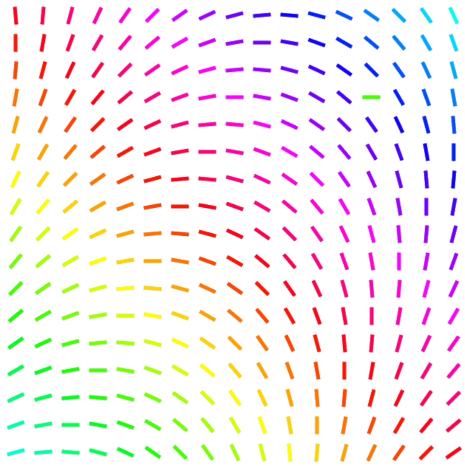
A Quantification of Visual Salience

by
Alexander Krüger



Submitted in partial fulfillment
for the
requirements of the degree
Doctor of Philosophy

Advised by
Prof. Dr. Ingrid Scharlau

Paderborn University
Faculty of Arts and Humanities
Psychology

2020

Erklärung und Bestandteile der Dissertation

Die vorgelegte Arbeit wurde von mir selbstständig und ohne Benutzung anderer als der in der Arbeit angegebenen Hilfsmittel angefertigt.

Ebenfalls wurde die Arbeit bisher weder im In- noch Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Desweitere habe ich weder früher noch gleichzeitig ein Promotionsverfahren bei einer anderen Hochschule oder bei einer anderen Fakultät beantragt.

Die Anteile meiner Koautoren an den Artikeln, die Teil diese kumulativen Dissertation sind, sind in der folgenden Tabelle 1 ausgewiesen.

Paderborn, 25. Februar 2020

	Titel	Beitrage der Autoren	Status
Article 1	Fast and Conspicuous? Quantifying Salience With the Theory of Visual Attention	**Alexander Krüger**: Idee, Literaturrecherche, Design, Durchführung, Modellierung, Analyse, Text; **Jan Tünnermann**: Betreuung der Analyse; **Ingrid Scharlau**: Betreuung	**Veröffentlicht** in Advances in Cognitive Psychology, 12(1), 20–38. https://doi.org/10.5709/acp-0184-1 *
Article 2	Measuring and modeling salience with the theory of visual attention	**Alexander Krüger**: Idee, Literaturrecherche, Design, Durchführung, Modellierung, Analyse, Text; **Jan Tünnermann**: Betreuung der Analyse; **Ingrid Scharlau**: Betreuung	**Veröffentlicht** in Attention, Perception, & Psychophysics, 1–22. https://doi.org/10.3758/s13414-017-1325-6 *
Article 3	Quantitative Explanation as a Tight Coupling of Data, Model, and Theory	**Alexander Krüger**: Idee, Literaturrecherche, Text; **Jan Tünnermann**: Beitrag des Beispiels unter “4 Application in Cognitive Psychology” inklusive Text, Experiment, Modell, Auswertung; **Katharina Rohlfing**: Betreuung; **Ingrid Scharlau**: Betreuung	**Veröffentlicht** in Archives of Data Science, Series A (Online First), 5(1), A10, 27 S. online. https://doi.org/10.5445/KSP/1000087327/10
Article 4	The time course of salience — not entirely caused by salience	**Alexander Krüger**: Idee, Literaturrecherche, Design, Durchführung, Modellierung, Analyse, Text; **Ingrid Scharlau**: Betreuung	**Im Review** bei Psychological Research *
			* Die jeweiligen Primärdaten, eine implementierte Version des TVA-TOJ-Modells, ein Datenanalyseskript und interaktive Erklärung zu deren Benutzung sind online publiziert (Krüger, 2020).

Tabelle 1: Artikel in Erstautorenschaft, die Teil der vorliegenden kumulativen Disseration sind.

Abstract

Visual conspicuousness — so-called salience — affects how we perceive the world. This is well known and congruent with everyday experiences. Surprisingly, today, no commonly accepted measure for the strength of salience exists. Consequentially, this dissertation aims at providing such a measure of visual salience. Visual salience arises from physical contrasts and may affect attention as well as the perception of the respective stimuli. This thesis is focused on salience's influence on attention. Besides methods from experimental psychology, also methods from cognitive modeling are applied to derive such a measure of salience. This cumulative dissertation encompasses four articles. Article 1 combines current salience research with a model based on Bundesen's Theory of Visual Attention and the experimental paradigm of temporal-order judgment. Article 2 adds a formal salience parameter to the model, investigates the growth of salience caused by physical contrast, and finally models and compares different interactions of two types of contrast as proposed in the literature. Unlike previously published results, this model shows that contrasts add up without a penalty. Bayesian Statistics are applied throughout all articles. Advantages of Bayesian Statistics include easy implementation of custom hierarchical models, their comparison, information-rich parameter estimation, and the generation of data based on the model. Article 3 is unique in the sense that it takes a bird's eye perspective by comparing a combination of Bayesian Statistics and modeling for established and new techniques in psychology. These are, namely, null-hypothesis testing and machine learning. Article 4 applies the developed model to the time course of salience. Additionally, this synopsis reports many more experiments in which the developed modeling and empirical approach is applied. To sum up, I propose yet another empirical measure of salience. However, because of the modeling, it is tightly coupled with a formal

theory of visual attention so that it has an explicitly defined meaning for visual selection.

Zusammenfassung

Visuelle Auffälligkeit, sogenannte Salienz, beeinflusst, wie wir die Welt wahrnehmen. Dies ist altbekannt und deckt sich mit Alltagserfahrungen. Überraschend dabei ist, dass es bis heute kein einheitliches Maß für die Stärke von Salienz gibt. Diese Doktorarbeit untersucht daher, wie genau die Stärke von Salienz, die durch physische Kontraste entsteht, gemessen werden kann. Dabei konzentriert sich die Arbeit auf visuelle Aufmerksamkeit im Gegensatz zu perzeptuellen Eigenschaften der Kontraste. Neben den Methoden der Experimentalpsychologie kommt die kognitive formale Modellierung zur Herleitung eines Salienzmaßes zum Einsatz. Insgesamt umfasst die kumulative Dissertation vier Artikel. Artikel 1 verbindet bisherige Salienzforschung mit einer auf Bundesens Theorie der visuellen Aufmerksamkeit basierten Modells mit dem experimentellen Paradigma des zeitlichen Reihenfolgeurteils. Artikel 2 fügt der Modellierung ein explizites Salienzmaß hinzu, modelliert das Wachstum von Aufmerksamkeitsvorteil durch physischen Kontrast und überprüft verschiedene Interaktionen für mehr als eine Sorte von Kontrast, die in der Literatur diskutiert werden. Auf Basis dieser Modellierung wird gezeigt, dass beide Kontrastarten sich ohne Einbuße ergänzen — im Gegensatz zu früheren Ergebnissen. Alle Artikel nutzen dabei die Möglichkeiten der Bayesischen Statistik. Zu diesen Möglichkeiten zählen die einfache Implementierung von individuell entwickelten hierarchischen Modellen, deren Vergleich, informationsreiche Parameterschätzung und das Generieren von Daten durch das Modell. Artikel 3 stellt eine Ausnahme dahingehend dar, dass die Modellierung und Bayesische Statistik aus der Vogelperspektive betrachtet und mit bekannten und neuen Methoden der Psychologie, Nullhypothesentests beziehungsweise maschinellem Lernen, vergleicht. Artikel 4 wendet die erarbeitete Modellierung auf den zeitlichen Verlauf von Salienz an. Ebenfalls werden kleinere Experimente berichtet, die alle eine Anwendung der entwickelten Model-

lierung und des entwickelten Experimentaldesigns darstellen. Zusammenfassend wird also ein weiteres empirisches Salienzmaß vorgeschlagen, was jedoch durch Modellierung und Bezug zu einer formalen Aufmerksamkeitstheorie in seiner Bedeutung für visuelle Selektion explizit dargelegt wird.

Preface

> Voici mon secret. Il est très simple: on ne voit bien qu'avec le cœur. L'essentiel est invisible pour les yeux.
>
> ---
>
> Antoine de Saint Exupéry
> (de Saint-Exupéry, 1946, p. 72)

I always thought writing the preface to a PhD thesis would evoke lofty feelings and a great sense of achievement. Shortly before writing these lines, I hauled a package home from the post office. Now, I am waiting for my private tutoring pupil. I drink lukewarm coffee and an extremely annoying fly buzzes around in my university office — I could not feel more ordinary.

During the last six years, I often spontaneously remembered my math lecturer Dr. Preis on some occasion saying something along these lines: "Für das Abitur reicht es aus, intelligent zu sein; an der Uni muss jeder arbeiten — aber für die Promotion brauchen Sie psychische Stärke." In fact, of all things I have attempted before, this dissertation is the most demanding — both intellectually and emotionally. It is only natural to ask and maybe doubt whether this effort is worth it.

When I have my doubts about my work and future, and contemplate my decisions and their consequences, and whether or not I am happy with my situation, I reach a point where the initial goal appears as a means to an end rather than an end in itself. Having finished this dissertation, for example, has much less value for me than the thinking, doubting, and experimenting involved while conducting the work. Making sense of the world is intrinsically valuable to me and a source of great enjoyment. It is the personal development that I value, the chance for being educated by others ("ausgebildet werden") and the freedom to educate myself ("sich bilden").

Put bluntly, education does not merely enable me to do something, but created the person that I am today [1]. Sure, I always possessed curiosity and some creativity but these initial drives need to be cultivated and developed. "If I have seen further it is by standing on the shoulders of giants." is a statement that I completely agree with. Besides my understanding of the world and competences, I feel that education affects my moral sensibilities, taste, and stance towards political topics.

[1] see Bieri (2017) for a more sophisticated elaboration on this idea.

In my experience, the giants are not luminaries but all the people that enriched my mental world. People that are not often thanked for their work and maybe did not even notice how strongly they have affected others. This is the reason why I am writing this preface: To say thank you to at least some people and for some moments I remember.

One of the first things I remember is that my mother taught me about morality in terms of reason. I remember that I stole a lego brick from another child once — a simple single brick that I did not have but wanted to build something in particular. I had to give it back and apologize because that is what I would have wanted had I been in his place. I am deeply grateful that she always provided reasons rather than mere rules ("Was du nicht willst, das man dir tu, ..."). My father taught me about natural science. I particularly remember biology: He explained the monads to me when I was in elementary school. Today, I know that "die feinen Unterschiede" become second nature and thus my preference to think about reason may well have its roots in this education. So, from a young age, they cultivated my interest in how things are, but also in why things are the way they are. Also, my parents always let me try my own way. Even in situations in which I may have liked more guidance at the time. For example, at the age of sixteen, they told me that I have to pick my job without any specific recommendation or advice because it would be an important long term decision. I want to thank my parents for these and all other occasions in which they laid a piece of the foundation for my whole life with love and reason — and for helping with countless relocations.

After completing the Realschule, it was only natural to apply for an apprenticeship. My parents — and their parents before them — did not go to university and so going for an Abitur in order to attend a university afterwards was not even considered a remote possibility. Besides, I was tired of school and happy to work in a local computer shop. However, after a few months, I realized that I would not be taught to the degree that I had expected. At this time, a teacher, Andreas Blomberg, gave me some important advice: Quitting is not a shame but it is much better to quit early than late — still valuable advice. Also, he suggested that further learning is still possible after the apprenticeship. So, I finished the apprenticeship and applied for an additional year of school to qualify for higher education and by lucky coincidence he became my class teacher. I remember him saying on multiple occasions "Die kochen auch nur mit Wasser." when we were worried about higher education. So, I want to thank him, because he did not only encourage my pursuit of higher education — he made me think of it as something that was not just for "the others".

Besides individuals, there are also institutions and systems for which I am very grateful. So, the Paderborn University — as a former Gesamthochschule — accepted me as a student although I did not possess an Abitur. Without the BAföG system, the perceived financial risk would have been much too high for me to try for higher education. Furthermore, I am thankful for the support I received from the DAAD that enabled me to visit the researchers that came up with the formal theory of attention that is central in this thesis. Also, I personally have benefited from the Bologna Process because I was able to pick a master's degree that corresponded best with my research interests after studying within a classical discipline[2].

[2]The advice to first study within a classical discipline and specialize afterward is another valuable piece of advice from my former teacher, Andreas Blomberg.

Ingrid Scharlau, the supervisor for this dissertation, is probably the single most influential person for my academic self and I am deeply grateful to her. Her influence did not start with the supervision but much earlier when I felt recognized by a professor for the first time when she employed me as a student assistant after reading my first texts on psychology. This subtle form of recognition encouraged me. I became so interested in psychology besides my major in computer science that I left Paderborn to study a cognitive science master's degree. To me, she is an example of some of the best properties a scientist can have. She is very aware of the fact that everybody has limited knowledge and secure in dealing with this situation. She is also very aware of discipline specific cultures. Her bird's eye perspective on psychology helped me understand how the discipline works. I am deeply grateful that she showed me these cultural aspects of science but also likewise cultivated an atmosphere of curiosity and independent thought. She encouraged me go beyond the classical methods if necessary, to learn Bayesian statistics and think critically about the limitations of accepted methods. Also, her interest in academic writing helped me understand and improve my own writing.

