intervals (Fig 4).

Based on previous research results by Huff et al. (2009), we surmised that object displacements are means to study the use of retinocentric reference frames. We assume that global and local processing (i.e. the extent of perceiving the objects as a group) will determine how much influence the rotation of the scene or the displacement of target objects has on tracking performance. To this end, we applied the "scene-over-objects" and the "objects-over-scene" logic for each grade and the adults separately. A side-by-side comparison of the results of each model by grade, as well as exact *p*-values, can be found in Table 1. Applying the scene-over-objects logic resulted in the acceptance of the model including scene rotation

Table 1
Generalized mixed-effects models for object displacement and scene rotation.

Grade	Scene-over-Objects χ^2	*p*	Objects-over-Scene χ^2	*p*
1	5.20	.023*	0.09	.783
3	0.49	.288	1.19	.503
5	4.77	.030*	0.38	.532
7	9.17	.003*	0.002	.959
Adults	8.97	.003*	1.89	.163
Reanalysis of Huff et al. (2009)	13.87	<.001*	0.36	.549

* Significant at the $p < 0.05$ level.

and object displacement. Thus, object displacement alone is not sufficient to explain the variance. In a second step, we applied the objects-over-scene logic. Results showed that the model with only scene rotation provided the best fit for grade 1 ($\chi^2(1) = 5.20$, $p = .023$), grade 5 ($\chi^2(1) = 4.77$, $p = .030$), grade 7 ($\chi^2(1) = 9.17$, $p = .003$) and adults ($\chi^2(1) = 13.87$, $p < .001$). Surprisingly, for grade 3, the scene-over-objects logic accepted the model with object displacement as fixed effect and the objects-over-scene logic accepted the model with scene rotation as fixed effect ($\chi^2(1) = 0.49$, $p = .288$). These findings stand in contrast to those of all other age groups. Thorough analysis neither revealed extreme outliers, nor an increased rate of guessing (calculated as proportion correct smaller than 0.5), or misunderstanding of the task (measured as participants pressing only one key, i.e. saying "Yes" or "No" constantly). Therefore, and based on the consistent picture of all other grades and adults demonstrated, we can only assume these effects to be due to random variation. Taken together, scene rotation was not only integrated into the tracking task but was a significant predictor of performance.

For the sake of completeness, we reanalyzed the original data published by Huff et al. (2009). The displacement range of an object and viewpoint change played a part in predicting performance in the original data set as well. Put in contrast by using the scene-over-objects/objects-over-scene logic, we found again, that the scene was superior over displacement of objects in predicting performance ($\chi^2(1) = 13.87$, $p < .001$), even when controlling for individual differences and varying speed—providing further support for the importance of the scene over an object during tracking.

4. Discussion

It has been the subject of considerable debate which representations visuospatial attention accesses during tracking processes (Huff et al., 2009; Seiffert, 2005; Liu et al., 2005). Huff et al. (2009) left allocentric coordinates intact and still found an impaired tracking performance. Although this points towards a retinocentric, viewer-based representation of dynamic scenes, other interpretations are possible. The focus of the current study was to replicate preceding results of studies concerning the usage of reference frames during tracking—concentrating in particular on the development of tracking abilities in younger participants in 3-D environments. The results presented here indicate that the impact of rotations is similar across all age groups tested—independent of the range of object displacements. These findings are in line with recent studies that linked global processing of objects to MOT as well. Evers et al. (2014) suggested a reduced global processing bias in participants with ASD compared to normally developing children (see also O'Hearn et al., 2013). Grouping of targets and distractors (paired by a connecting line) resulted in an interference of the tracking task, suggesting that forming object-based connections (grouping) is a tracking approach observable already in young children. If children and adults track multiple objects by utilizing scene-based processes such as grouping target objects to a higher-order object (e.g. a triangle), tracking across abrupt scene rotations should not only be influenced by the displacement of individual objects caused by the rotations but also by the extent of the scene rotations as such.

4.1. Does sole rotation of the whole scene or the extent of the displacement of a target object influence tracking performance?

The first part of the analysis addressed the question of which factors affect tracking performance. We found our results to replicate established findings in MOT research with main effects for number of targets, age, object displacement, and scene rotation. But which factor produces more tracking errors? For further examination, we introduced a new way of modeling tracking performance in relation to object displacement and scene rotation. Interestingly, scene rotation was a better predictor of tracking performance than object displacement (objects-over-scene logic). The finding that scene rotation influenced tracking performance more than object displacement leads us to speculate that humans not only rely on retinocentric changes, but also make use of scene-based, allocentric reference frames, especially during tracking tasks in 3-D environments.

4.2. Are viewer-based effects also observable in younger participants—and given that both rotation and displacement reveal an impact, which one is stronger?

All groups showed a similar pattern of performance drop due to object displacement and scene rotation, which may be continuous throughout development. Assuming that tracking processes are retinocentric in nature, larger object displacement should result in a higher number of errors. The current results only partially support this assumption. Scene rotation was a better predictor of performance than object displacements, suggesting a strong involvement of allocentric processes during tracking in 3-D environments. These results rather coincide with speculations by Liu et al. (2005) who surmised that a critical input needed for tracking multiple objects is a stable environment, not the objects themselves. The superior influence of scene rotations was present in almost all grades and conditions tested, leading to the conclusion that scene-based, not viewer-based effects are observable already in children of age 6.

4.1. Future research

Future research may determine (a) the extent of developmental effects on the use of reference frames in tracking and (b), the connection of reference frames and grouping. A field to apply this knowledge could be the design of perceptual-cognitive training games in dynamic, virtual reality environments, to help improve tracking-speed and tracking-capacity in order to reduce the risk of road accidents for children. It has been shown that age-related effects in tracking can be reduced by training for older participants (e.g. Legault, Allard & Faubert, 2013). Intelligently designed dynamic environments may be used to teach tracking and visuospatial skills.

Logan (1995) included linguistics in tasks of spatial representations and proposed linguistic cues to play a role in directing attention. Trick et al. (2003) suggested a relationship between tracking and enumeration. This could explain the considerably large difference found between primary school children and grades 5 and 7: language and enumeration skills, as well as tracking, all undergo huge improvement between childhood and young adulthood. Further research exploring these skill combinations in depth will be interesting with regard to the development of underlying cognitive skills and reference frames needed in tracking tasks.

But not only the application of reference frames should be specified in more detail. Generally, children are assumed to be less efficient in their deployment of attention (e.g. Plude, Enns, & Brodeur, 1994; Trick & Enns, 1998) whereas adults can make use of more than one reference frame simultaneously (Carlson-Radvansky & Jiang, 1998; Stein, 1992). Thus, the gradual improvement of sustained attention and longer periods of extended concentration may play a role, as well as visual working memory and attentional selection of items.

The most reasonable approach based on current literature and our recent results would be to assume that allocentric and retinocentric frames (or global and local processes) are at work simultaneously. It is possible that people develop a strategy to track objects in a global or local manner, possibly by activating different reference frames or using processing strategies that are applied depending on the situation. Whether multiple intrinsic representations are accessed in a top down manner, as well as when and if a strategy develops and why, this has yet to be determined.

5. Conclusion

By exploring developmental processes, we were able to show that the magnitude of age-related changes is consistent over different ages and depends on the stimulus complexity (number of targets, range of displacement, and extent of rotation). Concerning the hypotheses, the results indicated that (1) object-based effects are observable from early age on, but are less pronounced than scene-based effects, (2) scene rotation and displacement of targeted objects have an influence on tracking performance, and finally, we showed that (3) scene rotation had a stronger impact than object displacement, leading us to assume that tracking across abrupt viewpoint changes in 3-D environments relies more on allocentric than on retionocentric processes.

The findings of the presented experiment offer numerous theoretical and practical implications. Within the context of perceptual developmental theories on grouping processes, our measure of children's performance in situations of scene or object shifts brings us closer to understanding how attentionally-demanding visual-spatial skills mature with age. The limited tracking ability of children in grade 1, 3, and 5, relative to adults supports existing, findings suggesting that brain areas responsible for MOT develop and only become maximally efficient later in life (see Ryokai et al., 2013; Dye & Bavelier, 2010; Trick et al., 2003). The similar influence of scene rotation on all groups suggests that grouping (i.e. processing the presented objects in a global manner), is already present in children as young as 6 years. By documenting a specific window of time of the typical developmental trajectory of the use of reference frames during tracking, we can learn more about how children experience and structure their complex environments. Our results, and maybe even our version of the MOT task that was designed in a game-like manner, may guide parents, teachers, clinicians, and researchers in identifying developmental delays in scene-based motion processing.

Author contributions

Idea and research design: MH, GJ, TP, AB.
Operationalization: MH, FP, GJ, KW.
Data analysis: AB, FP, MH.
Writing: AB drafted the manuscript and all authors provided critical revisions.

Acknowledgement

This research was supported by German Research Foundation (DFG) Grants HU 1510/4-1, HU 1510/4-3, JA 1761/5-1, and JA 1761/5-3.

References

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track?: evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, *7*(13), 14.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: linear mixed-effects models using Eigen and S4. R package version 1. 1–6. http://CRANRproject.org/package=lme4.

Beaton, E. A., Stoddard, J., Lai, S., Lackey, J., Shi, J., Ross, J. L., et al. (2010). Atypical functional brain activation during a multiple object tracking task in girls with Turner syndrome: neurocorrelates of reduced spatiotemporal resolution. *Journal Information*, *115*(2)

Carey, S., & Xu, F. (2001). Infants' knowledge of objects: beyond object files and object tracking. *Cognition*, *80*(1), 179–213.

Carlson-Radvansky, L. A., & Jiang, Y. (1998). Inhibition accompanies reference-frame selection. *Psychological Science*, *9*(5), 386–391.

Dunbar, G., Hill, R., & Lewis, V. (2001). Children's attentional skills and road behavior. *Journal of Experimental Psychology: Applied*, *7*(3), 227.

Dye, M. W., & Bavelier, D. (2010). Differential development of visual attention skills in school-age children. *Vision Research*, *50*(4), 452–459.

Evers, K., de-Wit, L., Van der Hallen, R., Haesen, B., Steyaert, J., Noens, I., et al. (2014). Brief report: reduced grouping interference in children with ASD: evidence from a multiple object tracking task. *Journal of Autism and Developmental Disorders*, *44*(7), 1779–1787.

Farzin, F., Rivera, S. M., & Whitney, D. (2010). Spatial resolution of conscious visual perception in infants. *Psychological Science*.

Howard, I. P. (1982). *Human visual orientation*. Chichester: Wiley.

Huff, M., Jahn, G., & Schwan, S. (2009). Tracking multiple objects across abrupt. viewpoint changes. *Visual Cognition*, *17*(3), 297–306.

Jahn, G., Papenmeier, F., Meyerhoff, H. S., & Huff, M. (2012). Spatial reference in multiple object tracking. *Experimental Psychology*, *59*(3), 163–173. http://dx.doi.org/10.1027/1618-3169/a000139

Klingberg, T. (2006). Development of a superior frontal–intraparietal network for visuo-spatial working memory. *Neuropsychologia*, *44*(11), 2171–2177.

Legault, I., Allard, R., & Faubert, J. (2013). Healthy older observers show equivalent. perceptual-cognitive training benefits to young adults for multiple object tracking. *Frontiers in Psycholy*, *4*(323) http://dx.doi.org/10.3389/fpsyg.2013.00323

Liu, G., Austen, E. L., Booth, K. S., Fisher, B. D., Argue, R., Rempel, M. I., et al. (2005). Multiple-object tracking is based on scene, not retinal, coordinates. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(2), 235.

Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology*, *28*(2), 103–174.

O'Hearn, K., Landau, B., & Hoffman, J. E. (2005). Multiple object tracking in people with Williams syndrome and in normally developing children. *Psychological Science*, *16*(11), 905–912.

O'Hearn, K., Hoffman, J. E., & Landau, B. (2010). Developmental profiles for multiple object tracking and spatial memory: typically developing preschoolers and people with Williams syndrome. *Developmental Science*, *13*(3), 430–440.

O'Hearn, K., Hoffman, J. E., & Landau, B. (2011). Small subitizing range in people with Williams syndrome. *Visual Cognition*, *19*(3), 289–312.

O'Hearn, K., Franconeri, S., Wright, C., Minshew, N., & Luna, B. (2013). The development of individuation in autism. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 494–509. http://dx.doi.org/10.1037/a0029400

Plude, D. J., Enns, J. T., & Brodeur, D. (1994). The development of selective attention: a life-span overview. *Acta Psychologica*, *86*(2), 227–272.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 179–197.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, *80*(1), 159–177.

Seiffert, A. E. (2005). Attentional tracking across display translations. *Journal of Vision*, *5*(8), 643.

Stein, J. F. (1992). The representation of egocentric space in the posterior parietal cortex. *Behavioral and Brain Sciences*, *15*(04), 691–700.

Tabibi, Z., & Pfeffer, K. (2007). Finding a safe place to cross the road: the effect of distractors and the role of attention in children's identification of safe and dangerous road-crossing sites. *Infant and Child Development*, *16*(2), 193–206.

Trick, L. M., Audet, D., & Dales, L. (2003). Age differences in enumerating things that move: implications for the development of multiple-object tracking. *Memory & Cognition*, *31*(8), 1229–1237.

Trick, L. M., & Enns, J. T. (1998). Lifespan changes in attention: the visual search task. *Cognitive Development*, *13*(3), 369–386.

Trick, L. M., Jaspers-Fayer, F., & Sethi, N. (2005). Multiple-object tracking in children: the catch the spies task. *Cognitive Development*, *20*(3), 373–387.

Trick, L. M., Hollinsworth, H., & Brodeur, D. A. (2009). Multiple-object tracking across the lifespan: do different factors contribute to diminished performance in different age groups. *Computation, Cognition, and Pylyshyn*, 79–99.

Van der Hallen, R., Evers, K., de-Wit, L., Steyaert, J., Noens, I., & Wagemans, J. (2015). Multiple Object Tracking Reveals Object-Based Grouping Interference in Children with ASD. *J. Autism Dev. Disord.*, 1–9.

Wolf, K., & Pfeiffer, T. (2014). The development of attentional resolution. *Cognitive Development*, *29*, 62–80.

Yantis, S. (1992). Multielement visual tracking: attention and perceptual organization. *Cognitive Psychology*, *24*(3), 295–340.

Chapter 3

Study 2: Cognitive-perceptual processes and events

> This chapter covers the more complex perceptual organization of events. Here I present two experiments that explore an event completion effect elicited with soccer video clips. We tested three groups with different levels of soccer expertise on the event completion effect (introduced in the practical information in the beginning of the thesis). The effect was observed in all three groups independent of their level of expertise, that is, we observed causal gap filling independent of a priori experience based on the experts' extensive exposure to soccer videos and real-life soccer matches. A possible interpretation: the influence of higher cognitive processes in the perception of simple action events may be overruled by automatic processes that make sense of the world and map actions and features of an event onto a known structure (schema).

The results found in Study 1 can be explained with a global representational level, which would fit the idea of tracking as an automatic and preattentive mechanism. Pylyshyn and Storm (1988) argued that tracking of multiple objects requires more than one independent focus, thus, it can only operate preattentively (see also Scholl, 2009). Pylyshyn (1994) and Sears and Pylyshyn (2000) see the *effort*, that is, the involvement of attentive processes, as the observer's need to periodically refresh the representation of target objects (i.e. the targets' indexes as discussed later) as well as to rescue lost objects during motion. Accordingly, the perceptual grouping effects found in Study 1 were discussed as due to global representations in visuospatial *attention*.

However, the representation of objects based on their location in the observed scene may just as well be stored in memory. A design with sudden viewpoint changes (as applied in Study 1) may have activated an offline flexible memory that stored the whole scene, including the target (and maybe the distractor) objects in relation to each other. The offline mechanism allows for brief gaps during a stimulus observation. By that, the visual system can pick up tracking after the viewpoint change or gap by using the constellation of the scene stored in the offline memory. This may be facilitated by the visual system further storing a possible anticipated constellation of the objects, an anticipatory trace of the scene.

Some have argued that spatial working memory is the same as spatial attention (Awh & Jonides, 1998; Awh, Jonides, & Reuter-Lorenz, 1998). Horowitz, Birnkrant, Fencsik, Tran, and Wolfe (2006) proposed that the disappearance of objects may invoke an *offline flexible memory store* that retains the current state of the system *as a whole* and enables the reallocation of objects. In addition, tracking studies found that people could track at least four objects (e.g. Pylyshyn & Storm, 1988) which suggests an underlying capacity limit that matches mental storage limitations found in memory studies (see Cowan, 2001, for a re-evaluation of numerous competing views on "the magical number 4" in short-term memory). Some further assume that visual working memory and visual attention may rely upon a common pool of mental resource (Gazzaley & Nobre, 2012). The authors refer to "inward-directed" attention: selecting a target from visual working memory and from the environment may be the same as internally and externally directed selective attention (based on Kuo, Rao, Lepsien, & Nobre, 2009). Additionally, Howard and Holcombe (2008) observed a progressive decline in the precision of representations of tracked objects that was similar to, if not consistent with, the results of visual short-term memory tasks, in which Wilken and Ma (2004) observed an increase in errors made in reporting the features of multiple objects with increases in set size. Howard and Holcombe (2008) assumed that this similarity could be explained as due to attentional capacity that acts as a processing bottleneck or as a "manager" of information input that may reach visual short-term memory.

Further support for the idea that attentive visual tracking is at least connected to visual memory processes, if not the same, comes from findings by Huff, Meyerhoff, Papenmeier, and Jahn (2010) and Meyerhoff, Huff, Papenmeier, Jahn, and Schwan (2011) who suggested that abrupt viewpoint changes elicit spontaneous updating of visual properties, implicating an effect of visual inconsistency on the encoding of properties and, in general, on the perception of continuous events. In a similar vein, Baker and Levin (2015) proposed the relational trigger approach – however, they tested the perception of dynamic events. In five studies the authors showed that these triggers served as a heuristic mechanism that updates and compares event properties.

3.1 Cognition and perception

A similar automatic process to offline memory or relational triggers may be operating on a daily basis in order to make sense of the world around us; dealing with incomplete sensory information received from the visual system by processing or flexibly storing global aspects of a scene. The world around us, however, is not a display of identical moving stimuli. Objects may have typical characteristics (e.g. motion profiles), behave in a related or unrelated manner, and most living things act goal-oriented. A theory of mind is needed to predict the mental state, and accordingly the behavior, of others (Baron-Cohen, Leslie, & Frith, 1985; Premack & Woodruff, 1978). In addition, the observer needs to consider the position of his own eyes, head, or whether his body is moving. All these scene elements and factors require the observer to not only perceptually but also to *cognitively* process a real world event. A process described as arranging stimuli received from the environment and organizing experiences in the mind (Schultz & Schultz, 2015), as a categorical perception (Zacks, 2008), and as a continuous prediction of what will happen next – based on the previous influx of visual information and on prior knowledge – in order to facilitate perception (Zacks et al., 2007).

The interplay of perception and cognition has been extensively studied. It all started with Hermann von Helmholtz (1821–1894) who proposed that visual perceptions are *unconscious inferences* (Helmholtz, 1866). According to him, perception requires inferences based on knowledge to make sense of the fragmented and irrelevant data received from sensory signals. Gregory (1997) hypothesized that such an inferential perception (or perceptual interpretation: Gregory, 1980) bears the risk for cognitive illusions, a kind of illusion that, in contrast to physical and physiological illusions, can be explained by low-level physiological mechanisms, and arises from an interplay of perception and knowledge. Gregory (1997) proposed that cognitive visual illusions are due to advanced development of our visual system (neural learning). Learned simplified models help to speed up the interpretation processes, however, illusions arise when the unconscious analysis of the scene based on learned models conflicts *reasoned* considerations.

That learned models can even influence perception *independent* of displayed ambiguity (conflict) in the scene has been shown with the representational momentum (RM) effect. Freyd and Finke (1984) presented participants images of a rectangular that rotated in a given direction. When they asked them to indicate the final position of the stimulus, the participants showed a tendency to choose a position further rotated in the presented direction. The RM also interacts with prior knowledge (stored in memory) of fundamental Newtonian physics (e.g. movement of a linear path tested with infants: von Hofsten, Vishton, Spelke, Feng, & Rosander, 1998) or of the stimulus' typical motion (e.g. object-specific effects tested with typical and atypical pictures of rockets: Vinson & Reed, 2002). Vinson and Reed (2002) revealed in three experiments that context and prototypical appearance of an object assert a powerful influence on RM memory shifts: if an ambiguous stimulus was labeled "rocket", the participants showed a larger RM shift than when it was labeled "building". In addition, it was more important that the stimulus looked familiar (i.e. displayed a common prototype) than whether it pointed in a given direction. The authors suggest that observers attend, recognize, and record the identity of the stimulus object. Such a "recognition elicits conceptual knowledge from long-term memory that relates to that object's expected motion in the context of the displayed situations" (Vinson & Reed, 2002, , (p.14)).

3.2 Event perception

Jackendoff (1991) proposed that the semantics of events and their parts can be accounted for by models developed to represent objects and their parts. The basic ideas of Helmholtz (1866) and Gregory (1997) (simplified models that evolved from neural learning influence perception) are therefore also found in studies on event perception. Equivalently, according to Bartlett (1932), a "schema" is an active organization of past experience that can have a tremendous effect on memory (see also Verfaillie & d'Ydewalle, 1991), especially in early stages of processing (Biederman, Mezzanotte, & Rabinowitz, 1982; Friedman, 1979). Further, the term "situation model" is frequently used: it describes mental representations that capture the relations between components of a scene

(Johnson-Laird, 1983). Most of what we know about the perception of events comes from psycholinguistic research[1], investigating how people process written or narrative texts.

A technique for studying the perceived structure of an event involves asking observers to watch movies of everyday activities and to organize (”segment”) them into meaningful units of activity (Newtson & Engquist, 1973). By using segmentation techniques, Zacks et al. (2007) showed that humans perceive a continuous and dynamic event as a sequence of (sub)events. For each event, we identify discrete parts and attempt to find their relations. According to Zacks and Tversky (2001), this fragmented event processing reflects a deeper psychological reality, suggesting perception and knowledge to play an influencing role in how events are encoded and represented in memory. This is the core principle of the theory of event segmentation (EST; Zacks et al., 2007): it states that the processing of events forms sensory representations that are used to make predictions and to plan actions. EST is based on various theories on perception, neurophysiology and language processing (Biber, van Dijk, & Kintsch, 1986; Carpenter, Grossberg, & Arbib, 2003; Fuster, 1991; Neisser, 1967). Zacks et al. (2007) proposed further that processing is biased by event models; sets of working memory representations that integrate visual, auditory, and other sensory modalities. An observer will recognize a coherent movement pattern that is consistent with previous observations, realize that there is a goal to the movements and use this realization to predict future actions until the task is near completion and prediction becomes more uncertain. When the coherent movement pattern ceases, the statistical dependency is broken, and the inference of the goal no longer has predictive value. At this point, perceptual prediction declines, leading to the activation of the gating mechanism and to an updating of the event model.

Event schemata are constructed of semantic memory representations, capturing details of previously experienced events and storing specific information about features, objects and agents of the event. Event schemata affect the current content of the event models with top-down processes, expanding their effective capacity by assembling predictive information about the future relevance of certain features of events. This is similar to the idea of *Conceptual Semantics* proposed by Jackendoff (1991). According to his approach, there is a

[1] Note that, in general, any findings obtained from psycholinguistic studies were not applied primarily to specify theories of speech and text comprehension but rather to understand the construction of mental models in human cognition (see Johnson-Laird, 1980, 1983, 1994).

conceptual structure, a syntax, that guides the encoding of human understanding of the world. Similar to an event schema in EST, the formation of a conceptual structure is influenced by rules of inference, pragmatics, and heuristics (e.g. cause-effect relations). These rules are formed and applied individually, as Jackendoff (1991) insists, and to truly understand how a sentence/perception is connected to the world, one must consider *I-Semantics* (the internal structure) of the reader/observer. He states further that schematization is not an inherent property of reality, but it is part of the psychological organization that construes human reality. Therefore, *I-semantics* should cover psychological, social, and biological contexts.

3.3 Study 2: Internal structures and conceptual representations of events

The interaction of internal and external structures during the observation of an event has not been researched yet. In Paper (2) we wondered about the role of experience and expertise in the perception of soccer video material. In experiment 1 (of 2) published[2], we asked FIFA referees, soccer players, and novices to watch two-part soccer clips that were manipulated. The clips either made sense (the second part was connected causally to the first part of the clip) or not (second part did not match the first part). What the participants did not know: the moment in which the ball had contact with a player (i.e. the moment when the ball was thrown in or kicked off) was cut. After each clip the participants received pictures (either presented in the clip or not) and were asked to indicate which ones they had seen (yes or no).

Based on the text comprehension and event perception literature, we hypothesized the following: If perceptual processing is biased by event models and prediction mechanisms rely on experience, the perception of a dynamic event may involve postdictive reconstructions that are not always efficient. The observer may compare the current event to a previous event (learned model; event schema; conceptual structure) and weigh its plausibility to fill in perceptual

[2]The second experiment discussed in the paper explicitly informed the participants about the possible conditions and explicitly instructed them to pay attention to the contact moment only. It marks a transition between Part I and Part II of this thesis. Results will be discussed at the end of this chapter, preparing for controlled processes in the next one.

gaps with causal content (here based on learned cause-effect relations). This has been demonstrated by Strickland and Keil (2011) whose experimental design we partially replicated. Strickland and Keil (2011) showed their participants video clips of a person kicking a ball. Some of the videos were cut to eliminate the actual moment of ball contact, that is, participants only saw a person running towards a ball and a ball bouncing down the field. When asked to assign pictures to the video clip just seen, they often believed to have seen the contact moment in the clip. If they saw a person running towards a ball followed by an unrelated clip, they were less convinced to have seen the picture of the contact moment. The authors concluded that observers fill in gaps with meaningful events based on the causality of the clips, either with predictive (online) perceptual processes or with post hoc assumptions, but both based on similar events stored in memory.

However, there are quiet a few studies on the cognitive and perceptual skills of experts that would lead to a different hypothesis. Studies with chess players have shown that novices differ from experts in their perception of a chess scene (Chase & Simon, 1973; Reingold, Charness, Pomplun, & Stampe, 2001; Reingold, Charness, Schultetus, & Stampe, 2001). Gobet and Simon (1996) proposed that expert chess players store their knowledge in templates that contain the successive stages (patterns) of a possible game. Based on the idea that the visual system is influenced by knowledge and therefore processes meaningful objects in a scene selectively and semantically (see e.g. Chun & Jiang, 1998, 1999), Didierjean and Marmèche (2005) tested professional basketball players in their perceptual anticipation. In a first experiment, an immediate recognition task, the players compared two configurations of player positions presented one after another. When the second configuration was different, it was either an expected constellation or a constellation that would have preceded the first. In the second experiment, a long-term recognition task, the players studied game situations in the first part of the experiment. Afterwards, they were asked to recognize these constellations mixed-up with new ones that were not studied. Again, the interspersed new constellations were either a preceding constellation of a known one, or a constellation that would have come afterwards. In both experiments, experts differed from novices which allows the following conclusions: knowledge on successive stages of the game influences visual perception as soon as the constellation was processed and experts may store the perceptual input as well as an anticipatory memory trace. The idea of two competing traces is also addressed in the framework of the boundary extension effect (discussed in Paper (2)). Johnson, Hashtroudi, and Lindsay (1993) proposed that source-monitoring errors occur when the reality monitoring has too much influence (Johnson & Raye, 1981).

If we assume that the experts' experience with the stimulus material mediates visual perception and/or memory, we would expect performance differences in favor of the novices. The experts performance would suffer from a competition between perceptual and anticipatory (reality) traces (Didierjean & Marmèche, 2005), either online or posthoc, leading to (more) causal fillings.

In general, if the processing of complex events involves an interplay of perception and cognition we would expect to find performance differences between the groups tested – but if the global approach to the perception of dynamic stimuli also holds for complex events that require an understanding of semantic relations, we would expect to find no differences between the groups.

Brockhoff, A., Huff, M., Maurer, A., & Papenmeier, F. (2016). Seeing the unseen? illusory causal filling in FIFA referees, players, and novices. *Cognitive Research: Principles and Implications*, *1*(1). doi:10.1186/s41235-016-0008-5.

Brockhoff *et al. Cognitive Research: Principles and Implications* (2016) 1:7
DOI 10.1186/s41235-016-0008-5

Cognitive Research: Principles and Implications

ORIGINAL ARTICLE **Open Access**

Seeing the unseen? Illusory causal filling in FIFA referees, players, and novices

Alisa Brockhoff*, Markus Huff*, Annika Maurer and Frank Papenmeier

Abstract

Humans often falsely report having seen a causal link between two dynamic scenes if the second scene depicts a valid logical consequence of the initial scene. As an example, a video clip shows someone kicking a ball including the ball flying. Even if the video clip omitted the moment of contact (i.e., the causal link), participants falsely report having seen this moment. In the current study, we explored the interplay of cognitive-perceptual expertise and event perception by measuring the false-alarm rates of three groups with differing interests in football (soccer in North America) (novices, players, and FIFA referees). We used the event-completion paradigm with video footage of a real football match, presenting either complete clips or incomplete clips (i.e., with the contact moment omitted). Either a causally linked scene or an incoherent scene followed a cut in the incomplete videos. Causally linked scenes induced false recognitions in all three groups: although the ball contact moment was not presented, participants indicated that they had seen the contact as frequently when it was absent as in the complete condition. In a second experiment, we asked the novices to detect the ball contact moment when it was either visible or not and when it was either followed by a causally or non-causally linked scene. Here, instead of presenting pictures of the clip, the participants were give a two-alternative forced-choice task: "Yes, contact was visible", or "No, contact was not visible". The results of Experiment 1 indicate that conceptual interpretations of simple events are independent of expertise: there were no top-down effects on perception. Participants in Experiment 2 detected the ball contact moment significantly more often correctly in the non-causal than in the causal conditions, indicating that the effect observed in Experiment 1 was not due to a possibly influential design (e.g., inducing a false memory for the presented pictures). The theoretical as well as the practical implications are discussed.

Keywords: Expertise influences, Event perception, Dynamic scenes, Causal fillings

Abbreviations: CI, confidence intervals; EST, Event segmentation theory

Significance

The current work is, to our knowledge, the first to combine a study of perceptual-cognitive skills with event perception and it is, therefore, mainly of an explorative nature. We took theoretical research out into the real world and investigated the role of top-down factors on event completion by testing three groups with a differing level of interest and experience (novices, players, and FIFA referees) on a simple event-completion task (Strickland & Keil, 2011). Although there is considerable evidence that expertise in sports domains is connected to superior perceptual-cognitive skills, our results indicate no influence of these skills on event perception. They rather support a recent publication by Firestone and Scholl (2015b), who concluded that perception may be largely independent of top-down influences. Such a proposition not only challenges our theoretical understanding of event perception, but also has substantive practical implications for fairness in sports by strongly advocating the increased use of technology instead of perceptual training programs for match officials.

*Correspondence: alisa.brockhoff@uni-tuebingen.de; markus.huff@uni-tuebingen.de
Department of Psychology, Eberhard Karls Universität Tübingen, Schleichstr. 4, 72076 Tübingen, Germany

Background

During the FIFA World Cup tournament in 2010, the referees made many controversial calls that influenced the outcomes of matches so tremendously that the then-FIFA president apologized for the referees' mistakes. In response, the use of goal-line technologies was officially allowed in 2012, which since have become more and more common at the very top levels of the game. The current study was inspired by a controversial goal that happened in a Bundesliga match in 2013, a match in which no goal-line technology was used. The ball went through a hole in the side netting and everyone, including the referees, mistook it for an actual goal. This rare phantom goal demonstrated the limits and biases of human perception. Such a phantom goal is even more surprising in the light of numerous studies that reported experts to have superior domain-specific perceptual-cognitive skills (e.g., Williams, 2000), an expertise that even leads to an advantage in motion outside the expert's area (e.g., Romeas & Faubert, 2015). Vision and perception are shaped by one's individual experiences and knowledge: the mental representations of events. Such representations are reconstructed and updated through experience and knowledge and provide the basis for understanding the world around us (Zacks & Tversky, 2001). However, constant reconstruction and updating of mental representations make event perception effortful and, thus, fragile. Strickland and Keil (2011) reported a (possibly consequential) bias in event perception: the event-completion effect. Video clips that indicated a causal implication (example sequence: an athlete running towards a ball – cut – a flying ball) produced higher false-alarm rates for pictures displaying the athlete kicking the ball than video clips that did not imply any causation. The authors suggested that observers either confused online predictions (the ball will be kicked and will bounce down the field) with actually seen elements of the scene, or relied on schema- or principle-based post hoc inferences (a ball bouncing down a field must have been kicked).

Perceptual-cognitive expertise

A number of studies have reported that expert athletes show superior perceptual-cognitive skills compared to novices in sport-specific tasks, including visual cue usage (Abernethy, Gill, Parks, & Packer, 2001; Ward, Williams, & Bennett, 2002; Williams, 2000), visual search strategies (Vaeyens, Lenoir, Williams, & Philippaerts, 2007; Williams, 2000), and recall and recognition of meaningful patterns (Bell, Boshuizen, Scherpbier, & Dornan, 2009; Lesgold et al., 1988; Reingold & Sheridan, 2011; Smeeton, Ward, & Williams, 2004). In general, experts' demonstration of perceptual-cognitive expertise can go beyond the specific sports domain (Romeas & Faubert, 2015; Romeas, Guldner, & Faubert, 2016) and can help, for example, in learning complex neutral dynamic scenes (Faubert, 2013) or to outperform novices in everyday tasks (e.g., crossing a street as a pedestrian in a crowded inner city: Chaddock, Neider, Voss, Gaspar, & Kramer, 2011). While the majority of the reported studies intended to identify the exceptional perceptual-cognitive skills of experts by focusing on pattern recognition, decision-making, or biological motion perception, mainly aiming to create training programs or prevent incidents that result in injuries, the current paper is interested in a fundamental understanding of experts' perception, or memory, of events.

Hypotheses

In the current study, we conceptually replicated the design by Strickland and Keil (2011) and tested two expert groups (football players and FIFA referees) and a control group (students with no interest in football). We wondered whether the perceptual-cognitive skills of experts would prevent the event-completion effect when observing familiar motion. Based on the currently most prominent model of event perception, the event segmentation theory (EST; Zacks, Speer, Swallow, Braver, & Reynolds, 2007), prediction errors occur and an event boundary is perceived when certain event features change (e.g., situational features such as spatial location and characters: Zacks, Speer, & Reynolds, 2009). If online predictions of experts are more detailed, it may be more likely that the missing ball contact is actually reported to be perceived as a missing situational feature in the schema, and, thus, not perceptually filled in. More specifically, a more detailed representation would result in a lower false-alarm rate in referees and players.