A special mention is also reserved for Jan Tünnermann. I am deeply grateful for all discussions and exchanges we had. To me, Jan is a true polymath and one of the most remarkable personalities I have ever met. I remember the first TEAP we attended together: A fellow psychologist perceived Jan as a quiet person but there is an immense stock of knowledge. We talked about paleontology, astronomy, evolution and genetics, the power of formalism, artificial intelligence, physics, philosophy, and society. I find it surprising how much knowledge a person can acquire purely based on interest without external incentives. I have benefited much from his work in psychology on the connection between temporal-order judgment and Bundesen's theory of visual attention and we both started learning Bayesian methods together. I think the single most important thing Jan did for me was to lend an ear whenever I had some sort of idea or question, even if it was a spontaneous question for which I should have or could have known the answer. Jan has a particular way of listening without judging, but with a deep interest in the content and how it is menaingful. Having a good intellectual exchange based on my random ideas is much harder since he left Paderborn.

Also, Markus Hennig deserves a special mention: There were many days on which his calm advice during our

lunches together restored my sanity. It is good to be able to discuss the odd peculiarities of academia with somebody who knows them well. I will never forget how I became the Kicker-Azubi of Markus and Jan and consequentially was introduced to a new hobby.

Signe Vangkilde and Anders Petersen deserve a special mention as well. I thank them for their warm welcome in Copenhagen at the Center for Visual Cognition, their time for my questions and ideas, for taking me along on events, and for not only organizing a place for me to stay but also a bicycle — maybe the most important means of transportation in Copenhagen.

I would also like to thank Juliane Zelder for saving me from an increasing loneliness and isolation, for reintroducing many of life's aspects that I may have neglected in the pursuit of academic achievement, for believing in me when I could not, for all the love I sense in her presence, and — of course — for introducing the Flummitanz as a crucial means of emotional regulation to my life.

I could mention more people here, I could have tied up more loose ends in my research, I could have avoided some pitfalls in experimentation, I could definitively have written better texts — I could have done better in many aspects. As a student, I thought a PhD thesis would be close to perfection but in hindsight, looking at the sum of many learning experiences, all far from perfection. What you find here is my best attempt at showing that I have reached a level of education that allows me to work independently as a scientist. Whereas the intellectual content of this dissertation will be the focus of the following chapters, here I take the freedom to remind myself and the reader that scientists are embedded and embodied beings that do not exists apart from culture or emotion — and importantly other humans. And while I am convinced that education created the person and scientist I am today, I want to thank all the individuals that offered love, education, warmth, encouragement, friendship, inspiration, solace, and enabled the development that I have undergone.

Acronyms

DL difference limen.

FIRM fixed-capacity independent race model.

FIT feature integration theory.

HDI highest-density interval (in Bayesian statistics).

MAP maximum a posteriori probability.

MCMC Markov chain Monte Carlo.

NHST null-hypothesis significance testing.

PSS point of subjective simultaneity.

SOA stimulus-onset asynchrony.

TOJ temporal-order judgement.

TVA Bundesen's theory of visual attention.

VSTM visual short-term memory.

Contents

Synopsis

Chapter 1

Introduction

> In fact, of course, science is an unparalleled playground of the imagination, populated by unlikely characters with wonderful names (messenger RNA, black holes, quarks) and capable of performing the most amazing deeds: sub-atomic whirling dervishes that can be in several places — everywhere and nowhere — at the same time; molecular hoop-snakes biting their own tails; self-copying spiral staircases bearing coded instructions; miniature keys searching for the locks in which they fit, on floating odysseys in a trillion synaptic gulfs.
>
> Daniel C. Dennett (Hofstadter & Dennett, 2001, p. 458)

Physical contrast affects how the world is visually perceived. Visual contrasts, like edges and colors, are virtually everywhere, and arguably these contrasts are what the visual system is all about. Merely thinking about this claim provides a vivid thought experiment to show the reason why: A white wall only provides one information "everything is white". Successively adding contrasts like luminance contrast conveys the additional information "here is an edge" or "here is a shade" which increases the amount of information in the image such that a meaningful shape of maybe an animal or a plant can emerge. Much like an artist adding individual strokes of the paintbrush to express an abstract idea as a painting, the mind integrates information from individual contrasts into a coherent concept to make sense of the world. The importance of contrasts does of course not only arise from plausible thought about it:

Hubel and Wiesel (1959) demonstrated that contrasts are what the visual cortex is — loosely speaking — interested in. Individual points of light do not affect cells in the primary visual cortex but edges, orientations, and contrasts between on-neurons and off-neurons do. The neuronal processing discovered by Hubel and Wiesel (1959) is in line with the intuition evoked from the thought example that it is the contrasts that convey information.

Because visual contrasts are such a basic way of extracting information from the environment, they have an effect on many cognitive abilities. This thesis deals with their specific effect on visual attention. The term visual salience is used to describe that a physical contrast between stimulus and surrounding attracts attention. It should be noted that visual salience may also refer to the visual impression of a stimulus standing out, of a stimulus being conspicuous. It is likely that the two constructs — perceiving a stimulus as standing out because of physical contrasts and attending a stimulus because of physical contrasts — are related. Yet, because attention is known to affect perception (e.g., Carrasco, 2011; Kerzel, Schönhammer, Burra, Born, & Souto, 2011a) it may be prudent to distinguish both meanings. Thus, this thesis focuses on the effect of physical contrasts on attention rather than on perception.

Visual attention is not only affected by contrasts but also made necessary by visual contrasts. Contrasts make attention necessary in the sense that attention can be understood as a selection mechanism (Carrasco, 2011) that, in turn, originates from the fact that more information is available than can be processed to its full extent by the cognitive system. These limitations have been shown empirically, for example, in classical psychological experiments on short-term memory limits (Sperling, 1967) or experiments on inattentional blindness where apparently obvious changes in the environment are overlooked by the observer (Simons & Chabris, 1999). These limitations have also been central in early models of attention (e.g., Broadbent, 1958). Zhaoping (2014) gives a rough estimate in the introduction to her book stating that the bandwidth for information transfer reduces from $10^9 \frac{\text{bits}}{\text{s}}$ in the retina to $10^7 \frac{\text{bits}}{\text{s}}$ in the optic nerve further to $10^2 \frac{\text{bits}}{\text{s}}$ that are consciously perceived. Thus, she argues that looking somewhere is very different from seeing something because of attention-based selection.

In addition to empirical analysis and theoretical models, the limits of visual processing have also been a research topic in formal sciences: Tsotsos (1990) sets up the problem of visually searching for a predefined stimulus as a formal problem solvable by an algorithm. He concludes that such a problem is intractable, which means that there is no efficient algorithmic solution to finding the optimal solution. Thus, attention has the difficult role to select potentially relevant information without knowing what relevance a further analysis may yield. Accordingly, attention has to rely on readily available superficial features such as membership in a more or less relevant familiar category and, of course, physical contrasts.

In contrast to higher cognitive influences on selection like category membership, physical contrasts can easily be measured in common physical measures. Many contrasts that are highly likely to affect attention (Wolfe & Horowitz, 2004) can be quantified rather easily by physical measures. These contrasts include, e.g., luminance, color, motion, orientation, and size. So, it is natural to ask how these quantities influence attention. If certain properties affect attention more than others, this may be due to the way in which these properties are processed.

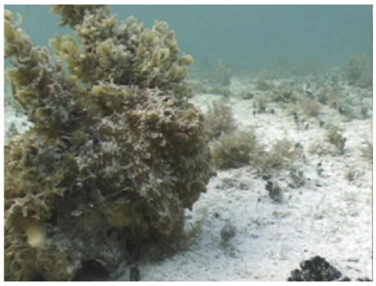
In his review on the neuronal underpinnings of salience, Treue (2003) describes two intertwined influences on visual processing: one driven by the stimulus and its salience and the other by task-relevance. Together, these influences yield a salience map that affects attention already in early visual processing. It is important to note that Treue understands the integration of task relevance and stimulus-driven influences on attention as the overall salience map. Some of the processes leading to this salience map are dynamic whereas others are hard-wired, e.g., as caused by the center-surround organization of receptive fields. Particularly contrasts are crucial for the stimulus-driven influences. The earlier the processing, the more it is based on physical contrasts. The degree of relevance-based modulation increases as the processing becomes more complex: Influences have been found as early as the thalamus. The primary visual cortex is affected by relevance-based modulation to a small degree whereas unattended stimuli are strongly suppressed in the parietal and prefrontal cortex.

After focusing on where attentional selection happens, Treue (2003) examines what is selected. In short, either locations, features, or objects are what is selected by attention. Treue ends with concluding that salience's function is to guide eye-movements to relevant stimuli, i.e., further, more detailed processing, that requires a higher visual resolution.

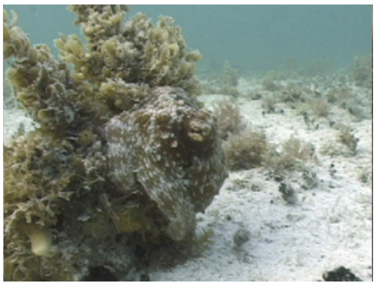
If you allow me to digress from the human cognitive system for a single paragraph, visual contrasts are also highly relevant for nonhuman animals. In fact, Treue's (2003) review is focused on the primate visual cortex in general. The behavior of animals relying on vision to find prey can be explained by assuming[1] a visual search on the image provided by their visual system (Tinbergen, 1960). Evidence for this theory comes from findings showing that predators distribute attention to a specific type of prey and switching prey types induces costs (Goto, Bond, Burks, & Kamil, 2014) which is similar to task-relevance induced influences on attention in the human visual system. Also, effects from salience have been shown, e.g., for pollinators: Goulson (2000) showed that bumblebees took twice as long to search for specific flowers when the surrounding was composed of similar-looking flowers in comparison to a neutral background. If the perspective is changed from the searcher to the prey animal, impressive mechanisms developed to defend against visual predators: Swarms of insects or schools of fish seem to be effective in predator defense because predators loose track of individuals in the group. This effect has been shown experimentally by dyeing some of the individuals which resulted in the predator being more successful while other factors remained constant (Landeau & Terborgh, 1986). Particularly stunning are animals that adapt their appearance dynamically to minimize the chance of being detected. This ability is impressive as it does not only fool predators, but also human observes.

[1] Assuming visual search is one of multiple plausible explanations of the observed animal behavior.

This camouflage is achieved by resembling the surrounding while simultaneously disguising the own shape by disruptive coloration and countershading (Stevens & Merilaita, 2011). Whereas for most readers, the chameleon will come to mind immediately, cephalopods are the true masters of camouflage (see Figure 1.1). They reproduce the perceived environment on their skin. This behavior even allows studying their perception experimentally (Hanlon, 2007). Thus, there is evidence that nonhumans' visual systems are also affected by a similar combination of relevance and stimulus-driven salience.



(a) 0 ms



(b) 270 ms

(c) 2070 ms

Figure 1.1: An octopus changing visual features including color and texture over the course of 2070 ms (Hanlon, 2007, Figure 1)

.

Treue (2003) introduces theoretical concepts and opinions; however, it is difficult to derive quantitative explanations or even testable predictions from the summary of results he presents. Arguably, science should be able to explain concrete behavior and provide theories. The entity that can provide quantitative explanations and testable prediction is models. Modeling of attention is not only undertaken by psychologists but also by engineers for artificial attention in technical systems. Central for many models is the idea of feature maps. These are independent retinotopic representations of a particular feature across the visual field. Koch and Ullman (1985) developed an architecture based on neurophysiological results on early visual processing in which different feature maps are integrated into a combined salience map. This work has been the basis for many algorithmic models with different aims. Some models are designed based on analogy with human information processing (Itti & Koch, 2001a), others are built based on theoretical considerations (Bruce & Tsotsos, 2009). Frintrop, Rome, and Christensen (2010) provide a survey of this type of model for further details and examples of this modeling approach. Remarkably, these algorithmic models produce salience maps for arbitrary input images or even video streams.

Salience has also been identified as a central component of attention in many models from cognitive psychology (Duncan & Humphreys, 1989; Theeuwes, 2010; Wolfe, Cave, & Franzel, 1989; Müller & Krummenacher, 2006). These models can explain and predict observable phenomena —

some of them even provide quantitative predictions. An example is Wolfe's (1989) Guided Attention model that aims to explain the phenomena encountered when searching visually for a particular item. Quantitative predictions include the search time but also parameters like error rates. Wolfe (2007) reports that it has been partially implemented and is indeed capable of quantitatively explaining and predicting many results of visual search experiments. However, Wolfe (2007)[2] reports that only the red-green axis for color and orientation contrast have been implemented; other dimensions cannot be used with the computational model.

[2] In personal communication in May 2016 at the ITVA meeting in Copenhagen, Wolfe stated that there were no major improvements in the quantitative predictions by the model.

Beyond symbolic models, there are subsymbolic models in which the salience representation emerges from the organization and behavior of individual (artificial) neurons. This approach has been developed by Li (2002): She presents an artificial neuronal network as a model of the primary visual cortex (Li, 2001). The main idea is that it does not have to represent salience explicitly in symbolic maps but explains how salience arises from local neuronal interactions in the architecture of the primary visual cortex. This model also aims at explaining salience not only on a conceptual but also on a quantitative level.

To sum up, there are many models that either aims at modeling salience directly or in which salience plays a central role. This might suggest that salience is sufficiently understood and the main question here is what an adequate model of salience is. However, when examined further, the literature yields that quantitative explanation and predictions are far from satisfactory because many conflicting results on quantitative aspects of salience exist.

Without a doubt, eye-movements and salience are strongly connected, yet salience's predictive power for eye-movements is low (Betz, Kietzmann, Wilming, & König, 2010). Thus, it has been suspected that to understand and predict eye movements other influences on attention are necessary (Schütz, Braun, & Gegenfurtner, 2011). Additionally, basic assumptions of algorithmic salience models have been refuted in experimental conditions: Einhäuser and König (2003) showed that the simple assumption that less local contrast would attract fewer fixations is not true for natural scenes. Also, it is far from clear which types of features contribute to salience at all (Engmann et al., 2009) and in which way the features that do contribute have to be summed up (e.g., Itti & Koch, 2001b) in symbolic algorithmic salience models. For visual search models, Wolfe and Horowitz (2004) ask whether orientation salience is not only about contrast, but also about absolute orientation values. For example, a 30° deviation for otherwise horizontal line fragments might be better detectable than the same deviation from a diagonal pattern. Li's (2002) model leads to questions particularly about the strength of the interaction of visual properties based on how these features are represented in primary visual cortex (Koene & Zhaoping, 2007). All of these questions have been answered to a degree. However, closer inspection of these salience studies reveals a high degree of heterogeneity in the quantification of salience, which is not without problems.

Most notably, Nothdurft (2000) published a comprehensive study involving luminance, motion color, and orientation contrasts. He presents a particular method for quantifying and comparing their impact on the cognitive system: To measure salience, he presented a contrast whose salience had to be quantified together with 11 steps of gradually increasing luminance contrast. Participants were presented with two contrasts and had to judge which one was more salient. If the first luminance contrast step was clearly less salient and the last step was clearly more salient than the physically different contrast in question, then the contrast in question could be matched to a luminance contrast of the same intensity. In this experimental design, salience was, of course, operationalized by the explicit salience judgment. As the effects on attention may differ from the perception of which contrasts stands out the most, the quantitative analysis may not directly be applied to attention. Yet, Nothdurft provides a quantification of salience and a conclusion about how individual contrasts combine. His analysis yields that the combined contrast is between 80 % and 30 % of the sum of the individual values. Nothdurft argues that these numbers may be related to the percentage of overlap between the neuronal mechanisms analyzing each of the two contrasts individually. Nothdurft's psychophysical experiments, however, are difficult to replicate, which may be explained by highly trained participants in the original study (Koene & Zhaoping, 2007; Krüger, Tünnermann, & Scharlau, 2016). Nothdurft's conclusion also entails the assumption — particular when cited in the context of models of attention — that the perceived salience is the same construct as the stimulus-driven effect of physical contrasts on attention. This is an assumption that should at least be checked carefully. For example, Kerzel et al. (2011a) showed that salience changes the appearance of otherwise non-salient properties. The authors conclude that after salience is computed on a salience map, this information may boost the representation of features on the feature maps and thus appearance may differ from salience.

To sum up, Nothdurft (2000) provides an indirect way of quantifying salience and concludes that the salience of two features does not add up perfectly although the salience clearly increases when two features are combined.

Huang and Pashler (2005) proposed a different method for measuring salience to mitigate the potential bias introduced by asking participants about their perception. They argued that a particular problem for measuring salience in attention-based tasks is that once a stimulus is selected because it stands out, an additional increase, in contrast, cannot make it more selected. Thus, an attention-based increase in performance is unlikely after the respective stimulus already stands out. It is important to note that this is based on an implicit model of attention that either selects or does not select, but that cannot boost visual processing gradually because of salience.

As a consequence, Huang and Pashler (2005) designed a procedure in which salience is used as a distracting stimulus while participants search for a predefined visual target. They presented displays with randomly positioned squares in them. There were three types of squares: Distractors that were neither bright nor big, a target that was bright and big, and a so-called key-distractor that was not as bright or as big as the target but bigger or brighter than the distractors.

Huang and Pashler reasoned that looking for the biggest and brightest distractor will divert part of the attention towards the key-distractor because of its salience: The more attention is directed towards the key-distractor, the longer the search will take on average. As soon as there were two kinds of features involved, they — like Nothdurft (2000) — compared the salience of one feature to the salience of a qualitatively different feature. Both methods related behavior caused by salience of one type of feature contrast to behavior caused by another type of feature. By systematically varying one feature while keeping the other constant, points of equal salience can be found, and by mapping different feature types to one reference type of feature contrast, a quantification of salience is achieved.

Contrary to the argument by Huang and Pashler (2005) — that the search time for unique elements is not a good dependent measure for salience, Koene and Zhaoping (2007) compared salience by varying the target of the visual search task. The authors did not use the time needed for searching as a direct measure of salience but used it to compare two models. One model treats the two features as independent cues by assuming statistical facilitation (e.g., Miller, 2016) because of two independent cues for the search target whereas the other model assumes an improved performance beyond statistical facilitation so that one could expect some form of interaction between both cues. In line with their expectations, only a color and motion combination did not show an interaction whereas color-orientation and orientation-motion combinations show an interaction according the the used race model. This result is, however, at odds with Nothdurft (2000) who reported the opposite: color and motion combinations showed more interaction than color-orientation and orientation-motion combinations. One explanation is that the statistical model comparison can be paraphrased as testing for the least bad model, which means that the validity of the result depends on the matches between assumed processes and used models. To evaluate this, however, one would have to spell out a quantitative model of attention for visual search tasks, including this experiment.

A point of criticism applicable to all mentioned experiments on salience is that salience is task-relevant. If salience's effect on stimulus-driven attention ought to be measured, the same salience should not be helpful to solve the task at hand. Otherwise, it would undoubtedly receive an additional attentional advantage because it is relevant in the respective situation. Although it is reasonable not to mix relevance and stimulus-driven attentional influences in a quantitative measure of salience it is quite difficult to achieve this: E.g., despite their attempt to rule out influences from task relevance, Huang and Pashler's (2005) key-distractor was not defined by features different from the features of the element to be searched. That is, if attention was directed towards these features, then also the key-distractor would draw attention because of the relevance of its features in this context.

To sum up this introductory chapter, physical contrasts that attract attention are called salient. Although salience is a central concept in theories and models of attention, methods for salience measurement are heterogeneous and sometimes yield conflicting results. One cause for these conflicting results are assumptions made about salience, e.g., whether it is affected by task relevance. These assumptions are partially implicit and can be explained by different intellectual judgments about salience which — in turn — affect both the design of the experiment and the analysis of data. As in the Dennett quote at the beginning of this chapter, creativity and the imagination of possible explanations thus are as vital in science as the empirical collection of data. However, the heterogeneity of salience measurement becomes problematic when models of salience are to be developed or quantitatively evaluated.

The approach to the quantification of visual salience presented in this dissertation aims to resolve ambiguities by explicitly the linking intellectual judgments and theory to experimental data by a formal model. In this way, I do not present the *right* way of quantifying salience but a way that resolves ambiguities in salience measurement by explicit appeal to theory while maintaining applicability to a broad array of stimuli.

Chapter 2

Psychology as a science

> The first principle is that you must not fool yourself — and you are the easiest person to fool.
>
> Richard P. Feynman (Feynman, 1974)

When I started my work as a young scientist, replication attempts were the first practical work I did. When these replications failed, I blamed the outcome on myself. Surely, the published work must be reliable — I am making mistakes, I told myself. In retrospect, I feel pity for my past self because I was unaware of the significance that replications bear for empirical science: On the one hand, replications are indeed an instructive way of learning the tools of the trade; on the other hand, the results of a well-made replication are actually of scientific value (Frank & Saxe, 2012). At that time, my attitude was completely different; I did not try to o convince others — myself included — that an empirical result is reliable, but to make an experiment "work." Making the experiment "work" entails trial and error and a lot of hidden flexibility both in experimental design and data analysis whereas convincing others is characterized by a well structured, systematic, and transparent process (e.g., guidelines see Asendorpf et al., 2013; Brandt et al., 2014).

In contrast to my anecdotal evidence about the reproducibility of a particular experiment, today, we have systematic evidence about the reproducibility in psychology as a discipline. The reproducibility project (Open Science Collaboration, 2015) attempted to replicate 100 experiments from three esteemed psychological journals and revealed that much of the published research is not as reliable as many scientists expected it to be — The reproducibility project evaluated the p-value, effect size, subjective assessment, and meta-analysis of effect size. Depending on the type of analysis that is preferred, between 39 % (subjective judgments) and 68 % (the combination of replication results and original) of replications were successful. As the large difference in these values suggests, there is discord about what counts as successful replication. Gilbert, King, Pettigrew, and Wilson (2016), e.g., reanalyzed the data by making different decisions in the analysis process. Their analysis revealed the opposite result of the (Open Science Collaboration, 2015), that is: Reproducibility is actually high. So, the discord on reproducibility warrants a closer look at the scientific methods of psychology.

Besides a discussion on how to estimate reproducibility for psychology as a whole, the initial results of the replication attempts motivated many more coordinated replication attempts. These replication attempts are aimed at individual well-known and evocative experiments — e.g., the experiment on the facial feedback hypothesis by Strack, Martin, and Stepper (1988). Strack et al. (1988) tested the hypothesis that not only emotion cause facial expressions but that facial expressions affect the perception of emotion. With a clever manipulation, either holding a pen with the teeth or with the lips, they produced a "smile" and "pout" condition. Subsequently presented comics were rated to be 0.82 points more funny on a ten items Likert scale for the "smile" condition. This experiment is literally a textbook example of how states of the body affect the perception of emotion. Acosta et al. (2016), however, showed that 17 independent direct replications conducted in 17 labs only found a difference of 0.03 instead of the original difference of 0.82 on the Likert scale. Only one of the 34 Bayesian statistical tests in the 17 labs decided against the null hypothesis. Overall, the replication attempt failed — however, it is important to note that the authors do not think that they falsified the facial feedback hypothesis. They merely conclude that it is possible that this particular experiment does not provide a strong test of the hypothesis. So, the issue here is the lacking reproducibility of a statistical result without immediate consequences for the theory.

Whereas the psychological theory on emotions stays largely untouched by the aforementioned replication attempt, so-called ego depletion — the idea the willpower draws from a limited resource — is a different case. Replication and some meta-analyses have raised serious doubts about the underlying theoretical claims. Ego depletion sounds plausible as a cause for behavior and as such it is not only a textbook example for concrete findings of motivation (e.g., Ryan, 2013) but also the basis for advice and guides (e.g., Baumeister, Heatherton, & Tice, 1994) for the public. And indeed, a meta-analysis finds convincing evidence for the existence of ego depletion (Hagger, Wood, Stiff, & Chatzisarantis, 2010). However, a coordinated replication attempt of the original study conducted in 23 labs (Hagger et al., 2016) was not able to replicate the results of the original experiments by Baumeister, Bratslavsky, Muraven, and Tice (1998). It is important to note that this was no direct replication because the original experiments were difficult to standardize for different labs. So the developed procedure included some carefully chosen design elements from other successful ego-depletion experiments (for details, see Hagger et al., 2016). Hagger et al. (2016) report a replication study that reveals that the effect size was highly overestimated. One reason may be that sample sizes are typically small in psychology and journals tend to publish only positive results. There are many ways to address bias in publications in meta-analyses. If this is done, the analyses, including many hundred ego-depletion studies, are inconclusive (Carter & McCullough, 2014; Vadillo, Gold, & Osman, 2016). Friese, Loschelder, Gieseler, Frankenbach, and Inzlicht (2018) give a summary of the arguments pro and contra the existence of ego depletion. They conclude that based on the evidence currently available, skeptics will not be convinced that ego depletion exists and a proponent of ego depletion will not be convinced of its nonexistence. The authors state that if this inconclusiveness is the result of two decades of research and hundreds of articles on ego depletion, then something must have gone seriously wrong.

Psychology is, by no means, the only empirical science that faces problems with reproducibility (Peng, 2015). In a nature online survey 88 % of the 1,576 scientists from different disciplines report at slight or a major reproducibility crisis (Baker, 2016). Among other disciplines where problems with reproducibility have been made evident are neuroscience (Button et al., 2013), cancer research (Begley & Ellis, 2012) and even artificial intelligence (Hutson, 2018).

What can go wrong in empirical science and cause grossly false conclusions and doubts about the reliability of results as in the cases mentioned above[1]? The obvious suspect is, of course, the theory and practice of statistical inference. Thus, I will give a survey of debates about statistical inference in psychology and related sciences. This will result in the assessment that all statistical inference needs models. What constitutes a good model, in turn, depends on the scientific inference to be made. Consequently, I will elaborate on scientific inference. At the end of the chapter, the perspective of Bailer-Jones (2009), a philosopher of science, is introduced in which statistical, as well as scientific inference, is used to link data and theory by models.

[1] Chambers (2017) provides an apt summary of the critique of scientific practice in psychology. Most of it is however beyond the point of this chapter.