We do have reason to hypothesize that the superior perceptual-cognitive skills of experts could prevent the event-completion effect since they may process visual information not only qualitatively but also quantitatively differently, but the opposite could be the case as well. Mann, Williams, Ward, and Janelle (2007) analyzed eye movements of experts and novices and revealed that the skilled performers required fewer fixations of longer duration to gather relevant information, compared to novices, who made many short fixations. Thus, novices consider the potential influence of all available visual information while experts concentrate on the relevant information by perceiving the multidimensional complexity of the situation (further examples are in Haider & Frensch, 1996; Hattie, 2003; North & Williams, 2008). Expertise was also shown to allow for a more efficient switch of attentional foci. Underwood, Chapman, Brocklehurst, Underwood, and Crundall (2003) observed that the scan paths during driving differ depending on the expertise of the driver. Novices were not able to switch their focus of attention as a response to potential hazards, while experts constantly monitored other road users. In the current study, the hardwired event schemata of experts could actually lead to a stronger bias if the ball contact is considered irrelevant information in the representation of the event. Or stated differently, novices may have a more detailed schema of the event (e.g., a ball kick) because, in their lives, there is no need for them to condense the schema for more efficient processing. Referees, however, have to make 3 or 4 decisions in each minute of the actual play time (Williams, 2013) and, thus, they benefit significantly from filtering visual information rigorously. On the other hand, experts may have a more detailed schema than a novice due to frequent exposure and the ability to switch their focus of attention if needed. However, based on the EST, this again would result in a stronger event-completion effect. If experts have a rather global observational approach to familiar scenes, they may even have event models that account for missing information and changes in visual information. The missing ball contact may then not be surprising; therefore, it may not be detected as an error, and will, thus, not result in the perception of an event boundary but in an event-completion effect. Finally, it is also possible that there are simply no top-down effects of cognition on perception as recently claimed by Firestone and Scholl (2015b). The two authors carefully reviewed hundreds of studies and extracted general (design) pitfalls of each approach to study the effect of cognition on attention. We will discuss our results with regard to the two disparate but interrelated systems of perception and memory.

Experimental overview

To ensure that we really tested perceptual-cognitive differences in event perception – and not declarative knowledge and analysis skills – we intentionally used video clips of dynamic events that did not require knowledge of the game, depicting actions that definitely have been observed by each participant before, independent of their level of interest in football. We cut out scenes from a real match, including corner kicks, kick-offs, free kicks, and throw-ins. In Experiment 1, we conceptually replicated the design of Strickland and Keil (2011) and presented the participants with (1) the complete sequences (i.e., including the contact moment), (2) an incomplete causal sequence (i.e., excluding the ball contact), or (3) an incomplete non-causal sequence (i.e., excluding the ball contact with a non-logical follow-up; example sequence: player about to throw the ball in – cut – a different player being fouled). However, note that our restricted sample of experts did not allow us to run a between-subject design as was done in the original study. To ensure that our design would not alert the participants to the purpose of the event, we left out one condition: a visible ball contact that was followed by a non-causal scene. In Experiment 2, we further controlled the design by showing video clips that either included or excluded a ball contact. Participants were fully informed about the probabilities of each clip type occurring (50 %) and were given a forced choice of the two alternatives (ball contact seen: yes or no). The latter inevitably brought in the aspect of attentional control by "knowing what to look for"; however, it helped us to understand further at which point of information processing the bias has its origin. We are aware, however, that our (or any) design may not be able to grasp the fine line between perception, memory, and post-perceptual judgment. Our results will be discussed with a focus on the event-completion effect and its occurrence in different groups. Any interpretation concerning perception or memory has to be regarded with caution.

Methods

Stimuli were presented on 15.4-inch notebooks using PsychPy (Peirce, 2008). The participants were seated at a distance of 60 cm from the screen. Footage of a soccer match of the Young Boys Bern against the Grasshoppers Zürich that took place on 23 March 2014 was used as stimulus material. The footage was compiled out of three camera perspectives. Clips of about 20 seconds each were created. Each clip consisted of two parts shot from different camera angles. The assignment of clips to conditions was balanced across participants in each experiment. In general, the two parts of each clip were causally linked or not (Fig. 1c or d), and the ball release or contact (kick) moment[1] (Fig. 1b) was visible or not. Figure 1 depicts example sequences.

In Experiment 1, we conceptually replicated the design by Strickland and Keil (2011) and used the following combinations of video clips (see Fig. 1): complete (A–B–C) vs incomplete causal (A–C) vs incomplete non-causal (A–D).[2] In Experiment 2, the basic idea of the design was similar; however, we measured only the detection rate of the contact moment and further added a condition in which the ball contact (B) was visible in non-causal sequences as well (A–B–D). In Experiment 1, each participant saw seven response pictures (see Strickland & Keil, 2011) after each clip. Three pictures were selected from the first part of each clip (a yes filler), three pictures were related to the yes-filler items but came from other parts of the game, such as other players preparing for a corner kick (a no filler), and the critical picture depicted the moment of ball contact or ball release (contact). The participants were asked whether they had seen the picture in the clip: Yes ("press 1") or No ("press 9"). See Fig. 2 for the response pictures for the example sequences (Fig. 1). Further, they were asked to rate how certain they were about their answer (on a scale from 1, not at all, to 5, extremely).

In Experiment 2, we showed the participants 40 clips and asked whether they had seen the ball contact moment

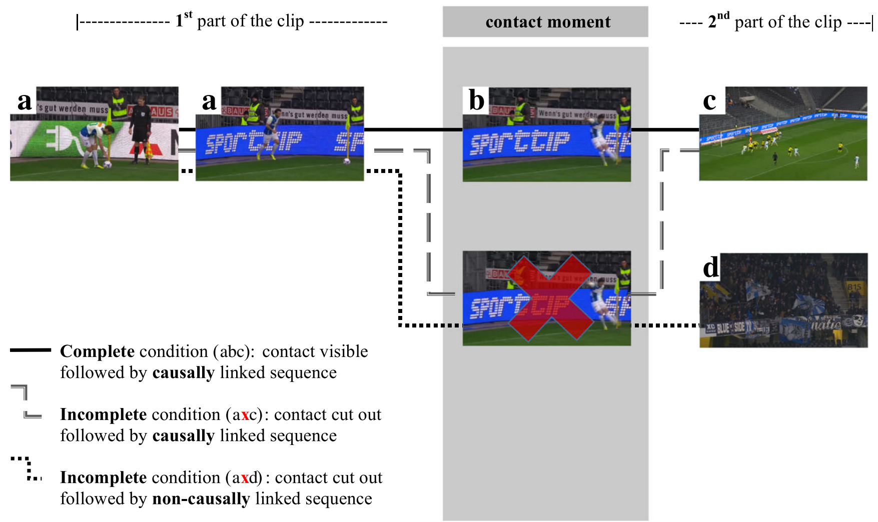

Fig. 1 Example sequences (pictures). **a** First part of the clip. **b** Ball release/contact moment. **c** Second part of the clip: causally linked scenes. **d** Second part of the clip: not causally linked scenes

(B in Fig. 1). Instead of response pictures (Fig. 2), we gave the participants forced-choice alternatives: "Yes, I have seen the ball contact" and "No, I have not seen the ball contact". The experiment was conducted as a mixed 2 (ball contact visible, within) × 2 (second part of the clip: causal or non-causal, between) subject design. We measured the sensitivity to the contact moments as d' and response criterion c (see Experiment 2 for further details).[3]

An expertise questionnaire tested basic declarative football knowledge using 11 questions, for example, "In which country did the last FIFA World Cup take place?" (see Additional file 1: Appendix for a complete list of questions).

Statistical analysis

In Experiment 1, we report expertise knowledge, proportion correct, proportion of yes answers, and confidence in the recognition test as separate dependent variables. Because of the binary response variable (yes or no), we analyzed effects on *proportion correct* and *proportion of yes answers* with a generalized mixed effect model (with a logit link), using the lme4 package (Bates, Sarkar, Bates, & Matrix, 2007; Pinheiro, Bates, DebRoy, & Sarkar,

Fig. 2 Example response pictures. 1: Contact, 2: yes filler (selected from the first part of the clip), 3: no filler (not in the clip)

2006) in the R environment (R Development Core Team, 2016). Participants were specified as the random factor to control for their associated intraclass correlation. We present the type II Wald χ^2 test results from GLMER. Further, we provide the results of planned contrasts (based on our hypotheses and the original study's results). Additionally, the credibility of the found null effect and the likelihood of the occurrence of the null and the alternative hypotheses are presented with Bayesian statistics and JASP (JASP Team, 2016). In Experiment 2, we report the sensitivity measure d'.

Experiment 1: Conceptual replication of the original study with groups with different expertise levels

Method

Participants Three groups of participants were tested on three different occasions. There were 42 novices (14 male and 28 female students, age $M = 25.76$, SD = 6.81 years), 16 football players of a seventh German football league (all male, age $M = 24.81$, SD = 3.64 years), and 18 referees from Switzerland appointed as officials for matches in competitions organized by the Fédération Internationale de Football Association (FIFA) (all male, age $M = 32.2$, SD = 4.93 years). Two referees were excluded because they retired from their active positions as official FIFA referees. The students tested participated in return for monetary compensation or course credits. The football players were students of the University of Tuebingen's department of sports science and their participation was a course requirement. The referees participated during one of their regular advanced training courses and were not compensated monetarily.

Design and procedure The first part of each clip was between 11.6 and 15.1 seconds long. A keeper during a kick-off was depicted in three clips, a throw-in in one clip, a corner kick in three clips, and a free kick in two clips. A clip was either shown completely or shortened by the removal of the moment of ball contact (kick) or ball release (throw-in). We deleted 1–4 frames; however, the deletion for causal and non-causal clips was always exactly the same. The second part of the clip lasted between 5.7 and 8.4 seconds. Each participant saw nine clips spread equally across three conditions: complete first part with causally linked second part (complete), shortened first part with causally linked second part (incomplete with causally linked sequence), or shortened first part with second part that was not causally linked (incomplete with non-causally linked sequence). See Fig. 1, combinations A–B–C, A–C, and A–D. The experiment reported here took 15 minutes. The participants received instructions and immediately started with the event-completion task. After each clip, seven response pictures (Strickland & Keil, 2011) were shown (see Fig. 2).

Results

Expertise knowledge We calculated the proportion of correctly answered questions. The football players' declarative football knowledge was significantly higher compared to the novices' ($M = .86$, SD = .34 and $M = .51$, SD = .50, respectively): $t(50.83) = 10.70$, $p < .001$. We regarded the referees' football knowledge as a precondition for their FIFA employment and did not test them on the questionnaire.

Proportion correct We analyzed participants' performance in the recognition test. Because the critical contact item was a target item in the complete condition and a distractor item in the remaining two conditions, we excluded this item from this analysis. We calculated the proportion of correctly answered questions and fitted a generalized mixed effect model with the binary dependent variable *yes/no answers*. Expertise was inserted as the fixed effect, and participants were specified as the random factor. The factor expertise was significant [$\chi^2(2) = 17.621$ and $p < .01$]. Post-hoc Tukey comparisons helped to specify the difference between the three groups of expertise. As can be seen in Fig. 3, the two expert groups outperformed the novices: for players vs novices $z = 3.33$ and $p < .01$, and for referees vs novices $z = 3.254$ and $p < .01$. We observed no differences between the players' and the referees' performance ($z = 0.06$, $p = .99$).

Proportion of yes answers. We analyzed the effects on the binary dependent variable (yes/no answers) with a generalized linear mixed model (with a logit link), using the lme4 package in the R environment. Participants were specified as a random factor to control for their associated intraclass correlation. We used the raw data and fitted a

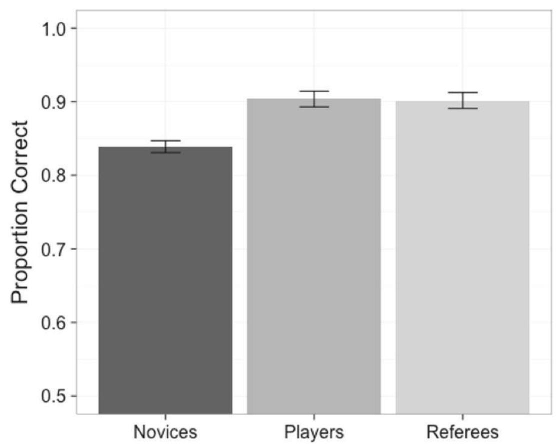

Fig. 3 Performance in the recognition test (excluding the critical contact item) as a function of expertise. Error bars represent the standard error of the mean

model including all main effects and interactions of expertise, item type, and condition as fixed effects. We analyzed the resulting model using type II Wald χ^2 tests.

Our main finding is a significant two-way interaction of condition and item type [$\chi^2(4) = 11.52$ and $p = .021$]. The three-way interaction of expertise, condition, and item type was not significant [$\chi^2(8) = 6.91$ and $p = .546$]. Further, there was a significant main effect of item type [$\chi^2(2) = 1262.00$ and $p < .001$], and a significant interaction of expertise and item type [$\chi^2(4) = 41.05$ and $p < .001$]. None of the other main effects and interactions reached significance , $p > .17$. While the proportion of yes answers in the non-causal condition was significantly lower (as expected), it should be noted that the false-alarm rate was still over chance level. However, our findings are in line with the results found in the original study by Strickland and Keil (Strickland and Keil 2011). See Fig. 4a for the analyzed proportions in each expertise group.

To investigate the interactive relationship of the two categorical variables *condition* and *item type*, we calculated contrasts. The underlying *glmer* model was now reduced (see Fig. 4b for the aggregated data used) and did not include expertise anymore, since the given expertise level (novice, player, or referee) did not interact with condition*item type (non-significant three-way interaction reported above). To prevent α inflation at this level of the analysis, a Bonferroni correction ($0.05/3 = 0.016$) for multiple comparisons was applied. Further insights into the variability of the (log) mean difference between the observed answers are given with 95 % confidence intervals (CI).

As expected, two of the three contrasts produced significant results. The number of yes answers (i.e., the number of reports indicating that the contact moment had been seen) in the condition with implied causation (causal) differed significantly ($z = 22.21$ and $p < .001$) from the number of yes answers in the condition *without* implied condition (non-causal), with an estimated (log) mean difference of 4.03, CI [3.60, 4.46]. The non-causal incomplete condition also differed significantly from the condition in which the ball contact was included (complete condition), $z = 16.51$ and $p < .001$ (estimated difference = 3.95, CI [3.52, 4.38]). The contrast of the causal vs the complete condition was not significant, $z = 0.73$ and $p = 0.75$ (estimated difference = 0.15, CI [−0.36, 0.68]).

Bayesian statistics We calculated a Bayes factor analysis for the proportion of yes answers to the contact items in the no causal implication and the conditions with causal implication. The Bayes factor evidence for the null hypothesis in a Bayesian repeated measures ANOVA comparing a model that included the main effects of condition (with causal implication or no causal implication) and expertise (novices, players, and referees) with a model including additionally the interaction of these factors amounted to 4.99, which is conventionally classified as substantial (Rouder, Morey, Speckman, & Province, 2012; Wetzels & Wagenmakers, 2012).

Confidence A repeated measures ANOVA was performed with confidence as the dependent variable (see Fig. 5). We observed a significant main effect of expertise [$F(2, 71) = 10.27$ and $p < .001$]. Players' and referees' confidence was significantly higher than novices' confidence , $p < .004$. Again, there was no difference between players and referees ($p = .501$). Further, we observed a significant main effect of item type [$F(2,142) = 27.20$ and $p < .001$], indicating that confidence was higher for the no-filler items compared to the contact items and the yes-filler items , $p < .003$. The interaction of item type and condition approached significance [$F(2,142) = 2.42$ and $p = .049$]. In this context, however, we observed no significant differences between the different conditions with regard to the contact item responses, $p >= .247$.

Discussion

To capture online perceptual performance errors, we presented video clips that implied causation (or not) and asked the participants afterwards whether they had seen certain pictures (or not). While overall performance (proportion correct) was higher for experts than for novices, all participants were prone to the event-completion effect (analyzed with the proportion of yes answers). Furthermore, we measured confidence rating to examine whether experts show illusionary superiority biases (observed as a coping mechanism for stress and self-esteem protection in referees; e.g., Wolfson & Neave, 2007). We observed higher confidence ratings in the referee and the player groups compared to the novices – however, they actually performed better, thus, showing an actual superiority instead of an illusionary superiority bias. This was expected based on the experts' superior recall and recognition of meaningful patterns and details (Bell et al., 2009; Lesgold et al., 1988; Reingold & Sheridan, 2011; Smeeton et al., 2004). The results of the present study replicate the event-completion effect measured in the original study by Strickland and Keil (2011). The results exemplify how the human information processing system struggles with perceiving and recalling details of an everyday life event. We found these difficulties to be independent of task-specific expertise, suggesting that on a certain basic perceptual level, if presented with a simple action event, humans equally chunk or segment continuous activity, resulting in the representation of a series of discrete events (Newtson, 1973) – a process that allows for online and post-hoc

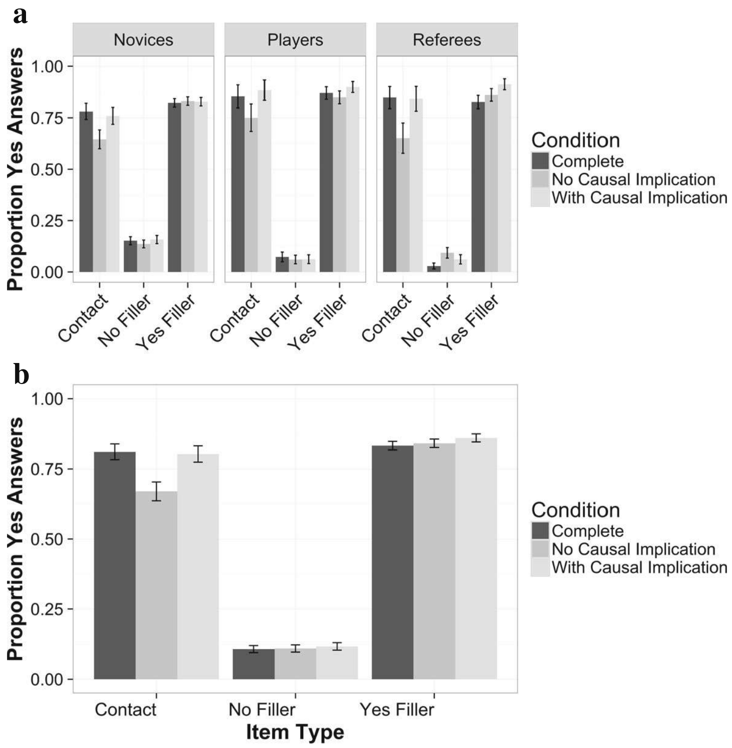

Fig. 4 Proportion of yes answers. **a** Proportion of yes answers for each expertise level as a function of condition (complete, no causal implication, with causal implication). **b** Aggregated proportions of yes answers as used in the contrast analysis. Error bars represent the standard error of the mean

inferences, and illusory causal fillings. However, before we interpret these results further, we need to ensure that the effect found is not due to the study instructions, which may have biased the participants to assume ball contact. Participants may have *assumed* they had seen contact because they did not know that omitted contact moments were an option.

The question remains whether the observed event-completion effect is a phenomenon based on online predictions or rather the result of backwards mapping, an effect known from text comprehension research (e.g., Potts, Keenan, & Golding, 1988). Although, backwards mapping was originally used to explain anticipation processes during text comprehension, its adaption to causal

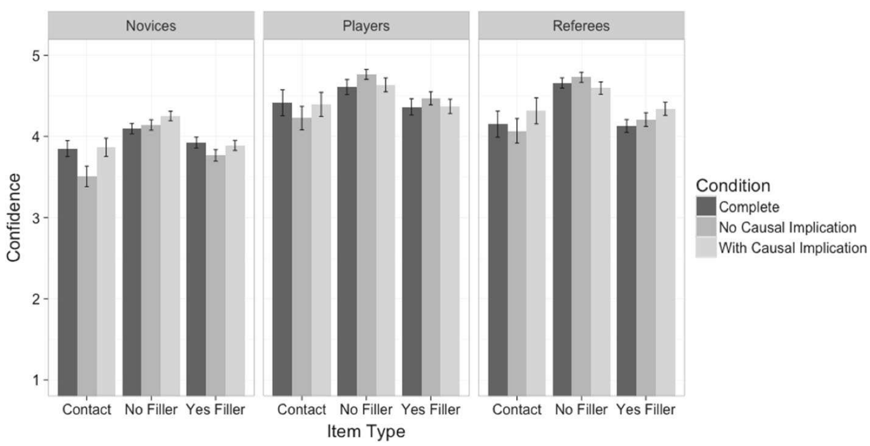

Fig. 5 Confidence rating as a function of condition and expertise. Error bars represent the standard error of the mean

fillings in event perception is straightforward: participants base their decisions of a recognition item at the very moment of presentation and check if the picture is a plausible cause of what they have already watched. In other words, a contact picture would be a plausible, and natural, cause of a video clip that showed a football player approaching a ball.

Experiment 2: contact – yes or no?

In a detection experiment, we presented participants with complete and incomplete stimuli with causal and non-causal continuation and asked them to indicate whether they had seen the contact moment or not. This may prevent backwards mapping because participants "know what to look for" before the presentation of the video clip. Further, without recognition items (pictures), the participants are less prone to picture-based biases, which allows us to measure participants' discrimination performance in the non-causal and causal conditions. If the event-completion effect is primarily a phenomenon based on online predictions, participants' discrimination performance should be lower in the causal compared to the non-causal condition.

Method

Participants Altogether, 32 students of the University of Tuebingen (7 male and 25 female students, age $M = 23.16$ years, SD = 4.61) participated in the experiment in return for course credits or monetary compensation. Of these, 17 participants were assigned to the causal and 15 to the non-causal condition. We excluded from the analysis one participant who did not understand the task. Thus, 17 in the causal and 14 participants in the non-causal condition entered the final analysis.

Design and procedure Then 40 video clips were shown either as complete clips or with the ball contact excluded. The first part of each clip was between 1.4 and 15 seconds long. The clips were either causally connected or not (see "General method"). We always deleted four frames before the ball contact frame, resulting in a deletion of 160 ms in both incomplete conditions (the presentation rate of each clip was 25 frames per second). The second part of the clip was between 1.2 and 6.3 seconds long.[4] Clips consisted of 14 kick-offs, 5 corners, 13 throw-ins, and 8 free kicks. Participants received specific information on the probability that the ball contact was visible (50 %). Further, they saw a process graphic of a matchstick man approaching a ball and kicking it so that they knew what "ball contact moment" or "moment of ball release" meant. The suggested experiment has been conducted as a mixed 2 (ball contact visible, within-subject manipulation) × 2 (second part of the clip: causal or non-causal, order was balanced between groups) design.

Results

We report sensitivity (d') and response criterion (c) from signal detection theory as dependent variables (Green & Swets, 1966). Yes answers to clips depicting the release moment (complete conditions) were counted as hits and yes answers to clips not depicting the release moment (incomplete condition) were counted as false alarms. Finally, we aggregated the data on the participant level and calculated separate independent sample t-tests for d' and c. Because d' and c are not defined for hit rates and false-alarm rates of 1.0 and 0.0, we adjusted such values to half a trial incorrect or half a trial correct, respectively.

Sensitivity Sensitivity (d') was well above chance ($d' = 0$) and was significantly higher in the non-causal ($M = 2.75$, SD = 0.59) compared to the causal condition ($M = 1.44$, SD = 1.15), $t(29) = 4.08$, $p < .001$. Thus, this supports the hypothesis that participants' online perception was distorted by the causal continuation of the scene.

Response bias We did not observe a statistically significant difference between the non-causal ($M = -0.09$, SD = 0.32) and the causal condition ($M = -0.20$, SD = 0.38) with regard to the response criterion (c), $t(29) = 0.88$ and $p = .388$.

Discussion

We observed lower discrimination performance in the group of participants who saw the causal sequel compared to the group of participants who were presented with non-causal sequences. Thus, these findings support the hypothesis that the causal continuation actually changed participants' perceptions. A (cautious) explanation of this finding refers to EST (Zacks et al., 2007). According to EST, participants' perceptions are based on predictions. For a non-causal continuation, these predictions fail and participants perceive an event boundary. As a consequence, participants' representations of this moment are more precise compared to the condition with non-causal continuation in which predictions were not violated and participants did not perceive an event boundary.

General discussion

The present study was interested in the interplay of cognitive and perceptual processes in experts compared to novices. The main objective was to study the appearance of the event-completion effect in groups with different cognitive-perceptual training. However, our results also give us an idea of how internal schema-based systems and external sensory input processing may result in an automatic completion of events. The results reported here allow a number of interesting implications.

Theoretical implications for event perception

The current most prominent model of event perception is EST (Zacks et al., 2007). EST is based on various theories of perception, neurophysiology, and language processing (Carpenter, Grossberg, & Arbib, 2003; Fuster, 1990; Van Dijk, Kintsch, & Van Dijk, 1983). A fundamental principle of EST states that the processing of events forms sensory representations that are influenced by experience and knowledge. Event schemata affect the current content of the event models with top-down processes, expanding their effective capacity by assembling predictive information about the future relevance of certain features of events. When certain event features change (e.g., situational features such as spatial location and characters; Zacks et al., 2009), prediction errors occur and an event boundary is perceived. Regarding EST, our results could be explained with an error-detection mechanism that operates on a temporal buffer holding a given number of causal snapshots (Wood, 2011). The error-detection mechanism constantly checks whether online predictions based on working memory representations of the ongoing event are fulfilled. Transient increases in the violation of predictions (Zacks et al., 2009) make the current event model useless and in need of an update. As our results suggest, one missing snapshot of an event (implied causation condition) does not automatically trigger an event update because enough predictions of the event are fulfilled. Clip sequences that did not imply causality may have activated an error-detection mechanism and triggered event boundary perception processes. The original EST model describes event models as a stable representation that can only be reset or updated based on the current perceptual information available when the error-detection mechanism opens the gate. Error detection may also play an important part in the actual perception of events: the comparable number of yes answers for contact items and causal yes-filler items in our data implies that the event-completion effect is nurtured by the sensitivity of the error-detection mechanism. In other words, the more prediction errors the error monitoring allows, the more illusory causal fillings will happen. Importantly, our data suggest that expertise does not influence event perception. That indicates that top-down processes do not influence the simple mechanisms of online prediction and error detection as much as is assumed in the EST (Zacks et al., 2007). This top-down component, however, is largely underspecified in EST. Zacks and colleagues write: "This claim is based largely on parsimony and may need to be revised in the future" (Zacks et al., 2007, p. 275). At least for our stimulus material with simple structured events, the idea of an unaffected gating mechanism is in line with Firestone and Scholl (2015b): there are no top-down effects of cognition on perception.

Top-down effects and the locus of contextual biases

Did the participants in our studies actually see (falsely perceive) or did they simply report to have seen (falsely remember) the ball contact? The presented studies applied a recognition and detection test to explore the event-completion effect. However, as recently suggested by Firestone and Scholl (in press), there is a great difference between seeing and recognizing. Any top-down effect measured can be due to an influence on front-end visual processing but equally likely be due to back-end memory. In the current paper, we communicated a tendency to define the event-completion effect as due to an error that occurs in perception rather than in memory. Although we do not have clear evidence for either involvement, the results of Experiment 2 (in which we decreased the possible memory biases due to backwards mapping) do indicate that the effect is partly due to online perceptional processes. We were further biased by the majority of results found in the literature that connect memory to experience. As memory fades due to brain damage or aging, representations become increasingly changed by preexisting knowledge. Especially popular is that patients with Alzheimer's tend to falsely remember details, words, or events that they actually did not experience (confabulation: e.g., Tallberg & Almkvist, 2001). However, experience and expertise did not influence the appearance of the event-completion effect. Thus, reversing the argument, our results could show that the event-completion effect cannot be an error in memory, because then we would have found differences between the expertise groups.

In a recent paper (Firestone & Scholl, 2015a), the authors discuss semantic (language) priming, universally understood as an effect on memory (Collins & Loftus, 1975) that may have been mistaken for top-down effects on visual processing in various studies. In semantic priming, reading a word such as "peach" lowers the threshold for related fruits in memory and they will be processed faster than an unrelated word. Language and event perception are closely related: much of what we know about our understanding of events comes from studies that asked participants to describe an event in their own words. For example, with such a linguistic account, Talmy (1975) was able to define the building blocks of motion events. However, it may be possible that the observed language structure does not only reflect how we perceive event units, but could be a general reflection of the preferred global-over-local approach of the human brain (e.g., Fink et al., 1996). If we assume that the activation of related words is comparable to the activation of related event models in memory (allowing for faster access to different scenarios and faster processing of related visual details), our null findings would again point towards a bias on the perceptual level. The wealth of experienced scenarios of the event should have activated a broader spectrum of experienced content in the experts, which should have resulted in differences between the three groups due to differences in memory activations. Firestone and Scholl (2015a) further proposed that it is possible to distinguish memory and perception clearly in practice. This seems to be a bold proposal since false memories (here, an error of commission) can be elicited within 1/20th of a second (Intraub & Dickinson, 2008). Intraub and Dickinson (2008) report a constructive error in scene representation, the boundary extension, in which observers falsely remember an image that is shown beyond the edges of the previously encountered view. When the first item is presented without a scenic structure, boundary extension does not occur (Intraub, Gottesman, & Bills, 1998). They propose that boundary extension is the result of a source-monitoring error (Johnson, Hashtroudi, & Lindsay, 1993) with a strong influence of a reality-monitoring error (Johnson & Raye, 1981). The error happens when the human brain has to distinguish between internally generated information (experience with certain structures) and externally generated information (sensory input). The authors suggest that the rapidity of such a boundary extension error is advantageous rather than harmful; it shows how the visual system incorporates fleeting views of images with spurious boundaries into a coherent representation of the world around us. The rapidity of the error may further imply that perception and memory are two processing systems of the same underlying cognitive mechanism.

Our data could be explained as the result of a distinction error between internally and externally generated information. Disregarding the traditional distinction of false memories and visual illusions and assuming an extraordinary fast engagement of both during the processing of visual input, the observed null effect of expertise in our experiment may be the result of an imbalance of weighted sources. The externally generated information processing of experts may be more efficient and more detailed; however, the internally generated information outperforms sensory input due to the system's need to embed the event into known reality. Experiment 2 further reflects the weight of the reality source. Here, participants knew precisely what would be tested in each trial and were prepared to answer a specific question. Conscious awareness is needed to be able to report whether the stimulus was visible or not (Lamme, 2003), but even in such an enhanced state of target processing, the internally generated source overruled the external sensory input, resulting in decreased sensitivity for the detection of the ball contact moment in causally linked scenes. For the current design, the ideas mentioned above are pure speculations and may be regarded as such. Future research could be concerned with whether expertise influences the level (global or local) of event processing. For example, Beaucousin et al. (2011) recorded event-related potentials and reported that the meaningfulness of an object influences global and local information. They assumed that knowledge about the world influences the global and local levels of processing. Comparing the performance of experts and novices on meaningful and non-meaningful patterns would help us to understand better the early stages of processing.

In addition, it would be interesting to see whether experts compared to novices structure events differently, measured as event segments indicated with a button press by participants. As memory distortions can happen within 50 ms (Intraub & Dickinson, 2008), behavioral measurements may not be able to grasp the difference between memory and perception (if there is any). To really answer such a question, functional neuroimaging procedures are advisable.

Practical implications

The present findings have a serious impact on the fairness of the game. A red card may be based simply on two single observations that perceptual processes have falsely interconnected in a causal manner: player A approached player B and player B got hurt. The match official may be absolutely certain that they had seen a contact, but it may have been an event-completion effect. The top Dutch football league (*Eredivisie*), therefore, employs a video referee who observes video replays of the game to help the referees on the field with tricky decisions. However, since many believe that the human element of sports is lost when technologies are used, eliminating, for example, the "enjoyment of debating mistakes" (Kelso, 2010), chances are rather low that other European football leagues will follow the example of the Dutch. Even in the presence of technology, the importance of the perceptual and cognitive skills of match officials is, thus, not reduced.

Limitations

It needs to be taken into account, however, that we aimed to test basic perceptual processes and can, thus, speak only about the organization of the mind when it is faced with simple events. The perception of complex events may nonetheless be influenced by domain-specific expertise. For example, when presented with a deliberate dive, novices may not be able to differentiate between whether it was a real foul or a fake fall by the player. The cognitive-perceptual excellence and the so-called intuitive skills of an expert to analyze such an incident may be based on a highly sensitive error-detection mechanism. Such ideas, however, will require theoretical and empirical development beyond the scope of this article. Left unknown is still whether memory or perception is responsible for the effect.

Conclusions

In conclusion, the results of the present study demonstrated a short-cut of the human information-processing system to deal with missing information when faced with causally linked video sequences: the event-completion effect. We explored basic processes that may be biased by an imbalance in external and internal source weighing, based on the similarity found for three groups of expertise. This indicates that the influence of higher cognitive processes in the observation of simple action events may be overruled by the human need to make sense of the world and the need to embed an event into a known structure.

Bearing in mind that we tested referees who had achieved the highest qualification level and who officiated at international FIFA matches, it is fair to surmise that the event-completion effect for simple events is hard-wired. Perceptual training programs that focus on *external sensory input* to prevent causal fillings of events will be difficult to design. Finally, the observed effect illustrates impressively that, without further game technology in the future, football players, fans, coaches, and journalists do not have to worry about losing the drama and the thrill of being defrauded by the human brain's biases.

Endnotes

[1] Note that we use the term contact moment throughout the rest of the paper to refer to both the kicking and releasing of the ball.

[2] Note that we present only one experiment event even though there were two. However, the hypotheses and the design were completely unrelated to the goal of this work and will be analyzed and published independently.

[3] Note that a between-subject design was reported in the original study but was not feasible in Experiment 1 due to the small sample of referees and time constraints during testing. In Experiment 2, we asked students to participate in the laboratory. Thus, using a between-design was possible and, additionally, allowed us to ensure that participants could not guess the purpose of the experiment when seeing both critical conditions.

[4] Note that the length of the second part of the clip was a natural consequence of the events happening in the footage of the match and not an intentional manipulation.

Additional file

Additional file 1: Appendix. (DOCX 85.6 kb)

Acknowledgments

The authors thank the Swiss television channel SF1 and the Swiss Football Association for providing the possibility to test FIFA referees, and also Oliver Höner and Florian Schultz from the Institute of Sports Science of the University of Tübingen for establishing contact with the football players. Further, the authors thank Carolin Riethmüller for her help in stimulus generation and testing the novices. Finally, we would like to thank Jeremy Wolfe and two anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions.