Statistical inference

The predominant tool for statistical inference is arguably null hypothesis significance testing (NHST). Despite its limitations — that have been presented clearly decades ago (e.g., Oakes, 1986; Cohen, 1990) and reiterated recently (e.g., Gigerenzer, 2018; Cumming, 2013) — This procedure is often used mechanically and without considering other approaches to statistical inference. It also creates the illusion of a "free lunch"compared to other more complex analysis methods (Rouder, Morey, Verhagen, Province, & Wagenmakers, 2016) — namely that one can generate evidence for a particular and often only implicitly stated alternative model by rejecting a null-model which naturally seems simpler that formalizing this alternative model and compare both models statistically. Gigerenzer (2004) describes the prevalent practice of applying NHST as follows:

> 1. Set up a statistical null hypothesis of 'no mean difference' or 'zero correlation.' Don't specify the predictions of your research hypothesis or of any alternative substantive hypotheses.
>
> 2. Use 5% as a convention for rejecting the null. If significant, accept your research hypothesis. Report the result as $p < 0.05$, $p < 0.01$, or $p < 0.001$ (whichever comes next to the obtained p-value).
>
> 3. Always perform this procedure. (p. 588)

Gigerenzer argues succinctly that statistical thinking is often replaced with statistical rituals that rather conform to superficial rules of orthodoxy. This problem can partially be alleviated by drawing attention to other useful tools in the statistical toolbox like effect size, confidence intervals, and meta-analyses, e.g., as explained by Cumming (2013). However, to decide which tool to use, it is mandatory to understand implications, strength, and weaknesses of the approaches — or as Gigerenzer (2018) calls it "statistical thinking" rather than the application of mechanical rules.

Besides Gigerenzer's (e.g., 2018) argument that the inferential power of orthodox statistical methods is overestimated and affected by wishful thinking, another position identifies the problem in the incentives of the scientific system. John, Loewenstein, and Prelec (2012) revealed that scientists apply questionable research practices — sometimes even against better knowledge, e.g., the rare but outrageous cases of proven fraud — because they yield publishable results.

The publication system itself is also problematic: New and "sexy" findings are eagerly published whereas conflicting evidence, e.g., in the form of a replication study is much harder to publish and will almost certainly not lead to academic recognition. This affects both the presentation in individual papers (Giner-Sorolla, 2012) and the literature at large (Song et al., 2010). Thus, the research literature provides an incomplete and biased view of the works of researchers. This is directly linked to the practice of controlling the type I errors, the false positives. Consequentially, if a researcher tests 20 “sexy” hypotheses that are not true, they will in almost two-thirds of the cases have at least one chance find. Thus, to assess the actual probability of a type I error, one would have to take all the nonsuccessful unreported experiments into account.

Similarly, orthodox statistics require knowing the sampling intentions of the experimenter: Did they collect data until they reached 30 participants, or until 30 days had passed or until they reached predetermined confidence in their estimation. Each intention leads to a different statistical test because it changes sampling distribution (Dienes, 2008, 2011).

Independently of whether the questionable practices of applying NHST rather stem from systematic incentives or from an overestimation of NHST’s inferential capabilities, it is fairly safe to assume that most researchers do not want to cheat. If methods are applied such that the error rates are not controlled, it is likely that this is related to problems in understanding the logic of NHST. Gigerenzer presents examples from textbooks that use an imprecise and at times contradictory verbal descriptions of NHST’s results. As Haller and Krauss (2002) showed with a six-item questionnaire, the inferential power of NHST is also misunderstood by scientists and even by teachers of methodology.

A reason for the difficulties in understanding NHST may be that it is used as an amalgamation of two not completely compatible approaches: the approach by Fisher and the approach to statistics developed by Neyman and Pearson. How today’s NHST deviates from the originally proposed procedures has been explained by, e.g., Gigerenzer (2004). If the reader — like myself — is interested in a comparison between both procedures, see Christensen (2005) for a presentation of the individual approaches with reference to an example. Whereas the distinction between Fisher’s approach and the Neyman-Pearson approach is often referred to, I seldom found the different conceptualizations outlined beside each other. Thus, I found the concise summary by Christensen (2005) especially helpful.

> The basic elements of a Fisherian test are: (1) There is a probability model for the data. (2) Multidimensional data are summarized into a test statistic that has a known distribution. (3) This known distribution provides a ranking of the 'weirdness' of various observations. (4) The p-value, which is the probability of observing something as weird or weirder than was actually observed, is used to quantify the evidence against the null hypothesis. (5) α level tests are defined by reference to the p-value.
>
> The basic elements of an NP test are: (1) There are two hypothesized models for the data: H_0 and H_A. (2) An α level is chosen, which is to be the probability of rejecting H_0 when H_0 is true. (3) A rejection region is chosen so that the probability of data falling into the rejection region is α when H_0 is true. With discrete data, this often requires the specification of a randomized rejection region in which certain data values are randomly assigned to be in or out of the rejection region. (4) Various tests are evaluated based on their power properties. Ideally, one wants the most powerful test. (5) In complicated problems, properties such as unbiasedness or invariance are used to restrict the class of tests prior to choosing a test with good power properties. (p. 125)[2]

[2]Such a long quote is unusual but this text does an exceptionally good job at boiling down the unique features of two fundamentally different approaches to statistics whose differences are much too often not recognized at all.

However, NHST may not only be difficult to understand because two approaches are mixed in NHST but as Oakes (1986) already points out, some misunderstandings are based on an intuitive Bayesian conceptualization of probability. The reader can check their own intuitions about how statistics ought to work with the three example questions by Dienes (2011). Arguably, results of Bayesian methods are more in line with what researchers are interested in: Namely, the probability of two or more hypotheses given the data rather than the probability of data given a null-hypothesis. Bayesian methods have been used to recreate well-known analysis techniques (e.g., the t-Test Kruschke, 2013; Rouder, Speckman, Sun, Morey, & Iverson, 2009; Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010) so that NHST can simply be swapped for another — a Bayesian — tool.

So should we all become believers in Bayesian statistics instead of doing NHST? Hardly. Simonsohn (2014) showed that many questionable research practices are still possible with Bayesian methods. Also, the comparison of 855 t-Tests revealed that by and large Bayesian methods yield similar results (Wetzels et al., 2011). Only in extreme cases that are often particularly constructed to show a discrepancy, Bayesian and NHST results differ.

If you want to be critical of Bayesian methods, you could even say that proponents Bayesian methods for the betterment of reproducibility have to have some sort of split personality disorder: They take a deficient control of error rates as a reason to switch to a statistical procedure that cannot control error rates at all. This might be as surprising to the reader — as it was to me, but if examined closely, Bayesian methods are neither designed nor guaranteed to control error rates (Mayo, 2016, e.g.,). It all boils down to the fact that Bayesian understand probability as a degree of belief that can be assigned to any propositions, whereas NHST is build on the assumption that a probability is a long-run frequency (Hacking, 2001).

This clash of different schools of thought might be resolved by practically-minded statisticians: It has been argued that both methods have strength and weaknesses and should be chosen with these properties in mind when a problem is tackled (Efron, 2005). Little (2006) in particular proposes to use NHST if little is known about the phenomenon at hand whereas Bayesian statistics is apter if the researcher has already a rough idea how a model of the phenomenon may look.

One conclusion of the surveyed debates may be that statistical tools are based on assumptions and that it depends on the match of the implied model and the situation whether the analysis is less useful (Box, 1976). Rodgers (2010) establishes that the awareness for the role of models for analyses has been raised within psychology. He even calls this process a "quiet revolution". It is revolutionary because a researcher himself or herself engages in explicitly designing a model that corresponds to his understanding of the phenomenon such that many implicit intellectual judgments are declared formally in the model and thereby made explicit. Whereas previously a model has been chosen from a toolbox, e.g., as linear regression. This approach is particularly interesting for cognitive psychology because many cognitive abilities entail nonlinear dependencies (Rouder & Lu, 2005). Particularly, in the domain of visual attention, many mechanisms are well documented such that it appears much more pressing to me how all the different findings can be subsumed under theories rather than finding yet another effect or interaction within a particular design. In fact, it has been shown that formal models of attention can be used to accumulate progress (Logan, 2004).

To sum up, statistical inference is central to many empirical research questions. There are, however, some problems associated with the use of statistics within psychology. On the one hand, there are problems with the practices of how statistics are actually used as opposed to how the methods were designed to be used. On the other hand, there is a rather new debate on whether to abandon NHST in favor of Bayesian statistics. To make informed decisions regarding both debates, it is important for a scientist to become aware of her own conceptualization of the phenomenon and the problem that ought to be solved by statistics. For example, Bayesian statistics are particularly apt for inference under custom models. Much less debated but nonetheless relevant for methodology is the increase in modeling. So, how does a researcher know which method to pick? This choice depends on the scientific inference to be made.

Scientific inference

Up until now, I reviewed psychology as a science from the perspective of its methods, which brought us roughly speaking from statistical tests to models and further to theory — like, e.g., laws of nature. Another perspective on psychological science is to start with its theories. A lack of theories explains problems with reproducibility well because, without a precise understanding of what results count as expected and in line with theory and what results are surprising, it is difficult to judge which findings need replications before publishing and which do not (Muthukrishna & Henrich, 2019). Unfortunately, theories in psychology are a complicated matter. While most researchers would state that theoretical progress is clearly the goal of their research, it is nearly impossible to find an all-encompassing theoretical framework to which all those researchers could commit (Green, 2015). Attempts have been made (for a review regarding evolutionary psychology, see Fitzgerald & Whitaker, 2010a), yet all have been met with resistance. To make the matters even worse, seven meanings of the term "theory" can be distinguished (Abend, 2008) — a further reason why describing theorizing merely as the search for laws is too simple. While not all of these conceptualizations are equally important in psychology, neither does psychology limit itself to one meaning of the word. Bolacchi (2004) even presents an analysis according to which psychology is a pre-science because scientific methods are used with pre-scientific philosophical conceptualizations like the search for essences as in Aristotelian thinking. Summing up this short look at the status of theories in psychology, one would be justified in saying that it is a complicated matter and that one has to be careful not to conduct a cargo-cult science (Feynman, 1974) — an undertaking that applies scientific methods from natural science that are bereft of their inferential power because of a gap between method and theory. Because there is no common theoretical framework in psychology, I will continue form the methods to what might be meaningful implications for theorizing.

In order to know which statistical method is meaningful, it is crucial to think about its role in the scientific inference process. Probably most empirical scientists would agree that they work according to some variant of the hypothetico-deductive (H-D) model. Godfrey-Smith (2003) describes it succinctly as

> (1) Gather some observations, (2) formulate a hypothesis that would account for the observations, (3) deduce some new observational predictions from the hypothesis, and (4) see if those predictions are true. If they are true, go back to step 3. If they are false, regard the hypothesis as falsified and go back to step 2. (p. 236)

Its core idea is that hypotheses cannot be proven but merely tentatively accepted. Superficial, there is a resemblance between falsification in scientific inference as introduced by Popper (1959), and NHST whereas Popper proposes that scientific inference is not achieved by proving propositions but rather by disproving the wrong propositions. NHST disproves a null hypothesis in order to create evidence for an arbitrarily chosen alternative. So, is there a direct correspondence in both lines of reasoning?

First of all, it is helpful to acknowledge that there are severe differences between (Popperian) scientific inference and statistical inference about chance experiments (Dienes, 2008, p. 71). The argument, why NHST is no application of falsification is that scientific reasoning and statistical reasoning works differently: Statistical inference introduces quantified uncertainty in the form of probability whereas scientific inference is about the truth or falsehood of a claim.

The difference between logical reasoning in scientific inference and statistical inference with uncertainty is nicely spelled out by Wagenmakers et al. (2017). They compare syllogistic reasoning with probabilistic reasoning. Take for example a classical syllogism

(Premise) All men are mortal;

(Premise) Sokrates is a man;

(Conclusion) Sokrates is mortal.

and an example adapted from Wagenmakers et al. (2017):

(Premise) If Socrates is a human, he is unlikely to be a famous philosopher (because only very few humans are famous philosopher);

(Premise) Socrates is a famous philosopher;

(Conclusion) Socrates is probably not a human.

From this example, you see that introducing uncertainty does not make reasoning a bit more uncertain but annihilates the validity of the proof by contradiction. Thus, reasoning with uncertainty is fundamentally different from bivalent logic, and it cannot be said that NHST follows the same logic as Popper's falsification. This property of NHST may not be obvious because statistical inference is only used to show that the data is highly unlikely under the null hypothesis — yet, not that the null is falsified in Popper's sense.

Thus, one could say that the imperfections of inductive reasoning are all shifted into statistical inference by the H-D model (e.g., as opposed to Bayesianism in the philosophy of science; see, e.g., Carrier, 2019, p. 107) But when scientists apply NHST, to produce evidence for an alternative hypothesis, they perform quite a stretch in terms of logic. This is because scientific inference usually takes such a result as evidence for whatever alternative was hypothesized without statistically testing it either individually or in direct comparison with the null hypothesis. This consideration of the alternative hypothesis is the central advantage for the Bayesian version of the t-Test (Rouder et al., 2009). This is the reason why the H-D model and Bayesian statistics, in fact, go along quite well (Gelman & Shalizi, 2013).

The reader may wonder if a null hypothesis as in the t-Test or ANOVA also includes some sort of model that, in turn, should be able to provide a link between data and theory. The mathematical formulation of the null hypothesis can indeed be understood as a model (for details on what to call a model in psychology, see Rodgers, 2010, Footnote 2); however, as Taagepera (2008) argues that these models are descriptive. He distinguishes them from logical models. A descriptive model provides a convenient description of the data in the form of parameter values. Descriptive models, however, do not necessarily represent the logic of the data at hand that is intellectual judgments about the phenomenon. This would, for example, include forbidden areas (a response time cannot be negative thus fitting a normal distribution does not provide a logic model of the data).

Interestingly, also Neyman and Pearson proposed that it is necessary to formulate the alternative hypothesis as part of the statistical analysis, however, this is the part that is left out when Fishers and Neyman-Pearson approaches are mixed. Arguably, the newly developed interest of psychologists in modeling, as documented by Rodgers (2010), may change this in the future such that individual models are compared rather than null hypothesis being rejected to provide evidence for an arbitrary alternative. Modeling becomes a central concept here because to test Hypothesis 0 against Hypothesis A, it is necessary to relate the hypotheses to the data explicitly. This explicit link is created by modeling.

For example, consider the Normal distribution: Usually, it goes without questions to assume that a property is distributed according to the normal distribution in a population. There are different arguments how to justify this that are described and used as an example in Section 3.

Taagepera (2008, p. 124) gives an impressive example of how crude some links between data and hypothesis actually are: Telephones per capita has been modeled with a normal distribution in social science publications leading to a negative number of telephones per capita for a substantial percentage of countries. He calls this ignorance-based modeling — not considering that there are reasons leading to a normal distribution but rather assuming that a normal distribution appears in the absence of any causal factors. If the link between data and hypothesis already entails logically impossible conclusions, the validity of the resulting scientific inference may be limited by the weak link between data and hypothesis.

Besides coupling scientific and statistical inference (e.g., as formulated by Rouder, Haaf, & Aust, 2017), modeling may also be important for explanations in psychology: Often it is assumed that discovering an effect suffices to explain a phenomenon. For Example, if I cannot read out the color word “green” quickly because it is drawn in yellow, then I could say: “It’s the Stroop effect!.” According to Cummins (2000), however, named effects like the Stroop effect are rather explananda than explanantia. Roughly speaking Cummins argues that although discovering effects is clearly central in psychology, they themselves cannot explain — so to speak — their own existence. The particular phenomenon that we understand as an instance of Stroop Effect is thus merely described by "Stroop effect". An explanation, i.e., explanans, would involve reasons for the existence of the Stroop effect. Cummins continues by elaborating that the classical model for an explanation in the philosophy of science is the deductive-nomological (DN) model. The general idea of this model is that the explanandum is explained by the explanans only if the explanans allows to deduce the explanandum form its premises and that at least one of these premises is a law of nature (thus deductive and nomological). This view is, however highly problematic in the philosophy of science because a law of nature is surprisingly difficult to pin down to a precise concept. This is, in particular, the case for the special sciences like psychology, sociology or geology whose subject are particular phenomena and not the general laws of nature.

The solution for Cummins is what he calls *functional analysis*. Roughly speaking, this means to break down a complex and problematic system into a set of less complex, less problematic systems. Modeling is relevant in this context because it makes this process of deconstruction into individual factors and their interaction explicit. Thus, a logical model can be understood as making a phenomenon explainable and providing the content for the act of explaining (Krüger, Tünnermann, Rohlfing, & Scharlau, 2018).

Although models are often presented as a link between data and theory, it is much harder to find a coherent position in philosophy of science that explains how models fit within the hypothetico-deductive understanding of science. Among the few sources, there is a particularly comprehensive one by Bailer-Jones (2009). She provides an overview of scientific models in the philosophy of science by reviewing current and historical perspectives on models. From this review, she develops an own account of the function of models within science. As she explains, this account is broad enough to cover the concept of "model" throughout the different scientific disciplines as well as thorough enough to explicate how "model" relates to other concepts and points of view in the philosophy of science. According to her account, models are positioned between phenomena and theories.

What is the difference between a theory and a model? Historically, before the 1950ies, theories have been thought to make models obsolete. Roughly speaking, a model was an inferior precursor of a theory that was not sufficiently abstract. Theories are abstract sets of propositions. They are abstract in the sense that they do not aim to capture the concrete properties of a phenomenon. In fact, they omit as many concrete properties as possible to be widely applicable. Models are needed to apply abstract theories to concrete phenomena. Models satisfy abstract logical constraints of theory and concrete empirical constraints of a phenomenon. Bailer-Jones (2009, p. 1) defines models as "interpretative descriptions of phenomena that facilitate access to that phenomenon." As an interpretative description, models can still be abstract but have to satisfy at least some empirical constraints of the phenomenon. Thus, models are always less abstract than theories. In other words, models are customizations of theories such that they become applicable to some of the concrete properties of the phenomena by filling in the gaps.

A good example of how different models may satisfy different empirical constraints is Marr's (1982) three levels of analysis. These levels are used to model complex information processing systems with distinct yet complementary levels of analysis. On the computational level, it is described what the system does, that is what problem it solves or why does it what it does. The algorithmic level is concerned with how the system achieves the outcomes, e.g., what representations are involved and what processes. The implementation level or physical level deals with how the system is realized in the physical world, e.g., is it made of transistors or neurons, how are they connected, and how are the algorithms realized on this substrate. A theory of, e.g., vision can now be tested and instantiated by a theoretic model for the individual levels.

For example, the theory that there is purely stimulus-driven attentional capture can be instantiated on the algorithmic level by modeling how the information is processed without referring to neuronal substrates or by a neuronal model that is explicitly about how such purely stimulus-driven process is implemented in the brain. By filling in the gaps for an application to an empirical phenomenon, these theoretic models make predictions testable — however, the phenomenon is not directly linked to the theoretic model.

The term phenomenon was not yet introduced properly. A phenomenon is a fact or event in nature. It arises from observation and is at least suspected to be stable and not a random. The discovery of a phenomenon can be theory-laden. Certain phenomena like search asymmetries would not be recognized without previous research in a field like visual search. What the phenomenon is can change throughout the investigation. So, the phenomenon may be debunked as a chance finding or as caused by a faulty measuring device (e.g., faster-than-light particles; Cho, 2012).

An important assertion is that a phenomenon is, however, not equal to data. Data arises according to Bailer-Jones (2009) view from a specific way of observation or experimentation. The collected data is, thus always a certain aspect of the phenomenon. A phenomenon, however, is assumed to be stable beyond the currently available data. The collection of data is not seen as theory-laden because in contrast to identifying the phenomenon, gathering data is understood as reading a measure of an instrument. In psychology, this may not be the case, e.g., studies including expert raters. However, for the experimental designs presented here, this view seems applicable: Automatically recorded participant judgments are arguably not influenced by theoretic views held by the experimenter. However, models and the discovery of phenomena can be influenced by theories.

Data does not directly allow to test hypotheses because of measurement errors and limited sample size. Data models, thus deal with these problems. They allow a statistical inference. A statistical inference does not directly test a theory but tests the prediction of the theoretic model; The theoretic model models an experimental situation and satisfies empirical constraints as well as theoretical constraints. That is, the theoretic model states what the theory means for a particular experimental or observational setup. Incorporating all these points into one frame, Bailer-Jones (2009) arrives at the relations depicted in Figure 2.1.

Bailer-Jones' (2009) account accommodates many points of view. Her frame allows reflecting practices in psychology because it accommodates central concepts like phenomenon, data, theory, and of course their interconnection with scientific models. Bailer-Jones' account is in accord with the hypothetico-deductive method as theoretic models are derived from theories and thus allow to test theories. The frame does not answer whether a theory or an auxiliary hypothesis was falsified but rather makes explicit at which points auxiliary hypotheses are introduced to satisfy both the empirical and the theoretic constraints.

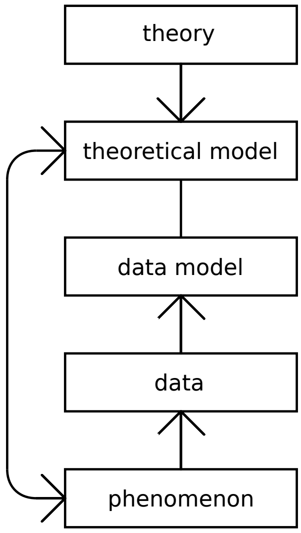


Figure 2.1: Bailer-Jones' 2009 complete picture of how models are a part of in statistical and scientific inference.

Chapter 3

From phenomenon to theory and back

> Finally we shall place the Sun himself at the center of the Universe. All this is suggested by the systematic procession of events and the harmony of the whole Universe, if only we face the facts, as they say, "with both eyes open."
>
> ---
>
> Nicolaus Copernicus as quoted by Kuhn (1957, p. 154)

In the last chapter, the scientific-philosophical framework by Bailer-Jones (2009) was introduced to explain how models close the gap between data and theory. Particularly, the framework shed light on how models fit into the hypothetico-deductive method that constitutes one of the foundation of scientific psychology (e.g., Dienes, 2008). This chapter takes the introduced framework and applies it to psychological research on salience. This is not a straightforward task because the framework is idealized and was not developed particularly for psychology but rather for empirical science in general — with a strong emphasis on physics. Also, Bailer-Jones states in her conclusion that the framework is still incomplete and, as it is typical for philosophy, competing positions exist. So, why bother to locate research results within this particular framework?

The answer is that by locating a research result within the framework, we gain an idealized view of what a particular research result should and should not be used for within the hypothetico-deductive method. Even if the reader does not agree with parts of this framework or its content, it helps to pinpoint critique to a particular aspect whereas the value of another part of the contribution can be evaluated independently. This line of reasoning has also been described in Krüger et al. (2018)[1]. To support this evaluation and provide the rationale for the reasoning behind this work, the current chapter will discuss several candidates or views for each of the positions to be filled within Bailer-Jones' framework. Certain theories, phenomena, and methods of scientific and statistical inference have been mentioned previously and will reoccur in this chapter because they have not yet reviewed in relation to Bailer-Jones framework. Thus, although I try to keep it to a minimum, there may be some overlap with the first two chapters.

[1]This chapter has partially been written before the article and is largely overlapping with its topic of linking data and theory quantitatively (Krüger et al., 2018). In addition to the present chapter, the article compares NHST, machine learning in psychology (for a review, see, Yarkoni & Westfall, 2017) and inference under Bayesian models with a particular focus on whether they provide a good explanation of the data.

The chapter starts by reviewing phenomena that are related to salience but also phenomena that ought to be distinguished from salience in a systematic investigation (for a more comprehensive overview on covert attention, see, e.g., Wright & Ward, 2008). From there on, the order might surprise the reader as Bailer-Jones' framework is not traversed from end to another. In contrast, the framework is filled in the order that is best to follow so that a topic does not presuppose a content from not yet introduced parts of the framework. So, after phenomena, I discuss theories that may be suitable for a quantitative investigation of visual salience. A focus is put on formal theories of attention, particularly on Bundesen's (1990) Theory of Visual Attention (TVA) that provides the basis for the present work. In congruence with Bailer-Jones, theoretical models link theory and concrete situations. In the case of experimental psychology, these concrete cases correspond to experimental paradigms. In Section 3, I discuss approaches to statistics that comprise and compare methods from the Bayesian school, Nayman-Pearson school, and likelihood-based approaches. The chapter ends with a decision in favor of Bayesian methods for the present work.

Thus, after this chapter, the phenomenon of salience is distinguished from related phenomena and embedded in a theoretical context.

Phenomenon

Salience can be distinguished from other sources of attention with the help of taxonomies. Salience affects attention; this much is already known from the introductory examples of Chapter 1. However, it is important to clearly describe this phenomenon, its boundaries, and phenomena ascribed to other causes to ensure that whatever quantitative measure of salience is not confounded by related yet different influences on attention.

As a first systematic approach to salience and attention research as a whole, it is useful to situate salience within a taxonomy of attention. Already James (1890) distinguished different phenomena under the term attention by means of two central categories. Particularly, James distinguished between "passive" and "active" attention. Whereas "passive" attention is a reflex-like, nonvoluntary, and effortless attraction of attention, "active" attention is the voluntary direction of attention towards a particular stimulus or thought. In over a hundred years, these distinction is still central although "passive" is usually called bottom-up, exogenous, or stimulus-driven whereas James' "active" attention is called top-down, endogenous, or goal-directed. In congruence with a current taxonomy of attention by Chun, Golomb, and Turk-Browne (2011), I will use the terms *stimulus-driven* and *goal-directed* in this work. This dichotomy has been established in experimental psychology, e.g., by the seminal work of Posner (1980). Another important contribution that emphasizes this decision is Corbetta and Shulman (2002) influential review on the partially-segregate neuronal substrates of these two mechanisms. According to Corbetta and Shulman (2002), the processing of stimulus-driven events in the ventral frontoparietal network works as a "circuit breaker" for the goal-directed system that is located dorsally. However, this conceptualization as a dichotomy of influences may not capture all attentional selection biases and as such fail to provide a true dichotomy (e.g., Awh, Belopolsky, & Theeuwes, 2012). In his recent review, Wolfe and Horowitz (2017) distinguishes three additional factors: history, perceived value, and scene guidance. Nevertheless, it is undisputed that stimulus-driven influences like salience are distinguishable from goal-directed influences.

Attentional phenomena caused by salience clearly belong to stimulus-driven attention. However, sooner or later, salience-based influences get integrated with goal-directed influences. This is a central point in Treue's (2003) already discussed comprehensive review of the underpinnings of salience. In particular, Treue notes that the computation of salience is at least partially hard-wired because of the center-surround organization of many neurons. Furthermore, the neuronal processing can be understood as composed of different stages as processing in the retina is stimulus-driven with a special emphasis on discontinuity. Whereas the bandwidth decreases, the goal-directed influence increases along the path through the visual system and cannot be segregated into a late modulation of a largely stimulus-driven process. In Treue's (2003) view, this system has — roughly speaking — a common currency[2] of attention that "equates the absence of attention with low stimulus power" (p. 430). Thus, different influences on attention can be understood as contributing to a common currency that determines how much attention is attracted.