Authors' contributions

MH developed the research idea. MH, FP, and AM designed Exp. 1. MH and FP programmed Exp. 1. Data collection of Exp. 1 was done by all authors (FP only referees and novices, AB only football players). Data analysis was performed by MH (Bayes) and AB (glmer, contrasts). The design of Exp. 2 was proposed by the editor. Programming, stimulus generation, and data collection of Exp. 2 was done by FP. MH and FP analyzed the data of Exp. 2. Literature research and integration of the data into existing theory and practice was done by AB. AB drafted the manuscript and received critical revisions by MH, FP, and AM. All authors approved the final version of the paper.

Competing interests

None of the authors had any conflict of interest. We are not involved in any profitable use of video technology, nor do we have any business relations with the Swiss television or other television broadcasters.

Ethics approval and consent to participate

The research was conducted in accordance with APA standards for the ethical treatment of participants and supported by Swiss television (educational program).

Received: 8 February 2016 Accepted: 27 July 2016
Published online: 22 September 2016

References

Abernethy, B., Gill, D.P., Parks, S.L., & Packer, S.T. (2001). Expertise and the perception of kinematic and situational probability information. *Perception*, *30*(2), 233–252. doi:10.1068/p2872.

Bates, D., Sarkar, D., Bates, M.D., & Matrix, L. (2007). The lme4 package. *R package version*, *2*(1), 74.

Beaucousin, V., Cassotti, M., Simon, G., Pineau, A., Kostova, M., Houdé, O., …Poirel, N. (2011). ERP evidence of a meaningfulness impact on visual global/local processing: When meaning captures attention. *Neuropsychologia*, *49*(5), 1258–1266. doi:10.1016/j.neuropsychologia.2011.01.039.

Bell, K., Boshuizen, H., Scherpbier, A., & Dornan, T. (2009). When only the real thing will do: junior medical students' learning from real patients. *Medical Education*, *43*(11), 1036–1043. doi:10.1111/j.1365-2923.2009.03508.x.

Carpenter, G.A., Grossberg, S., & Arbib, A. (2003). *The Handbook of Brain Theory and Neural Networks*, (pp. 87–90). Cambridge, MA.

Chaddock, L., Neider, M.B., Voss, M.W., Gaspar, J.G., & Kramer, A.F. (2011). Do athletes excel at everyday tasks? *Medicine & Science in Sports & Exercise*, 1. doi:10.1249/mss.0b013e318218ca74.

Collins, A.M., & Loftus, E.F. (1975). A spreading-activation theory of semantic pro cessing. *Psychological Review*, *82*(6), 407–428. doi:10.1037/0033-295x.82.6.407.

Faubert, J. (2013). Professional athletes have extraordinary skills for rapidly learning complex and neutral dynamic visual scenes. *Scientific Reports*, *3*. doi:10.1038/srep01154.

Fink, G.R., Halligan, P.W., Marshall, J.C., Frith, C.D., Frackowiak, R.S.J., & Dolan, R.J. (1996). Where in the brain does visual attention select the forest and the trees? *Nature*, *382*(6592), 626–628. doi:10.1038/382626a0.

Firestone, C., & Scholl, B.J. (2015). Enhanced visual awareness for morality and pajamas? Perception vs. memory in 'top-down' effects. *Cognition*, *136*, 409–416. doi:10.1016/j.cognition.2014.10.014.

Firestone, C., & Scholl, B.J. (2015). 'Top-down' effects where none should be found: The El Greco fallacy in perception research. *Psychological Science*, *25*(1), 38–46. doi:10.1017/s0140525x15000965.

Fuster, J.M. (1990). Prefrontal cortex and the bridging of temporal gaps in the perception-action cycle. *Annals of the New York Academy of Sciences*, *608*(1), 318–336. doi:10.1111/j.1749-6632.1990.tb48901.x.

Green, D., & Swets, J. (1966). Signal detection theory and psychophysics. *Society*, *1*, 521.
Haider, H., & Frensch, P.A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology*, *30*(3), 304–337. doi:10.1006/cogp.1996.0009.
Hattie, J. (2003). *Teachers Make a Difference: What Is the Research Evidence?* Melbourne: Australian Council for Educational Research Annual Conference on Building Teacher Quality.
Intraub, H., & Dickinson, C.A. (2008). False memory 1/20th of a second later: What the early onset of boundary extension reveals about perception. *Psychological Science*, *19*(10), 1007–1014. doi:10.1111/j.1467-9280.2008.02192.x.
Intraub, H., Gottesman, C.V., & Bills, A.J. (1998). Effects of perceiving and imagining scenes on memory for pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(1), 186–201. doi:10.1037/0278-7393.24.1.186.
JASP Team. (2016). [Computer software].
Johnson, M.K., Hashtroudi, S., & Lindsay, D.S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28. doi:10.1037/0033-2909.114.1.3.
Johnson, M.K., & Raye, C.L. (1981). Reality monitoring. *Psychological Review*, *88*(1), 67–85. doi:10.1037/0033-295x.88.1.67.
Kelso, P. (2010). England v Germany: Frank Lampard's "goal" reignites goal-line technology debate. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/sport/football/teams/england/7857951/England-v-Germany-Frank-Lampards-goal-reignites-goal-line-technology-debate.html.
Lamme, V.A. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, *7*(1), 12–18. doi:10.1016/s1364-6613(02)00013-x.
Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: diagnosing X-ray pictures. In M TH Chi, R Glaser, M J Farr (Eds.) *The nature of expertise* (pp. 311–342). Hillsdale. Lawrence Erlbaum Associates, Inc.
Mann, D.T., Williams, A.M., Ward, P., & Janelle, C.M. (2007). Perceptual-cognitive expertise in sport: A meta-analysis. *Journal of Sport and Exercise Psychology*, *29*(4), 457.
Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*(1), 28–38. doi:10.1037/h0035584.
North, J.S., & Williams, A.M. (2008). Identifying the critical time period for information extraction when recognizing sequences of play. *Research Quarterly for Exercise and Sport*, *79*(2), 268–273. doi:10.1080/02701367.2008.10599490.
Peirce, J.W. (2008). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*. doi:10.3389/neuro.11.010.2008.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2006). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1–128, http://CRAN.R-project.org/package=nlme.
Potts, G.R., Keenan, J.M., & Golding, J.M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language*, *27*(4), 399–415. doi:10.1016/0749-596x(88)90064-2.
R Core Team. (2016). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna. URL https://www.R-project.org/.
Reingold, E.M., & Sheridan, H. (2011). Eye movements and visual expertise in chess and medicine. In S P Liversedge, I D Gilchrist, S Everling (Eds.) *The Oxford handbook on eye movements* (pp. 528–550). Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199539789.013.0029.
Romeas, T., & Faubert, J. (2015). Soccer athletes are superior to non-athletes at perceiving soccer-specific and non-sport specific human biological motion. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.01343.
Romeas, T., Guldner, A., & Faubert, J. (2016). 3D-multiple object tracking training task improves passing decision-making accuracy in soccer players. *Psychology of Sport and Exercise*, *22*, 1–9. doi:10.1016/j.psychsport.2015.06.002.
Rouder, J.N., Morey, R.D., Speckman, P.L., & Province, J.M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. doi:10.1016/j.jmp.2012.08.001.
Smeeton, N.J., Ward, P., & Williams, A.M. (2004). Do pattern recognition skills transfer across sports? A preliminary analysis. *Journal of Sports Sciences*, *22*(2), 205–213. doi:10.1080/02640410310001641494.
Strickland, B., & Keil, F. (2011). Event completion: Event based inferences distort memory in a matter of seconds. *Cognition*, *121*(3), 409–415. doi:10.1016/j.cognition.2011.04.007.
Tallberg, I.-M., & Almkvist, O. (2001). Confabulation and memory in patients with Alzheimers disease. *Journal of Clinical and Experimental Neuropsychology (Neuropsychology, Development and Cognition: Section A)*, *23*(2), 172–184. doi:10.1076/jcen.23.2.172.1215.
Talmy, L. (1975). Semantics and syntax of motion. *Syntax and Semantics*, *4*, 181–238.
Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., & Crundall, D. (2003). Visual attention while driving: Sequences of eye fixations made by experienced and novice drivers. *Ergonomics*, *46*(6), 629–646. doi:10.1080/0014013031000090116.
Vaeyens, R., Lenoir, M., Williams, A.M., & Philippaerts, R.M. (2007). Mechanisms underpinning successful decision making in skilled youth soccer players: An analysis of visual search behaviors. *Journal of Motor Behavior*, *39*(5), 395–408. doi:10.3200/jmbr.39.5.395-408.
Van Dijk, T.A., Kintsch, W., & Van Dijk, T.A. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.
Ward, P., Williams, A.M., & Bennett, S.J. (2002). Visual search and biological motion perception in tennis. *Research Quarterly for Exercise and Sport*, *73*(1), 107–112. doi:10.1080/02701367.2002.10608997.
Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, *19*(6), 1057–1064. doi:10.3758/s13423-012-0295-x.
Williams, A.M. (2000). Perceptual skill in soccer: Implications for talent identification and development. *Journal of Sports Sciences*, *18*(9), 737–750. doi:10.1080/02640410050120113.
Williams, A.M. (2013). *Science and Soccer: Developing Elite Performers*. Oxon, UK: Routledge.
Wolfson, S., & Neave, N. (2007). Coping under pressure: Cognitive strategies for maintaining confidence among soccer referees. *Journal of Sport Behavior*, *30*(2), 232.
Wood, J.N. (2011). A core knowledge architecture of visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 357–381. doi:10.1037/a0021935.
Zacks, J.M., Speer, N.K., & Reynolds, J.R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, *138*(2), 307–327. doi:10.1037/a0015305.
Zacks, J.M., Speer, N.K., Swallow, K.M., Braver, T.S., & Reynolds, J.R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293. doi:10.1037/0033-2909.133.2.273.
Zacks, J.M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21. doi:10.1037/0033-2909.127.1.3.

3.4 Explaining the results with the event horizon model

The discussion of Paper (2) does not reflect the diversity of approaches that were developed to understand event perception. At this point, one should at least mention the event horizon model (Radvansky & Zacks, 2011) as a possible answer to the remaining question: why do participants in our study remember details on the course of actions better in non-causal sequences even though the non-causal part elicited the perception of an event boundary? The EST proposes an attentional gating mechanisms that can enhance processing at event boundaries. However, many studies showed that experts outperform novices in (trained) attentional tasks (e.g. Bellenkes, Wickens, & Kramer, 1997; Green & Bavelier, 2003). For example, highly skilled soccer and basketball players benefited from fixating on fewer locations than novices (Helsen & Pauwels, 1993); also learned fixation patterns and a focused gaze that can draw more information from fewer fixations are thought to characterize experts (Henderson & Hollingworth, 1999), as well as a greater breadth of attention that allows soccer players to encode additional information (Williams, Davids, Burwitz, & Williams, 1994). Hence, an *attentional* gating mechanism may not catch the true reason for the null effect found for level of expertise in our experiment.

Pettijohn and Radvansky (2016) encountered the same phenomenon in one of their experiments. Instead of a declined memory due to global updating processes at an event boundary, participants' performance improved. They refer to the Event Horizon Model (see Radvansky & Zacks, 2011) that describes how the structure of an event influences the availability of information. The model's following components explain our findings as well, and do so better than the EST: (1) events are segmented and different event models are stored as *separate* traces in memory (e.g. perceptual and anticipatory traces), (2) causal relations among the events are stored as well, (3) if the attribute retrieval is noncompetitive between the stored traces, memory is facilitated, and finally, (4) attribute retrieval is interfered when the traces are competitive. It becomes clear now, that the *anticipatory* trace "won", when it received enough support from the available information structure (here: causal). In contrast, in non-causal sequences, attribute retrieval was non-competitive. There were clearly two different non-competitive events happening and the *perceptual* trace was facilitated. Components (3) and (4) make it clear that both directions are possible. The design of Study 1 did not include a control group (we did not test a condition in which causally linked sequences *without* a visible ball-contact were presented), thus we cannot say with 100% accuracy whether memory was enhanced in non-causal sequences, or whether performance was increased in causal sequences.

What have we learned? First of all, the processes involved in event perception are fundamental and based on global representations of event segments. At least for a basic scene that typically follows clear cause-effect rules that everyone has encountered before, the inferential heuristics of conceptual structures do not differ in experts and novices. Second, while the EST explains these fundamental processes well, and even allows for memory improvements based on an attentional gating mechanism, it is not elaborated enough to consider (1) the internal structures (or I-semantics) of the perceiver, and (2) the role of how the available information is structured and presented (e.g. allowing for causal relations among the stored events and, accordingly, for competition of stored traces).

Part II

CONTROLLED PERCEPTUAL PROCESSES

Perception can be directed or enriched by intentionally focusing attention on certain areas (spatial attention: Bressler, Tang, Sylvester, Shulman, & Corbetta, 2008) or on features (Beauchamp, Cox, & Deyoe, 1997). Such processes are thought to be controlled by an individual who is consciously aware, and who effortfully draws on attentional resources (Schneider & Shiffrin, 1977). Because anticipatory traces can influence perceptual reports, it is difficult to isolate attentional processes in such dynamic scenes that may comprise semantic relationships or trigger cause-effect heuristics (i.e. real-life video clips). Therefore, Study 3 used the "simple" tracking environment again to explore how intentional control of attention enriches dynamic mental representations with locally perceived object features.

Chapter 4

Study 3: Attention, intention, and perception

> In this chapter I argue that the interplay of intention and perception elicits attentional processes. The idea came from literature on conscious and unconscious awareness, which I present shortly, and I discuss two basic ideas on attention (visual search and attention capture). Study 3 presented within this framework involved four experiments that explored the effect of a task on the allocation of attention during object tracking. Obtained findings will be discussed in light of existing assumptions about the role of attention in MOT with regard to conscious and unconscious visual perception.

In order to understand the world and navigate through it, humans form mental representations. So far I have discussed how these are built with automatic processes of perception – unintentional and effortless processes that occur outside the awareness of the observer (Uleman & Uleman, 1990), following a global approach in preliminary stages of perception. However, a cognitive process that can select stimuli to be presented in mental representations is *attention*. In contrast to James (1890, p.261) who was sure that "everyone knows what attention is", the long history of the study of attention is full of debate and disagreement and if I would had to draw conclusions from my extensive literature research, I would rather agree with Pashler (1998): no one knows what attention is. The study of attention is not only about finding an adequate explanation of the phenomena, it is still about what attention actually is, and what its theories ought to explain (see Fernandez-Duque & Johnson, 2002). For example, Posner and Boies (1971) proposed three components of attention: the maintenance of a vigilant or alert state, the detection of signals for focused processing, and the orientation to sensory events. Pashler (1998) saw three different core aspects of attention: the selectivity of attention that allows us to process some stimuli more than others, the capacity limitations we experience when trying to do more than one task simultaneously and, the effort we have to invest in sustained processing of visual stimuli. Effort and capacity are only two of the terms often used synonymously for attention in literature: one also finds awareness, consciousness, control, perceptual set, and arousal – just to name a few. Most of these terms are used interchangeably in literature, but since attention covers bottom-up as well as top-down processes, they can have different meanings in different attention contexts.

Lamme (2003) focused on awareness and proposed different models (two of them are recreated in Figure 6) that theorize the terms to be related but not synonymous. Both models assume that visual awareness is limited and that attention and conscious awareness play a role in what stimulus can be *reported*. In Model A, it is assumed that attention determines what becomes conscious; attention is the same as consciousness. Model B however assumes that consciousness and unconsciousness may be entirely separated from attentional selection, that is, compared to Model A, more stimulus input reaches consciousness – but attention is needed for them to enter awareness.

The separation of consciousness and unconsciousness, and the role of spatial and feature attention inspired a new line of research that occupied a major share of my doctoral phase. I wondered whether such a division of processing levels as depicted in Figure 6[1] could explain the contradicting results in tracking literature concerning the ability of the observer to process features during multiple object tracking. It is a prominent opinion in tracking research that the attentive processes during tracking are feature-blind visual indices that stick to target objects (Pylyshyn & Storm, 1988; Shapiro, 2009); they are even described as a pre-attentive "cognitively impenetrable" mechanism (Pylyshyn, 1999) (but see Oksama & Hyönä, 2004). Recent studies considered the role of surface features during tracking and contradicted previous findings by showing an influence of feature information, but favored automatic, pre-attentive processes (e.g. Papenmeier, Meyerhoff, Jahn, & Huff, 2014). In Study 3, I attempted to connect findings from MOT, visual search, and attention capture, in order to

[1]Note that Figure 6 receives further attention in Part III (Discussion and Conclusion) in section 5.3.3, page 113.

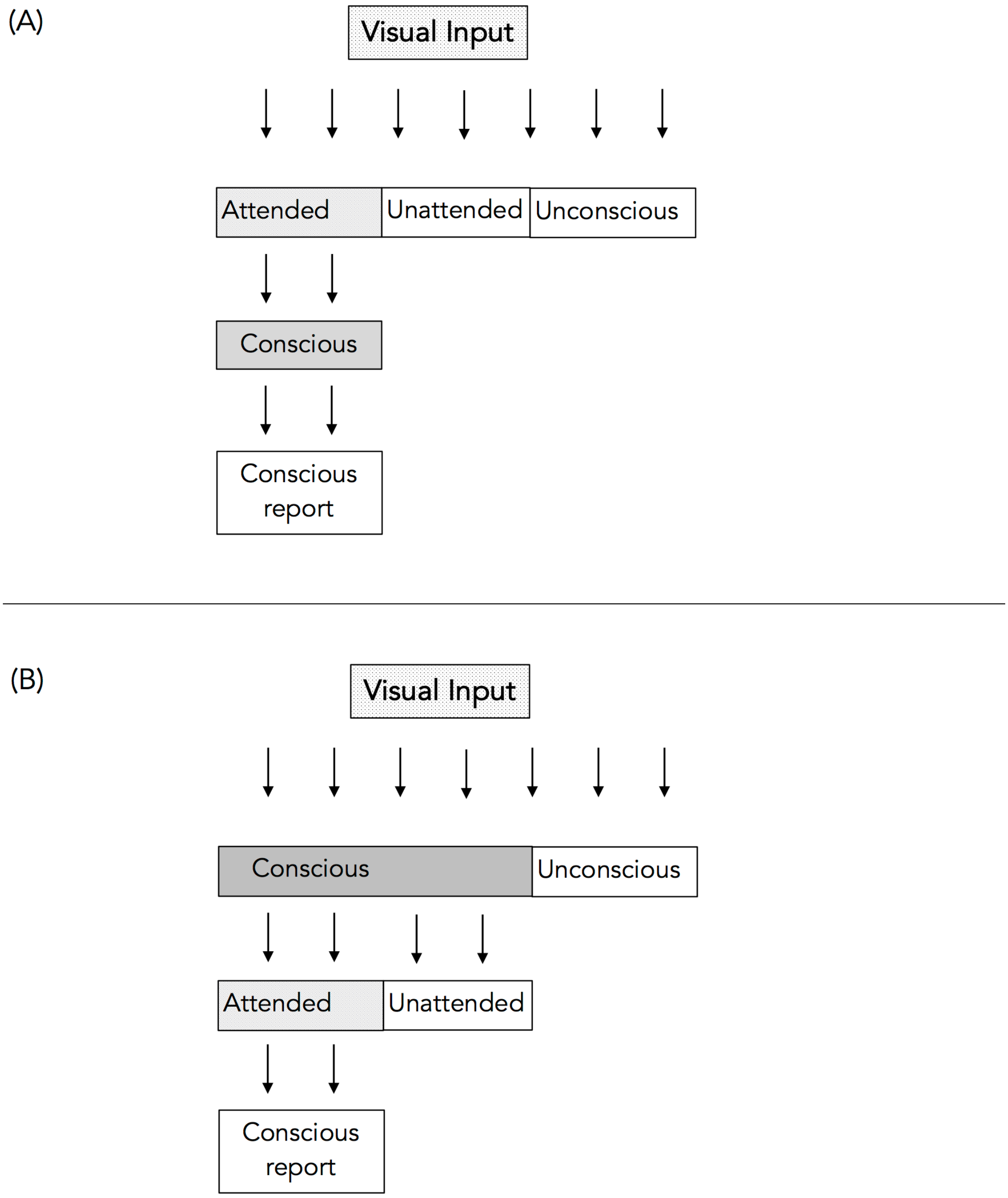

FIGURE 6: Conscious Awareness and Attention: 2 Models adapted from Lamme (2003).

examine whether tracking might also utilize a top-down component that regulates selective feature processing based on its relevance for the task. I tried to reconcile different aspects of attention that have been researched in less complex, mainly static environments, in one dynamic paradigm. That this was not an easy task to perform is reflected in the wide-ranging introduction of Paper (3). Here, I would like to introduce Paper (3) by adding even more detail to two of the main ideas.

4.1 Visual Search and Features

Treisman and Gelade (1980) proposed the "Feature Integration Theory" (FIT) which became one of the most prominent models in visual search. Based on FIT, the purpose of (spatial) attentional selection is to bind different features of an object into a single representation. In their studies, they asked participants to perform a visual search within arrays of colored letters of varying set-sizes. When objects that differed from distractors by one unique feature were targets, the search rate did not increase as a function of the set-size. However, if the searched target was a combination of features, the search rate increased the more items there were in the array. This suggests that feature search operates in parallel while conjunction search processes need to operate on items serially. Accordingly, the FIT assumes a two-stage processing architecture in which the first stage is preattentive and registers basic stimuli features in parallel across the visual field, the second stage is attentional and processes items serially, one by one. The FIT is widely accepted but has been adapted based on performance in search tasks that could not be explained with the strict dichotomy of parallel and serial processing. For example, Egeth, Virzi, and Garbart (1984) observed that participants were able to restrict their search to a subset of items (here: searching the red "O" only among the "O"s and exclude red "N"s). Further, results of Wolfe (1992) indicated that observers can exploit multiple features, searching a red vertical target among red and green horizontal distractors very efficiently even though the target is apparently not defined by a single feature. The Guided Search (Wolfe, 1994; Wolfe, Cave, & Franzel, 1989) was developed, building upon the FIT architecture but adding that serial attention allocation can be guided by information.

What happens to object processing when task demands are added? In a visual search task, for example, Lavie (1995) measured less activation of brain regions when task demands increased – reflecting decreased processing of non-selected information (Lavie, 1995). The author further observed that distractors' interference only happened in *low*-load conditions, suggesting that an actual *selection* of information is only required when the capacity limit is reached. Lavie (1995) used the Eriksen paradigm (e.g. Eriksen & Eriksen, 1974) and asked participants to search for a target letter in a multiletter display, measuring their reaction time (RT). In contrast to earlier studies that manipulated display size (e.g. Yantis & Johnston, 1990), Lavie (1995) manipulated load without affecting the salience or the quality of the stimuli. He maintained an identical display but manipulated load by making the target search more complex: a correct response was also dependent on the color feature (red or blue) of a closely flanking shape, or on the combination of both color and shape (circle or square), and additionally on the identification of the exact position. Participants only succeeded in the rejection of irrelevant information in the high-load conditions (interpreted as decreased processing of incompatible distractors resulting in less interference) – which indicates that the order of priority in the allocation of selected attention is modifiable (see also Yantis & Johnston, 1990). In other words, cognitive load determines how efficient the attentional selection will be and, most importantly, the degree of which irrelevant information is processed. Other demonstrations of attentional modulation of competing stimuli were found when the observer was required to divide attention among perceptual features. Attending to color or shape of an object resulted in the activation of many visual areas; and even more regions, including the parietal cortex, were found to be active when the observer was asked to *switch* between feature and shape (Le, Pardo, & Hu, 1998).

4.2 Attentional Capture

Wolfe (1992) came to the following conclusion (p. 762):

"[...] there appear to be two aspects to parallel processing in visual search: a bottom-up, stimulus-driven aspect and a top-down, cognitively- or strategically-driven aspect. The bottom-up component directs attention to unusual loci including isolated, unique items and borders. We may conjecture that this component is responsible for the subjective experience of 'pop-out' [...]. The top-down component allows for parallel selection of all items with a given attribute [...]. It allows for goal-directed use of preattentively processed information."

Attention capture (including "pop-out" effects) has been studied extensively and it has been proposed that it is not necessarily a bottom-up effect. The attentional control setting (top-down) of the observer plays an important role. Monsell and Driver (2000) reviewed the hotly debated problem of whether attentional capture by task-irrelevant stimuli is determined by bottom-up, stimulus-driven mechanisms or by top-down factors. Over the past two decades proponents of both sides cumulated evidence that support each of their theories. Supporters of the stimulus-driven hypothesis (e.g. Theeuwes, 1992; Yantis &

Jonides, 1990) showed for example that a task-irrelevant onset stimulus captured attention in spite of intentions to the contrary (e.g. Franconeri, Simons, & Junge, 2004; Theeuwes, 1995; Turatto & Galfano, 2000). A salient stimulus captures attention regardless of whether it shared properties with the target or not. In contrast, supporters of the contingent capture hypothesis (e.g. Bacon & Egeth, 1994; Folk, Remington, & Johnston, 1992; Gibson & Kelsey, 1998) claim that such a salient distractor object can only capture attention when the distractor's feature matches the attentional control setting of the observer. It was shown that the onset of a task-irrelevant stimulus failed to capture attention when the goal was to find a color target (e.g. Folk, Leber, & Egeth, 2002; Folk & Remington, 1998, 1999; Folk et al., 1992).

That is, under given circumstances, only a color stimulus, consistent with the attentional control setting, captures attention. Whether a stimulus captures attention highly depends on the observer's attentional readiness, or his attentional control set (in the following referred to ACS).

Another area of prominent studies that integrate both bottom-up and top-down aspects is the area of spatial cueing. Spatial cues (e.g. gazing eyes or pointing arrows) with no predictive value attract attention to a spatial location when presented briefly and prior to a target presentation (see also Friesen & Kingstone, 1998; Yantis, 1993; Yantis & Hillstrom, 1994; Yantis & Jonides, 1990). In contrast, the same spatial cueing studies also support the notion of endogenous control of spatial attention in which input is prioritized in accordance to goals and intentions of the observer. When given correct information about the location of an upcoming target, participants detected the presence of a stimulus faster than when given incorrect information (Carrasco et al., 2004; Eriksen & Hoffman, 1973; Posner, 1980; Posner, Snyder, & Davidson, 1980; Remington, 1980; Remington, Johnston, & Yantis, 1992).

4.3 Study 3: Feature-based tasks in multiple object tracking

In Study 3, I explored whether one can actively change perceptual processes by changing the level of *attentional* representation. In the tracking environment used, I presented the observer with cartoon eyes that displayed (spatial) gaze cues: seven (moving) eyes looked at one (moving) singleton. By combining an ACS based on features (gazing vs singleton) and status (target or distractor object) with a tracking task, I showed successfully that features can indeed be processed when they are actively attended. The design further included two different gazing behaviors. The eyes either stared (stalking) or looked at the single object twice (flirty). This was implemented to differentiate between attention capture and feature processing. The allocation of attention was investigated with a novel approach. The dependent variable was not how successful the participants tracked the objects (mean proportion), but how often participants marked the singleton (the neutral object) compared to its related group (either gazing targets or gazing distractors).

Brockhoff, A. & Huff, M. (2016). All eyes on relevance: strategic allocation of attention as a result of feature-based task demands in multiple object tracking. *Attention, Perception, & Psychophysics*, *78*(7), 2090–2109. doi:10.3758/s13414-016-1129-0.

Original article published in: Attention, Perception, & Psychophysics
October 2016, Volume 78, Issue 7, pp 2090–2109, doi:10.3758/s13414-016-1129-0

All eyes on relevance: strategic allocation of attention as a result of feature-based task demands in multiple object tracking

Alisa Brockhoff & Markus Huff
University of Tübingen

Multiple object tracking (MOT) plays a fundamental role in processing and interpreting dynamic environments. Regarding the type of information utilized by the observer, recent studies reported evidence for the use of object features in an automatic, low- level manner. By introducing a novel paradigm that allowed us to combine tracking with a noninterfering top-down task, we tested whether a voluntary component can regulate the deployment of attention to task-relevant features in a selective manner. In four experiments we found conclusive evidence for a task-driven selection mechanism that guides attention during tracking: The observers were able to ignore or prioritize distinct objects. They marked the distinct (cued) object (target/distractor) more or less often than other objects of the same type (targets /distractors) – but only when they had received an identification task that required them to actively process object features (cues) during tracking. These effects are discussed with regard to existing theoretical approaches to attentive tracking, gaze-cue usability as well as attentional readiness, a term that originally stems from research on attention capture and visual search. Our findings indicate that existing theories of MOT need to be adjusted to allow for flexible top-down, voluntary processing during tracking.

Keywords: Multiple object tracking, Multiple identity tracking, Attentional control, Attentional resolution, Attention allocation, Goal-directed attention

In dynamic real-life environments, keeping track of relevant objects can be a rather challenging task. For example, in crowded inner-city pedestrian areas, multiple salient cues (ads, blinking lights, etc.) attract visual attention that might distract from keeping track of family members (e.g., children or the dog): stimulus-driven processes interact with goal-directed tasks during attentive tracking. In the present manuscript, we explore how these processes affect the distribution of visual attention in dynamic scenes and contribute to the human ability of multiple object tracking (MOT).

Attention and focus

It is a well-known fact that attention is influenced actively or passively (James, 1890). We can actively exert control over the allocation of attention (top-down; e.g., Yarbus, 1967) or the deployment of attention happens passively as a result of an event in the environment (bottom-up; e.g., Pratt, Radulescu, Guo, & Abrams, 2010). The question of how narrowly attention can be focused has been characterized as a question of attentional resolution. Based on different metaphors that described visual attention (including attention as a spotlight: Posner, 1980; as a nonlinear filter: Cutzu & Tsotsos, 2003; as a zoom lens: Eriksen & James, 1986), researchers wondered about the limitations of spatial extent of focused attention and suggested that the area of selection has a facilitatory-center-inhibitory-surround profile (e.g., Cutzu & Tsotsos, 2003). A detailed discussion of all approaches and findings in the area of attention selection, facilitation, and inhibition would go beyond the scope of this work. However, a finding that is important for the studies we report here is that the shape of the focus varies depending on task and characteristics of stimuli (e.g., Eriksen & James, 1986; Laberge & Brown, 1986). Eriksen and James (1986) built upon existing research methods in which participants were asked to discriminate between the letters S and C as quickly as possible. In their static display they manipulated the number of cued positions and the discriminative difficulty with incompatible noise letters, and measured reaction times. Their gathered data indicated that participants were able to change the size of the attentional focus and drew upon additional attentional resources when the number of cued positions increased. The authors interpreted their findings as support for the zoom lens model.

Similar theoretical approaches were postulated later. Morgan, Ward, and Castet (1998), and Morgan and Watt (1997)

Alisa Brockhoff, Faculty of Science, Department of Psychology, University of Tübingen.
Correspondence concerning this article should be addressed to Alisa Brockhoff, Faculty of Science, Department of Psychology, University of Tübingen, Schleichstr. 4, 72076 Tübingen.
E-mail: alisa.brockhoff@uni-tuebingen.de

suggested that cues do not elicit a switch to a higher spatial frequency passband, but act upon the scale of spatial frequency analysis. Thus, focusing attention (e.g., with a cue) allows processing with a smaller spatial range, resulting in a decreased influence of distractors and increased target acuity.

Attention and cognition

Observed effects of selective and focused attention in several cognitive tasks also reflect a combination of different mechanisms such as noise reduction, signal enhancement, and decisional as well as intentional factors (Folk, Remington, & Johnston, 1992; Lu & Dosher, 1998; Morgan et al., 1998; Yeshurun & Carrasco, 1998, 1999). The compelling demonstrations of both early and late selection imply that attention is located flexibly (see also Vogel, Woodman, & Luck, 2005).

Several authors have argued further that passive/exogenous signals can only capture attention when the observer has an optimal attentional control set (ACS; Folk et al., 1992). Yantis and Egeth (1994) reported that response times to singleton targets are highly sensitive to the relevance of the singleton (see also Folk & Annett, 1994; Hillstrom and Yantis, 1994; Jonides & Yantis, 1988). That is, it highly depends on the observer's "attentional readiness" (top-down or goal-directed control, as described in Egeth & Yantis, 1997) or its equivalent, the ACS (as used as the generic concept for attentional readiness throughout the rest of this paper; e.g., Folk et al., 1992; Folk, Remington, & Wright, 1994) to find an L in an array of Ts (Joseph & Optican, 1996). Others also recently argued that, although exogenous cues might work to orient attention in space, the strength of the effect may be endogenously modulated by an ACS that the observer adopted in line with a given goal or task (Lupiáñez, Milliken, Solano, Weaver, & Tipper, 2001).

Attention and spatial resolution

The widely accepted view is that attention is used to control the allocation of limited perceptual processing resources in various ways. The described studies present evidence that attention can increase or decrease its resolution as a result of spatial occurrences (e.g., crowding). Furthermore, attention changes the range of spatial analysis (e.g., cueing). And finally, attention itself is modified by task and intention of the observer (e.g., attention capture and control settings). Since the first statement concerning spatial resolution has already been explored in dynamic environments (e.g., Franconeri, Jonathan, & Scimeca, 2010; Intriligator & Cavanagh, 2001), the current paper is concerned with the effects of a flexible attention allocation and cue usability that result from feature-based task demands in a multiple object tracking paradigm (MOT; Pylyshyn & Storm, 1988).

Interestingly, while the common understanding of resolution (e.g., of a TV or a picture) rather describes how many details the observer can see, attentional resolution during tracking is usually associated with locations and the selection and separation of objects (see Intriligator & Cavanagh, 2001), and not with the perception of object features. From a technical point of view, resolution is measured as how closely lines can be resolved in an image, and the clarity of an image depends in fact on its spatial resolution (and not as, often erroneously believed, on the number of pixels per inch, ppi). The current paper therefore adopts, and by that extends, the term attentional resolution to describe how a task changes attentional resolution to represent object features with more clarity during MOT. Following Luck, Hillyard, Mouloua, and Hawkins (1996)—who proposed that there are multiple selection mechanisms that operate at different processing levels to control different types of interference and attentional overload (see Cave & Bichot, 1999, for a review)—we expect the pattern of attentional resolution to change, depending on the mechanism triggered by the demands of the task. That is, we propose that a featural task demand will trigger and activate a feature level of attention that influences the allocation and resolution of attention for object features in a dynamic tracking environment.

Attention and MOT

A typical MOT experiment presents the subject with a number of identical objects of which some are briefly flashed to indicate that they are the targets to be tracked. Then, the objects become indistinguishable from each other and move. When the motion stops, the observer is asked to click on the objects that were marked as targets before. The MOT task has been originally invented to show that observers track objects in parallel following a preattentive mechanism (Pylyshyn, 1999). As described by Pylyshyn (2007), visual attention uses indexes that stick to the moving targets, and these indexes are attracted to moving objects in a bottom-up manner.

However, the described indexes serve the purpose of providing a structure for guiding focused attention needed in specific situation, for example, in situations of object crowding in which the observer has to prevent confusions. That is, the basic MOT mechanism is a preattentive "cognitively impenetrable" mechanism (Pylyshyn, 1999) but may pave the way for top-down influences. Still, the role of visual attention in tracking has been approached in several different theoretical ways, and no agreement has been reached on the issue. Cavanagh and Alvarez (2005) proposed that multifocal attention covers the objects simultaneously and independently. Alternatively, Alvarez and Franconeri (2007) suggested that a limited attentional resource is allocated flexibly toward objects, increasing local attentional resolution and enhancing tracking performance. Iordanescu, Grabowecky, and Suzuki (2009) supported the flexible idea with their study on the dynamic adjustment of the spatial distribution of attention in an MOT environment based on spatial demands (e.g., crowding). They blanked out the MOT task and asked participants to localize the position of targets on the blank screen. Their results, namely that the accuracy of correctly marked targets increased when the distance of the nearest distractor decreased, indicated that the attentional allocation to individually tracked target objects changes dynamically. That is, close distractors lead to an increased attentional resolution (e.g., to prevent confusions), or to a modulation of the local attentional resolution, possibly by a top-down component, as you will. Increasing or decreasing the attentional resolution based on current demands has only been investigated for spatiotemporal occurrences, like crowding or overlapping objects, but has never been investigated for goal-related processing of feature information.

Location, features, and MOT

Concerning top-down effects for object features during tracking, Feria (2012) was the first to study the effects of distinct object features on tracking performance. The utilized distractors were either identical or featurally distinct (in one or two dimensions) from the targets; the number of distractors per trial was varied. Her findings indicated that the effect distractors have on tracking is top-down in nature: differently colored or shaped, or motionless distractors impaired tracking less than target-identical distractors—still, this was only the case when tracking load was low. With her study she generalized previous findings from visual search to MOT: The effect of a distractor object is dependent on sharing the features of the target, indicating that the role of distractors as well as distractor features may have been underestimated in its influence on tracking performance.

Still, most theories on MOT have a strong focus on spatiotemporal information as the only source used (see, e.g., Cavanagh & Alvaraz, 2005; Oksama & Hyönä, 2008; Pylyshyn, 1989). Nonetheless, a considerable amount of tracking studies has found effects that can be attributed to information access to features during tracking (e.g., Cohen, Pinto, Howe, & Horowitz, 2011; Drew, Horowitz, & Vogel, 2013; Makovski & Jiang, 2009a, b; Oksama & Hyönä, 2008). While typically irrelevant for tracking (Makovski & Jiang, 2009a, b; Pylyshyn, 2004), features are used when spatiotemporal information is(made) unreliable (Bae & Flombaum,2012; Papenmeier, Meyerhoff, Jahn, & Huff, 2013). This indicates again, that attention to different sources of information (here: object's feature or location) is somewhat flexibly.

MOT, features, and attentional resolution

With the present research, we explore the role of top-down modulation of attention allocation and feature processing in MOT. We hypothesize that an increased or decreased attentional resolution can also happen as a response to a featural demand, shifting attention toward or away from a feature singleton within an otherwise homogenous crowd of objects. Following the idea of an ACS, we hypothesize that based on its relevance for the tracking task, target and distractor singletons will lead to changes in the allocation of attention; that is, an intentional focus on feature singletons will lead to a strategic use of helpful target singletons and an inhibition of tracking-harming distractor singletons. While being aware however that our study does not explicitly test attention capture, we are convinced that a task-related manipulation of the relevance for features in a dynamic environment bares a practical share to the induction of an ACS used in attention capture settings. To understand the factors that control attentional resolution and possible effects of accessing object features, it is important to be able to systematically vary the locus of selection within a single paradigm. We present here a new method of manipulating the attentional resolution toward a single object feature in an MOT task. By using dynamic gaze cues (cartoon eyes with moving pupils), that cued either a single target or a single distractor among the objects by looking at them, we integrated minimal featural gaze cues within an MOT task (see Fig.1).

The main objective was to enhance the attentional resolution to allow for feature processing by influencing the allocation of attention. In three of the four experiments presented here, seven dynamic cartoon eyes cued one neutral pair of eyes by gazing at it. Such a "dual function" of objects—that is, the observer can process the objects by following the spatial cues, or by processing featural distinctiveness, or both—was inevitable in a dynamic cueing design. The combination of gaze cues and MOT had an exploratory character with two possible outcomes: Either participants use the cue itself reflexively, or they only focus on the different featural aspects of the objects, or both.

We implemented two gazing behaviors in order to control for the possible difference between feature encoding and cue processing, and to avoid difficulties in the later interpretation of the results. The gaze cue we presented was either constant ("stalking") or intermittent ("flirty"). The stalking condition presented stimuli with constant orientation toward the cued object and thus gave the participants 8 full seconds to encode object features. The flirty condition however presented a short gaze twice during a trial and we assumed that feature encoding would be very difficult, that is, we hoped to observe reflexive gaze-following effects.

Attention and gaze

The rational behind introducing the gazing behavior of the cartoon eyes was that gaze perception can direct attention reflexively, performed by an innate module (Eye Direction Detector; Baron-Cohen, Campbell, Karmiloff-Smith, Grant & Walker, 1995) and also known as joint attention, mainly studied in infants (e.g., Farroni, Massaccesi, Pividori, & Johnson,

Figure 1. Example of the stimuli. Left: Eyes with neutral gaze. Middle: Eyes cueing one of the objects. Right: Eyes with color cue (used in Experiment 4).

2004; Scaife & Bruner, 1975). Findings indicated that infants followed the eyes of the speaker and turned their attention to the looked-at object, a behavior that promotes the acquisition of language (e.g., Baldwin, 1992, 1993). The spatial cueing paradigm and its variations (Posner, 1980; Posner & Cohen, 1984; Posner, Nissen, & Ogden, 1978) showed that gaze keeps playing a strong role for the orientation of attention throughout the life span. Friesen and Kingstone (1998) studied the effects of gaze cues on the orienting of attention in adults. Presenting faces in the center of the screen whose eyes either gazed right or left resulted in facilitated reaction times when the target appeared in the gaze-cued area.

Gaze and attentional control

However, Driver et al. (1999) informed participants in one of their experiments that the target was four times more likely to appear at the noncued side. In contrast to the assumption of an automatic, reflexive effect of gaze, they found that participants eventually shifted their attention voluntarily in the opposite direction of the cue. Furthermore, a recent study by Macdonald and Tatler (2013) provided insights into how gaze cues are actually used in real-life scenarios. Participants had to build a given structure with colored blocks, receiving either ambiguous or unambiguous instructions. They found that participants only used the gaze cues of an attending experimenter when instructions were ambiguous, that is, when the gaze cue provided information that was helpful to solve the task. This suggest a strong influence of task demands on the effects of gaze cues.

Experimental overview

In three of the four studies presented here we exploit the property that spatial cues focus attention on an area in visual space and that the allocation of attention has a large effect on feature detection and encoding (Treisman & Sato, 1990; Theeuwes, Kramer, & Atchley, 1999; Treisman & Gormican, 1988). However, this has only been tested with static displays and visual search so far.

In Experiment 1, participants tracked one round without further information. Before the second round of tracking, they received the task to identify the objects' behavior (cueing constantly or intermittently) and, additionally and more importantly, to identify the cued object (target or distractor singleton). Instead of measuring proportion correct, we measured the selection preference (proportion marked) for the different object types displayed: targets, distractors, and the singleton (cued distractor or cued target), to provide evidence that facilitated feature processing based on task demands influences the allocation of attention flexibly among the objects. This approach is then used in Experiments 2 and 3 to demonstrate that the effect is truly task-based and not due to inevitable learning. Using the same methodological design with different stimuli, we show in Experiment 4 that the use of featural gaze-cues is basically identical to color-cues, indicating that cue usability in MOT relies upon the same flexible and intentional attentional enhancement for object features.

General Method

Overview

In the experiments reported here, we used the MOT task developed by Pylshyn and Storm (1988) in combination with an identification task. In three of the four experiments, there were two rounds of tracking. Round 1 was with no task or further information. In Round 2 participants received information on the objects' behavior and the task to identify it after each tracking trial. In Experiment 2, we added a third round in which participants still knew about the eyes' behavior but were explicitly informed that there will be no identification task.

Stimuli

Two types of stimuli were used. In the first three experiments, we used cartoon eyes that gazed at one of the target, or one of the distractor objects (see Fig.1). In Experiment 4, we used the same eyes but instead of a gaze all objects except the single target or the single distractor were colored grey (see Fig.1). The cue was either displayed constantly (stalking) or intermittently (flirty); 500 ms before the motion stopped, all objects went back to their neutral initial position.

Phase 5 (neutral eyes) was displayed until the end of the trial in flirty trials. In stalking trials, the eyes switched to the neutral position 500 ms before they stopped moving. Thus, in both conditions it was not sufficient to pay attention to the cues at the end of the trial only. Correct identification of the objects' behavior and the cued object required sustained attention throughout the entire trial.

Procedure

At the start of each session, participants received the Reading the Mind in the Eyes test (Baron-Cohen et al., 2001; Bölte, 2005). This was followed by instructions for one round of tracking (64 trials). Except for Experiment 3 (tracking without an additional identification task), participants then received specified instructions on the eyes' behavior, and a multiple-choice question appeared on the screen after each trial. In Experiment 2, participants tracked an additional round without a task. The tracking procedure was a standard procedure, which is depicted in Fig.2, which also depicts a flirty trial.

Exclusion criteria and data analysis

In each of the experiments described in this work, we tested participants on their eye-reading skills with the Read the Mind in the Eyes test by Baron-Cohen et al. (2001). It seems counterintuitive that a test that measures the ability to correctly judge an emotion from gaze is appropriate to use as a baseline for the sensitivity to gaze cues that are directional in nature. However, recent findings suggest that gaze direction and facial expression are not independently processed; in fact, gaze direction is considered as an important cue in the perceptual processing of facial displays of emotions (e.g., Adams & Kleck, 2003). Further, the Read the Mind in the Eyes test is frequently used (as part of a test battery) in clinical research and considered to be equally reliable for participants in treatment as well as for healthy participants in control groups. Reflexive orienting to gaze cues can be reduced or absent in people who suffer from psychiatric disorders (e.g., Asperger's or autism; see Nation & Penny, 2008, for a review). Thus, with relatively simple means, we attempted to ensure a rough comparability of sensitivity to gaze cues among participants. Additionally, data of trials in which the participant did not correctly identify the condition (see Appendix A for chi-square tests of the differences in proportion correct), that is, failed to process the task, were excluded from the analysis. Such an exclusion of data was justifiable from a theoretical point of view: we needed to ensure that attention is actually controlled by top-down processes (task).

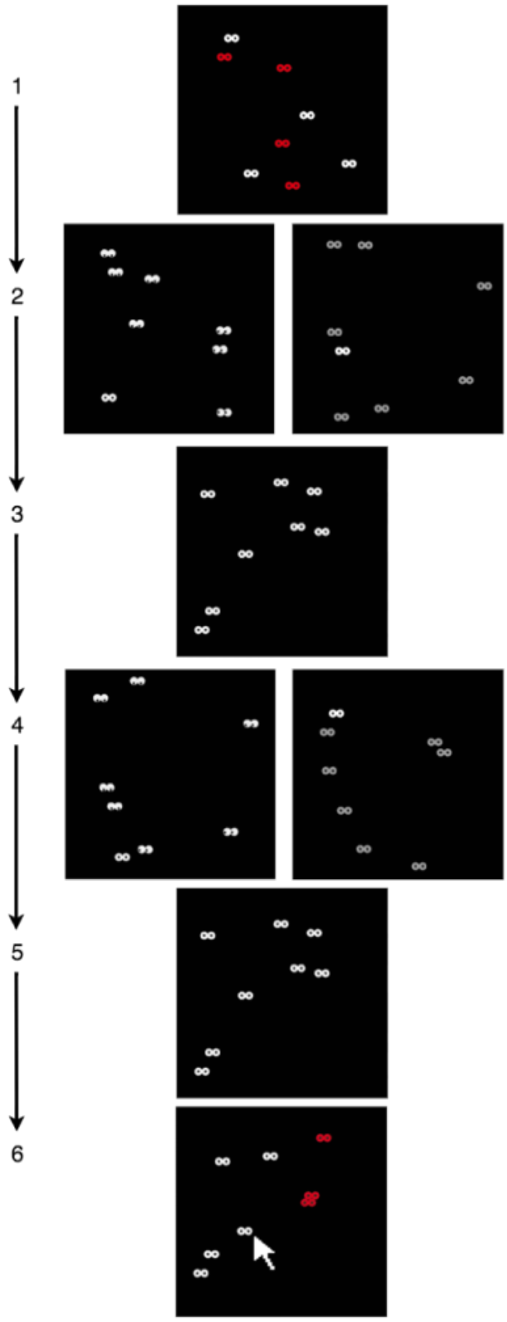

Figure 2. Example of a Flirty trial. Left: gazing eyes. Right: Shaded eyes. Numbers indicate the sequential differences displayed. Note that the flirt moments (2 and 4) appeared randomly at 2 of 3 predefined moments during the trial. In Stalking trials, the cueing phase was constant and lasted through sequences 2 to 4.

Considering the goal of the study, namely to explore the influence of "active" task-related attention, we believe that an inclusion of trials in which it is impossible to tell what actually happened during tracking (e.g., the participant only tracked the objects without trying to identify the condition and guessed, or the participant accidentally tracked a distractor and came to wrong conclusions when asked which object has been cued) bares the risk to lead to false conclusions based on unknown and uncontrolled factors in the analysis.

The planned contrasts were kept within one gazing behavior (flirty and stalking), that is, we only compared the proportion of the selection of noncued objects versus cued objects in flirty or stalking trials. In other words, we tested a subset of possible main effects for the type of object (cued vs. noncued), and none of our comparisons involved both the flirty and the stalking trials, nor Round 1 and Round 2 trials. In the fifth contrast, though we analyzed the difference of the differences of each behavior. However, due to differences in proportions correctly identified we did not give this calculation too much weight in the result section. Based on our specific a priori expectations that noncued objects would be marked less often than cued objects when the participants' attention was influenced by task-demands, an omnibus F test like a repeated-measures ANOVA and subsequent pairwise comparisons would have resulted in an inflated Type I error and were thus not the most appropriate analysis (see Maxwell & Delaney, 2004; Ruxton & Beauchamp, 2008). We chose planned contrasts that derived from specific hypotheses (see Table 1). We further applied Bonferroni corrections to reduce the risk of type I errors and to compensate for the fact that our set of chosen contrasts was not orthogonal.

EXPERIMENT 1

Method

Participants. Forty students (28 female, 12 male; mean age M = 22.63 years, SD = 3.32), of the University of Tübingen participated in return for course credit or monetary compensation. All subjects had reportedly normal or corrected-to-normal sight. Informed consent was obtained after the experimental procedures were explained to the subjects.

Stimuli and design. PsychoPy (Peirce, 2009) was used to present stimuli on a 15.4-in. notebook. The experiment was divided into three parts. The participants started with a revised version of the Reading the Mind in the Eyes test (Baron-Cohen et al., 2001; Bölte, 2005) that took them approximately 10 minutes to complete. Stimuli consisted of 37 photographs of human eye-regions, the first one being a practice trial. These photographs were taken from the revised version of the Reading the Mind in the Eyes test (Baron-Cohen et al., 2001) translated into German by Bölte (2005). Four different adjectives were given. The participant indicated the adjective that described the mental and emotional state of the person displayed best.

Next, participants were given instructions for a MOT task. Each trial started with the appearance of rectangular outlined space against a black background. After 0.5 seconds, the stimuli appeared. After another 0.5 seconds, four of the stimuli were marked as targets by flashing red for four times. Each flash and each pause in between lasted for 0.2 seconds, while the last flash lasted for 0.4 seconds. The stimuli were eight artificial, cartoon eyes with a white sclera and black pupils and had a diameter of 2.6°x1.3° (see Fig.1). They moved at a constant speed of 8°. The duration of a trial was 8 s, and participants were instructed to select the target items by marking them after each trial. All objects were able to overlap during each trial and they bounced off of the box in a way that was physically correct. Gaze direction was manipulated by moving the black pupils as follows: The eyes either constantly stared ("stalking") at a specified object or looked at it shortly ("flirted") for twice during each trial.

The two "flirt moments" were randomly chosen out of three possible time points (in order to avoid predictability) and lasted about 1.5 s each (that included movement of the pupil toward the object, moment of glance, and the aversion of the eyes). The specified object either belonged to the defined set of targets or was one of the distractors. Each of the four conditions (i.e., the four ways in which the eyes could behave) occurred twice in each of 16 blocks, with two blocks of practice (which were not further analyzed). Participants tracked two rounds of 64 trials, so 128 trials in total. Round 1 was with no task or further information. In Round 2, participants received information on the objects' behavior and the task to identify it after each tracking trial. The first round of tracking took approximately 20 minutes to complete; the second round took 25–30 minutes. Before the second round of tracking, participants were actively engaged in attending to the objects' behavior during tracking (as in Huff, Papenmeier, & Zacks, 2012). The research assistant handed out a written description of the behavior and the resultant answer options (translated):

In the third part of the experiment, you will track pairs of cartoon eyes again. This time you not only have to track the predefined targets, but also have to pay attention to how the eyes behave. There are four options that will be given as a multiple-choice after each tracking trial:

1. The eyes stared constantly at a target object.
2. They eyes stared constantly at a distractor object.
3. The eyes looked twice quickly at a target object.
4. The eyes looked twice very quickly at a distractor object.

The research assistant stayed in the room to answer individual questions of the participants in case necessary and left

ALL EYES ON RELEVANCE

Table 1

Research Hypotheses for target and distractor trials (Experiment 1, applicable to 2 and 4 as well)

Type of Influence		Selection preferences	Contrasts
Round 1, Stalking Round 1, Flirty	H_1 H_2	*Cued objects will not be marked more often than non-cued objects.*	S: CT vs NCT S: CD vs NCD F: CT vs NCT F: CD vs NCD
Round 2, Stalking Round 2, Flirty	H_3 H_4	*If participants received information and task, they will be attentionally ready to process features. Cued objects will be marked more often than non-cued objects when they are relevant for the identification task.*	S: CT vs NCT S: CD vs NCD F: CT vs NCT F: CD vs NCD
Round 2: Flirty vs Stalking	H_5	*We explore the effects of two cue intensities. Differences between the two conditions may be explained with the observers reacting to the cue (*Flirty*) or relying on the distinct object feature constantly available (*Stalking*).*	S: CT - NCT vs F: CT - NCT S: CD - NCD vs F: CD - NCD

Notes. S= Stalking, F = Flirty, CT= Cued Target, NCT= Non-cued Targets, CD = Cued Distractor, NCD = Non-cued Distractors

before she or he started the second tracking round. The participants were further informed that they would have to rate how confident they were about their answers (5: maximally confident–1: not at all). The tracking task with questions took about 25 minutes.

Results

With an average of 25.07 correctly identified gaze expressions in the Reading the Mind in the Eyes test, subjects in the present study scored only slightly lower than the normal control group suggested by Baron-Cohen et al. (2001; 26.2–30.9 of 36 gaze expressions in total). No participant was excluded from the analysis. Overall, participants tracked 2.87 of 4 objects correctly. There was no significant effect for round; that is, the participant did not track less or more objects in the second round with the identification task than in the first round without an additional task, $t(39) = 0.53$, $p = .598$, $d = 0.08$.

The main analysis focused on the mean proportions of object selection preferences (object marked) on the two levels of object status and the two levels of cueing behavior. We operationalized the ACS as trials in which participants identified the eyes' behavior correctly, resulting in the inclusion of 64 % of the total number of trials of Round 2. The planned contrasts were based on a linear mixed-effects model, fit via maximum likelihood, with participants as the random effect and the variables condition (Round 1, Round 2), object status (Cued, Non-Cued) and cueing behavior (Flirty, Stalking). A priori hypotheses were conducted using Bonferroni adjusted alpha levels of .01 per test (.05/5). We report Cohen's d as effect size measures. The planned contrasts for target trials were not significant in Round 1, neither in the Stalking condition, $|z| = 0.697$, $p = .928$, $d = 0.01$, nor in the Flirty condition, $|z| = 0.22$, $p = .999$, $d = 0.01$. In Round 2, both

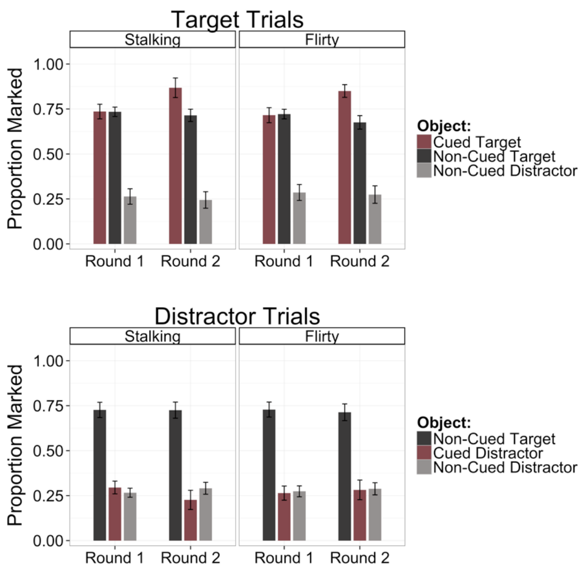

Figure 3. Proportion marked of the different object types displayed in both rounds for cued target and cued distractor trials depending on cueing behavior. Error bars indicate 95% within-subject confidence intervals.

contrasts reached significance, Stalking $|z| = 5.40$, $p < .001$, $d = 0.99$; Flirty $|z| = 6.14$, $p < .001$, $d = 1.37$. The distractor trials showed a similar picture for Round 1, but showed no significant differences between the selection of cued distractor and non-cued distractors. See Appendix B for complete tables. Refer to Fig.3 for a graphical display of the analyzed target distractor trial data.

In the first round, the participants did not process object features, i.e. they did not show a selection preference for the cued target. In contrast, in the second round in which the cued single target was of relevance for both tasks (tracking and answering the cueing behavior question), they selected it significantly more often compared to the non-cued targets. The same analysis was done for distractor trials. Here, we would expect that the cued single object (one of the distractors) is either ignored strategically (i.e. marked less often than the non-cued distractors) or not processed differently at all, since both, the tracking task and the question, could be solved without guiding attention away from the four target objects. Statistically, there is no significant difference between the compared means.

Discussion

The first experiment introduced an ACS manipulation before the second round of tracking. Information and task showed strong effects on the selection preference for the cued target but not for the cued distractor. Whether the objects were flirted with or constantly stared at did not influence the strength of the selection preference for cued- target trials. It is important to note, that the results of Round 2 cannot be explained by strategically paying attention only to the very end of each trial. First, 500 ms before the end of a trial all objects were indistinguishable (i.e. the pupils turned into the neutral position). Second, even if there might have been some strategic processing in the Stalking condition, the randomly chosen cueing intervals in the Flirty condition presumably impeded such a strategy. As the results of the Flirty and Stalking conditions in Round 2 were comparable, we propose that the ACS did not just trigger a strategy that selects the relevant information when asked for but rather changes the distribution of spatially distributed visual attention during the whole trial. For the cued-distractor trials, we only observed a close to significant inhibition effect for the cued distractor: it was marked less often than the non-cued distractors.

The research assistant stayed in the room to answer questions after the participant had received the instructions for Round 2 and reported that the majority of participants had questions about the explanations on distractor trials. Even though participants scored well over the multiple-choice guessing rate of 25 % (about 65 % of trials correctly identified), we wondered whether the influence of the ACS would change with revised instructions. Considering the indistinct effect for cued-distractor trials and taking the concerns about the instructions into account, we conducted a second experiment in which we handed out an illustrated (as opposed to the former written version) instruction to the participants. Furthermore, a third round of tracking was added in which no identification of the cueing behavior was required. That is, in Round 3 the participants still knew about the behavior of the different object types. If results of Round 3 resemble those of Round 2, we could derive that knowledge (without the task) is capable to amend the distribution of attention. But if results of Round 3 rather resemble those of Round 1, the difference in processing of cued objects owes its existence to the task demands that decreases relevance for cued distractors, and heightens relevance for cued target objects. Experiment 2 was thus not only conducted to replicate results of Experiment 1 with modified instructions, but also to consider and extract the different roles task and information play in eliciting the previously observed effect.

EXPERIMENT 2

Method

Participants. Thirty-two new students of the University of Tübingen participated in return for course credit (24 female and 8 male, aged $M = 22.42$ years, $SD = 3.64$). All subjects had reportedly normal or corrected-to-normal sight. Informed consent was obtained after the experimental procedures were explained to the subjects.

Stimuli and design. The design of the experiment was similar to Experiment 1, except that we added a third tracking round. Participants tracked 192 (3 x 64) trials in total. The first and the second round were identical to the former experiments: the participants tracked the first round without a task, and the second round with a task. Before the third round started, the participants were informed that the eyes will move and gaze in the exact same way, but that they will not be asked to identify the eyes' behavior at any point during Round 3. The participants also did the "Read the Eyes in the Mind Test". We excluded no participant based on the test performance. The stimuli were not changed. However, we revised the instructions that were now illustrated with a series of screenshots and detailed descriptions of the real stimuli, one for cued-target and one for cued-distractor trials.

Results

With an average of 25.35 correctly identified gaze expressions in the Reading the Mind in the Eyes test, subjects in the present study scored only slightly lower than the normal control-group suggested by Baron-Cohen et al. (2001; 26.2 - 30.9 of 36 gaze expressions in total). No participant was excluded from the analysis. There was no significant effect for round that is the participant did not track less objects in the second round with the identification task than in the first round without an additional task, $t(31) = 0.55$, $p = .58.$, $d = 0.09$. Neither differed the second from the third round, $t(31) = 1.05$, $p = .30.$, $d = 0.18$. Based on participants' correct answers, we included 66.2 % of the total trials of Round 2. The planned contrasts were based on a linear mixed-effects model, fit via maximum likelihood, with participants as the random effect and the variables condition (Round 1, Round 2, Round 3), object status (Cued, Non-Cued) and cueing behavior (Stalking, Flirty). A priori hypotheses were conducted using Bonferroni adjusted alpha levels of .006 per test (.05/8). The planned contrasts for target trials were not significant in Round 1, neither in the Stalking condition, $|z| = 0.77$, $p = .97$, $d = 0.18$, nor in the Flirty condition, $|z| = 0.44$, $p = .99$, $d = 0.12$. Also in Round 3 there were no significant effects, neither in the Stalking condition, $|z| = 0.847$, $p = .96$, $d = 0.20$, nor in the Flirty condition, $|z| = 1.10$, $p = .87$, $d = 0.26$. In Round 2, both target contrasts were significant, Stalking $|z| = 5.08$, $p < .001$, $d = 1.21$; Flirty $|z| = 3.87$, $p < .001$, $d = 0.75$. The distractor trials showed a similar picture for Round 1 and Round 3 (no significant effects). However, in Round 2 the difference in the selection of the cued distractor and the non-cued distractors was significantly different from zero in the Stalking condition, $|z| = 3.81$, $p < .001$, $d = 0.95$. See tables in Appendix C. Refer to Fig.4 for a graphical display of the analyzed target and distractor trial data.

Discussion

The second experiment replicated the design of Experiment 1 and added a third round of tracking. The findings of the preceding results are reflected in the current data. Under the ACS manipulation, the single cued target-objects were marked more often than the other targets when the participant was attentionally ready. This leads us to assume that the demands of the identification task play a stronger role here than knowledge about features alone.

In Experiment 1, we found a slight hint of an inhibition of the single object when it was a distractor and constantly cued (Stalking condition), which is why we introduced revised instructions in Experiment 2 to ensure that participants understood the concept of distractor objects completely. And in fact, the effects found for the cued-distractor objects in Experiment 1 reached statistical significance in the current data. Yet be aware that the distractor object was marked less often (than the three equally available other distractors) only in the Stalking but not in the Flirty condition. While Experiment 2 showed that the effects of the ACS on object selection is only observable when it is activated but is disregarded when the participant is no longer asked to focus on features, it may be possible that the effects observed in Experiment 1 have a learning component. That is, experience in tracking may result in better integration of object features and possible learning for helpful cues without further information. Experiment 3 replicated the design of Experiment 1 but participants were not given any information or an identification task.

EXPERIMENT 3

Method

Participants. Twenty students (12 female, 8 male; mean age $M = 21.75$ years, $SD = 3.18$), of the University of Tübingen participated in return for course credit or monetary compensation. All subjects had reportedly normal or corrected-to-normal sight. Informed consent was obtained after the experimental procedures were explained to the subjects.

Stimuli and design. Stimuli and design was an exact replication of Experiment 1 with the exception that participants did not receive any information or task before Round 2.

Results

With an average of 26.05 correctly identified gaze expressions in the Reading- the-Mind-in-the-Eyes test, subjects in the present study scored minimally below the range of what is suggested as a normal control-group average by Baron-Cohen et al. (2001; 26.2 - 30.9 of 36 gaze expressions in total). No participant was excluded from the analysis.

Overall, participants tracked 2.87 of 4 objects correctly. In contrast to the previous experiments, there was a significant

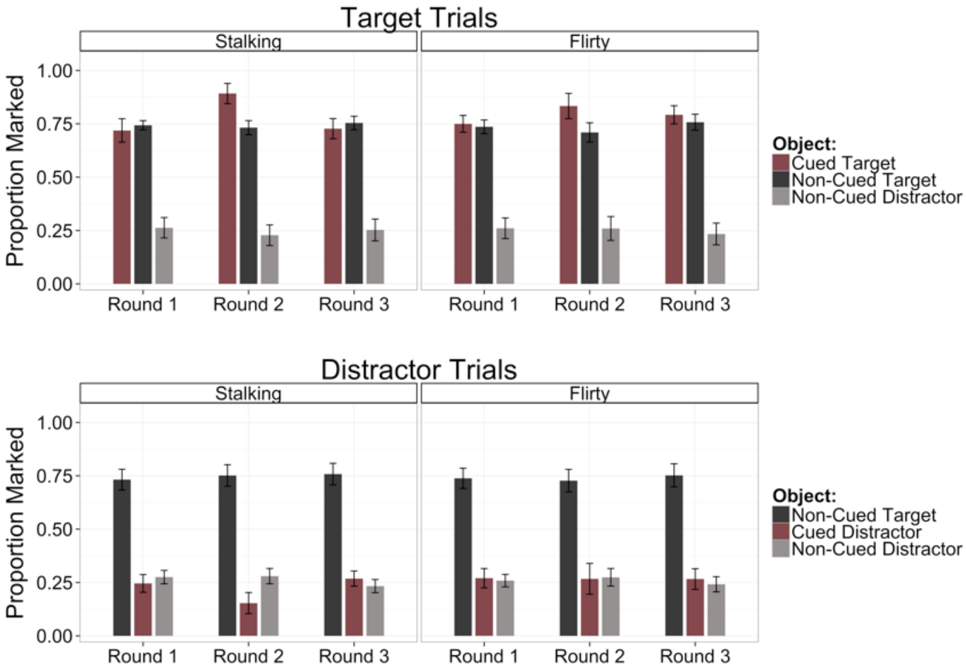

Figure 4. Proportion marked of the different object types displayed in all three rounds for cued target and cued distractor trials depending on cueing behavior. Error bars indicate 95% within-subject confidence intervals.

effect for round. Participants tracked more targets correctly in the second round, $t(19) = 3.11$, $p = .006$., $d = 0.70$. The included contrasts were based on a linear mixed-effects model, fit via maximum likelihood, with participants as the random effect and the variables condition (Round 1, Round 2), object status (cued, noncued) and cueing behavior (flirty, stalking) as fixed effects. A priori hypotheses were conducted using Bonferroni adjusted alpha levels of .0125 per test (.05/4). Refer to Fig.5 for a graphical display of the analyzed target and distractor trial data. See Appendix D for full results.

All planned comparisons for target trials were nonsignificant (see Appendix D). Results indicate that the selection preference of the participants was not influenced by gaze cues, neither in the first, nor in the second round of tracking: they did not mark the cued targets more often than the noncued target. This was also independent of the presented cueing behavior (Flirty, Stalking). Equal to the target trials, the gaze cue did not give rise to a selection preference for the cued distractor. Whether the cue was provided as flirty or as stalking did not make a difference without an ACS.

Discussion

The third experiment was conducted to measure possible effects due to learning, that is, a possible interaction of tracking "experience" (Round 1, Round 2) and cue usability. None of the planned comparisons yield significant results, indicating that the participants did not process the gaze-cue as an object feature during tracking. The data reported are in agreement with previous findings stating that feature processing is disregarded during MOT when spatiotemporal information is available constantly (Papenmeier et al., 2013) and that singletons (in this case, the cued object) are not processed differently when they are of no relevance to the given task (in this case, the tracking task) (e.g., Yantis & Egeth, 1994). We conclude that the use of features truly depends on our previously introduced manipulation (ACS). In the final experiment, we tested whether the ACS manipulation could be generalized to color cues by simply coloring neutral cartoon eyes for the length of a Flirt or the length of a trial (Stalking). The cued object remained white, while other objects were gray. We intended to replicate the results produced by the ACS effect.

EXPERIMENT 4

Method

Participants. Twenty new students of the University of Tübingen participated in return for monetary compensation (mean age M = 24.15 years, SD = 3.48). All subjects had reportedly normal or corrected-to-normal sight. Informed consent was obtained after the experimental procedures were

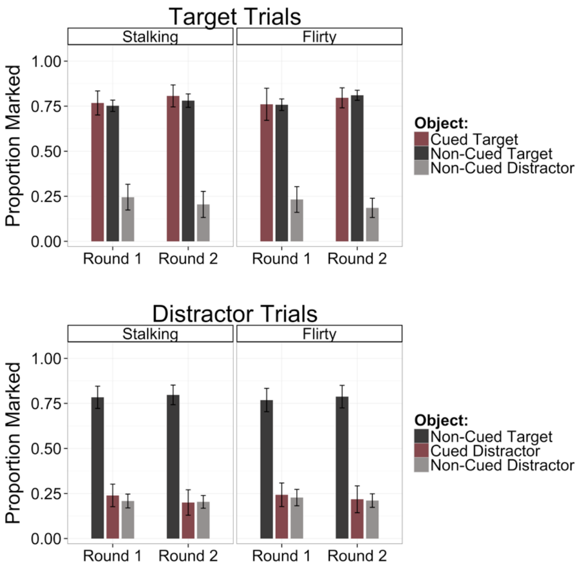

Figure 5. Proportion marked of the different object types displayed in both rounds for cued target and cued distractor trials depending on cueing behavior. Error bars indicate 95% within-subject confidence intervals.

explained to the subjects.