[2] A common currency as the Euro is for many EU countries so that qualitatively different things like goods or services can be quantitatively compared.

The information reduction throughout the visual processing indicates stimuli competing for limited resources. As the representations become more complex in the later stages, the amount of stimuli that can be represented diminishes. Desimone and Duncan (1995) describe this as a parallel biased competition because only a few of simultaneously visible stimuli are available for higher cognitive processing including conscious perception and the chance to reach such a representation is biased towards stimuli that are relevant for behavior. This bias corresponds well with Treue's (2003) common currency of attention in which stimulus-driven and goal-directed influences are factored in.

Whereas the distinction of stimulus-driven and goal-directed influences shed light on how a stimulus is selected, one can also ask what is selected. As summarized by several reviews (Treue, 2003; Corbetta & Shulman, 2002;

Carrasco, 2011) attentional selection is either based on location, feature, or object. Also, points in time can be selected (Egeth & Yantis, 1997; Los, 2010). It is important to note that these phenomena can occur without shifting the gaze (covert attention; Carrasco, 2011). The rationale for a keeping the gaze and eccentricity of the salient stimuli constant is that spatial resolution changes depending on the eccentricity (for orientation, see Westheimer, 1998; for color, see Hansen, Pracejus, and Gegenfurtner, 2009). Thus, a quantitative investigation of salience manipulations can be operationalized with what is selected when external factors like visual resolution are controlled.

In summary, stimulus-driven influences like salience cannot be separated cleanly from goal-directed on the substrate level where selection mechanisms are implemented in the brain. Both mechanisms are, however, partly segregated and contribute to an overall currency of attention. This currency, in turn, is used to distribute limited processing resources. For a quantitative measure of salience, salience has to be manipulated while goal-directed and more basic influences like visibility have to be kept constant. A quantitative measure of salience would be particularly meaningful for the functional analysis (Cummins, 2000) of attention if it were formally distinguished from and explicitly related to goal-directed influences and the selection mechanism itself.

Whereas the previous paragraphs were about properties of phenomena that are shared by many attentional phenomena, now properties specific to salience will be reviewed. To this end, I start with a thought experiment and review empirical evidence in this direction afterward.

Figure 3.1: Salience example: The more difficult to predict a stimulus from its surrounding, the more salient it is.

Salience is understood in the context of this work as local physical contrasts that affect attention as mentioned in the introduction. This should not be taken for granted as salience may also refer to the perceived conspicuousness of a stimulus standing-out from its surrounding. Kerzel, Schönhammer, Burra, Born, and Souto (2011b) showed that salience changes perception which indicates the possibility that salience may have distinct effects on perception and attention. Although the perceived conspicuousness is not focused in the present work, it provides an evocative thought experiment: Suppose you have a photo and what a rough estimate of the salience of a particular region, e.g., one that shows a particular object. Now, cover this region with a coin and imagine a reconstruction of the coin-covered part purely from the rest of the image. Figure 3.1 shows an example. You will notice that some parts are more easily reconstructed that others — or more precisely, these parts deviate less from a reconstruction based on the local environment. These regions are of low salience. Other parts may be so specific in their contrasts, colors, and edges that they can hardly be inferred from their surroundings. Such regions are of high salience. If you limit your mental reconstruction of the covered regions to physical properties, you are doing something comparable to a neuron with a center-surround organization that exhibits a high fire rate of center and surrounding deviate.

The idea that salience is connected to the likelihood of reconstruction a visual stimulus from its surrounding is taken from a purely information-theoretic approach to salience and the respective model by Bruce and Tsotsos (2009). This model does not analyze the semantics but information-theoretic properties of pixel images not wholly unlike a statistical analysis. The model does not claim to model empirical findings in particular but the functional properties of salience. And indeed, in congruence with empirical findings, this model shows that according to information theory, the more physically unique, the more likely a stimulus attracts attention. This is, however, only every rough description because the interesting aspect is not that the model tells us that a unique thing stands out in an image but why — it contains the most information.

With this example, I do not want to discuss the model (Bruce & Tsotsos, 2009) in depth; after all this section is about the phenomenon to study. However, the model is supposed to prompt the intuition that the more basic properties a spot in an image contains that cannot be inferred from its surrounding, the more salient it should be according to information theory. Of course, the view presented by Bruce and Tsotsos is idealized and developed from an engineering perspective. To detect in how far the actual human visual system deviates from this idealized principle, I will review research on the influence of visual properties in the following.

Already James (James, 1890) identifies individual features that attract attention. Among his examples are "strange things, moving things, wild animals, bright things, pretty things, metallic things, words, blows, blood." (p. 417) Current research does not confirm most of these examples as sources of stimulus-driven attention because attention is not drawn by a feature per se but by its contrast to the surrounding (Duncan & Humphreys, 1989). Still today, different types of features and feature contrasts are distinguished into feature dimensions as, e.g., motion or color. This view is likely founded in the early research on visual attention in which the processing of visual features was seen as building blocks from which more complex representations are pieced together. This view, however, showed to have limited explanatory power. Nevertheless, and as explained by Treue (2003) salience is at least partially "hard-wired" so that independent dimensions based on neurophysiology are reasonable idea to start with.

As reviewed by Wolfe and Horowitz (2004), there is a continuum of feature dimensions from being either highly likely to affect stimulus-driven attention to highly unlikely to do so. In congruence with the assumption of basic properties driving visual salience, properties like motion, orientation, luminance, and color contrast are all effective at manipulating attention.

Another important aspect of salience is conveyed by the work on visual contrast by (Duncan & Humphreys, 1989). Duncan and Humphreys used the paradigm of visual search to show that search efficiency is affected gradually by the degree of similarity between target and distractor stimuli (which makes the search less efficient) and the degree of similarity amongst the distractor stimuli (the more homogenous the distractors, the more efficient the search). So, salience is a gradual property that depends on the similarity between an object and its surrounding with respect to certain salience-relevant feature dimensions.

This also fits well with the rescinding of the dichotomous distinction between parallel and sequential search in which a stimulus was believed to be either immediately found or found later by linearly scanning the whole stimulus array. Instead, today, a continuum of efficient to inefficient search is assumed (Wolfe & Horowitz, 2017).

Whereas much research has been done on what physical contrasts can affect attention (Wolfe & Horowitz, 2004), much less research has been done to investigate how salience strength and physical contrasts are related. Roughly speaking, instead of asking whether a feature dimension can affect attention at all, here one would ask if its impact is strong or weak and whether it contributes independently of other feature dimensions to a common currency of salience. Moreover, the few results stem from different research methods and yield contradictory results. Related works have already been introduced in Chapter 1. The next paragraph provides a short summary.

Nothdurft (2000) provides a comprehensive comparison of single-dimension contrasts, e.g., a stimulus only unique because of a color difference, and contrasts combined from two contrast, e.g., a stimulus unique because of a color and an orientation contrast to its surrounding. He concludes that the estimated salience from the contrast combinations is always higher than the salience of the individual contests. However, the salience of this combination is always less that the salience estimates of the individual contrasts added up. Whereas Nothdurft (2000) explicitly asked to judge the salience, Huang and Pashler (2005) proposed method to reduce goal-directed influences on the salience estimate by estimating the salience of a distracting element rather than on an element that has to be found to solve the experimental task. Again, a salience value for the combination of two contrasts was found that was between the maximum of the individual contrasts' salience estimate and the sum of both estimates. Koene and Zhaoping (2007) report that they were not able to replicate Nothdurft (2000). Similarly to Huang and Pashler (2005), Koene and Zhaoping (2007) used a search task but did not use the degree of distraction but the search time for the salient target stimulus. Koene and Zhaoping (2007) did not estimate salience directly but decided whether salience dimensions interact by a model comparison. This model comparison contradicts Nothdurft (2000) findings who reported a stronger interaction for the color-motion combination for the color-orientation and orientation-motion combinations whereas Koene and Zhaoping (2007) found no interaction at all for color-motion but interactions for the other two combinations.

Whereas this review clearly shows that there is no obvious or simple way how questions about the strength of salience should be translated into an experimental design or statistical analysis, some critical points can be observed: Salience was part of the task in all designs because salience was necessary to solve the task at hand. Thus, goal-directed influences may have been kept constant but are difficult to discern from the influence of salience; Although Koene and Zhaoping (2007) use a model, they do not actually apply statistical model comparison, instead their argument is based on the nonsignificant deviation from model prediction for the color-motion combination. So, for the validity of the argument, the absence of a statistical effect has to be show. Although it is possible to argue for the null hypothesis, a high power is needed such that possible effects are not overlooked. Their sample size of eight and the absence of a power analysis makes their argument look vague. Nothdurft's (2000) sample size comprised even less participants; There were five, probably highly trained individuals including the author himself.

Salience also has a distinct time course such that it peaks at around 100 ms to 150 ms after stimulus onset and decays afterward. Again, like in the case of possible interactions, it is tightly connected to the question of how to access the strength of salience at all. In this pursuit different designs have been used: Saccadic selection (Donk & van Zoest, 2008; Silvis & Donk, 2014), saccadic trajectories (van Zoest, Donk, & Van der Stigchel, 2012; Tudge, McSorley, Brandt, & Schubert, 2017), variants of cueing (Donk & Soesman, 2010; Dombrowe, Olivers, & Donk, 2010), visual search (Couffe, Mizzi, & Michael, 2016), temporal order judgements (Donk

& Soesman, 2011). It is unclear in how far a deviation for a straight saccadic trajectory measures the same construct as for example response time in a cueing experiment. An additional difficulty with assessing salience's time course is that whatever paradigm is used, the presentation duration of the salient stimulus has to be manipulated largely independent of other factors. Also, the transient influence of salience has to be distinguished from potentially confounded temporal phenomena like expectation (Vangkilde, Coull, & Bundesen, 2012) and alertness (Matthias et al., 2010).

To sum up the relevant properties of the phenomenon of visual salience, I would like to come back to the initial example. In the model by (Bruce & Tsotsos, 2009), a stimulus is salient when it is difficult to "predict" it from is surround that is it carries much information because of its high contrast to the surrounding. The example also illustrates the point that although salient stimuli may appear as conspicuous, it is unclear whether this is the same as the effect that physical contrasts have on attention. Visual salience affecting attention is effectively manipulated using basic visual contrast types. These types are categorized in dimensions. These dimensions are likely not entirely independent neither do they add up perfectly. As hypothesized by Koene and Zhaoping (2007) and Nothdurft (2000) alike, this is likely linked to the underlying neurophysiology. Yet, there are many competing measures of visual salience such that different findings may equally likely be contributed to the different methods than to true effects of the independent variables. All of these methods must deal with the problem that salience is sooner or later integrated with goal-directed influences on attention. They also must account for the fact that different salience dimensions must be manipulable to measure and compare the effect of salience. Additionally, time is a crucial factor when measuring the quantitative strength of salience as salience is not stable in time. Thus, questions about the quantitative strength of salience are unlikely to be solved by yet another experimental design alone. Instead, it is advisable to look at existing theories and models to develop a measure of salience such that its quantification can derive meaning from its relationship to theoretical constructs and functions.

Theories

Theories are particularly difficult to discriminate from theoretical models. According to Bailer-Jones (2009), the idealized version of a theory is maximally abstract. That is, for example, the case for a set of mathematical axioms. To apply a theory, however, a connection has to be made to concrete instances in which the theory is applied. For example, this would be formulas derived from these axioms that describe a certain experimental situation so that a particular outcome is expected based on the theory. Such a formulation, however, hold the status of a formal model rather than theory.

Bailer-Jones (2009) further summarizes that abstract theories arise form descriptions of concrete instances and are then formulated independently of the these instances when much evidence has shown to support the idea expressed by the theory. So, it is not surprising that during their development, many theories are incorporated into models but may not be formulated independently of the models describing certain situations. However, in this section, works are reviewed that provide a set of abstract assumptions about salience and attention in general.

In contrast to early concepts of attention, e.g., in James' (1890) work, attention is not understood as a unitary construct (Poth & Schneider, 2013). Carrasco (2011) adds that, rather than to simply infer attention's presence, it is necessary to define and manipulate it while keeping the stimuli and task constant. She also summarizes that attention research is conducted on many levels of analysis and is advanced by their integration. The climate for a unified theory of attention is rather bad as there are many rather independent research processes that additionally are studied on different levels of analysis.

One way of dealing with this complexity is to focus theorizing on a specific paradigm. This has been done in the case of visual search. Visual search is closely connected to the phenomenon of visual salience because stimulus-driven attention is one factor that can guide attention towards the target stimulus. Some examples are shown, e.g., by Wolfe and Horowitz (2004), Wolfe and Horowitz (2017). So, salience plays an important part in whether a stimulus is found among others efficiently. Theoretical work aims at how this search works in general. Thus, theories from this particular paradigm are worth a look. In this pursuit, it is apt to start with Treisman's (1980) Feature Integration Theory (FIT). FIT, however, assumed a sharp distinction between features that can be pre-attentive and that others require attention. Wolfe's Guided Search Theory was a successor to FIT. This theory proposes an early but shallow parallel analysis of all locations that selects some stimuli based on features to make them accessible to higher cognitive processing (Wolfe, 1994). The core idea is that features are represented in maps. Without the intention to mock this approach one may say that introducing a map is basically just claiming that features are not represented independently of the space which, if put plainly, seems quite obvious because, e.g., eye movements need a target location. Thus, it is highly likely that each theory of attention will deal with features and locations. The interesting question is how this works. Wolfe (1994) hypothesizes different possible organizations of feature maps. Central to this question is which features are independently analyzed. So, it is questionable whether there is a map for color in general or red-green and blue-yellow contrasts. Independently of how the feature analysis is organized, the respective dimensions are multiplied with a goal-directed influence.