Stimuli and design. The design of the experiment was equal to the preceding Experiment 1. The participants tracked two rounds of 64 trials and received questions after each trial in the second round. They were also tested on their sensitivity to eyes (Read the Mind in the Eyes test) even though the stimuli used for the tracking task were changed to neutral eyes with motionless pupils (see Fig.1). The participants were asked to track the objects that could, without further description, be perceived as resembling the figure 8 on its side (infinity sign) rather than cartoon eyes. Instead of taking a glance at the cued object, the noncued eyes were colored in a light gray tone while the cued object stayed white. In the stalking condition, noncued objects were colored during the whole trial (to be more specific, until 500 ms before the end of the motion), while in flirty trials the corresponding objects changed color twice, for the duration of 1.5 seconds (the exact duration of flirt moments in preceding experiments). The answer options were adapted (translated):

1. One of the target-objects was white; the other objects were gray.
2. One of the distractor-objects was white; the other objects were gray.
3. One of the target-objects changed its color to gray twice.
4. One of the distractor-objects changed its color to gray twice.

The participants received instructions with illustrations of the real tracking situations, similar to the instructions of Experiment 2 and 3.

Results

With an average of 26.75 correctly identified gaze expressions in the Reading the Mind in the Eyes test, subjects in the present study scored within the normal control-group range suggested by Baron-Cohen et al. (2001; 26.2–30.9 of 36 gaze expressions in total). The test was kept in the current experiment to ensure that differences do not arise due to a different course of the experiment.

Overall, participants tracked 2.85 of 4 objects correctly. We found a significant effect for round. Participants tracked more targets correctly in the second round, $t(19) = 2.18$, $p = .04.$, $d = 0.49$.

Only correctly identified trials were included in the analysis, resulting in the inclusion of 70.9 % of the total trials of Round 2. The planned contrasts were based on a linear mixed-effects model, fit via maximum likelihood, with participants as the random effect and the variables condition (Round 1, Round 2), object status (cued, noncued) and cueing behavior (flirty, stalking). Refer to Fig.6 for a graphical display of the analyzed target and distractor trial data. The according means and the results of the planned comparisons are presented in Appendix E. There were no significant differences between the selection of the cued-target objects and the noncued targets in Round 1, target trials stalking, $|z| = 0.03$, $p = .99$, $d = 0.01$, flirty, $|z| = 1.61$, $p = .43$, $d = 0.62$; distractor trials stalking, $|z| = 1.07$, $p = .77$, $d = 0.30$, flirty, $|z| = 1.04$, $p = .18$, $d = 0.58$. This reinsures that the color feature of the objects alone was not weighted as a source relevant enough to benefit tracking (in line with Papenmeier et al, 2013). Since color is known to be a preference for selective attention, this was especially remarkable with regard to the constantly visible white singleton among gray objects in the stalking trials. In Round 2, the interaction of ACS and the processing of color cues showed significant effects as seen before in Experiment 1, 2, and 3; target trials: stalking, $|z| = 5.39$, $p < .001$, $d = 1.11$, flirty, $|z| = 1.4.01$, $p < .001$, $d = 1.51$; distractor trials: stalking, $|z| = 4.59$, $p < .001$, $d = 1.47$, flirty, $|z| = 2.12$, $p = .15$, $d = 0.63$.

The results replicate what we have reported so far: The cued distractor was marked significantly less often. Whereas the cues on the target object in Round 2 worked independently of cueing behavior, the cued distractor was only marked less often than the rest of the distractors in the stalking condition compared to the flirty condition. Compared to the graphical inspection of the previous experiments, the current color experiment showed some slight indications for

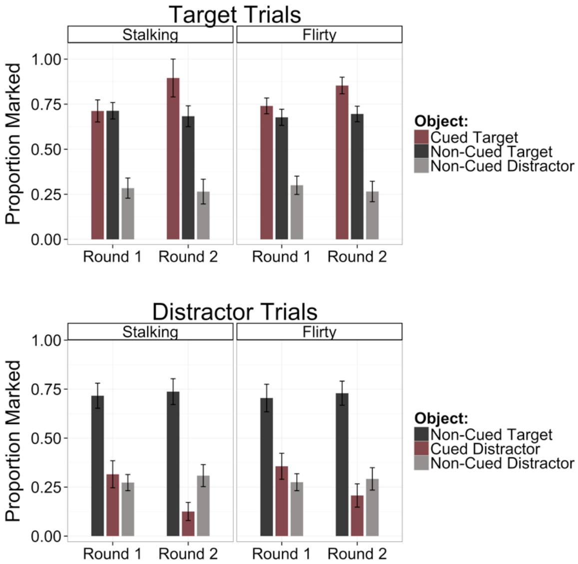

Figure 6. Proportion marked of the different object types displayed in both rounds for cued target and cued distractor trials depending on cueing behavior. Error bars indicate 95% within-subject confidence intervals.

an automatic processing of abruptly occurring color cues in Round 1 (see tables in Appendix E and Fig.6).

Discussion

The final experiment meant to generalize the effect found for gaze cues to color cues: Single cued target objects are marked more often than the other targets during tracking when the participant had an ACS (Round 2). Single cued distractor objects are marked less often, but this was only the case when the distractor was cued constantly compared to only twice. The results of Experiment 4 were remarkably similar to those of the previous experiments. They demonstrate that an ACS effect for single cued objects during a tracking task can be obtained with gaze cues as well as with color cues. The reasons why we shy away from claiming complete interchangeability between the two types of cueing (color and gaze) are the graphical results of Experiment 4. The results of the flirty trials in Round 1 hint to a different processing of color cues, which could have been due to its rather abrupt nature. As Yantis (1993) suggested, those kinds of visual onsets may capture attention independent of an attentional state of feature readiness.

A note on learning and task effects

We wondered about the effect of the identification task on general tracking performance. To that end we calculated the mean difference of proportion marked for the three noncued targets of Round 2 minus Round 1 (leaving out Round 3 of Experiment 2) independent of condition and behavior and applied two-sided t tests. The calculated differences were significantly different from zero in the first three experiments (see Fig.7), Experiment 1: $t(19) = -2.38$, $p = 0.03$; Experiment 2: $t(31) = -2.58$, $p = 0.01$; Experiment 3: $t(19) = 2.58$, $p = 0.02$. The negative t values as well as the graphical presentation suggest that the identification task had a detrimental influence on the overall tracking performance in Experiment 1 and Experiment 2. In Experiment 3, when participants were not given the identification task in Round 2, they showed a (positive) learning effect. Interestingly though, Experiment 4 showed neither, $t(19) = 0.09$, $p = 0.93$. This could suggest that the task to identify a specific color "behavior" does not interfere with the tracking task at all. However, with regard to the learning effect in Experiment 3, it may well be assumed that the effect of learning is simply canceled out by the detrimental effect of the identification task.

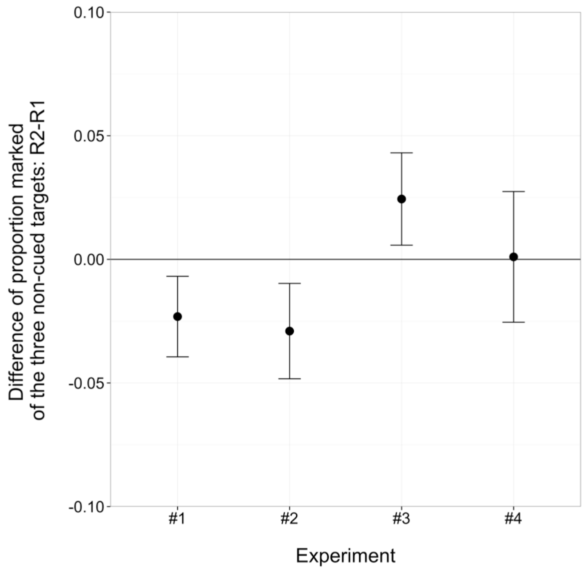

Figure 7. Mean difference of proportion marked of the three non-cued targets of (Round 2 minus Round 1) compared to zero (H0: no difference between Round 1 and Round 2). Error bars indicate 95% within-subject confidence intervals.

General discussion

In the experiments presented, we explored how manipulating the relevance of object features modifies the allocation of attention in a MOT task. While it has been theorized that the attentional resolution is increased and decreased based on spatiotemporal occurrences like crowding or overlapping objects, we assume that an intentional, task-based component can influence attentional resolution to access feature objects as well. Our research design has a practical share with the concept of attentional capture and control settings (e.g., Egeth & Yantis, 1997; Folk et al., 1992) which lead us to hypothesize that changes in attentional resolution will happen in a strategic manner, inhibiting distractor objects and preferring target objects flexibly based on task-demands.

Theoretical accounts on tracking and features

The first interesting finding is reflected in the null-effects in the first round of each experiment: While it is well known that in the MOT paradigm it is easier to track the targets if they are easily distinguishable from the distractors, for example by being of a different color (Howe & Holcombe, 2012; Makovski & Jiang, 2009a; Makovski & Jiang, 2009b; similar: Oksama & Hyönä, 2008), our analysis of the selection preference for singletons and nonsingletons does not support this claim. However, Makovski and Jiang (2009a) also stated that features are not properly conjoined during attentive tracking because feature-location bindings are not necessary to track successfully. They concluded that the attentional tracking system has no access to bound target representations. Nonetheless, the model of multiple identity tracking (MOMIT; Oksama & Hyönä, 2008) predicts greater accuracy for distinctive targets because the observer can recover from tracking errors more easily. Based on the MOMIT, we would have expected that the target singleton is automatically (i.e. Round 1 of the experiments) marked more often, while the distractor singleton is confused less with a target than the other distractors—but that was not the case, not even in Experiment 4 with color singletons. This does not necessarily contradict previous findings, since for example Howe and Holcombe (2012) showed that the greatest advantage of distinct features is observed when distractor objects share absolutely none of the target features.

Our target and distractor objects shared the exact same features, that is, white spheres with pupils that moved along with the items and the singleton they cued (or as in Experiment 4: colored gray) with the exception that the singleton displayed "neutral" pupils (or stayed white). Our results suggest that focusing attention toward minimal featural differences during tracking is, first of all, possible, and more importantly, probably leads to a comparable advantage as if tracking targets and distractors that are different in all featural aspects.

If we disregard the difference between identity and feature and assume that the MOMIT (Oksama & Hyönä, 2004, 2008) applies to our stimuli, it would not explain our results entirely either. The MOMIT proposes that an automatic, parallel low-level system collaborates with a serial high-level attentional spotlight. When a target is at risk to fall below an activation threshold, the spotlight is alarmed by the low-level system, initiating an exhaustive search by the spotlight to reactivate the representation of the particular object in the low-level system. In order to explain our results, the model would need a modification: The spotlight is not only controlled by a low-level system but should also work in a top-down manner, even inhibiting stimuli with features that have no relevance.

The FLEX account by Alvarez and Franconeri (2007) describes a mechanism for the reallocation of resources in specific situations, for example, when a distractor comes close to a tracked target. They hypothesized that such interfering events would enhance the attentional resolution at the specific location. While we did not control for eye-movements in the current study, we found a way to approximate the proportion of attention each target and each distractor received, and observed attention shifts in favor of the identification task (see also Yarbus, 1967). That means, participants were able to find the singleton and/or follow the gaze cues (use the color cues) and used the information to avoid confusions with the cued distractor, or to have a "save" target (in the cued-target trials). However, whether attention was enhanced for targets and suppressed for distractors (as seen in Bettencourt & Somers, 2009, or Doran & Hoffman, 2010) cannot be determined with the current design. What we can propose is that, in principal, the shifted selection preferences may reflect attentional resolution. Observers can modulate the attentional resolution not only in accordance with spatial demands (as previously suggested by, e.g., Franconeri et al. 2008; Intriligator & Cavanagh, 2001; Pylyshyn, 2004) but also to access object feature information that based on task demands.

Underestimated objects: Distractors

As in Feria (2012), the current study not only emphasizes the role of distractors but also the role of distractor features during tracking. Especially interesting is the finding, that the singleton distractor is marked less often compared to other, noncued distractors in stalking conditions compared to the flirty conditions. In terms of existing tracking research, this observation does not support the majority of findings and could not reasonably be foreseen. In general, tracking capacity is greater when no other distracting objects are present (Horowitz & Cohen 2008; Horowitz et al. 2007) and decreases when the number of distractors increases (Bettencourt & Somers, 2009). Furthermore, by using a secondary probe detection task, researchers found that the detection of probes was less reliable on distractors compared to probes on targets and on an empty background (see Pylyshyn, 2006; Pylyshyn, Haladjian, King, & Reilly, 2008; Huff et al., 2012; but see Drew et al., 2009).

However, while these results point towards distractor inhibition during tracking, others reported that the role of distractors has been underestimated in its influence on tracking performance. Supporting findings of Alvarez and Oliva (2008), who showed that the locations of distractors are represented above chance level, Meyerhoff and team (2015) demonstrated that distractor displacements impair tracking performance. In a recent study, Meyerhoff, Papenmeier, Jahn, and Huff (in press) compared different speed profiles and reported an affected tracking performance even when only distractor objects varied in speed. They suggested that the spatial configuration of targets and distractors, in contrast to tracking for example only a virtual polygon of target objects (Yantis, 1992), is encoded in order to enhance the allocation of visual attention toward target objects. While they propose further that the distractor location is represented to an extent that allows for the detection of crowding events (see also Iordanescu et al., 2009), our data indicate that not only distractor location is represented but also a specified distractor feature.

In our data, the feature representation on the distractor object was at least pronounced enough to detect approaching events and to allocate attention accordingly in order to prevent confusions. Our findings nicely fit the results of Drew et al. (2009). In their tracking study with moving and stationary objects, they measured the electrophysiological responses (ERP) of participants to task-irrelevant probes that were located on targets, distractors, stationary objects, or in empty space. The authors report the response to distractors as located between the (greatest) response to targets and the (weakest) response to background and empty space probes. They concluded that distractors are not suppressed, at least not on an early level of perceptual processing. The idea of a hierarchy of attentional allocation (Drew et al., 2009)—with distractors being secondary to the dominant targets, but nonetheless processed—could help to explain why the singleton distractor is marked less often compared to other, noncued distractors in stalking conditions but not in flirty conditions. Given that the task triggered a feature level of attention, which changed the distribution of attention on the moving objects, distractors still received less attention than targets due to the tracking task demands. The intermittent flirty cue was simply too weak to be processed within the limited amount of attention allocated to a distractor, while the constantly visible stalking cue was stronger and thus needed less resources to have an attentional effect.

Intentional attentional control

However, it should be noted that we did not measure automatic effects that could be contributed to the featural singletons in Round 1. Features were only represented or at least measurably used during tracking in conditions in which we induced an ACS. Approaching the results in terms of attentional control and goal- related processing of singletons, it is easier to explain why targets were prioritized and distractor inhibited. Folk and team (1992) proposed that a task-driven selection mechanism guides attention and by that, the observer is able to ignore or prioritize distinct objects. Ignoring the distractor was harder to accomplish (i.e., demanded more cognitive resources) in the flirty distractor condition because participants had to keep track of four targets and figure out whether a cue that only appeared twice was beneficial or not. In the stalking distractor condition, the cue was constantly visible, and confusions with the cued distractor were less likely to occur. The same logic applies for tracking Round 1 (and Round 3 in Experiment 2). Without the task, no ACS was activated and thus cued single objects had no relevance. The spatiotemporal information was sufficient to track the objects successfully. This is what has been observed in previous studies and what has led to the conclusion that MOT is a feature-independent, preattentive and low-level task (e.g., Huff, Jahn, & Schwan, 2009; but see Papenmeier et al., 2013).

Costs of the additional task for the overall tracking performance

Concerning the overall tracking performance, we were certain that the additional identification task would draw upon cognitive resources, resulting in a lower performance in the dual-task conditions. However, based on our results of the control experiment compared to the experiments in which the manipulation was applied (Experiment 3 vs. 1, 2, and 4), we can only assume that the costs of the second task are basically rather small, and, in case of the Experiment 4 (color), congruent with small learning effects in tracking. Considering Cohen et al. (2011), who maintained that feature and location processing during tracking draw on the same, single cognitive resource—a claim further supported by neurophysiological and functional neuroimaging studies (e.g., Corbetta & Shulman, 2002; Sàenz, Buraĉas, & Boynton, 2003), that found brain regions for attention to feature and location to overlap—this is a surprising finding. In contrast to conclusions made by Cohen et al. (2011), the small costs of the additional identification task found in our data would rather support the notion that tracking is either handled by an entirely separate, encapsulated system from feature processing, or some sharing of resources is possible without much decrement. That is, when tracking a group of children of which one is your own offspring on the playground, you will be able to (attentionally) prioritize your own while at the same time your tracking ability for the others will not be interfered tremendously. Especially when the children's features, for example clothes, are colored differently.

Our findings do not necessarily contradict the notion that there is a trade-off between locations and identities completely. Our findings simply bring us a step further to the identification of the scope and limits of the involved resources, suggesting that the identity of only one object can be processed with negligible decrement to the overall location-based tracking performance of the other objects.

Closely related to Luck et al. (1996), who proposed that different processing levels control different types of attentional overload and interferences, Cohen et al. (2011) also presented evidence that mental resources can be voluntarily distributed across targets depending on task-demands (identity tracking or location tracking). The current study provides additional evidence that attention allocation during MOT has, or can be influenced by, a top-down component as well. This was reflected in the strategically suppression of distractor objects to avoid errors.

Flexible gaze-cue usability

A final word on our choice for the stimuli is needed. First and foremost, we used gazing eyes because we expected to observe reflexive attention shifts to the gazed-at object, even when the gaze cue was counterpredictive of the intended saccade direction (which would have been especially disturbing in cued-distractor trials; Kuhn & Kingstone, 2009). Gaze following is supposed to be reflexive and independent of cognitive load (e.g., Driver et al., 1999; Friesen & Kingstone, 1998). This equally applies for objects within dynamic displays, indicating that attention induced by such cues can be attached to moving objects and not only cue a spatial area (Marotta, Casagrande, & Lupiáñez, 2013).

Yet, while there is compelling evidence for a highly automatic behavior as a response to gaze cues, recent studies, including the presented one, raise some doubts (e.g., Koval, Thomas, & Everling, 2005). Numerous studies suggest that gaze cues can be used with a degree of flexibility (e.g., depending on the observers goal: Bayliss, Frischen, Fenske, & Tipper, 2007; Brooks & Meltzoff, 2005; Johnson, Slaughter, & Carey, 1998; Macdonald & Tatler, 2013; Ricciardelli, Carcagno, Vallar, & Bricolo, 2013). Particularly Böckler, Knoblich, and Sebanz (2011) use the gaze-cue paradigm to show that attention sharing (operationalized as mutual gaze (i.e., at least two pairs of eyes/faces shift their gazes simultaneously) can modulate joint attention. They not only propose that observing others sharing attention increases the significance of an ensuing gaze, but with a crucial condition in which they tested whether gaze following was modulated by the relevance of the looked-at-target to the observer's current goal/task, they concluded that certain contextual conditions and top-down mechanisms affect gaze-following behavior.

Our study provides further support for the possibility that gaze cues are not necessarily followed reflexively independent of cognitive load. Based on the non significant Round 1 results, we conclude that participants may have used the gaze-cue flexibly and strategically during tracking in our study. Still, it is difficult to determine whether they actually followed the gaze cue or concentrated on featural differences only. One finding that would speak in favor of gaze-cue processing is that the stalking and the flirty target trials showed the same effects. However, while this argues in favor of parallel processes one could also defend a serial account in combination with the hierarchy of attention as proposed by Drew et al. (2009) and discussed before. The participants would have scanned the targets first before scanning the distractors. The short duration of feature visibility in flirty trials (3 seconds in total) may have been not enough to process features and identify the behavior in distractor trials, since by the time the participant reached the first distractor object, the featural cue would have already disappeared.

On the other hand, the Experiment 4 (color cues) showed that the same attentional control applies to color stimuli—but still does little to clarify the specific issue of cue—versus feature use. While replicating the results of Experiments 1–3, in Experiment 4 with color cues we found some indication, even though not reaching statistical significance, of reactions to an abrupt onset of the color cue in flirty trials in Round 1. As Yantis (1993) suggested, those kinds of visual onsets may capture attention independent of an attentional state of feature readiness. Therefore, it is difficult to arrive at a definite conclusion as to whether gaze cues were used or object features were compared during tracking. We may have simply observed here that gaze cues were used intentionally but did not work reflexively, while color cues elicited bottom-up reactions that were actively suppressed and channeled in the ACS condition. Possible future research could be concerned with abrupt occurrences of salient features during tracking. By further connecting MOT to other fields of research (e.g., attention capture), we may be able to solve some of the riddles and misunderstandings that tracking studies could have not disentangle up to now.

Benefits, drawbacks, and further applications

The presented novel variation of the standard MOT task (Pylyshyn & Storm, 1988) kept the general structure of the task but presented a feature singleton among identical objects, that was either among the targets or among the distractors—a design that has not been applied before. This allowed us to explore various effects at the same time, which may simultaneously represent a benefit as well as a drawback. Here, we focused on activating an ACS and observed the prioritization or inhibition of the feature singleton. Nonetheless, additional beneficial information could have been gained from including a focus on the social aspect and/or including a clinical sample. We believe that our modification of the paradigm can be applied with clinical populations—for example, through testing the ability to switch between parallel and serial processing, or testing patients with autism or Asperger's on their processing skills of dynamic gazes. With this being said, an eye-tracking study will be of tremendous use to further understand the modified paradigm. One could determine whether color and gaze cues produce the same results but are processed differently. In other words, gaze cues may produce involuntary saccades to the cued object but could be actively suppressed by the observer in order to successfully track the objects. Furthermore, we may find parallel processing (focus on the centroid of the targets) in Round 1, and serial processing (target jumping) in Round 2. Another option could be the use of single-pulse transcranial magnetic stimulation (TMS) on the superior lateral temporal cortex that is known to interfere with gaze direction tasks (Pourtois, Grandjean, Sander, & Vuilleumier, 2004). This interference was also found to be task-specific.

Although we did not find direct evidence for an automatic distinctiveness effect for the feature singletons, it is possible that we failed to measure the effect due to our experimental design. Future experiments concerned with automatic attentional shifts in MOT environments should consider contrasting trials with and without singletons, that is, including trials without distinct objects. Furthermore, the slightly different results found in Experiment 4 (color), however not statistically significant, could be an indicator of reflexive attention shifts in case of abrupt appearing cues. Future studies could be concerned with bottom-up effects of abrupt and gradual appearing features in MOT studies, and its dependency on an ACS.

Finally, we would like to highlight our decision for adapting the standard analysis of tracking capacity to our hypotheses. In contrast to the majority, if not all, MOT studies, we measured how often each of the eight objects was marked and compared this selection mean of cued and noncued objects by condition. This gave us a far more vivid picture of the attention distribution than a simple mean value of correctly tracked items. In fact, proportion correct showed no effects in our data (i.e., was not affected by the cueing behavior displayed). We believe this approach to be promising for future research (e.g., for studies concerned with the different role of targets and distractors). It is our hope that our newly modified analysis will be in use in future studies in order to gain more refined insight. However, it is important to mention that our decision to include only trials in which the participants correctly identified the eyes behavior and the type of object that was cued, was a theoretical advantage (excluding trials in which we cannot be sure that the participants was actively involved in the task), but a slight disadvantage for interpretative purposes. Because the proportion of correctly identified trials differed for the two gazing behaviors, we only analyzed data within each behavior and each round, and not across behaviors and rounds. Any statements concerned with the strength of the effect found in flirty compared to stalking conditions are thus purely speculative. Future studies focusing on ACS and MOT should find a way to control the participant's attentional engagement with less consequences for data analysis. Regardless, the present results indicate a consistent pattern of attentional resolution in tracking tasks that changes due to task demands.

Conclusion

Our results reveal a striking cued-target selection preference and a cued- distractor inhibition when participants received an identification task that engaged them actively in the processing of object features during tracking. These effects were attributed to the activation of an attentional control set, a term that originally stems from research on attention capture and visual search. We propose here that the allocation of attention and a flexible attentional resolution is not only an automatic reaction of the visual system to prevent confusions when interobject spacing decreases but also managed by a goal-related, top-down component. The introduced modification of the MOT paradigm, as well as the unusual type of analysis, offer various new options for future research in different areas.

Author Note

We thank Dominik Frommeyer and Lia Heubner for conducting the experiments. We are also grateful to Julian Hofmeister and Steffen Hildebrandt from the computer science department. We thank Linda Kersten for proofreading and Annika Maurer for general support. We sincerely thank Trafton Drew and two anonymous reviewers for constructive criticisms and valuable comments that have helped tremendously in the improvement of the overall presentation and clarity of this work.

This research was supported by grants from the German Research Foundation (DFG 1510/4-3). It has been partially presented (poster) at the annual meeting of the Vision Science Society in Florida, 2015, and at the Psychonomic Society's 56th Annual Meeting held in Chicago, 2015.

References

Adams, R. B. & Kleck, R. E. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychological Science*, *14*(6), 644–647. doi:10.1046/j.0956-7976.2003.psci_1479.x

Alvarez, G. A. & Franconeri, S. L. (2007). How many objects can you track?: evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, *7*(13), 14. doi:10.1167/7.13.14

Alvarez, G. A. & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*(4), 392–398.

Bae, G. Y. & Flombaum, J. I. (2012). Close encounters of the distracting kind: identifying the cause of visual tracking errors. *Attention, Perception, & Psychophysics*, *74*(4), 703–715.

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, *20*, 395–418. doi:10 . 1017 / s0305000900008345

Baldwin, M. W. (1992). Relational schemas and the processing of social information. *Psychological Bulletin*, *112*(3), 461. doi:10.1037//0033-2909.112.3.461

Baron Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*(2), 241–251.

Baron-Cohen, S., Campbell, R., Karmiloff-Smith, A., Grant, J., & Walker, J. (1995). Are children with autism blind to the mentalistic significance of the eyes? *British Journal of Developmental Psychology*, *13*(4), 379–398. doi:10.1111/j.2044-835x.1995.tb00687.x

Bayliss, A. P., Frischen, A., Fenske, M. J., & Tipper, S. P. (2007). Affective evaluations of objects are influenced by observed gaze direction and emotional expression. *Cognition*, *104*(3), 644–653.

Bettencourt, K. C. & Somers, D. C. (2009). Effects of target enhancement and distractor suppression on multiple object tracking capacity. *Journal of Vision*, *9*(7), 9.

Böckler, A., Knoblich, G., & Sebanz, N. (2011). Observing shared attention modulates gaze following. *Cognition*, *120*(2), 292–298. doi:10.1016/j.cognition.2011.05.002

Bölte, S. (2005). *Reading the Mind in the Eyes Test für Erwachsene (dt Fassung) [Reading the Mind in the Eyes test for adults (German version) by S. Baron-Cohen]* (J. G. U. F. am Main, Ed.). Eigendruck.

Brooks, R. & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*(6), 535–543. doi:10.1111/j.1467-7687.2005.00445.x

Cavanagh, P. & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, *9*(7), 349–354. doi:10.1016/j.tics.2005.05.009

Cave, K. R. & Bichot, N. P. (1999). Visuospatial attention: beyond a spotlight model. *Psychonomic Bulletin & Review*, *6*(2), 204–223. doi:10.3758/bf03212327

Cohen, M. A., Pinto, Y., Howe, P. D., & Horowitz, T. S. (2011). The what-where trade-off in multiple-identity tracking. *Attention, Perception, & Psychophysics*, *73*(5), 1422–1434.

Corbetta, M. & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215.

Cutzu, F. & Tsotsos, J. K. (2003). The selective tuning model of attention: psychophysical evidence for a suppressive annulus around an attended item. *Vision Research*, *43*(2), 205–219.

Doran, M. M. & Hoffman, J. E. (2010). The role of visual attention in multiple object tracking: evidence from erps. *Attention, Perception, & Psychophysics*, *72*(1), 33–52.

Drew, T., Horowitz, T. S., & Vogel, E. K. (2013). Swapping or dropping? electrophysiological measures of difficulty during multiple object tracking. *Cognition*, *126*(2), 213–113. doi:10.1016/j.cognition.2012.10.003

Drew, T., McCollough, A. W., Horowitz, T. S., & Vogel, E. K. (2009). Attentional enhancement during multiple-object tracking. *Psychonomic Bulletin & Review*, *16*(2), 411–417. doi:10.3758/pbr.16.2.411

Driver, J., Iv, D., G., R., P., K., P., M., & E., S., & Baron-Cohen. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition*, *6*(5), 509–540. doi:10.1080/135062899394920

Egeth, H. E. & Yantis, S. (1997). Visual attention: control, representation, and time course. *Annual Review of Psychology*, *48*(1), 269–297. doi:10.1146/annurev.psych.48.1.269

Eriksen, C. W. & James, J. D. S. (1986). Visual attention within and around the field of focal attention: a zoom lens model. *Perception & Psychophysics*, *40*(4), 225–240.

Farroni, T., Massaccesi, S., Pividori, D., & Johnson, M. H. (2004). Gaze following in newborns. *Infancy*, *5*(1), 39–60. doi:10.1207/s15327078in0501_2

Feria, C. S. (2012). The effects of distractors in multiple object tracking are modulated by the similarity of distractor and target features. *Perception*, *41*(3), 287–304.

Folk, C. L. & Annett, S. (1994). Do locally defined feature discontinuities capture attention? *Perception & Psychophysics*, *56*(3), 277–287. doi:10.3758/bf03209762

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1030. doi:10.1037//0096-1523.18.4.1030

Folk, C. L., Remington, R. W., & Wright, J. H. (1994). The structure of attentional control: contingent attentional capture by apparent motion, abrupt onset, and color. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 317.

Franconeri, S. L., Jonathan, S. V., & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological Science*, *21*(7), 920–925. doi:10 . 1177 / 0956797610373935

Franconeri, S. L., Lin, J. Y., Enns, J. T., Pylyshyn, Z. W., & Fisher, B. (2008). Evidence against a speed limit in multiple-object tracking. *Psychonomic Bulletin & Review*, *15*(4), 802–808.

Friesen, C. K. & Kingstone, A. (1998). The eyes have it! reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490–495. doi:10.3758/bf03208827

Hillstrom, A. P. & Yantis, S. (1994). Visual motion and attentional capture. *Perception & Psychophysics*, *55*(4), 399–411. doi:10.3758/bf03205298

Horowitz, T. S. & Cohen, M. A. (2008). How do distractors distract in multiple object tracking? *Perception*, *37*, 98–99.

Horowitz, T. S., Klieger, S. B., Fencsik, D. E., Yang, K. K., Alvarez, G. A., & Wolfe, J. M. (2007). Tracking unique objects. *Perception & Psychophysics*, *69*(2), 172–184. doi:10.3758/bf03193740

Howe, P. D. & Holcombe, A. O. (2012). The effect of visual distinctiveness on multiple object tracking performance. *Frontiers in Psychology*, *3*(307), 1–7. doi:10.3389/fpsyg.2012.00307

Huff, M., Jahn, G., & Schwan, S. (2009). Tracking multiple objects across abrupt viewpoint changes. *Visual Cognition*, *17*(3), 297–306. doi:10.1080/13506280802061838

Huff, M., Papenmeier, F., & Zacks, J. M. (2012). Visual target detection is impaired at event boundaries. *Visual Cognition*, *20*(7), 848–864. doi:10.1080/13506285.2012.705359

Intriligator, J. & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, *43*(3), 171–216. doi:10.1006/cogp.2001.0755

Iordanescu, L., Grabowecky, M., & Suzuki, S. (2009). Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. *Journal of Vision*, *9*(4), 1–1. doi:10.1167/9.4.1

James, W. (1890). *A saliency-based search mechanism for overt and covert shifts of visual attention*. Harvard University Press, Cambridge, MA (1890/1980: Principles of Psychology.

Johnson, S., Slaughter, V., & Carey, S. (1998). Whose gaze will infants follow? the elicitation of gaze-following in in 12-month-olds. *Developmental Science*, *1*(2), 233–238.

Jonides, J. & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics*, *43*(4), 346–354. doi:10.3758/bf03208805

Joseph, J. S. & Optican, L. M. (1996). Involuntary attentional shifts due to orientation differences. *Perception & Psychophysics*, *58*(5), 651–665. doi:10.3758/bf03213098

Koval, M. J., Thomas, B. S., & Everling, S. (2005). Task-dependent effects of social attention on saccadic reaction times. *Experimental Brain Research*, *167*(3), 3. doi:10.1007/s00221-005-0206-8

Kuhn, G. & Kingstone, A. (2009). Look away! eyes and arrows engage oculomotor responses automatically. *Attention, Perception, & Psychophysics*, *71*(2), 314–327. doi:10.3758/app.71.2.314

Laberge, D. & Brown, V. (1986). Variations in size of the visual field in which targets are presented: an attentional range effect. *Perception & Psychophysics*, *40*(3), 188–200. doi:10.3758/bf03203016

Lu, Z. L. & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research*, *38*(9), 1183–1198. doi:10.1016/s0042-6989(97)00273-3

Luck, S. J., Hillyard, S. A., Mouloua, M., & Hawkins, H. L. (1996). Mechanisms of visual-spatial attention: resource allocation or uncertainty reduction? *Journal of Experimental Psychology: Human Perception and Performance*, *22*(3), 725. doi:10.1037//0096-1523.22.3.725

Lupiáñez, J., Milliken, B., Solano, C., Weaver, B., & Tipper, S. P. (2001). On the strategic modulation of the time course of facilitation and inhibition of return. *The Quarterly Journal of Experimental Psychology: Section A*, *54*(3), 753–773.

Macdonald, R. G. & Tatler, B. W. (2013). Do as eye say: gaze cueing and language in a real-world social interaction. *Journal of Vision*, *13*(4), 6.

Makovski, T. & Jiang, Y. V. (2009). Feature binding in attentive tracking of distinct objects. *Visual Cognition*, *17*(1/2), 180–194. doi:10.1080/13506280802211334

Marotta, A., Casagrande, M., & Lupiáñez, J. (2013). Object-based attentional effects in response to eye-gaze and arrow cues. *Acta Psychologica*, *143*(3), 317–321. doi:10.1016/j.actpsy.2013.04.006

Maxwell, S. E. & Delaney, H. D. (2004). *Designing experiments and analyzing data: a model comparison perspective (vol. 1)*. New York, NY: Psychology Press.

Meyerhoff, H. S. [H. S.], Papenmeier, F., Jahn, G., & Huff, M. (2015). Distractor locations influence multiple object tracking beyond interobject spacing. *Experimental Psychology*, *62*, 170–180. doi:10.1027/1618-3169/a000283

Meyerhoff, H. S. [H S], Papenmeier, F., Jahn, G., & Huff, M. (2015). Not flexible enough: exploring the temporal dynamics of attentional reallocations with the multiple object tracking paradigm. doi:10.1037/xhp0000187

Morgan, M. J., Ward, R. M., & Castet, E. (1998). Visual search for a tilted target: tests of spatial uncertainty models. *The Quarterly Journal of Experimental Psychology: Section A*, *51*(2), 347–370. doi:10.1080/027249898391666

Morgan, M. J. & Watt, R. J. (1997). The combination of filters in early spatial vision: a retrospective analysis of the mirage model. *Perception*, *26*(9), 1073–1088.

Nation, K. & Penny, S. (2008). Sensitivity to eye gaze in autism: is it normal? is it automatic? is it social? *Development and Psychopathology*, *20*(1), 79–97.

Oksama, L. & Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? an individual difference approach. *Visual Cognition*, *11*(5), 631–671.

Oksama, L. & Hyönä, J. (2008). Dynamic binding of identity and location information: a serial model of multiple identity tracking. *Cognitive Psychology*, *56*(4), 237–283. doi:10.1016/j.cogpsych.2007.03.001

Papenmeier, F., Meyerhoff, H. S., Jahn, G., & Huff, M. (2013). Tracking by location and features: object correspondence across spatiotemporal discontinuities during multiple object tracking. journal of experimental psychology: human perception and performance. *40*(1), 159.

Peirce, J. W. (2009). Generating stimuli for neuroscience using psychopy. *Frontiers in Neuroinformatics*, *2*(10). doi:10.3389/neuro.11.010.2008

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. doi:10.1080/00335558008248231

Posner, M. I. & Cohen, Y. (1984). Components of visual orienting. *Attention and Performance X: Control of Language Processes*, *32*, 531–556. doi:10.1162/jocn.1991.3.4.335

Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: the role of set for spatial location. *Modes of Perceiving and Processing Information*, *137*, 158.

Pourtois, G., Grandjean, D., Sander, D., & Vuilleumier, P. (2004). Electrophysiological correlates of rapid spatial orienting towards fearful faces. *Cerebral Cortex*, *14*(6), 6. doi:10.1093/cercor/bhh023

Pratt, J., Radulescu, P. V., Guo, R. M., & Abrams, R. A. (2010). It's alive!: animate motion captures visual attention. *Psychological Science*, *21*(11), 1724–1730. doi:10.1177/0956797610387440

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: a sketch of the finst spatial-index model. *Cognition*, *32*(1), 65–97. doi:10.1016/0010-0277(89)90014-0

Pylyshyn, Z. (1999). Is vision continuous with cognition?: the case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*(3), 341–365. doi:10.1017/s0140525x99002022

Pylyshyn, Z. W. (2004). Some puzzling findings in multiple object tracking: i. tracking without keeping track of object identities. *Visual Cognition*, *11*(7), 801–822.

Pylyshyn, Z. W. (2006). Some puzzling findings in multiple object tracking (mot): ii. inhibition of moving nontargets. *Visual Cognition*, *14*(2), 175–198.

Pylyshyn, Z. W. (2007). *Things and places: how the mind connects with the world*. Cambridge, MA: MIT Press.

Pylyshyn, Z. W., Haladjian, H. H., King, C. E., & Reilly, J. E. (2008). Selective nontarget inhibition in multiple object tracking. *Visual Cognition*, *16*(8), 1011–1021. doi:10.1080/13506280802247486

Pylyshyn, Z. & Storm, R. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 179–197. doi:10.1163/156856888x00122

Ricciardelli, P., Carcagno, S., Vallar, G., & Bricolo, E. (2013). Is gaze following purely reflexive or goal-directed instead? revisiting the automaticity of orienting attention by gaze cues. *Experimental Brain Research*, *224*(1), 93–106. doi:10.1007/s00221-012-3291-5

Ruxton, G. D. & Beauchamp, G. (2008). Time for some a priori thinking about post hoc testing. *Behavioral Ecology*, *19*(3), 690–693. doi:10.1093/beheco/arn020

Sàenz, M., Buraĉas, G. T., & Boynton, G. M. (2003). Global feature-based attention for motion and color. *Vision Research*, *43*(6), 629–637. doi:10.1016/s0042-6989(02)00595-3

Scaife, M. & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*, 265–266.

Theeuwes, J., Kramer, A. F., & Atchley, P. (1999). Attentional effects on preattentive vision: spatial precues affect the detection of simple features. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(2), 341. doi:10.1037//0096-1523.25.2.341

Treisman, A. & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological Review*, *95*(1), 15.

Treisman, A. & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 459.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2005). Pushing around the locus of selection: evidence for the flexible-selection hypothesis. *Journal of Cognitive Neuroscience*, *17*(12), 1907–1922.

Yantis, S. (1992). Multielement visual tracking: attention and perceptual organization. *Cognitive Psychology*, *24*(3), 295–340. doi:10.1016/0010-0285(92)90010-y

Yantis, S. (1993). Stimulus-driven attentional capture and attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 676–681. doi:10.1037//0096-1523.19.3.676

Yantis, S. & Egeth, H. E. (1994). Visual salience and stimulus-driven attentional capture. *Investigative Oph-*

thalmology & Visual Science, *35*(4), 1619–1619. doi:10.1037//0096-1523.25.3.661

Yarbus, A. L. (1967). *Eye movements and vision* (T. B. Haigh, Ed.). New York, NY: Plenum Press. doi:10.1007/978-1-4899-5379-7

Yeshurun, Y. & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, *396*(6706), 72–75.

Yeshurun, Y. & Carrasco, M. (1999). Spatial attention improves performance in spatial resolution tasks. *Vision Research*, *39*(2), 293–306. doi:10.1016/s0042-6989(98)00114-x

APPENDIX A

Chi-Square tests for differences in Flirty and Stalking identifications with Yate's continuity correction for small samples:

Flirty vs Stalking	χ^2	*p*	CI (%)	Difference (%)
Experiment 1	31.15	> .001	[.07; .14]	9.56
Experiment 2	6.43	.001	[.01; .10]	5.31
Experiment 4	9.46	.002	[.04; .16]	7.94

APPENDIX B

Table 1a, Experiment 1

Proportion marked of the different object types (cued, non-cued) in the two conditions (R1, R2) and the two cueing behaviors (F, S) and results of the planned simultaneous comparisons ***for cued target trials.***

	M (*SD*)		|Estimate|(*SD*)	|*z*|	*p*
R1, S, CT	.74 (.15)	H_1	.02 (.03)	.697	.928
R1, S, NCT	.73 (.10)				
R1, F, CT	.72 (.14)	H_2	.00 (.03)	.223	.999
R1, F, NCT	.72 (.10)				
R2, S, CT	.87 (.19)	H_3	.15 (.03)	5.40	< .001
R2, S, NCT	.71 (.11)				
R2, F, CT	.85 (.12)	H_4	.17 (.03)	6.14	< .001
R2, F, NCT	.68 (.13)				
R2, S, CT - NCT vs R2, F, CT - NCT		H_5	02 (.04)	.524	.973

Notes. A priori hypotheses were conducted using Bonferroni adjusted alpha levels of .01 per test (.05/5). R1 = Round 1, R2 = Round 2, F = Flirty, S= Stalking, CT= Cued Target, NCT= Non-cued Targets.

Table 1b, Experiment 1

Proportion marked of the different object types in the two conditions (R1, R2) and the two cueing behaviors (F, S) and results of the planned simultaneous comparisons ***for cued distractor trials****.*

	M (*SD*)		|Estimate| (*SD*)	|*z*|	*p*
R1, S, CD	.29 (.13)	H_1	.03 (.03)	1.04	.774
R1, S, NCD	.26 (.10)				
R1, F, CD	.26 (.13)	H_2	.01 (.03)	.351	.995
R1, F, NCD	.27 (.11)				
R2, S, CD	.23 (.18)	H_3	.03 (.03)	2.26	.101
R2, S, NCD	.29 (.11)				
R2, F, CD	.28 (.13)	H_4	.01 (.03)	.212	.999
R2, F, NCD	.29 (.11)				
R2, S, CT - NCT vs R2, F, CT - NCT		H_5	.06 (.04)	1.45	.491

Notes. A priori hypotheses were conducted using Bonferroni adjusted alpha levels of .01 per test (.05/5). R1 = Round 1, R2 = Round 2, F = Flirty, S= Stalking, CD= Cued Distractor, NCD= Non-cued Distractors.

APPENDIX C

Table 2a, Experiment 2

*Proportion marked of the different object types (cued, non-cued) in the three conditions (R1, R2, R3) and cueing behavior (Stalking, Flirty) and results of the planned simultaneous comparisons **for cued target trials**.*

	M (*SD*)		|Estimate|(*SD*)	|*z*|	*p*
R1, S, CT	.72 (.17)	H_1	.02 (.03)	.767	.974
R1, S, NCT	.74 (.09)				
R1, F, CT	.75 (.13)	H_2	.01 (.03)	.441	.999
R1, F, NCT	.74 (.11)				
R2, S, CT	.89 (.15)	H_3	.16 (.03)	5.08	< .001
R2, S, NCT	.73 (.11)				
R2, F, CT	.83 (.17)	H_4	.12 (.03)	3.873	< .001
R2, F, NCT	.71 (.14)				
R3, S, CT	.73 (.15)	H_5	.03 (.03)	.847	.958
R3, S, NCT	.75 (.11)				
R3, F, CT	.79 (.14)	H_6	.03 (.03)	1.10	.870
R3, F, NCT	.76 (.12)				
R2, S, CT - NCT vs R2, F, CT - NCT		H_7	.28 (.04)	6.32	< .001
R3, S, CT - NCT vs R3, F, CT - NCT		H_8	.01 (.04)	.178	.999

Notes. A priori hypotheses were conducted using Bonferroni adjusted alpha levels of .00625 per test (.05/8). R1 = Round 1, R2 = Round 2, R3 = Round 3, S = Stalking, F= Flirty, CT= Cued Target, NCT= Non-cued Targets.

Figure 8

Table 2b, Experiment 2

Proportion marked of the different object types (cued, non-cued) in the three conditions (R1, R2, R3) and cueing behavior (Stalking, Flirty) and results of the planned simultaneous comparisons ***for cued-distractor trials****.*

	M (*SD*)		|Estimate|(*SD*)	|*z*|	*p*
R1, S, CD	.24 (.14)	H_1	.03(.03)	.926	.936
R1, S, NCD	.28 (.11)				
R1, F, CD	.27 (.15)	H_2	.01 (.03)	.347	.999
R1, F, NCD	.26 (.10)				
R2, S, CD	.15 (.15)	H_3	.13 (.03)	3.81	< .001
R2, S, NCD	.28 (.11)				
R2, F, CD	.27 (.23)	H_4	.01 (.03)	.225	.999
R2, F, NCD	.27 (.12)				
R3, S, CD	.27 (.12)	H_5	.04 (.03)	1.12	.858
R3, S, NCD	.23 (.12)				
R3, F, CD	.26 (.16)	H_6	.02 (.03)	.775	.973
R3, F, NCD	.23 (.12)				
R2, S, CD - NCD vs R2, F, CD - NCD		H_7	.13 (.05)	2.90	< .003
R3, S, CD - NCD vs R3, F, CD - NCD		H_8	.06 (.04)	1.34	.725

Notes. A priori hypotheses were conducted using Bonferroni adjusted alpha levels of .00625 per test (.05/8). R1 = Round 1, R2 = Round 2, R3 = Round 3, S = Stalking, F= Flirty, CT= Cued Distractor, NCT= Non-cued Distractors.

APPENDIX D

Table 3a, Experiment 3

Proportion marked of the different object types in the two conditions (R1, R2) and the two cueing behaviors (F, S) and results of the planned simultaneous comparisons ***for cued target trials****.*

	M (*SD*)		|Estimate| *(SD)*	|*z*|	*p*
R1, S, CT	.76 (.16)	H_1	.02 (0.04)	.389	.992
R1, S, NCT	.75 (.08)				
R1, F, CT	.76 (.22)	H_2	.00 (0.04)	.060	1.00
R1, F, NCT	.76 (.09)				
R2, S, CT	.81 (.16)	H_3	.03 (0.04)	.658	.943
R2, S, NCT	.78 (.11)				
R2, F, CT	.80 (.15)	H_4	.01 (0.04)	.359	.994
R2, F, NCT	.81 (.07)				

Notes. A priori hypotheses were conducted using Bonferroni adjusted alpha levels of .0125 per test (.05/4). R1 = Round 1, R2 = Round 2, F = Flirty, S= Stalking, CT= Cued Target, NCT= Non-cued Targets

Table 3b, Experiment 3

Proportion marked of the different object types in the two conditions (R1, R2) and the two cueing behaviors (F, S) and results of the planned simultaneous comparisons ***for cued distractor trials****.*

	M (*SD*)		|Estimate| (*SD*)	|*z*|	*p*
R1, S, CD	.24 (.15)	H_1	.03 (.04)	.798	.891
R1, S, NCD	.21 (.10)				
R1, F, CD	.24 (.14)	H_2	.02 (.04)	.399	.991
R1, F, NCD	.23 (.11)				
R2, S, CD	.20 (.16)	H_3	.00 (.04)	.092	1.00
R2, S, NCD	.20 (.09)				
R2, F, CD	.22 (.18)	H_4	.01 (.04)	.184	1.00
R2, F, NCD	.21 (.10)				

Notes. A priori hypotheses were conducted using Bonferroni adjusted alpha levels of .0125 per test (.05/4). R1 = Round 1, R2 = Round 2, F = Flirty, S= Stalking, CD= Cued Distractor, NCD= Non-cued Distractors.

APPENDIX E

Table 4a, Experiment 4

*Proportion marked of the different object types (cued, non-cued) in the two conditions (R1, R2) and the two object shapes (E, A) and results of the planned simultaneous comparisons **for cued target trials**.*

	M (*SD*)		|Estimate|(*SD*)	|*z*|	*p*
R1, S, CT	.71 (.14)	H_1	.00 (.04)	.026	1.00
R1, S, NCT	.71 (.10)				
R1, F, CT	.74 (.11)	H_2	.06 (.04)	1.61	.43
R1, F, NCT	.67 (.10)				
R2, S, CT	.90 (.23)	H_3	.21 (.04)	5.39	< .001
R2, S, NCT	.68 (.13)				
R2, F, CT	.86 (.18)	H_4	.16 (.04)	4.01	< .001
R2, F, CT	.85 (.11)				
R2, E, CT - NCT vs R2, A, CT - NCT		H_5	.03 (.06)	.519	.990

Notes. A priori hypotheses were conducted using Bonferroni adjusted alpha levels of .01 per test (.05/5). R1 = Round 1, R2 = Round 2, S = Stalking, F= Flirty, CT= Cued Target, NCT= Non-cued Targets.

Table 4b, Experiment 4

*Proportion marked of the different object types in the two conditions (R1, R2) and the two cueing behaviors (F, S) and results of the planned simultaneous comparisons **for cued distractor trials**.*

	M (*SD*)		|Estimate| (*SD*)	|*z*|	*p*
R1, S, CD	.32 (.17)	H_1	.04 (.04)	1.07	.772
R1, S, NCD	.27 (.10)				
R1, F, CD	.35 (.16)	H_2	.08 (.04)	1.036	.177
R1, F, NCD	.28 (.11)				
R2, S, CD	.13 (.12)	H_3	.18 (.04)	4.59	< .001
R2, S, NCD	.30 (.13)				
R2, F, CD	.21 (.15)	H_4	.08 (.04)	2.12	.146
R2, F, NCD	.29 (.13)				
R2, E, CT - NCT vs R2, A, CT - NCT		H_5	.25 (.06)	4.45	< .001

Notes. A priori hypotheses were conducted using Bonferroni adjusted alpha levels of .01 per test (.05/5). R1 = Round 1, R2 = Round 2, E = Eyes, A= Arrows, CD= Cued Distractor, NCD= Non-cued Distractors.

Part III

Review

The last part of this thesis reviews and discusses the results in relation to each other, particularly with regard to their significance for our understanding of dynamic representations of objects and events. The discussion is partially based on a process flow diagram that depicts the different levels of perception discovered as well as the different influential factors discussed.

Chapter 5

General discussion and conclusion

5.1 Summary of results

The three studies presented in this thesis produced the following main results:

1) Children as young as 6 years process multiple dynamic objects based on the scene and the relation of objects. The results found may have been intensified using an experimental design that involved abrupt viewpoint changes – consequently, participants may have adopted a scene-based strategy to recover tracked targets after the rotation of the (3-D) board more easily. The preliminary processing stage of dynamic objects seems to be congruent with the initial approach to perceive static objects: global over local.

2) Processing dynamic scenes globally, that is, choosing core aspects instead of fine details of the scene to be mentally represented, can also explain results of Study 2 that looked at event perception with omitted visual input. In scenes that involve relations of stimuli and goal related behavior, one needs to consider not only the visual input, but also the complexity of the scene as well as how the available information is structured and presented. Results are explained as a "victory" of the anticipatory trace over the perceptual trace. As shown in studies on representational momentum as well as on basketball and chess constellations, an anticipatory trace is automatically represented and stored along with the perceptual trace when the object (e.g. typical motion) or the scene (e.g. typical sequence in chess) is (well) known. In recognition tasks, the competition is biased

towards the anticipatory trace when the observed scene is logical, for example, fitting a known cause-effect heuristic (a ball that is kicked will move somehow). In contrast, when the scene does not make sense, the perceptual trace wins and the observer can report correctly the missing information in the observed scene. The overwrite of the perceptual trace by the anticipatory trace can happen extremely fast as indicated by the results of an unpublished control experiment in which participants were asked to segment the clips by pressing a button while watching. Surprisingly, the anticipatory trace was even fast and/or robust enough to overwrite perceptual traces when the participants were asked to actively pay *attention* to the contact moment. Theories of (event) perception should integrate how information is presented and structured to account for the competition of the two stored traces.

3) In order to observe attentional processes that are not "blurred" by information structures (inevitable in real-world video clips), we relied upon the MOT task in Study 3 again. Here, I manipulated the way in which information was filtered and weighted by asking the participants to identify object behavior (gazing and color) and object status (target or distractor). (a) Results replicated findings of numerous studies showing that features are not processed during tracking. (b) When participants did the identification during tracking, they were able to weigh the incoming sensory information based on their current task, that is, they integrated and maintained feature information in their representations. (c) When features were valued, it helped the participants to *suppress* identified distractors during tracking. (d) When features were not of interest any more, participants immediately put full weight on spatiotemporal aspects. However, I assume that many features still reach consciousness but they simply do not reach the threshold of awareness when they are not actively attended to, and consequently, not valued (see Figure 6). (e) The strategic input weighing effect during tracking has been observed to be the strongest with color cues, but it was also possible for the participants to identify the object status even when only a minimal gaze cue (two small dots moving in two white spheres) was presented shortly.

5.2 A multi-process approach to visual perception?

The question mark in the title reflects the general uncertainty that is quite prominent in literature: it is hardly possible to define any definite processes. Gordon (2004) summarizes the problem as follows (p. 217):

"To conclude (...), it can be asserted that there is as yet no satisfactory general theory of visual perception. For example, no theory has adequately united a full analysis of the environment and the cognitive aspects of seeing. No general theory has thoroughly incorporated and explained the motor aspects of seeing. The extent to which perception is determined by stimulation (involving bottom-up processes) or knowledge (top-down processes) has not been agreed upon."

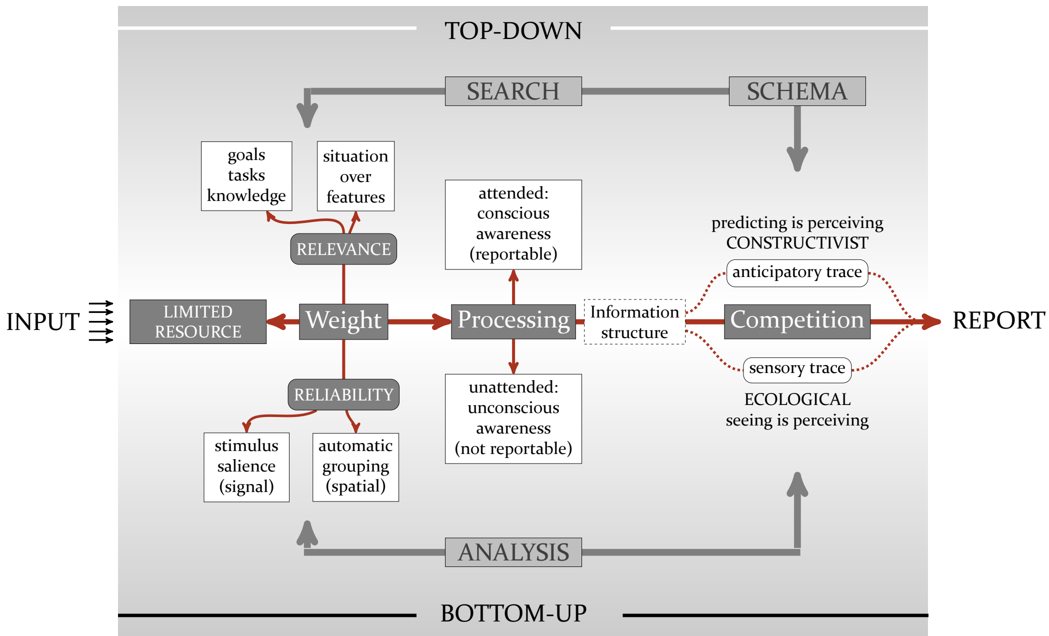

FIGURE 7: Visual perception as a dual process: Perceptual processing on different perceptual levels.

Figure 7 incorporates the different ideas and theoretical considerations addressed in this thesis in one diagram. Basically, it follows the general order that lays the foundations to understand perceptual processes in different theories (see Gordon, 2004, for an overview): sensory input, filtering through a limited resource, information processing, and action (or report). Based on the current work, one needs to include the following processes to account for the obtained results[1]: input weighing, awareness, and competition for representation. Furthermore, the included integral parts have a top-down and bottom-up component that may be given more or less priority based on the given situation or strength of the signal. In this work, I have argued that both, bottom-up and top-down processes, can be important to represent a dynamic scene. Such an integration the constructivist and ecological approach is possible when one considers perception as a cyclical activity. Therefore, Figure 7 also includes the cyclical approach of Neisser (1967), who assumed that perception is a continuous process with no definite starting or finishing point.

However, I do not dare to claim that Figure 7 automatically generalizes to situations and conditions outside the specific experimental designs in this work. Perception of dynamic objects and events is not that simple: one needs to account for numerous internally and externally influencing factors; different conditions result in different perceptual approaches. It is possible that the human brain has numerous ways (useful or useless) to represent objects and events; ways that evolved through a trial-error principle. Still, there are neurobiological constraints that should not be disregarded and that result in resource limitations or give rise to competition of representations. In the following, I will describe Figure 7 with regard to each component, as well as dorsal/ventral and bottom-up/top-down systems and attentional processes, along with their role in each of the presented papers and future research.

5.2.1 Limited resources

The sudden appearance of a gorilla can go unnoticed when the observer focuses all of his attention on a challenging primary task like counting basketball throws (e.g. Simons & Chabris, 1999). The effect was named *inattentional blindness* and is thought to reflect the limited capacity of the perceptual-visual resource. This and similar experiments reveal that the human perceptual system continuously monitors only relatively few, specific visual properties. One may say that such an experiment defines the *quality* of the limited capacity. The paradigms presented in this work, however, rather *quantify* the limitations by suggesting that tracking capacity reflects a fixed number of object-based "slots" (e.g. Awh, Barton, & Vogel, 2007) and by suggesting that scene encoding processes are limited to a number of meaningful (sub)events (Zacks et al., 2007).

[1]Note that definitive statements in respect to the order of individual perceptual processes can not be made based on the experiments presented. Further research is needed to determine the order of processes, especially for the competitive traces.

Limited object tracking

In MOT, people can track at least four objects e.g., Pylyshyn and Storm (1988) – a limit similar to the mental storage capacity found in memory studies (see Cowan, 2001) discussed in Chapter 3. I will briefly present the different approaches that can be found in literature to describe the relationship of a limited mental resource and tracking.

With a neurological approach, one can explain the nature of the resource-limiting tracking performance with an assigned pool of neurons that reacts to targets (e.g., Culham et al. (1998), see Horowitz and Cohen (2010) for a complete list of relevant references). When target load increases, the number of neurons that can react to each target decrease which reduces the quality of the representation (tracking as a parallel process: Pylyshyn & Storm, 1988; Yantis, 1992).

d'Avossa, Shulman, Snyder, and Corbetta (2006) proposed that mental resources may, alternatively, be constrained by a temporal, that is, a serial or "a limited capacity, time-multiplexing processor rather than multiple, parallel processes" (p. 3409). The idea: a single attentional focus has to be moved among the targets (serially), reducing the quality of the representation when the number of targets increases. The results of a study by Howard and Holcombe (2008) support the application of serial and parallel processes during MOT tasks. The authors assume that some attributes are updated in parallel, while others (e.g. object identities: Oksama & Hyönä, 2004) are processed serially.

While there is widespread agreement that some kind of limited mental resource is involved in tracking processes (see for example Alvarez & Franconeri, 2007), no agreement can be reached on the nature of perceptual, attentional, or cognitive processes underlying tracking. A main share of theories considers limited *attention* resources to be the cause of tracking errors. MOT and attention is discussed extensively later in section 5.3.3.

In the current work, I especially focused on the debate concerning the role of spatiotemporal aspects and objects features (see section **??** and Study 3 for a review of the debate) and hypothesized that there must be a secondary finer filter that is responsible for the ambiguous results found in literature.

Limited event perception

In event perception, the information in our environment almost always exceeds the capacity of the working memory system. It is assumed that the limited working memory capacity influences the representational process by picking only informative moments to be retained (Levin & Saylor, 2008). Based on the EST, the capacity of the working memory can be further augmented by the use of previously stored knowledge (Zacks et al., 2007), called *event schemata*. These schemata consist of semantic memory representations of shared features and events, and also store information about the sequential structure of an activity. The EST states that event schemata influence a current event model that in turn ensures a stable representation of a current event by perceptual processing accordingly. In Study 2, we observed illusionary gap fillings that cannot be explained sufficiently by the EST. If event schemata integrate experience and knowledge, we would have expected to find differences between referees, players, and novices. For our results, it seems rather plausible that, due to the limited mental capacity, only informative moments are processed. However, this approach also does not explain our results completely because this would imply that the ball-contact moment was an informative moment in some conditions but not in others. In section 5.2.4, I therefore propose that the *information structure* influences which perceived event moments can be reported. That is, whether the event segments fit in a causal or non-causal manner influence what can be reported after watching them.

Little is known about the nature of the limitation of the mental capacity in event perception. Future research may test individuals' ability to process and segment different events simultaneously. For example, one could expose participants to a situation in which they follow a conversation and observe a card game (different streams), or expose them to two conversations or two activities (similar streams). It may be possible that observers are limited to processing only one stream at a time, or only different or only similar streams in parallel.

5.2.2 Weight

Weighted object tracking

The "magic number four" that is often assumed to reflect limited resources does not appear to be fixed. With the right display conditions (e.g. reduced object speed) in MOT tasks, Alvarez and Franconeri (2007) were able to raise the target number up to eight objects at once. As shown in Study 1 in Chapter 2, multiple objects can be compressed by higher-order structures. Such an effect has been observed in literature before; targets were organized and automatically grouped by symmetry or common fate (Wang, Zhang, Li, & Lyu, 2016), feature (Erlikhman, Keane, Mettler, Horowitz, & Kellman, 2013), or simply structured as a rigid polygon (Yantis, 1992). Papenmeier et al. (2014) observed that the **reliability** of spatiotemporal information determines whether feature information is processed or used (in an automatic manner).

While such automatic grouping or preferential processes could still be explained as having evolved as short-cuts due to limited resources, the results of Study 3 cannot. Here, observers were able to integrate knowledge about the objects' properties and motion behavior to strategically enhance tracking performance for the cued target and reduce the confusion of targets with the cued distractor. Participants were given highly reliable spatiotemporal and feature information, that is, they were not forced to rely on one or the other due to external circumstances (e.g. reduced frame rate) but due to internal control (i.e. the identification task). Still, the **relevance** of the processing object features played an influential role in whether participants actually processed them. Even when participants tracked one round *with* the identification task and knew about the helpful properties, they did not continue to process them in addition to the tracking task when no identification task was given.

As others before, I conclude that MOT is a task that involves both attention-bound and pre-attentive processes that in turn constitute the object-tracking capacity (see also Bello, Bridewell, & Wasylyshyn, 2016). The current work adds, however, that the capacity is not fully exhausted with a simple tracking task; the task still allows the participants to weigh the information input and process objects and features strategically. Attention and weighing are discussed in more detail in 5.3.3 where I refer to the biased competition model of selective attention in effect models (Desimone, 1998).

Weighted event perception

In real life scenarios with more complex stimuli and more detailed situational information, as presented in Study 2), such a weighing filter must be running constantly in order for us to function properly. However, less is known about the influence of expertise (i.e. top-down processes) on event perception. In Study 2), we showed that the vulnerability of cognitive-perceptual processes to illusionary event fillings is similar in experts and novices when observing well-known events. Future research should work with stimuli and properties that specifically trigger knowledge in experts in order to observe weighing differences. I hypothesize that experts "know where to look" in more complex tasks, thus, they will process stimuli based on their relevance, while novices will rather depend on the reliability of presented information.

Concerning the idea of situation and feature, or object- and scene-based processing, one possible experiment to contextualize our results could be to disrupt a higher-level form of event organization (e.g. with misordered events). It may be possible that the **reliability** of the coarse event structure automatically determines the weight assigned to smaller subparts of events like the ball contact moment (idea inspired by face perception research of Young, Hellawell, & Hay, 2013).

Left unanswered is the question whether the failure to report the gaps arises from perceptual or memory processes, nor did Study 2 help to determine *when* exactly event perception involves prediction and comparison, and when it does not. Hymel, Levin, and Baker (2016) tested the event perception of participants with disordered event sequences and deduced that viewers can perceive elements of events without necessarily testing expectations and hypotheses about the observed sequences (as would have been predicted by the EST, Zacks et al., 2007) – unless task-specific demands (like detecting misorderings) require them to do so. Such an automatic "default" processing mode of events may follow similar, or the same, principles as the automatic scene-based tracking of multiple objects observed in Study 1, while the **relevance** of task-specific demands can influence the hierarchical organization of the perceptual levels in order to detect misorderings or, as observed in Study 3, to process object properties.

5.2.3 Processing

Dual-mechanisms frameworks as described in section 5.2.2, namely, (1) event perception depends on the awareness of prediction errors (Zacks et al., 2007) and (2) processing sequences is an *inherent* dimension of event perception (Raisig, Welke, Hagendorf, & van der Meer, 2010), are frequently found in literature. As participants in our lab were always asked to *report* what they had seen, it makes sense to go beyond the area of perception and look at executive functions as well. An interesting dual-mechanisms concept for executive functions that nicely fits the ideas discussed is a division in proactive and reactive processing. Braver (2012) proposed a dual-mechanisms of control (DMC) framework in which individuals can *proactively* organize their cognitions in order to inhibit distractions and maintain a goal. In contrast, the *reactive* control mode rather works as a "just-in-time" mechanism that mobilizes the identification and execution of goal-relevant actions only when absolutely needed, probably relying on bottom-up short-cuts. Assuming a similar functioning to the DMC framework in event perception, the perception of an event may involve two types of processing: (1) a default or reactive mode in which parts and subparts of the ongoing event are not actively attended, expectations are not tested, and details cannot be reported, and (2) a proactive mode in which participants are consciously aware of the current state of the on-going event, continuously maintain a representation of goal-relevant information, and mobilize resources to identify the event and relevant parts. Whether participants can actually *report* a seen event correctly afterwards may still be independent of their applied processing framework. The overall global information structure could affect the competition of perceived sensory traces and predicted anticipatory traces (discussed in the next section), leading to biased reports.

The DMC framework of Braver (2012) also mirrors the concept of attentional control settings described in Study 3. The applied design in the tracking study is assumed to evoke a top-down filtering of sensory input due to the identification task. The participants still processed spatiotemporal information (in order to solve the tracking task), but put weight on the feature information as well. By proactively *attending* to the objects' behavior and status, they were able to report consciously their observations after each trial. As demonstrated in the second experiment of Study 3, in which the participants were told that the objects would show the exact (possibly helpful) behavior pattern but that there would not be an identification task after each trial, the significance of object features is as important as the proactive or reactive control setting of attention.

In general, together with findings of (Papenmeier et al., 2014), results of Study 3 support Model B presented in Figure 6, page 61. Many stimuli reach the mind but the eye sees only what the mind is prepared to comprehend (Davies, 1951), that is, what is conscious. Attending to the conscious stimuli (or features) allows them to reach the threshold of awareness, thus, they form a representation (for competition) and can be reported.

One may assume that the weighing of relevance and reliability provides the structure for the actual processing part, similar to what has been called "bottleneck" or filtering in literature. But, as I see it, weighing is more advanced than a simple filter since it does not only provide bottom-up and automatic "tricks" like grouping, but also includes a top-down structure that adapts flexibly to current needs. Next to internal needs, like proactively and strategically attending to object features, the weighing process can also be influenced by external circumstances like speed of moving objects and their spatial distance to each other (inter-object spacing, e.g. Franconeri, Jonathan, & Scimeca, 2010). More research could specify the relation of limited resources, weighing and processing: for example, when and how the weighing process intervenes, whether weighing is also influenced by processing output (initiating a cyclical process), or how many components (like reliability and relevance) can be assigned to the weighing process. Clearly defined goals and tasks may be important for the weighing to occupy limited resources efficiently – as we have observed in Study 3, Experiment 3, in which a better instruction led to improved transparency and decreased confusions of cued distractors. Optimistically assuming an interaction of flexible weighing and proactive/reactive (or conscious/unconscious) processing provides more feasible and more intuitive explanations and may help to explain ambiguous findings in the dynamic-perception literature. Future studies in tracking may eventually move beyond questions of *how* resources are limited and start discussing fundamental questions on what the resources are and how we can use them for maximum benefit in everyday life.

5.2.4 Competition

Baker and Levin (2015) propose that relational triggers induce comparisons between currently visible visual properties and those encountered and encoded in the past. This enables the observer to intelligently allocate limited resources to new features and to withdraw resources from stable, well-known features. The authors further state that the observer not only associates the ongoing perception and activates previously seen visual properties, but also activates a predictive process that anticipates future properties and events. In Chapter 3, section 3.3, I presented a study with basketball novices and experts by Didierjean and Marmèche (2005) who explained the errors made in the recognition task by assuming two competitive traces that are stored simultaneously. They explained that experts made more errors because they immediately predicted the follow-up constellation of the depicted actual constellation and had trouble afterwards to distinguish between the actual perceptual trace and the anticipated trace. Similarly, Johnson et al. (1993) and Johnson and Raye (1981) assumed a competition of source and reality monitoring to explain the boundary extension effect.

In Study 2 we observed illusory gap fillings when participants watched a well-known, simple activity: the kick or throw of a ball. While one could explain their reporting errors in the causal condition with the selection of only informative moments to reduce the amount of incoming information for the limited mental resource, such an approach does not explain why the participants were able to report to have seen the contact moment in the non-causal condition. How could they know *when* the ball-contact moment was an informative moment to select for storing? Following the idea of Didierjean and Marmèche (2005) and Johnson et al. (1993), I propose that we may have observed a competition of stored traces and identified an influential factor: the information structure. The outcome of the competition was influenced by the information structure: When the information was presented in a causal manner, the participants' recognition report was biased towards the stored anticipatory trace – but when the observed scene was linked in a non-causal manner, the participants relied on the stored sensory/perceptual trace. Such an approach is somewhat connected to the conceptual semantics proposed by Jackendoff (1991): a syntax, that is, a structure guides the encoding of human understanding of the world, making inferences based on heuristics, and is connected to the internal structure of the perceiver.