The theory part of Guided Search provides a qualitative account of how the search process is structured and which type of representations are used. To research quantitative aspects of salience, it would be crucial to know in advance how the maps are organized and how other influences like goal-directed feature-based attention interacts with these stimulus-driven influences. In fact, if this information were known, the present research question would be obsolete. Also, it is important to note that visual search theories are by their nature focused on the particular task of search so one may argue that a measure of salience should be based on a theory that is at least not by design focused on one paradigm. Thus, visual search theories elicit important structural properties of attention but are not a suitable basis for quantitative analysis of salience.

If theories are not focused on a particular paradigm, another approach is to focus on the phenomenon of salience itself by providing a computational explanation. Koch and Ullman (1985) provide such a theory. Basically, they propose independent maps as in Wolfe's (1994) Guided Search. However, this architecture is not limited to a particular paradigm and specifies how salience might arise from neuronal representations of contrasts. The core idea in this architecture is retinotopic maps that encode independent contrasts and are combined only after the contrasts have been computed. A master map determines where to attend. Combined with the inhibition of return mechanisms, this theory provides the basis for many computational models of salience and attention. Arguably the independence of a certain task makes it even harder to discern stimulus-driven and goal-directed influences quantitatively and it is often not clear which dependent measure is actually predicted (Koehler, Guo, Zhang, & Eckstein, 2014). As in the case of visual search theories: If these theories of the neuronal origin of salience were already certain about quantitative aspects of salience, this work would be obsolete.

Li (2002) approaches a computational explanation of salience based on modeling the neuronal structure of the primary visual cortex. This model focuses on the implementation level of analysis (in contrast to computational or algorithmic as discussed in detail in Section 3) that is how is salience actually been computed by neurons. Li proposes that the structure of the artificially recreated neuronal connections in the primary visual cortex (Li, 2001) suffice to compute salience without a separate and explicit master map. However, again, this model is not explicit in how goal-directed influences affect these computations.

Instead of the paradigm specific theories, these theories are salience-specific by which I mean that they are not specific about other influences of attention or provide a specific prediction for particular paradigms. Thus neither paradigm-specific nor salience-specific theories provide a sufficient basis for a quantitative investigation. Existing quantitative modeling on the basis of these theories will be reviewed in the next Section 3 on theoretical models to provide additional support for this claim.

The third approach to theories that I present is formal theories of attention. These are rather abstract, but their mathematical form ensures that influences are quantitatively discernible. In comparison to the paradigms-specific or salience-specific theories, they are better suited to accumulate progress in attentional research as they have a more general scope (Logan, 2004). Logan reviews two approaches: signal-detection theory and similarity-choice approaches, including Bundesen's theory of visual attention that is discussed in the next section.

Viewed from a more abstract vantage point, one could say that attention is always about selection. Desimone and Duncan (1995) describe this selection as parallel processing of stimuli until a bottleneck is reached. In this case, a stimulus is selected according to its attentional bias. This becomes evident in the phenomenon that attended stimuli are perceived earlier than otherwise equal stimuli. It has already been formulated by James (1890) and is currently well established within experimental psychology under the term prior entry (for reviews, see Shore, Spence, & Klein, 2001; Spence & Parise, 2010). These particular implications will be relevant to the theoretical model of the present approach. Salience, thus, can be understood as merely one "trick" the visual system uses to select information without full knowledge about its implications.

Summing up, there are usually more visual stimuli than can be processed by limited cognitive resources. Thus, stimuli must be selected for processing. This is done by the attention, which includes salience as the effect of physical contrasts. These do not have a status as exclusive as initially assumed in FIT, i.e., privileged pre-attentive processing. Instead, contrasts guide attention. Theories of salience admit that it is affected by goal-directed influences but also assume a rather fixed “hard-wired” computational process. Although theories differ in whether they assume an explicit neuronal representation of salience, all agree that some sort of retinotopic mapping must happen in which the surrounding plays a crucial role in determining the amount of contrast. Different types of contracts are distinguished, based on neurophysiology. Whereas these computational models focus the process of how salience arises, there are also formal mathematical theories of attention that aim at relating abstract concepts as attentional bias and selection by mathematical functions. These different approaches may seem quite alike and can, in fact, they can overlap. However, they differ in the focus of what to explain, which may become more clear when the theoretical models are discussed that such theories may entail.

Theoretical models

Theoretical models are what links abstract principles or theoretical axioms to actual situations. In actual theoretical work, this distinction is, however, seldom made and maybe a seem a bit artificial. However, as Bailer-Jones (2009) argues, theories cannot be tested directly: One has to specify what exactly they entail for a particular concrete situation. This is done by the theoretical model. Theoretical models can be differentiated in how general or situation-specific they are; Some may deal with classes of instances whereas others deal with one particular experimental design.

Visual salience can be approached by different disciplines from different angles — as other phenomena that emerge in complex cognitive systems. These approaches have different and sometimes competing foci: Computer scientists are interested in computational properties of salience models, neuroscientists are concerned with identifying and locating the neuronal substrate of salience, and psychologists are interested in observable behavior. In all of these different approaches to salience, models are present although they provide — borrowing form Bailer-Jones definition — a different interpretative view that provides access to a phenomenon. One way to bring order to these different models and their different perspectives are Marr's (1982) levels of analysis that are divide models into either dealing with the computational analysis that is what problem does the system solve; algorithmic how is the problem solved which includes processes, mechanisms and representation; and finally the implementation level that analyses how a system is physically implemented, e.g., on a neuronal substrate.

There are a large number of computational models that are built around Koch and Ullman's (1985) theory of map-like representations that are combined to a master map (for a survey, see Frintrop et al., 2010) these models are not necessarily based on empirical research. For example, Bruce and Tsotsos (2009) use information theory, a purely theoretical approach, to implement a model that attends to the spots that convey the most information. The impressive achievement of these models is that "attention" can be predicted for every digital image or even digital video.

From a psychological perspective, however, these models are often not very precise in terms of which psychological construct they predict. As reviewed earlier attention is not a unitary construct thus predicting "attention" for all stimuli, all tasks, etc. may be an approach that will never reach a precise prediction of human behavior because it may be too ignorant of the individual processes and properties involved in attention. One use of these models is to predict eye-movements. As Schütz et al. (2011) review, these models usually explain only a fraction of the variance, and their performance depends on the task. More generally, Koehler et al. (2014) ask "What do salience models predict?" by comparing model predictions to different behavior measures. It turned out that the best models where better in predicting explicit judgments of salience than eye-movements in free-viewing, object search, and salience search tasks.

Also, with Bailer-Jones (2009) in mind, the linking of theory and data is somewhat unidirectional in these models as they allow to predict data for an arbitrary input but do not shed light on how the model or the theory should be changed in case predictions not deemed good enough.

The reason for this is that the theory is quite explicit how salience arises but remains vague in terms of other psychological phenomena like selection or visual search speed. Thus for applying the hypothetico-deductive method to the theories is difficult, its auxiliary hypotheses are not precisely formulated, yet needed to apply it to concrete behavior as eye-movements.

Models that have a more precise focus have other difficulties to deal with: Wolfe (2007) reports that parts of his guide search model can describe and predict the quantitative properties of empirical results. Yet, to this day, the model is not fully implemented, and the details of the implementation are based on assumptions and phenomena that are related to but not exactly salience, e.g., detectability of difference between contrasts which are a necessary but not sufficient criterion for salience. Also, Li's (2002) neuronal salience model is implemented as an artificial neuronal network based on a simulation of primary visual cortex (Li, 2001). Yet to this point, it cannot be used for precise quantitative prediction of salience.

Whereas up until now models particularly focused on how salience arises in general or in particular situations or by particular means of computation, formal theories of attention are — as mentioned in the previous section — concerned with formal relationships between abstract concepts like selection and attentional bias. As such, they aim at a functional analysis, which Cummins (2000) proposed as a primary mode of explanation in psychology. Furthermore, Moore (2015) formulates how the activities of mathematical modeling and theorizing are overlapping activities, which both contributing to the gain of knowledge.

Two formal approaches to attention are reviewed by Logan (2004). These are similarity-choice based theories and signal-detection-theory based works. Similarity-choice theory aims to predict choice probabilities based on similarity and bias whereas signal-detection theory's goal is to separate sensitivity from bias (for a detailed review, see Logan, 2004). He concludes that similarity-choice based theories may be better at providing insight into the cognitive processes, or in Cummins' (2000) terms provide a better functional analysis. Bundesen's theory of visual attention (TVA) which is used in the present work is a well worked out formal theory following the similarity-choice thinking (Bundesen, 1998, 1990). It is a promising candidate to make explicit how salience affects attention and processing of visual stimuli because TVA is a formal and quantitative account for visual selection and recognition. TVA, in particular, allows estimating attentional parameters with high precision (Habekost, Petersen, & Vangkilde, 2014).

Data models

According to Bailer-Jones (2009), data models deal with the problem that, whatever data are collected, they provide only a limited perspective on the phenomenon to be measured. She provides the example of measuring the melting point of lead: The exact melting temperature may not be read of the thermometer once during an experiment. The mean of repeated measurements, however, will be very close to the true melting point.

From Bailer-Jones' (2009) perspective, data are a limited perspective on a phenomenon that is persistent beyond the data. Thus, intellectual judgments are needed to infer general properties from individual observations. These assumptions are formalized in the data model. Data models include intellectual judgments to infer general properties form limited observation.

Such intellectual judgments, for example, are the assumption that a certain property is normally distributed within a population. As mentioned in Chapter 2, Taagepera (2008) notes that some scientists believe that a normal distribution tends to happen exactly when causal factors are absent which he calls ignorance-based modeling. However, there are of course proper arguments for the assumption of a normal distribution; This assumption can be justified by the central limit theorem if there are many independent factors affecting the property in question. Another justification is based on observation; many natural properties are roughly normally distributed. Even if the normal distribution is nearly omnipresent when making statistical inferences, both arguments — theoretical necessity and empirical fact — can be disputed effectively.

The assumptions necessary for the central limit theorem as the cause of the distribution are in practice often violated (Lyon, 2013). Lyon illustrates this with baking loaves of bread: He tries to justify the normal distribution of the baked loves by the central limit theorem step by step. This exercise shows that the argument is far from trivial an does not work for many practical examples where the number of factors is limited (in his example, the weights of flour, water, sugar, salt, yeast, and water lost during the baking process). Additionally, the central limit theorem does not only support the argument for the normal distribution; If the influences rather interact than being independent, the same argument would yield the log-normal distribution.

The log-normal distribution plays also a role in disputing the claim that the normal distribution is simply best for describing most data of natural phenomena, i.e., an empirical fact. Limpert, Stahel, and Abbt (2001) challenge this claim by showing that the log-normal distribution provides a better fit for many datasets. Considering that the central limit theorem argument can also be used to supports the log-normal distributions, the argument for the normal distribution seam at least far from obvious.

Maybe Taagepera (2008) is right that the normal distribution reflects ignorance. This assessment does not have to be negative, however. Lyon (2013) shows a third argument for the preferential use of the normal distribution: The normal distribution may be "normal" in science because of all distribution with the same variance the normal has maximum entropy, that is it has the minimum amount of information. Thus, applying the normal distribution may simply be the optimal way of complexity reduction for values of a continuous random variable.

This discussion of the normal distribution should not be seen a general argument for or against its use but rather as an example that intellectual judgments are crucial for checking the validity of a model — even thought "methodologists and statisticians might be uncomfortable with the mean and variance being defined as 'models.'" (Rodgers, 2010, p. 4). Also, the normal distribution serves a an example here because its omnipresence and the usually assumed obviousness of its justification.

Taagepera (2008) distinguishes two types of models: *descriptive models* and *logical models*. Whereas descriptive models describe the data, logical models capture the intellectual judgements about the data. For example, describing telephones per capita or response time as a normal distribution does not fit the intellectual assessment that these quantities cannot be negative. Thus, a normal distribution may be able to describe the data but does not capture basic logical properties of these quantities. Additional logical properties are forbidden areas (like being nonnegative), anchor points, modeling of nonlinear relations.

Also, variability may stem from different sources. Hierarchical models can model these sources separately which is particularly apt for inferences under nonlinear cognitive models (Rouder & Lu, 2005; Rouder et al., 2009). For example, such a hierarchical model can represent that each participant is assumed to have individual attentional parameters that is drawn from the same distributions but may also varying in trial tor trial to performance when taking repeated measurements are taken.