Somewhat far-fetched but still worth speculating about is the connection of visual information structure in event perception to lexical structure in sentence processing. Pustejovsky (1991) showed that the lexical specification of a verb's event-type (e.g. "she hammered the metal" (a process) vs "she hammered the metal flat"(a transition)) can be overruled "as a result of syntactic and semantic compositionality of the verb with other elements in the sentence" (p. 56). That is, the internal structure of the event not only changes the meaning of a word but is also an important representation for general lexical semantics. Applied to Study 2, this would mean, that the causal/non-causal information structure overruled the "meaning" of the ball-contact moment and by that may have changed the overall representation reflected in the report of the recognition task.

Future research in event perception should clarify how the perception, or the report of perceived events, changes due to the information structure. First, a two-part experiment in which participants first watch event sequences and do the recognition task and second, watch the event sequences and segment them while doing so, would provide clarity concerning the debate of "report vs perception" and by that may be able to show *when* the overriding due to information structure happens exactly . Second, further research is needed to verify the influence of stored anticipatory and perceptual/sensory trace and its relation to schemata and event models suggested in the EST.

The current work does not provide any indication of active competitive traces in tracking. For one thing, most of the researchers argue that extrapolation is not applied in MOT because observers rely solely on "location-only", that is, only on the positions of the tracked objects (see Lukavský & Děchtěrenko, 2016, for an up-to-date summary of studies and accounts). Hence, there would be no anticipated trace to compete with the perceptual trace. On the other hand, findings on extrapolation MOT are mixed and it would be interesting to explore whether a given higher-order information structure influences tracking performance and/or the report of an object-constellation.

Is the competition of traces the same as weighing? The way I understood the concept of competing traces by Didierjean and Marmèche (2005) is, that traces can only happen when the stored traces are fully "equipped" or fully anticipated. Thus, the traces are a result of extensive filtering, weighing, and processing, a procedure that may be cyclic in nature to find the best possible interpretation. The next section describes the outer circle of Figure 7 that is based on Neisser (1967) and, in a broader sense, on the EST by Zacks et al. (2007).

5.2.5 Bottom-up and top-down levels

What and where?

In Chapter 1 I compared the constructivist (inferential) approach (e.g. Gregory, 1980) to the ecological (direct) approach (e.g. Gibson & Gibson, 1955) of perception. Data presented here advocate that these two seemingly contradictory approaches can co-exist if a current goal or the structure of the available information allows or demands it.

Such a possible dual-process approach is in line with Norman (2002) who reviewed the two theoretical approaches with regard to their parallelism to dorsal and ventral systems. The idea of two visual systems stems from research by Schneider (1967) in which ablation of the cortical visual system of a group of golden hamsters lead to incapability of pattern discrimination but left the ability to orient towards objects intact, while disconnecting the superior colliculus lead to the opposite effect. According to the researcher and quite a few other studies on cortical and subcortical systems (e.g. Ingle, 1967; Trevarthen, 1968), the ventral pathway answers the question “What is it?”, the dorsal pathway answers the question “Where is it?” (see e.g. Mishkin & Ungerleider, 1982; Mishkin, Ungerleider, & Macko, 1983). Thus, the ventral system is responsible for recognition and identification by comparing input with stored representations, while the dorsal pathway’s primary function is to individuate and analyze visual stimuli to guide the observer’s behavior in the given environment.

Norman (2002) contrasted the ventral and the dorsal system and identified their differences. For example, the dorsal pathway responds to low contrasts at coarse spatial frequencies (contrast sensitivity) and to static shapes, but the perception of complex motion is processed by the ventral system (Ferrera, Rudolph, & Maunsell, 1994). Further, the ventral system is responsible for memory-based processes and uses stored representations to identify objects and events. Additionally, Norman (2002) also explained how different reference frames arise due to the different purposes of the two pathways. The dorsal one aims to (inter)act with the environment, that is, it needs absolute metrics (e.g. to grasp something). As a result, an egocentric reference frame is the primary choice of the dorsal system. On the other hand, the ventral system focuses on object-centered information to organize, recognize, and identify objects and events. Therefore, an allocentric reference frame, or scene-based processing with relative metrics is applied based on purposes of the ventral system.

Scholl (2009) compares MOT to visual search. Visual search can interfere with scene encoding because both processes rely on ventral pathways, that is, “an identification-based form of attention” (see p. 71). In contrast, MOT may rely heavily on dorsal (individuation-based) attention because visual search and tracking hardly interfere with each other (Alvarez, Horowitz, Arsenio, DiMase, & Wolfe, 2005). Accordingly, we observed an automatic scene-based tracking approach in Study 1 that was probably executed without conscious awareness, thus via the dorsal system. The participants “filtered” the sensory input through automatic grouping which seems to be a fundamental perceptual strategy. The dorsal pathway picks up visual information quickly and mainly unconsciously, presumably being able to carry out the performance of well-ingrained actions or behaviors on its own (i.e. many of our daily activities).

Such an automatic processing of visual information may also apply to Study

2 in which participants processed a well-known event/motion pattern independent of (presumably top-down) expertise. However, here, it was important "how" the information was presented and structured, a component that has been overlooked so far. Since the ventral pathway relies on long-term memory to compare current visual information with experienced representations, it is possible that the observed cause-effect relationship triggered the ventral pathway to increase processing speed by supplying heuristics. Norman (2002) proposed that the ventral and dorsal pathway can act in parallel. Similarly, Figure 7 presents different subcomponents that are roughly divided into bottom-up and top-down processing that act in parallel and may interact when needed.

Scholl (2009) ascribed the inability of the observer to encode surface features of objects during tracking to the reliance on the dorsal pathway. Yet, the data presented in Study 3 point towards the use of feature information during a spatiotemporal-based task – thus, favoring rather a cooperation of both, the dorsal *and* the ventral pathway that is triggered by specific situational demands.

Unidirectional or iterative?

Most theories of perception suggest some form of interaction between seeing and interpreting or perception and prediction (including Helmholtz, 1866; Neisser, 1967). Di Lollo, Enns, and Rensink (2000) summarized existing approaches and considered two options:

On the one hand, one could see the communication between lower and higher levels as an unidirectional process. The perceived attributes of a stimulus are gathered and maintained while processing ascends to higher levels. That is, when all bottom-up stimulus information is gathered, the process is completed and the outcome is transferred to the next higher processing level. On the other hand, the process could be iterative through constant exchange of neuronal signals among different levels. Many ascending and descending pathways allow for an iterative-loop system that can reduce noise and helps to verify hypotheses. Di Lollo et al. (2000) argued that many current theoretical approaches could be augmented with notions of iterative reentrant processing – that is, theories could profit from neurological findings that propose that the brain has a back-projecting (i.e. reentrant) architecture that connects sensory modalities with unimodal associations (Pandya, Seltzer, & Barbas, 1986). If memory processes operate early to organize sensory data into meaningful percepts based on current expectancies and experiences, perception requires such a reentrant mechanism that consolidates memory while linking multiple other networks in the brain (see Luu et al., 2010, for more neurophysiological details).

One such cyclical approach to perception comes from Neisser (1967). According to the "perceptual cycle", perception is a cyclical process with no start or finish point. Observers constantly analyze and sample the perceptual world (bottom-up) and react to stimuli that are relevant to them. These reactions activate given models (or schemata) that observers try to verify with a search for expected features. Accordingly, our perception is led by what we expect to find as well as what we have already found. The schema directs the perception to certain parts of the information that is available to us, and the found information itself continually modifies and changes the schema. Neisser sees perception as a complex process at all levels: cognitive, physiological, interpersonal and socio-cultural.

How would such an approach work for the presented data? Di Lollo et al. (2000) stated that many findings can be explained with a unidirectional process, and so can the findings of Study 1 and Study 2. Participants relied heavily on the bottom-up part and sampled the given information. However, the tracking study in Study 3 involved a top-down task that required to adopt a schema (or an attentional control set) that influenced the perceptual process to sample a specified part of the available information and to constantly search for a verification of the schema (e.g. is the target I am tracking a cued object?). While I agree with the authors that many findings follow a simple unidirectional process, I can easily imagine that participants went through cycles during each tracking trial to ensure that they could identify the objects' nature afterwards.

Neisser (1976) also had an interesting view on attentional processes. He regarded attention as a skilled activity that is *not* limited in its capacity. The selection of stimuli would then be based on the immediate situation and on the anticipation of what may be relevant. Neisser did not state it explicitly, but his view rather sees attention as an *effect* triggered by demands of the current situation, than as the *cause* for a selective perceptual output. While event perception is thought to rely mainly on memory processes – event models are basic units of long-term memory (Ezzyat & Davachi, 2011) and are represented in working memory to make relevant information to the immediate context available (Speer & Zacks, 2005; Zacks, 2008) – MOT is often discussed as an attentional process. In the next section I will shortly present cause and effect theories of attentional selection and speculate on their use to explain tracking results. Very briefly, I will comment on the role of attention in event perception.

5.3 Theories of Attentional Selection

As already indicated in the beginning of Part II of this thesis: attention has many denotations, connotations, and subtypes. The lack of clarity in the literature may have caused researchers to rather propose their own theories than to extend or falsify existing work. With regard to the two models that represent the most contrasting concepts of the many models proposed in the literature (Figure 6 on page 61, Lamme, 2003), it seems plausible that the numerous views and theoretical approaches to understand attentional processes classify as either "cause-oriented" or "effect-oriented". Viewing attention as the "cause" of various cognitive phenomena assumes that attention is a specific mechanism, made of interacting subcomponents in the brain, and a pool of resources that is allocated to a task (order in line with Figure 6, model A). Viewing attention as an "effect" conceptualizes attention as an epiphenomenon of multiple independent cognitive systems, that is, as a byproduct of information processing among multiple systems (order in line with Figure 6, model B).

5.3.1 Cause Theories

An often used metaphor in cognitive psychology that is a good example for a "cause-oriented" theory is the attentional spotlight metaphor (e.g., Cave & Bichot, 1999; Posner, 1980; Posner et al., 1980). Posner et al. (1980) suggested that a spotlight scans the environment serially and draws attention to stimuli within the field of view in an endogenous manner. Stimuli outside the central view are brought into the center via an exogenous system (e.g., reflexive orientation to a color singleton or gaze-cues). The delay between onset of a cue and the increase of an electrophysiological response at the cued location has been thought to measure the time needed to re-allocate the attentional spotlight to the new location (Müller, Teder-Sälejärvi, & Hillyard, 1998). However, by attempting to measure the maximum width of the attentional spotlight, it was found that unattended stimuli also produced brain activity (less than attended stimuli though) (see e.g., Castiello & Umiltà, 1992; Tong, 2004; Wojciulik, Kanwisher, & Driver, 1998). Attention could thus be split, or divided, urging the adoption of a multifocal metaphor that describes how multiple spotlights scan localized regions of the visual field.

Closely related to the spotlight metaphor is the idea that attention depends on a limited resource that is allocated to selected perceptual units in accordance to a general purpose (task or goal) (e.g., Hasher & Zacks, 1979; Kahneman, 1973). The amount of resource allocated is thought to affect attentional quality, measured as task performance (Pashler, 1998). Thus, the attentional resource actively modulates information processing. Attentional phenomena can then be explained as a flexibly graded sharing of a single resource or capacity, for example dual-task interference (Christie & Klein, 1996), or mental rotation (Carpenter, Just, Keller, Eddy, & Thulborn, 1999).

The limited resource idea explained phenomena that the spotlight metaphor could not, because it allowed to modulate the attention intensity to account for a second task. However, some tasks do not interfere with each other (e.g., difficulty insensitivity: Navon, 1984), while some modality changes lead to interferences in the other task even when the difficulty stays exactly the same (e.g., structural alteration effect: Wickens, 1991). Empirical findings like these led to models that favor multiple resources instead of a single one. Accordingly, an interference between tasks only occurs when both tasks draw upon the same resource "reservoir".

5.3.2 Effect Theories

Effect theories of attention are all about the competition for limited perceptual processing. Instead of focusing on the properties of a central executive system responsible for the modulation of spotlights and the management of resources, effect theories attempt to answer which stimulus will win the competition, that is, which stimulus will rise above the threshold and reach awareness. Attention and awareness are seen as a result of emerging brain processes (see Dennett & Kinsbourne, 1992) and thus are only byproducts of information processing. Krauzlis, Bollimunta, Arcizet, and Wang (2014) proposed that attention arises as a byproduct of value-based decision making that requires to properly estimate the current state of the environment and their actors – a complex task that involves the interpretation of many sources, including features of the external world as well as internal status aspects like goals, needs, and prior knowledge. The central premise of this idea is: attention's selective filtering is due to weighing of input (a process that could be based on Bayesian inferences, similar to what is described in Rao, 2010). Krauzlis et al. (2014) further says that such context-dependent decision making is based on a competition between possible interpretations of the current scene. Each weighted possible interpretation can be seen as a candidate template that competes to be the "best-matching" template.

Support for the "effect-view" comes from studies with patients suffering from parietal lesions and hemispheric neglect (see Fernandez-Duque & Johnson, 2002). They react more slowly to targets in non-cued locations, a finding that would be explained with deficits in the disengagement of the attentional spotlight, thus it would be a problem of *spatial* information processing - which is not located in the parietal cortex. The effect-oriented approach would argue for a competition of weights that have preferences for stimuli in the undamaged area compared to neural representations in the damaged (parietal) area. This would also explain the finding that a better than normal perception for the unaffected area was observed (Seyal, Ro, & Rafal, 1995): there is no competition of the lesioned area. Figure 8 shows the distinction of findings that can be associated with attention as a cause or as an effect.

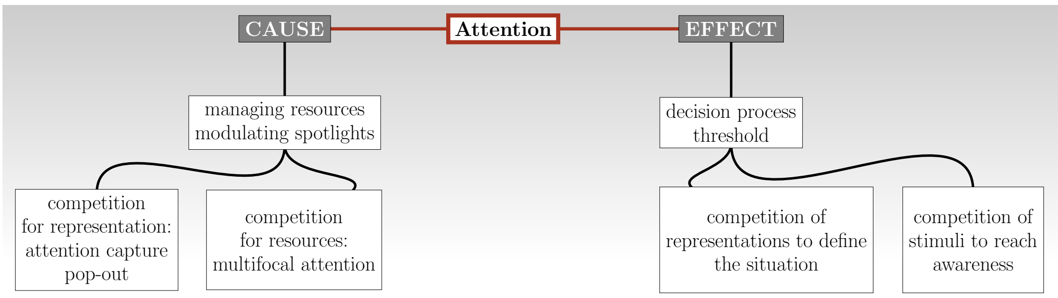

FIGURE 8: Attention as a cause and as an effect: an overview.

It has been criticized that effect theories describe attention as a pure bottom-up byproduct while humans can indeed suppress salient stimuli and assign priorities to less salient stimuli intentionally. A *biased competition model* of selective attention (Desimone, 1998) evolved that allows top-down factors to modulate bottom-up processes. That is, the observer holds feature and spatial information of an object in mind and can bias the competition of which one, that ist, which representation (feature or location), reaches the threshold of awareness individually, based on task or goal (top-down modulation: DeWeerd, Peralta, Desimone, & Ungerleider, 1999). Studies with preparatory attention to "good" stimuli (e.g., expected features) revealed an enhanced activation of brain areas for the expected features, that is, to "winning" (Reynolds, Chelazzi, & Desimone, 1999), while attending to a "poor" stimulus lead to response suppression, that is, to “losing” (Chawla, Rees, & Friston, 1999).

5.3.3 Attention in MOT: mixed cause-effect models?

Performance limitations in tracking tasks are usually attributed to a common underlying mechanism that has spatial and temporal limitations. Scimeca and Franconeri (2015) added limitations ascribed to the shape recognition system to the discussion. A main share of theories considers limited attention resources as the *cause* of limited performance. However, in recent work by Papenmeier et al. (2014), performance limitations were attributed to a competitive weighing system (position vs color) – an idea similar to the competitive traces approach described earlier in connection to event processing. The paper by Papenmeier et al. (2014) laid the groundwork for the above presented Study 3 that took the idea a step further: observers do not only use the trace that provides the most reliable information (e.g. Papenmeier et al., 2014, manipulated the reliability of information with frame rate reductions), they can also strategically put weight on the trace that helps to solve the task (here: identify the objects’ behavior and status). It may be time to integrate both effect and cause theories of attention to fully understand dynamic tracking. In this section I will shortly explain the different accounts and propose theoretical extensions.

Spatial, temporal, and shape: limits of tracking

The first paper that presented the MOT paradigm (Pylyshyn & Storm, 1988) assumed that the underlying tracking mechanism was based on a parallel set of object location pointers (called Fingers of Instantiation or “FINSTs”) that adhere to the moving objects like “sticky pointing fingers”. FINST only allows for encoding the spatial location of the object, not properties or features (Pylyshyn, 2004). FINST is broadly described as a fixed-resource theory of visual capacity limits (e.g. Bae & Flombaum, 2012): attention has slot-like representations; processing more objects than available slots is not possible (Drew & Vogel, 2008). The FINST-theory matches a more recent idea that has been called attentional priority map (see Scimeca & Franconeri, 2015; Serences & Yantis, 2007): a set of locations in the visual field are marked for enhanced processing via cortical representations. Franconeri et al. (2010) suggested spatial interference to be the only factor that limits performance. The decrease in tracking performance when the number of targets was increased is assumed to reflect interfering cortical representations of two nearby targets (Franconeri, 2013). While this may just as well be interpreted as two competing representations and thus may be closer to the effect than the cause account of attention, Franconeri (2013) proposed that two close targets are tracked less well because the suppressive surround of one target overlaps with the *spotlight* oriented on the other target and vice versa.

Close to the FINST, but more flexible, is the account by Alvarez and Franconeri (2007). How many targets can be tracked is determined by flexible indexes (FLEXs instead of FINSTs). The flexible part: there is no limit to the number of FLEXs; still, a finite resource sets the spatial and the temporal resolution of each FLEX. That is, the finite resource is divided between the infinite FLEXs, resulting in decreased performance (less resolution) for each target (FLEX) when the number of targets (FLEXs) increases (similar account: the normalization model of attention proposed by Reynolds & Heeger, 2009).

Oksama and Hyönä (2004, 2008) tried to explain their findings on "Multiple Identity Tracking" (MIT; objects were not identical but drawings of animals and things) with only *one* available representation slot, or accordingly, one spotlight. In their MOMIT model they propose a serial switching theory: each target is attended one by one, that is each target position must be updated frequently which gets more difficult with more targets (Howe, Cohen, Pinto, & Horowitz, 2010). Target positions are registered and when the targets are re-attended (in a serial manner), objects that are closest to the last-registered positions are assumed to be the target. That is, more targets mean that the (one) attentional spotlight must cycle through the objects faster.

Instead of an attentional resource, some have proposed the visual system's shape recognition system as an alternative (see Scimeca & Franconeri, 2015, for a review of relevant studies). Observers may encode the target position to form a polygon (see Chapter 2: scene-based processing). Encoding more than 4-5 targets into a polygon when the objects move may decrease performance. Further, the maintenance of the shape may become difficult in some dynamic situations (e.g. crowding).

The MOT paradigm is completely saturated with explanations on how it may work in specific situations. What is missing is an overarching approach that allows object tracking to be a flexible and penetrable process. For me, there is no doubt that tracking can be, and usually is, low-level motion processing. The brain always seeks the path of least resistance and, as shown in Study 1 and 2, the first stage of visual processing is guided by global mechanisms that automatically groups objects in order to save resources and/or energy. Switching to local/feature-processing requires additional efforts. However, if you ask someone to follow a car on the motorway, he will be able to tell the color of the car while tracking others cars to avoid accidents. The existing MOT theories that deny the observer's ability to track features may be too strongly worded and may be based on too many simple and isolated experiments in laboratories. Study 3 shows that observers can strategically chose which path they give more weight. That is, in situations in which it is (for whatever reason) important to process features, the observer *can*. This would be in line with the biased competition model of selective attention in effect models (Desimone, 1998). The observer himself represents both feature and spatial information and can decide, strategically, which representation reaches the threshold of awareness. MOT may be best solved by relying on spatiotemporal information – which is why putting weight on non-helpful feature information decreases performance – but that does not mean per se that the observer cannot process features.

Which model of conscious awareness and attention (Figure 6, page 61) explains feature processing in MOT? While model A describes that only attended visual input reaches consciousness, results by Papenmeier et al. (2014) suggest that most of the given information reached consciousness and that the "sorting by reliability" comes later, thus favor model B. Study 3 shows that the observers can actively chose whether they attend to object features or not (see especially Experiment 2), that is, while the spatial and feature visual input reaches consciousness, the observer chooses whether he attends to the feature objects or not. Attended information streams can be reported, unattended cannot. Study 3 adds a new aspect to the existing MOT literature: the relevance of representations.

5.3.4 Attention in Figure 7: a multi-layered process?

I surmise that attention plays a determining role in the interplay of input weighing, awareness, and competition for representation. In my opinion, allowing attention to be both, a cause and an effect, solves many confusions found in the literature. Here, attention as an effect rather than a cause supports the understanding of top-down processes in visual perception. While, for example, automatic grouping can be explained as caused by managing resources through attentional modulation, feature processing in tracking cannot. A more suitable explanation for top-down processes is that different available representations are weighed and later compete to reach conscious awareness in accordance with the given situation. In Study 2, it is the competition of anticipatory and perceptual representations to define the situation, where the top-down component arises from cause-effect heuristics. In Study 3, participants emphasized features intentionally due to task demands, hence boosted the "winning potential" of a representation that united both feature and location information.

A competition for representation may always take place – whether it is a competition for resources (attention as a cause) or for awareness and definition (attention as an effect) – but while the anticipation of the paths of more than two objects in tracking has not been observed yet (see Luu & Howe, 2015)[2], a competition of anticipatory and perceptual traces can happen in the observation of static images (e.g. Didierjean & Marmèche, 2005) and real-world video clips (Study 2), as long as the given object or scene can assume different modes, is easy to recognize, and well-known (a rocket can be expected to move, a building cannot: Vinson & Reed, 2002). Study 2 added the role of logic and causality, that is, how the available visual information is structured, as a factor that influences the outcome of a competition between anticipatory and perceptual traces.

5.4 Some philosophical final remarks

General opinions have changed many times in the past seventy-five years between the position that perception is "encapsulated" (Pylyshyn, 1999), that is, independent of cognitive processes (bottom-up) and one that sees perception as a cognitive-penetrable process (top-down). Considering experiments outside psychology, including literature, visual arts, and humanities, the rate of observed "opinion tendencies" would be multiplied (see Konecni, 2015).

In this work, I propose that cognitive, higher-level processes in perception are possible but are not an all-or-none phenomenon. There is no clear borderline

[2]Note that a recent line of research proposed that observers shift their attention to the retinal position an object is expected to occupy after a saccade (Rolfs, Jonikaitis, Deubel, & Cavanagh, 2011; Rolfs & Szinte, 2016). That is, the focus of attention may be anticipatory, but the eyes may lag behind (but see Gallagher, 2015). This could explain why some studies did not find anticipation in tracking: the design may have been inexpedient (e.g. Atsma et al., 2012, who asked paricipants to detect a probe that was presented on or away from the objects future path). An in-depth discussion of this approach would go beyond the scope of this work.

between memory, perception, and cognition, and the involvement of top-down effects depends on many factors that are yet to be explored. In these last two sections I try to give the reader an idea why it is so difficult to find a one-for-all theory of perception – and why it may be fruitful for psychological research that there is none yet.

5.4.1 Utilitarian perception

With regard to the obtained results and after my extensive literature research, I side with the utilitarian theory of Ramachandran. Ramachandran (1990) proposed (p. 24):

"*(...) (P)erception is essentially a "bag of tricks"; that through millions of years of trial and error the visual system has evolved numerous short-cuts, rules-of-thumb and heuristics which were adopted not for their aesthetic appeal or mathematical elegance but simply because they worked (hence the "utilitarian" theory).*"

Perceptual processing may apply many different strategies to encode and understand a dynamic scene. Observing effects ascribed to a given manipulation may depend on numerous factors, each triggering a different "trick in the bag". Instead of finding one overarching theory that explains different perceptual processes (shape, pattern, motion, static, etc.), Ramachandran (1990) suggested that one should first identify the *problems* that perception was designed to solve. He mentions six important aspects that should be integrated in an approach to understand perception:

1) A biological system, like the visual system, works because it is based on computing approximate solutions quickly. The solutions are adequate for the given perceptual problem, but usually not optimal.
2) Due to the environmental (natural constraints), there are often too many ways of solving a problem and using a theoretical (i.e. computational) approach cannot distinguish between them.
3) Biological constraints (the "neural machinery) are often disregarded in theoretical approaches.

4) Identifying constraints (e.g. number of targets to track) is an incomplete account. A constraint does not tell us much about the working of the mechanism.

5) Nature is opportunistic and adopts ad hoc solutions.

6) The brain solves perceptual problems by applying multiple mechanisms in parallel. The authors mention two reasons: (a) by using multiple systems, none of the systems needs to be "perfect" and is therefore easy to implement in neural hardware, and (b) simultaneous use of different mechanisms ensures rapid processing and an easier suppression of noise.

Of course, this is a pessimistic view: it implies that we may never fully understand each and every perceptual process or that, at a minimum, we have to deal with countless contradictions and inconsistencies until we eventually do. Ramachandran (1990), however, finds that confusion and contrary approaches are beneficial; they usually precede an understanding of a function.

This work further defined fine grains of existing constraints. As one may deduce after reading Ramachandran (1990), this has not been simple. For example, in Study 3, I quickly found a working combination of number of targets, speed of objects, length of trial, number of trials, and wording of instructions. In each of the published experiments, I asked participants to track one block without an additional task before they had to identify the objects and their behavior in a second block. However, in an unpublished control experiment, I skipped the first block and directly asked participants to do the identification task during tracking – and found no significant differences between the allocation of attention to the cued target or distractor object. It seems that one factor for the observation of ACS effects in dynamic perception is a certain amount of experience with the environment. In reality, this may be comparable to novice drivers who cannot mentally represent the traffic around them and see special features (e.g. a street name sign) simultaneously – however, after a while, a driver can suppress or prioritize such additional information based on given needs, goals, or tasks. The latter is important, because as shown in Experiment 2 of Study 3: simply being familiar with the environment does not automatically result in feature processing if the tracking task can be solved without it. A possible interpretation: MOT may be based on the dorsal pathway (low-level processing), but when it is learned, it is *possible to consult* the ventral pathway in order to represent details. The ventral and dorsal pathway can work together. For example, while learning a given action, for example in playing tennis, the ventral system may work together with the dorsal system (i.e. conscious activation of muscle movement). When the action is well-learned, the dorsal system is sufficient (Williams, Davids, & Williams, 1999). In addition, both pathways can overlap in their roles, for example in the perception of size (Norman, 2002) – the two system simply solve the task in a different manner. Similar ideas of dual-processing could also apply to MOT.

On the other hand, when we asked the participants in Study 2 to watch out for the ball contact moment, a similar concept to the ACS idea in Study 3, we still observed causal fillings. As Ramachandran (1990) assumed in the second aspect listed above, this may have happened due to environmental (natural) constraints. One finding cannot be generalized to similar fields, because the brain has too many options to choose from and it may do so in a non-logical or "quick and dirty" manner.

5.4.2 The human need for simplexity

Although at first glance the three studies do not seem to describe the same thing, they do: they all reflect the multidimensional nature of perceptual processing. What my research taught me is that we may have to disagree in order to agree. There is not one theory for human perception. There is not one theory for human attention. And if there were, "a general theory of vision may indeed be so abstract and complex that few workers now alive would be able to understand it" (Gordon, 2004, p.226). Even the definitions of cognitive mechanisms, the proposed organization of the visual system in terms of independent parallel systems (as I myself have done in Figure 7), or assuming that the brain works like a computer (Marr, 1982) are only desperate attempts to squeeze countless perceptual tricks and processes that co-exist, interact, and overlap into a construction of thought, an artificial environment comparable to language and symbols, that humans create in order to make sense of the complexity.

Stewart and Cohen (2000) called such a tendency of simple rules to emerge from underlying disorder and complexity "simplexity". The authors contrast simplexity to "complicity", a more subtle system that applies completely different rules but produces the same structural patterns. Complicity emerges from a dynamic and interactive environment with too many factors to grasp with a reductionist approach. Even worse, the authors claim that, if the human brain would be simple enough for us to understand it in its whole complexity, humans would be too simple to do so (a paradox worth thinking about). The combination of simplexity and complicity produces a constant picture of features from the wealth of complexity and randomness, or a “collapse of chaos”. They further wonder why humans try to find a fundamental generalizing principle, a simplexity theory, but do not apply a reductionist analysis in other areas. As an example they mention the neck of a giraffe and the trunk of an elephant: both body parts serve the same purpose (not to kneel down for feeding and drinking), but we still gave them different names.

Why do we accept multidimensionality in other biological areas but not for theories of visual perception? Still, the search for a general theory of perception led to the discovery of many general properties of vision and real progress has been made concerning the preciseness and the scientific rational of theories.

Unfortunately, the more we learn, the less we know.

Bibliography

Agam, Y. & Sekuler, R. (2007). Interactions between working memory and visual perception: an erp/eeg study. *Neuroimage*, *36*(3), 933–942. doi:10.1016/j.neuroimage.2007.04.014. (Cit. on p. 14)

Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics*, *16*(1), 3–14. doi:10.1016/0304-4076(81)90071-3. (Cit. on p. 6)

Alvarez, G. A. & Franconeri, S. L. (2007). How many objects can you track?: evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, *7*(13), 14–14. doi:10.1167/7.13.14. (Cit. on pp. 99, 100, 114)

Alvarez, G. A., Horowitz, T. S., Arsenio, H. C., DiMase, J. S., & Wolfe, J. M. (2005). Do multielement visual tracking and visual search draw continuously on the same visual attention resources? *Journal of Experimental Psychology: Human Perception and Performance*, *31*(4), 643. doi:10.1037/0096-1523.31.4.643a. (Cit. on p. 107)

Atsma, J., Koning, A., & van Lier, R. (2012). Multiple object tracking: anticipatory attention doesn't "bounce". *Journal of Vision*, *12*(13), 1–1. doi:10.1167/12.13.1. (Cit. on pp. 8, 116)

Attention. (n.d.). *In Merriam Webster Online Dictionary, retrieved September 25, 2016, from http://www.merriam-webster.com/dictionary/attention*. (Cit. on p. 15).

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–628. doi:10.1037/e527352012-294. (Cit. on p. 98)

Awh, E. & Jonides, J. (1998). *Spatial working memory and spatial selective attention.* The MIT Press. (Cit. on p. 34).

Awh, E., Jonides, J., & Reuter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 780. doi:10.1037//0096-1523.24.3.780. (Cit. on p. 34)

Bacon, W. F. & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics*, *55*(5), 485–496. doi:10.3758/bf03205306. (Cit. on p. 64)

Bae, G. Y. & Flombaum, J. I. (2012). Close encounters of the distracting kind: identifying the cause of visual tracking errors. *Attention, Perception, & Psychophysics*, *74*(4), 703–715. doi:10.3758/s13414-011-0260-1. (Cit. on p. 113)

Baker, L. J. & Levin, D. T. (2015). The role of relational triggers in event perception. *Cognition*, *136*, 14–29. doi:10.1016/j.cognition.2014.11.030. (Cit. on pp. 35, 104)

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46. doi:10.1016/0010-0277(85)90022-8. (Cit. on p. 35)

Bartlett, F. C. (1932). *Remembering: an experimental and social study*. Cambridge: Cambridge University. (Cit. on p. 36).

Beauchamp, M. S., Cox, R. W., & Deyoe, E. A. (1997). Graded effects of spatial and featural attention on human area mt and associated motion processing areas. *Journal of Neurophysiology*, *78*(1), 516–520. (Cit. on p. 57).

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: the role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine*, *68*, 569–579. (Cit. on p. 54).

Bello, P., Bridewell, W., & Wasylyshyn, C. (2016). Attentive and pre-attentive processes in multiple object tracking: a computational investigation. In *Proceedings of the thirty-eighth annual meeting of the cognitive science society. philadelphia, pa*. (Cit. on p. 101).

Bhatt, R. S., Hayden, A., & Quinn, P. C. (2007). Perceptual organization based on common region in infancy. *Infancy*, *12*(2), 147–168. doi:10.1111/j.1532-7078.2007.tb00238.x. (Cit. on p. 20)

Bhatt, R. S. & Quinn, P. C. (2011). How does learning impact development in infancy? the case of perceptual organization. *Infancy*, *16*(1), 2–38. doi:10.1111/j.1532-7078.2010.00048.x. (Cit. on p. 20)

Biber, D., van Dijk, T. A., & Kintsch, W. (1986). Strategies of discourse comprehension. *Language*, *62*(3), 664. doi:10.2307/415483. (Cit. on p. 37)

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*(2), 143–177. doi:10.1016/0010-0285(82)90007-x. (Cit. on p. 36)

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127–135. doi:10.1016/j.tree.2008.10.008. (Cit. on p. 6)

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. doi:10.1016/j.tics.2011.12.010. (Cit. on p. 103)

Bressler, S. L., Tang, W., Sylvester, C. M., Shulman, G. L., & Corbetta, M. (2008). Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *The Journal of Neuroscience*, *28*(40), 10056–10061. doi:10.1523/jneurosci.1776-08.2008. (Cit. on p. 57)

Brockhoff, A. & Huff, M. (2016). All eyes on relevance: strategic allocation of attention as a result of feature-based task demands in multiple object tracking. *Attention, Perception, & Psychophysics*, *78*(7), 2090–2109. doi:10.3758/s13414-016-1129-0. (Cit. on p. 66)

Brockhoff, A., Huff, M., Maurer, A., & Papenmeier, F. (2016). Seeing the unseen? illusory causal filling in FIFA referees, players, and novices. *Cognitive Research: Principles and Implications*, *1*(1). doi:10.1186/s41235-016-0008-5. (Cit. on p. 41)

Brockhoff, A., Papenmeier, F., Wolf, K., Pfeiffer, T., Jahn, G., & Huff, M. (2016). Viewpoint matters: exploring the involvement of reference frames in multiple object tracking from a developmental perspective. *Cognitive Development*, *37*, 1–8. doi:10.1016/j.cogdev.2015.10.004. (Cit. on p. 23)

Carpenter, G. A., Grossberg, S., & Arbib, A. (2003). *The handbook of brain theory and neural networks*. doi:10.1037/000547. (Cit. on p. 37)

Carpenter, P. A., Just, M. A., Keller, T. A., Eddy, W., & Thulborn, K. (1999). Graded functional activation in the visuospatial system with the amount of task demand. *Cognitive Neuroscience, Journal of*, *11*(1), 9–24. doi:10.1162/089892999563210. (Cit. on p. 111)

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience*, *7*(3), 308–313. doi:10.1038/nn1194. (Cit. on pp. 16, 64)

Carrasco, M., Williams, P. E., & Yeshurun, Y. (2002). Covert attention increases spatial resolution with or without masks: support for signal enhancement. *Journal of Vision*, *2*(6), 4–4. doi:10.1167/2.6.4. (Cit. on p. 16)

Castiello, U. & Umiltà, C. (1992). Splitting focal attention. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 837. doi:10.1037//0096-1523.18.3.837. (Cit. on p. 110)

Cave, K. R. & Bichot, N. P. (1999). Visuospatial attention: beyond a spotlight model. *Psychonomic Bulletin & Review*, *6*(2), 204–223. doi:10.3758/bf03212327. (Cit. on p. 110)

Chase, W. G. & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. doi:10.1016/0010-0285(73)90004-2. (Cit. on p. 39)

Chawla, D., Rees, G., & Friston, K. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience*, *2*(7), 671–676. doi:10.1038/10230. (Cit. on p. 113)

Christie, J. & Klein, R. M. (1996). Assessing the evidence for novel popout. *Journal of Experimental Psychology: General*, *125*(2), 201–207. doi:10.1037/0096-3445.125.2.201. (Cit. on p. 111)

Chun, M. M. & Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71. doi:10.1006/cogp.1998.0681. (Cit. on p. 39)

Chun, M. M. & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*(4), 360–365. doi:10.1111/1467-9280.00168. (Cit. on p. 39)

Cowan, N. (2001). Metatheory of storage capacity limits. *Behavioral and Brain Sciences*, *24*(01), 154–176. doi:10.1017/s0140525x0161392x. (Cit. on pp. 34, 99)

Culham, J. C., Brandt, S. A., Cavanagh, P., Kanwisher, N. G., Dale, A. M., & Tootell, R. B. (1998). Cortical fmri activation produced by attentive tracking of moving targets. *Journal of Neurophysiology*, *80*(5), 2657–2670. (Cit. on p. 99).

Davies, R. (1951). *Tempest-tost*. New Canadian Library. (Cit. on p. 103).

d'Avossa, G., Shulman, G. L., Snyder, A. Z., & Corbetta, M. (2006). Attentional selection of moving objects by a serial process. *Vision Research*, *46*(20), 3403–3412. doi:10.1016/j.visres.2006.04.018. (Cit. on p. 99)

Dennett, D. C. & Kinsbourne, M. (1992). Time and the observer: the where and when of consciousness in the brain. *Behavioral and Brain Sciences*, *15*(02), 183–201. doi:10.1017/s0140525x00068229. (Cit. on p. 111)

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *353*(1373), 1245–1255. doi:10.1098/rstb.1998.0280. (Cit. on pp. 101, 112, 115)

DeWeerd, P., Peralta, M. R., Desimone, R., & Ungerleider, L. G. (1999). Loss of attentional stimulus selection after extrastriate cortical lesions in macaques. *Nature Neuroscience*, *2*(8), 753–758. doi:10.1038/11234. (Cit. on p. 112)

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General*, *129*(4), 481. doi:10.1037/0096-3445.129.4.481. (Cit. on pp. 108, 109)

Didierjean, A. & Marmèche, E. (2005). Anticipatory representation of visual basketball scenes by novice and expert players. *Visual Cognition*, *12*(2), 265–283. doi:10.1080/13506280444000021. (Cit. on pp. 39, 40, 104–106, 116)

Drew, T. & Vogel, E. K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. *The Journal of Neuroscience*, *28*(16), 4183–4191. doi:10.1523/jneurosci.0556-08.2008. (Cit. on p. 113)

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(1), 32. doi:10.1037//0096-1523.10.1.32. (Cit. on p. 62)

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*(1), 143–149. doi:10.3758/bf03203267. (Cit. on p. 62)

Eriksen, C. W. & Hoffman, J. E. (1973). The extent of processing of noise elements during selective encoding from visual displays. *Perception & Psychophysics*, *14*(1), 155–160. doi:10.3758/bf03198630. (Cit. on p. 64)

Erlikhman, G., Keane, B. P., Mettler, E., Horowitz, T. S., & Kellman, P. J. (2013). Automatic feature-based grouping during multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(6), 1625. doi:10.1037/a0031750. (Cit. on p. 101)

Ezzyat, Y. & Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychological Science*, *22*(2), 243–252. doi:10.1177/0956797610393742. (Cit. on p. 109)

Fernandez-Duque, D. & Johnson, M. L. (2002). Cause and effect theories of attention: the role of conceptual metaphors. *Review of General Psychology*, *6*(2), 153. doi:10.1037//1089-2680.6.2.153. (Cit. on pp. 59, 112)

Ferrera, V. P., Rudolph, K. K., & Maunsell, J. (1994). Responses of neurons in the parietal and temporal visual pathways during a motion task. *The Journal of Neuroscience*, *14*(10), 6171–6186. (Cit. on p. 107).

Folk, C. L., Leber, A. B., & Egeth, H. E. (2002). Made you blink! contingent attentional capture produces a spatial blink. *Perception & Psychophysics*, *64*(5), 741–753. doi:10.3758/bf03194741. (Cit. on p. 64)

Folk, C. L. & Remington, R. W. (1998). Selectivity in distraction by irrelevant featural singletons: evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 847. doi:10.1037//0096-1523.24.3.847. (Cit. on p. 64)

Folk, C. L. & Remington, R. W. (1999). Can new objects override attentional control settings? *Perception & Psychophysics*, *61*(4), 727–739. doi:10.1037/e536982012-031. (Cit. on p. 64)

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1030. doi:10.1037//0096-1523.18.4.1030. (Cit. on p. 64)

Franconeri, S. L. (2013). The nature and status of visual resources. *Oxford Handbook of Cognitive Psychology*, *8481*, 147–162. doi:10.1093/oxfordhb/9780195376746.013.0010. (Cit. on p. 114)

Franconeri, S. L., Jonathan, S., & Scimeca, J. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological Science*, *21*(7), 920–925. doi:10.1177/0956797610373935. (Cit. on pp. 104, 113)

Franconeri, S. L., Simons, D. J., & Junge, J. A. (2004). Searching for stimulus-driven shifts of attention. *Psychonomic Bulletin & Review*, *11*(5), 876–881. doi:10.3758/bf03196715. (Cit. on p. 64)

Freyd, J. J. & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 126. doi:10.1037//0278-7393.10.1.126. (Cit. on p. 36)

Friedman, A. (1979). Framing pictures: the role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, *108*(3), 316. doi:10.1037//0096-3445.108.3.316. (Cit. on p. 36)

Friesen, C. K. & Kingstone, A. (1998). The eyes have it! reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490–495. doi:10.3758/bf03208827. (Cit. on p. 64)

Fuster, J. M. (1991). The prefrontal cortex and its relation to behavior. *Progress in Brain Research*, *87*, 201–211. doi:10.1016/s0079-6123(08)63053-8. (Cit. on p. 37)

Gallagher, K. (2015). Commentary: attentional tradeoffs maintain the tracking of moving objects across saccades. *Frontiers in Systems Neuroscience*, *9*. doi:10.3389/fnsys.2015.00188. (Cit. on p. 116)

Gazzaley, A. & Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends in Cognitive Sciences*, *16*(2), 129–135. doi:10.1016/j.tics.2011.11.014. (Cit. on p. 34)

Ghim, H.-R. & Eimas, P. D. (1988). Global and local processing by 3- and 4-month-old infants. *Perception & Psychophysics*, *43*(2), 165–171. doi:10.3758/bf03214194. (Cit. on p. 20)

Gibson, B. S. & Kelsey, E. M. (1998). Stimulus-driven attentional capture is contingent on attentional set for displaywide visual features. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 699. doi:10.1037//0096-1523.24.3.699. (Cit. on p. 64)

Gibson, J. J. (1966). The problem of temporal order in stimulation and perception. *The Journal of Psychology*, *62*(2), 141–149. doi:10.1080/00223980.1966.10543777. (Cit. on p. 21)

Gibson, J. J. & Gibson, E. J. (1955). Perceptual learning: differentiation or enrichment? *Psychological Review*, *62*(1), 32. doi:10.1037/h0048826. (Cit. on pp. 13, 106)

Gobet, F. & Simon, H. A. (1996). Templates in chess memory: a mechanism for recalling several boards. *Cognitive Psychology*, *31*(1), 1–40. doi:10.1006/cogp.1996.0011. (Cit. on p. 39)

Goldstone, R. L. (2003). Learning to perceive while perceiving to learn. In *Perceptual organization in vision: behavioral and neural perspectives* (pp. 233–278). Psychology Press. (Cit. on p. 20).

Gordon, I. E. (2004). *Theories of visual perception*. Informa UK Limited. doi:10.4324/9780203502259. (Cit. on pp. 97, 119)

Green, C. S. & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*(6939), 534–537. doi:10.1038/nature01647. (Cit. on p. 54)

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *290*(1038), 181–197. doi:10.1007/978-1-349-02110-9_9. (Cit. on pp. 14, 35, 106)

Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *352*(1358), 1121–1127. doi:10.1098/rstb.1997.0095. (Cit. on pp. 35, 36)

Harrison, S. A. & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635. doi:10.1038/nature07832. (Cit. on p. 14)

Hasher, L. & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, *108*(3), 356. doi:10.1037//0096-3445.108.3.356. (Cit. on p. 111)

Hatfield, G. (1998). Attention in early scientific psychology. In R. D. Wright (Ed.), *Visual attention* (pp. 3–25). New York: Oxford University Press. (Cit. on p. 15).

Hayden, A., Bhatt, R. S., & Quinn, P. C. (2008). Perceptual organization based on illusory regions in infancy. *Psychonomic Bulletin & Review*, *15*(2), 443–447. doi:10.3758/pbr.15.2.443. (Cit. on p. 20)

Helmholtz, H. v. (1866). Concerning the perceptions in general. *Treatise On Physiological Optics*, *3*. doi:10.1037/11304-027. (Cit. on pp. 35, 36, 108)

Helsen, W. & Pauwels, J. (1993). The relationship between expertise and visual information processing in sport. *Advances in Psychology*, *102*, 109–134. doi:10.1016/s0166-4115(08)61468-5. (Cit. on p. 54)

Henderson, J. M. & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*(1), 243–271. doi:10.1146/annurev.psych.50.1.243. (Cit. on p. 54)

Hespos, S. J. & Rochat, P. (1997). Dynamic mental representation in infancy. *Cognition*, *64*(2), 153–188. doi:10.1016/s0010-0277(97)00029-2. (Cit. on p. 21)

Hochberg, J. (1988). Visual perception. *Stevens' Handbook of Experimental Psychology*, *1*, 195–276. doi:10.1037/020073. (Cit. on p. 13)

Horowitz, T. S., Birnkrant, R. S., Fencsik, D. E., Tran, L., & Wolfe, J. M. (2006). How do we track invisible objects? *Psychonomic Bulletin & Review*, *13*(3), 516–523. doi:10.3758/bf03193879. (Cit. on p. 34)

Horowitz, T. S. & Cohen, M. A. (2010). Direction information in multiple object tracking is limited by a graded resource. *Attention, Perception, & Psychophysics*, *72*(7), 1765–1775. doi:10.3758/app.72.7.1765. (Cit. on p. 99)

Howard, C. J. & Holcombe, A. O. (2008). Tracking the changing features of multiple objects: progressively poorer perceptual precision and progressively greater perceptual lag. *Vision Research*, *48*(9), 1164–1180. doi:10.1016/j.visres.2008.01.023. (Cit. on pp. 34, 99)

Howe, P. D., Cohen, M. A., Pinto, Y., & Horowitz, T. S. (2010). Distinguishing between parallel and serial accounts of multiple object tracking. *Journal of Vision*, *10*(8), 11–11. doi:10.1167/10.8.11. (Cit. on p. 114)

Huff, M., Jahn, G., & Schwan, S. (2009). Tracking multiple objects across abrupt viewpoint changes. *Visual Cognition*, *17*(3), 297–306. doi:10.1080/13506280802061838. (Cit. on pp. 21, 22)

Huff, M., Meyerhoff, H. S., Papenmeier, F., & Jahn, G. (2010). Spatial updating of dynamic scenes: tracking multiple invisible objects across viewpoint changes. *Attention, Perception, & Psychophysics*, *72*(3), 628–636. doi:10.3758/app.72.3.628. (Cit. on p. 34)

Hymel, A., Levin, D. T., & Baker, L. J. (2016). Default processing of event sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(2), 235–246. doi:10.1037/xhp0000082. (Cit. on p. 102)

Iglói, K., Zaoui, M., Berthoz, A., & Rondi-Reig, L. (2009). Sequential egocentric strategy is acquired as early as allocentric strategy: parallel acquisition of these two navigation strategies. *Hippocampus*, *19*(12), 1199–1211. doi:10.1002/hipo.20595. (Cit. on p. 21)

Ingle, D. (1967). Two visual mechanisms underlying the behavior of fish. *Psychologische Forschung*, *31*(1), 44–51. doi:10.1007/bf00422385. (Cit. on p. 107)

Jackendoff, R. (1991). Parts and boundaries. *Cognition*, *41*(1), 9–45. doi:10.1016/0010-0277(91)90031-x. (Cit. on pp. 36–38, 105)

James, W. (1890). The consciousness of self. *The Principles of Psychology*, *8*. doi:10.1037/10538-010. (Cit. on p. 59)

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28. doi:10.1037/0033-2909.114.1.3. (Cit. on pp. 40, 105)

Johnson, M. K. & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*(1), 67–85. doi:10.1037/0033-295x.88.1.67. (Cit. on pp. 40, 105)

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, *4*(1), 71–115. doi:10.1207/s15516709cog0401_4. (Cit. on p. 37)

Johnson-Laird, P. N. (1983). *Mental models: towards a cognitive science of language, inference, and consciousness*. Harvard University Press. doi:10.2307/414498. (Cit. on p. 37)

Johnson-Laird, P. N. (1994). Mental models and probabilistic thinking. *Cognition*, 50(1-3), 189–209. doi:10.1016/0010-0277(94)90028-0. (Cit. on p. 37)

Kahneman, D. (1973). *Attention and effort*. Citeseer. doi:10.2307/1421603. (Cit. on p. 111)

Keane, B. P. & Pylyshyn, Z. W. (2006). Is motion extrapolation employed in multiple object tracking? tracking as a low-level, non-predictive function. *Cognitive Psychology*, *52*(4), 346–368. doi:10.1016/j.cogpsych.2005.12.001. (Cit. on p. 8)

Konecni, V. J. (2015). The significance of encapsulation of visual perception for philosophy of mind and aesthetic analysis. *Clinical and Experimental Psychology*, *01*(01). doi:10.4172/2471-2701.1000102. (Cit. on p. 116)

Krauzlis, R. J., Bollimunta, A., Arcizet, F., & Wang, L. (2014). Attention as an effect not a cause. *Trends in Cognitive Sciences*, *18*(9), 457–464. doi:10.1016/j.tics.2014.05.008. (Cit. on pp. 111, 112)

Kuo, B.-C., Rao, A., Lepsien, J., & Nobre, A. C. (2009). Searching for targets within the spatial layout of visual short-term memory. *The Journal of Neuroscience*, *29*(25), 8032–8038. doi:10.1523/jneurosci.0952-09.2009. (Cit. on p. 34)

Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, *7*(1), 12–18. doi:10.1016/s1364-6613(02)00013-x. (Cit. on pp. 60, 61, 110)

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 451. doi:10.1037//0096-1523.21.3.451. (Cit. on p. 62)

Le, T. H., Pardo, J. V., & Hu, X. (1998). 4 t-fmri study of nonspatial shifting of selective attention: cerebellar and parietal contributions. *Journal of Neurophysiology*, *79*(3), 1535–1548. (Cit. on p. 63).

Levin, D. & Saylor, M. (2008). *Shining spotlights, zooming lenses, grabbing hands, and pecking chickens: the ebb and flow of attention during events*. Oxford University Press New York. doi:10.1093/acprof:oso/9780195188370.003.0022. (Cit. on p. 100)

Liu, G., Austen, E. L., Booth, K. S., Fisher, B. D., Argue, R., Rempel, M. I., & Enns, J. T. (2005). Multiple-object tracking is based on scene, not retinal, coordinates. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(2), 235. doi:10.1037/0096-1523.31.2.235. (Cit. on p. 21)

Lu, Z.-L. & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research*, *38*(9), 1183–1198. doi:10.1016/s0042-6989(97)00273-3. (Cit. on p. 16)

Lukavský, J. & Děchtěrenko, F. (2016). Gaze position lagging behind scene content in multiple object tracking: evidence from forward and backward presentations. *Attention, Perception, & Psychophysics*, *78*(8), 2456–2468. doi:10.3758/s13414-016-1178-4. (Cit. on p. 106)

Luu, P., Geyer, A., Fidopiastis, C., Campbell, G., Wheeler, T., Cohn, J., & Tucker, D. M. (2010). Reentrant processing in intuitive perception. *PloS One*, *5*(3), e9523. doi:10.1371/journal.pone.0009523. (Cit. on p. 109)

Luu, T. & Howe, P. D. (2015). Extrapolation occurs in multiple object tracking when eye movements are controlled. *Attention, Perception, & Psychophysics*, *77*(6), 1919–1929. doi:10.3758/s13414-015-0891-8. (Cit. on pp. 8, 116)

Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. San Francisco: WH Freeman and Company. (Cit. on p. 119).

McCulloch, C. E. & Neuhaus, J. M. (2001). *Generalized linear mixed models*. Wiley Online Library. doi:10.1002/9781118445112.stat07540. (Cit. on p. 7)

Memory. (n.d.). *In Farlex Partner Medical Dictionary, retrieved September 25, 2016, from http://medical-dictionary.thefreedictionary.com/memory*. (Cit. on p. 14).

Meyerhoff, H. S., Huff, M., Papenmeier, F., Jahn, G., & Schwan, S. (2011). Continuous visual cues trigger automatic spatial target updating in dynamic scenes. *Cognition*, *121*(1), 73–82. doi:10.1016/j.cognition.2011.06.001. (Cit. on p. 34)

Mishkin, M. & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural Brain Research*, *6*(1), 57–77. doi:10.1016/0166-4328(82)90081-x. (Cit. on p. 107)

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, *6*, 414–417. doi:10.1016/0166-2236(83)90190-x. (Cit. on p. 107)

Monsell, S. & Driver, J. (2000). *Control of cognitive processes: Attention and performance XVIII*. MIT Press. doi:10.1002/acp.849. (Cit. on p. 63)

Moore, D. S., McCabe, G. P., & Craig, B. A. (2012). *Introduction to the practice of statistics*. WH Freeman New York. doi:10.2307/3619120. (Cit. on p. 6)

Müller, M. M., Teder-Sälejärvi, W., & Hillyard, S. A. (1998). The time course of cortical facilitation during cued shifts of spatial attention. *Nature Neuroscience*, *1*(7), 631–634. doi:10.1038/2865. (Cit. on p. 110)

Navon, D. (1977). Forest before trees: the precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383. doi:10.1016/0010-0285(77)90012-3. (Cit. on pp. 13, 15, 20)

Navon, D. (1984). Resources—a theoretical soup stone? *Psychological Review*, *91*(2), 216. doi:10.1037/0033-295x.91.2.216. (Cit. on p. 111)

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts. doi:10.2307/1421282. (Cit. on pp. 37, 98, 106, 108, 109)

Neisser, U. (1976). *Cognition and reality: principles and implications of cognitive psychology.* WH Freeman/Times Books/Henry Holt & Co. (Cit. on p. 109).

Newtson, D. & Engquist, G. (1973). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, *12*(5), 436–450. doi:10.1016/0022-1031(76)90076-7. (Cit. on p. 37)

Norman, J. (2002). Two visual systems and two theories of perception: an attempt to reconcile the constructivist and ecological approaches. *Behavioral and Brain Sciences*, *25*(01), 73–96. doi:10.1017/s0140525x0200002x. (Cit. on pp. 106–108, 119)

Oksama, L. & Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? an individual difference approach. *Visual Cognition*, *11*(5), 631–671. doi:10.1080/13506280344000473. (Cit. on pp. 60, 99, 114)

Oksama, L. & Hyönä, J. (2008). Dynamic binding of identity and location information: a serial model of multiple identity tracking. *Cognitive Psychology*, *56*(4), 237–283. doi:10.1016/j.cogpsych.2007.03.001. (Cit. on p. 114)

Pandya, D. N., Seltzer, B., & Barbas, H. (1986). *Input-output organization of the primate cerebral cortex*. AR Liss. (Cit. on p. 108).

Papenmeier, F., Meyerhoff, H. S., Jahn, G., & Huff, M. (2014). Tracking by location and features: object correspondence across spatiotemporal discontinuities during multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 159. doi:10.1037/a0033117. (Cit. on pp. 60, 101, 103, 113, 115)

Pashler, H. E. (1998). *Attention*. Psychology Press. doi:10.1037/0096-1523.16.4.8 26. (Cit. on pp. 59, 60, 111)

Perception. (n.d.). *In Merriam Webster Online Dictionary, retrieved September 25, 2016, from*
http://www.merriam-webster.com/dictionary/perception. (Cit. on p. 13).

Pettijohn, K. A. & Radvansky, G. A. (2016). Walking through doorways causes forgetting: environmental effects. *Journal of Cognitive Psychology*, *28*(3), 329–340. doi:10.1080/20445911.2015.1123712. (Cit. on p. 54)

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., et al. (2007). Linear and nonlinear mixed effects models. *R Package Version*, *3*, 57. (Cit. on p. 6).

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. doi:10.1093/acprof:oso/9780199791217.003.0003. (Cit. on pp. 64, 110)

Posner, M. I. & Boies, S. J. (1971). Components of attention. *Psychological Review*, *78*(5), 391. doi:10.1037/h0031333. (Cit. on p. 59)

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160. doi:10.1037//0096-3445.109.2.160. (Cit. on pp. 64, 110)

Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(04), 515–526. doi:10.1017/s0140525x00076512. (Cit. on p. 35)

Pustejovsky, J. (1991). The syntax of event structure. *Cognition*, *41*(1), 47–81. doi:10.1016/0010-0277(91)90032-y. (Cit. on p. 105)

Pylyshyn, Z. W. (1994). Some primitive mechanisms of spatial attention. *Cognition*, 50(1-3), 363–384. doi:10.1016/0010-0277(94)90036-1. (Cit. on p. 33)

Pylyshyn, Z. W. (1999). Is vision continuous with cognition?: the case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*(03), 341–365. doi:10.1017/s0140525x99002022. (Cit. on pp. 60, 116)

Pylyshyn, Z. W. (2004). Some puzzling findings in multiple object tracking: i. tracking without keeping track of object identities. *Visual Cognition*, *11*(7), 801–822. doi:10.1080/13506280344000518. (Cit. on p. 113)

Pylyshyn, Z. W. & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 179–197. doi:10.1163/156856888x00122. (Cit. on pp. 8, 33, 34, 60, 99, 113)

Radvansky, G. A. & Zacks, J. M. (2011). Event perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(6), 608–620. doi:10.1002/wcs.133. (Cit. on p. 54)

Raisig, S., Welke, T., Hagendorf, H., & van der Meer, E. (2010). I spy with my little eye: detection of temporal violations in event sequences and the pupillary response. *International Journal of Psychophysiology*, *76*(1), 1–8. doi:10.1016/j.ijpsycho.2010.01.006. (Cit. on p. 102)

Ramachandran, V. S. (1990). Visual perception in people and machines. *AI and the Eye*, 21–77. (Cit. on pp. 117–119).

Rao, R. P. (2010). Decision making under uncertainty: a neural model based on partially observable markov decision processes. *Frontiers in Computational Neuroscience*, *4*, 146. doi:10.3389/fncom.2010.00146. (Cit. on p. 112)

Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual span in expert chess players: evidence from eye movements. *Psychological Science*, *12*(1), 48–55. doi:10.1111/1467-9280.00309. (Cit. on p. 39)

Reingold, E. M., Charness, N., Schultetus, R. S., & Stampe, D. M. (2001). Perceptual automaticity in expert chess players: parallel encoding of chess relations. *Psychonomic Bulletin & Review*, *8*(3), 504–510. doi:10.3758/bf03196185. (Cit. on p. 39)

Reisberg, D. (Ed.). (2013). *The oxford handbook of cognitive psychology*. Oxford University Press (OUP). doi:10.1093/oxfordhb/9780195376746.001.0001. (Cit. on p. 15)

Remington, R. W. (1980). Attention and saccadic eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(4), 726. doi:10.1037//0096-1523.6.4.726. (Cit. on p. 64)

Remington, R. W., Johnston, J. C., & Yantis, S. (1992). Involuntary attentional capture by abrupt onsets. *Perception & Psychophysics*, *51*(3), 279–290. doi:10.3758/bf03212254. (Cit. on p. 64)

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas v2 and v4. *The Journal of Neuroscience*, *19*(5), 1736–1753. doi:10.1016/S0896-6273(00)80821-1. (Cit. on p. 112)

Reynolds, J. H. & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, *61*(2), 168–185. doi:10.1016/j.neuron.2009.01.002. (Cit. on p. 114)

Rolfs, M., Jonikaitis, D., Deubel, H., & Cavanagh, P. (2011). Predictive remapping of attention across eye movements. *Nature Neuroscience*, *14*(2), 252–256. doi:10.1038/nn.2711. (Cit. on p. 116)

Rolfs, M. & Szinte, M. (2016). Remapping attention pointers: linking physiology and behavior. *Trends in Cognitive Sciences*, *20*(6), 399–401. doi:10.1016/j.tics.2016.04.003. (Cit. on p. 116)

Ruxton, G. D. & Beauchamp, G. (2008). Time for some a priori thinking about post hoc testing. *Behavioral Ecology*, *19*(3), 690–693. doi:10.1093/beheco/arn020. (Cit. on p. 6)

Schneider, G. E. (1967). Contrasting visuomotor functions of tectum and cortex in the golden hamster. *Psychologische Forschung*, *31*(1), 52–62. doi:10.1007/bf00422386. (Cit. on p. 106)

Schneider, W. & Shiffrin, R. M. (1977). Controlled and automatic human information processing: i. detection, search, and attention. *Psychological Review*, *84*(1), 1–66. doi:10.1037/0033-295x.84.1.1. (Cit. on p. 57)

Scholl, B. J. (2009). What have we learned about attention from multiple object tracking (and vice versa). In *Computation, cognition, and pylyshyn* (pp. 49–78). MIT Press Cambridge, MA. (Cit. on pp. 33, 107, 108).

Schultz, D. P. & Schultz, S. E. (2015). *A history of modern psychology*. Cengage Learning. (Cit. on p. 35).

Scimeca, J. M. & Franconeri, S. L. (2015). Selecting and tracking multiple objects. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(2), 109–118. doi:10.1002/wcs.1328. (Cit. on pp. 113, 114)

Sears, C. R. & Pylyshyn, Z. W. (2000). Multiple object tracking and attentional processing. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *54*(1), 1. doi:10.1037/h0087326. (Cit. on p. 33)

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*(2), 207–214. doi:10.1111/j.1467-9280.2009.02276.x. (Cit. on p. 14)

Serences, J. T. & Yantis, S. (2007). Spatially selective representations of voluntary and stimulus-driven attentional priority in human occipital, parietal, and frontal cortex. *Cerebral Cortex*, *17*(2), 284–293. doi:10.1093/cercor/bhj146. (Cit. on p. 113)

Seyal, M., Ro, T., & Rafal, R. (1995). Increased sensitivity to ipsilateral cutaneous stimuli following transcranial magnetic stimulation of the parietal lobe. *Annals of Neurology*, *38*(2), 264–267. doi:10.1002/ana.410380221. (Cit. on p. 112)

Shapiro, L. (2009). *Things and places: how the mind connects with the world, by zenon pylyshyn*. Oxford University Press (OUP). doi:10.1093/mind/fzp128. (Cit. on p. 60)

Shiffrin, R. M. & Czerwinski, M. P. (1988). A model of automatic attention attraction when mapping is partially consistent. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 562. doi:10.1037//0278-7393.14.3.562. (Cit. on p. 20)

Simons, D. J. & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074. doi:10.1068/p2952. (Cit. on p. 98)

Speer, N. K. & Zacks, J. M. (2005). Temporal changes as event boundaries: processing and memory consequences of narrative time shifts. *Journal of Memory and Language*, *53*(1), 125–140. doi:10.1016/j.jml.2005.02.009. (Cit. on p. 109)

Stewart, I. & Cohen, J. (2000). *The collapse of chaos: discovering simplicity in a complex world*. Penguin UK. doi:10.5860/choice.32-2090. (Cit. on p. 119)

Stoffer, T. H. (1993). The time course of attentional zooming: a comparison of voluntary and involuntary allocation of attention to the levels of compound stimuli. *Psychological Research*, *56*(1), 14–25. doi:10.1007/bf00572129. (Cit. on p. 20)

Strickland, B. & Keil, F. (2011). Event completion: event based inferences distort memory in a matter of seconds. *Cognition*, *121*(3), 409–415. doi:10.1016/j.cognition.2011.04.007. (Cit. on pp. 8, 39)

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, *51*(6), 599–606. doi:10.3758/bf03211656. (Cit. on p. 63)

Theeuwes, J. (1995). Abrupt luminance change pops out; abrupt color change does not. *Perception & Psychophysics*, *57*(5), 637–644. doi:10.3758/bf03213269. (Cit. on p. 64)

Tong, F. (2004). Splitting the spotlight of visual attention. *Neuron*, *42*(4), 524–526. doi:10.1016/j.neuron.2004.05.005. (Cit. on p. 110)

Treisman, A. M. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. doi:10.1016/0010-0285(80)90005-5. (Cit. on p. 62)

Trevarthen, C. B. (1968). Two mechanisms of vision in primates. *Psychologische Forschung*, *31*(4), 299–337. doi:10.1007/bf00422717. (Cit. on p. 107)

Turatto, M. & Galfano, G. (2000). Color, form and luminance capture attention in visual search. *Vision Research*, *40*(13), 1639–1643. doi:10.1016/s0042-6989(00)00061-4. (Cit. on p. 64)

Uleman, J. S. & Uleman, J. K. (1990). Unintended thought and nonconscious inferences exist. *Behavioral and Brain Sciences*, *13*(04), 627–628. doi:10.1017/s0140525x00080602. (Cit. on p. 59)

Verfaillie, K. & d'Ydewalle, G. (1991). Representational momentum and event course anticipation in the perception of implied periodical motions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(2), 302. doi:10.1037//0278-7393.17.2.302. (Cit. on p. 36)

Vinson, N. G. & Reed, C. L. (2002). Sources of object-specific effects in representational momentum. *Visual Cognition*, *9*(1-2), 41–65. doi:10.1080/13506280143000313. (Cit. on pp. 36, 116)

von Hofsten, C., Vishton, P., Spelke, E. S., Feng, Q., & Rosander, K. (1998). Predictive action in infancy: tracking and reaching for moving objects. *Cognition*, *67*(3), 255–285. doi:10.1016/s0010-0277(98)00029-8. (Cit. on p. 36)

Wang, C., Zhang, X., Li, Y., & Lyu, C. (2016). Additivity of feature-based and symmetry-based grouping effects in multiple object tracking. *Frontiers in Psychology*, *7*. doi:10.3389/fpsyg.2016.00657. (Cit. on p. 101)

Wertheimer, M. (1938). Laws of organization in perceptual forms. In *A source book of gestalt psychology.* (pp. 71–88). American Psychological Association (APA). doi:10.1037/11496-005. (Cit. on p. 13)

Wickens, C. D. (1991). *Processing resources and attention*. American Psychological Association (APA). doi:10.1037/10520-148. (Cit. on p. 111)

Wilken, P. & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 11–11. doi:10.1167/4.12.11. (Cit. on p. 34)

Williams, A. M., Davids, K., Burwitz, L., & Williams, J. G. (1994). Visual search strategies in experienced and inexperienced soccer players. *Research Quarterly for Exercise and Sport*, *65*(2), 127–135. doi:10.1080/02701367.1994.10607607. (Cit. on p. 54)

Williams, A. M., Davids, K., & Williams, J. G. P. (1999). *Visual perception and action in sport*. Taylor & Francis. doi:10.4324/9780203979952. (Cit. on p. 119)

Wojciulik, E., Kanwisher, N., & Driver, J. (1998). Covert visual attention modulates face-specific activity in the human fusiform gyrus: fmri study. *Journal of Neurophysiology*, *79*(3), 1574–1578. (Cit. on p. 110).

Wolfe, J. M. (1992). "effortless" texture segmentation and "parallel" visual search are not the same thing. *Vision Research*, *32*(4), 757–763. doi:10.1016/0042-6989(92)90190-t. (Cit. on pp. 62, 63)

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. doi:10.3758/bf03200774. (Cit. on p. 62)

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 419. doi:10.1037//0096-1523.15.3.419. (Cit. on p. 62)

Yantis, S. (1992). Multielement visual tracking: attention and perceptual organization. *Cognitive Psychology*, *24*(3), 295–340. doi:10.1016/0010-0285(92)90010-y. (Cit. on pp. 13, 20, 99, 101)

Yantis, S. (1993). Stimulus-driven attentional capture. *Current Directions in Psychological Science*, *2*(5), 156–161. doi:10.1037/e665412011-521. (Cit. on p. 64)

Yantis, S. & Hillstrom, A. P. (1994). Stimulus-driven attentional capture: evidence from equiluminant visual objects. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(1), 95. doi:10.1037//0096-1523.20.1.95. (Cit. on p. 64)

Yantis, S. & Johnston, J. C. (1990). On the locus of visual selection: evidence from focused attention tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(1), 135. doi:10.1037//0096-1523.16.1.135. (Cit. on pp. 62, 63)

Yantis, S. & Jonides, J. (1990). Abrupt visual onsets and selective attention: voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(1), 121. doi:10.1037//0096-1523.16.1.121. (Cit. on pp. 63, 64)

Young, A. W., Hellawell, D., & Hay, D. C. (2013). Configurational information in face perception. *Perception*, *42*(11), 1166–1178. doi:10.1068/p160747. (Cit. on p. 102)

Zacks, J. M. (2008). Event perception. *Scholarpedia*, *3*(10), 3837. doi:10.1093/oxfordhb/9780195376746.013.0006. (Cit. on pp. 35, 109)

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychological Bulletin*, *133*(2), 273. doi:10.1037/0033-2909.133.2.273. (Cit. on pp. 14, 35, 37, 98, 100, 102, 106)

Zacks, J. M. & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3. doi:10.1037//0033-2909.127.1.3. (Cit. on p. 37)