Although there are different philosophies behind statistical inference, all of them need model assumptions for their inferences (Rodgers, 2010). Like in the introductory quote, data does not speak for themselves but has to be evaluated with respect to intellectual judgements on the matter. Bayesian Statistics is particularly apt when particular models are assumed (Little, 2006) and is able to handle the nonlinearity of many cognitive processes more easily than frequentist methods (Rouder & Lu, 2005). It is important to note that Bayesian methods go along well with the hypothetico-deductive method (Gelman & Shalizi, 2013). Thus, Bayesian methods are used within this work. A further argument pro Bayesian methods is that its properties may be more in line with the behavior expected and desired by psychologists (Dienes, 2011). The reader does not have to adopt a Bayesian approach to epistemology in order to accept Bayesian inference as valid. Furthermore, I find it important to note that all the work presented here could have been done without Bayesian data models. For example, maximum likelihood estimation can be used for parameter estimation. Thus, Bayesian inference is not used as a part of the theoretical model (in contrast to assuming cognitive processes to work according to the principles of Bayesian Inference, e.g., as in the assumption that the brain works according to Bayesian principles Friston, 2012) but merely as a convenient tool for statistical inference.

Chapter 4

Bundesen's Theory of Visual Attention

> Philosophy is written in that great book which ever lies before our eyes — I mean the universe — but we cannot understand it if we do not first learn the language and grasp the symbols, in which it is written. This book is written in the mathematical language, and the symbols are triangles, circles, and other geometrical figures, without whose help it is impossible to comprehend a single word of it; without which one wanders in vain through a dark labyrinth.
>
> Galileo Galilei in The Assayer (1623), as translated by Thomas Salusbury (1661), p. 178, as quoted in The Metaphysical Foundations of Modern Science (2003) by Edwin Arthur Burtt, p. 75.

In his comprehensive book, Bundesen (2008) introduces TVA as an attempt to move from metaphorical explanations of attention to process-based explanations. He compares TVA's approach to the theory by Shiffrin and Schneider (1977), who hypothesize an attentional director. A valid pint of criticism of this approach might be that postulating an intelligent agent as part of explaining another intelligent agent's internal processes merely shifts the problem and in doing so creates a strange loop (Hofstadter, 2007): The explanation of how the attentional director chooses which stimulus to attend ends up right where it started.

TVA, on the other hand, provides a formal mechanistic description of cognitive processes that belong to attention.

TVA by no means discards the view that attention is influenced by an intelligent agent. However, what TVA aims at is identifying mechanisms on which the agent relies to direct its attention. As such, TVA is the result of functional analysis as described by Cummins (2000), see Section 3: Instead of postulating axioms and laws, a complex system is decomposed into smaller, less problematic systems. These interacting simple systems explain different attentional phenomena with high precision. The decomposition can be tracked step by step by reading Bundesen and Habekost's (2008) historical introduction to TVA: The authors explain individual experiments on recognition and selection and how the processes involved were first modeled independently, and finally by a single integrated model that was developed into TVA. This chapter provides a general introduction to TVA and explains current developments in TVA research that are related to the quantification of salience as described in Chapter 6.

The basic TVA model

TVA belongs to the formal theories of attention (Logan, 2004), that is, TVA provides a formal description of cognitive processes involved in visual attention. Nevertheless, it may be helpful to introduce an analogy before the formalism. Imagine a horse race. Such a horse race — like all analogies — has certain aspects that map nicely to TVA's properties whereas others may evoke inadequate ideas about TVA[1]. So, let me point out the properties of a horse race that map nicely to TVA's formal properties. First of all, all horses start at the same point in time. Also, horses may race at different speeds that are independent of the other racers. Although one horse will finish first, there are also prices for the second, third, and maybe even more runners-up, though not for all competitors. Also, there may be favorites in each race that are likely to win, yet before the race actually happens it is not predetermined that they will win, i.e., there is uncertainty about the outcome of each race individually but tendencies over many races.

The analogy fits visual attention if you imagine the horses

[1]Analogical reasoning is in itself an highly interesting topic and is related to both scientific explanation and modeling (Bartha, 2013). Interestingly, Bailer-Jones (2009) provides one of the few distinctions between analogy and model: Whereas analogies establish plausibility, models aim at predicting and explaining. Following this distinction, please note that the horse race analogy is not supposed to model attention — like the attentional director — but merely seeks to establish plausibility of TVA's processes.

to be visual stimuli, all revealed at the same time and for a brief moment. Our visual system will process these stimuli. However, the system has limited resources: It takes a certain time to process a stimulus. This speed is analogous to the speed of an individual racing horse. Also, only a limited number of stimuli can be stored in the visual short-term memory (VSTM). This property is analogous to multiple winners, e.g., gold, silver, and bronze medals with other runners-up not getting a medal, meaning only a few visual stimuli, usually up to four or five, can be reported from VSTM after their brief presentation. Thus, there are two capacity limitations: One for processing stimuli and another one for storing visual stimuli.

The formal equivalent of the analogy is a fixed-capacity independent race model (FIRM). In their seminal work, Shibuya and Bundesen (1988) described this type of processing for TVA. It can be imagined as a fixed capacity for visual processing that is distributed to the stimuli in the visual field. This property does not map well to a horse race because the sum of the horses' speed is not fixed in advance, i.e. one horse may increase speed without another horse becoming slower. This overall visual processing capacity is formalized as a processing rate, C, that describes how many stimuli per second can be processed. Because a rate (units per time) can easily be transformed into speed (time per unit), processing rate and processing speed are sometimes used interchangeably. The processing rate parameter for process i is denoted as v_i. The overall processing capacity of the visual system is equal to the sum of all processing rates $C = \sum_i v_i$.

Figure 4.2 shows how the process can be imagined for two stimuli. The overall capacity, C, is distributed to the stimuli in the visual field. Each stimulus i is processed at its individual rate, v_i. How much of the overall capacity is received by a stimulus depends on its attentional weight, w. The higher the rate v_i, the more likely it is for the stimulus i to finish first and arrive in VSTM. Because VSTM is limited in its memory capacity, stimuli finishing late may not be represented in VSTM and thus be unavailable for higher cognitive processes. It is important to note, however, that the FIRM is based on stochastic processes — meaning a higher rate means higher odds to win the race — but does not determine a particular outcome in each individual situation. This is illustrated by Figure 4.1: With increasing time (x-axis), the chance of being encoded to VSTM increases (y-axis). According to the logic of the stochastic processes, a process with a low rate parameter can sometimes overtake a process with a high parameter.

To fully describe the Figure 4.1, another variable has to be introduced: The maximum ineffective exposure duration, t_0, describes the maximum duration for which a stimulus can be presented without a chance for its processing to finish. Or, roughly speaking, t_0 denotes the time it takes until complete processing starts.

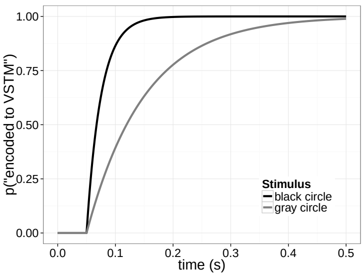


Figure 4.1: Distribution functions, Equation 4.1, of FIRM arrival times for a stimulus processed at a high rate (shown in black) and at a low rate (shown in gray; the same exemplary stimuli as in Figure 4.2).

The distribution of arrival time of stimulus x can thus be written, as shown in Equation 4.1.

$$F(t) = \begin{cases} 1 - e^{-v_x(t-t_0)}, & \text{if } t > t_0 \\ 0, & \text{else} \end{cases} \tag{4.1}$$

Critical thinking reveals that TVA does not model how evidence for a stimulus is accumulated over time as it may happen, e.g., in a neuronal network. It is not possible to, so to speak, look into the race at a particular moment: Only the outcome is modeled in the present FIRM. This is, however, not a fundamental limitation of TVA because TVA does not depend on exponentially distributed arrival times. In the terms of Bailer-Jones' framework, this means that the present FIRM is part of the theoretic model but not a part of the theory. As such, it can be replaced by a different mathematical model and still be consistent with TVA (e.g., for modeling perceptual confusion and blind guessing; and response time modeling, respectively, see Kyllingsbæk, Markussen, & Bundesen, 2012; Blurton, Nielsen, Kyllingsbuk, & Bundesen, 2016). This work limits itself to FIRM, however. Also, I will omit to discuss the neuronal or connectionist version of TVA, the nTVA (Bundesen, Habekost, & Kyllingsbæk, 2005, 2011) — even though the theory is strong in the sense that it makes testable predictions leading to discoveries; (Li, Kozyrev, Kyllingsbæk, et al., 2016) — because this work focuses on symbolic modeling and these aspects, like the evidence accumulation models, assume a more complex theoretical model.

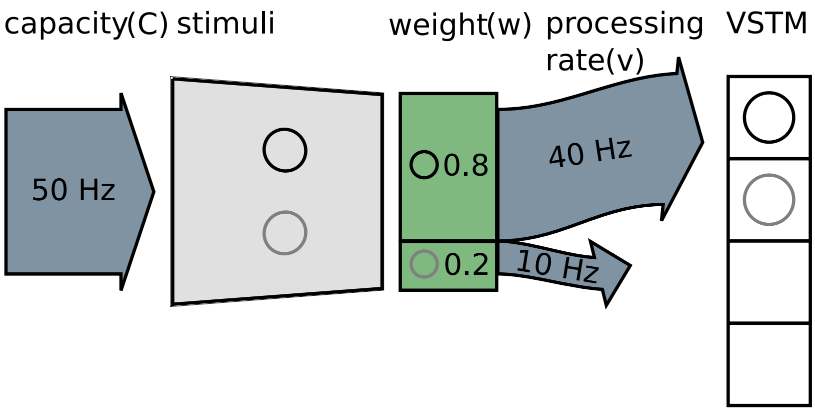


Figure 4.2: Sketch of the fixed-capacity independent race model for visual processing of two stimuli.

So far, I introduced FIRM; however, TVA is about recognition as well as selection. The mechanisms that enable recognition and selection are called *pigeonholing* and *filtering* in TVA literature. The interplay between both mechanisms is what enables TVA to explain a range of experimental findings with high quantitative precision. This interplay is also what determines processing speed and attentional weight of stimuli. TVA's *rate equation*, Equation 4.2, and *weight equation*, Equation 4.3, formally express how both mechanisms interact. The previous example of stimuli racing being represented in VSTM was simplified in that a stimulus does not necessarily have to be encoded correctly. If you imagine handwritten letters, some letters are easier to recognize than others. In TVA, this ease of recognition is expressed by the sensory evidence, $\eta(x, i)$, that stimulus x provides for belonging into category i.

As the *rate equation*, Equation 4.2, shows, the processing rate is further affected by the perceptual decision bias β_i, and the relative attentional weight for the stimulus X. The perceptual decision bias, β_i, describes a bias towards making categorizations of the category i. The fraction containing the attentional weights expresses that the attentional advantage of stimulus x does not only depend on its own effects on attention but on its attentional advantage relative to the set of all stimuli in the visual field S.

$$v(x,i) = \eta(x,i)\beta_i \frac{w_x}{\sum_{z \in S} w_z} \quad (4.2)$$

How the attentional weights are computed is described by TVA's *weight equation*, Equation 4.3. This equation states that the attentional weight for stimulus x, w_x, depends on the sensory evidence, $\eta(x,j)$, which stands for the evidence of x belonging to a particular category j and the pertinence, π_j, of the respective category. Of course, all possible categorizations, R, are considered to calculate the attentional weight. These properties are expressed by the sum. The pertinence can be understood as the task-relevance of a particular category.

$$w_x = \sum_{j \in R} \eta(x,j)\pi_j \quad (4.3)$$

Together, these equations form the basic TVA model. This model is used — together with partial and whole report designs — to estimate attentional parameter in clinical (e.g., Finke et al., 2005) as well as basic research (e.g., Schubert et al., 2015). I call this model basic because many additions have been made in the past 30 years (for a recent review, see Bundesen, Vangkilde, & Petersen, 2015). In the following, I will introduce work concerning salience, and the use of TVA in the temporal-order judgment paradigm.

Salience in TVA

The basic TVA model, as described by Bundesen (1990, 1998), does not explicitly deal with salience. Initially, the phenomena of visual contrasts attracting attention could only be modeled in terms of relevant features, that is feature-based attention: The contrast must be modeled as part of a categorization and the respective category must be assigned a pertinence value larger than 0. Roughly speaking, salience is a perceptual category that is in some way important for the task. Conceptualization of the impact of salience is necessary because an attentional weight could only be attained through the evidence for categories that are relevant. However, physical contrasts that are not directly task-relevant are known to affect attention and performance in experiments as well[2] . This phenomenon has also been researched with TVA formalisms and paradigms.

[2]The ongoing discussion on the nature of attentional capture is related (e.g., Theeuwes, 2013; Folk, Remington, & Johnston, 1992).

In this pursuit, Nordfang, Dyrholm, and Bundesen (2013) introduced a salience manipulation in the partial report design often used with the basic TVA model. Their design extends the partial-report design: Letters and digits were shown, only letters ought to be reported after masking. Masking was done after different intervals which is necessary to estimate the processing rate in this design. Nordfang et al. (2013) added a task-irrelevant color to one of the elements. This is not trivial because the color must not affect sensory evidence for the identity of the respective element, see Chapter 6 for details and an example. After careful experimentation, they introduced a new parameter to TVA to capture the stimulus-driven influence of visual attention caused by the color. The new parameter is named κ. This parameter is supposed to capture the influence of contrasts on attention that should have a pertinence value of 0 but affect attention in experiments nonetheless. The resulting new version of the weight equation is shown in Equation 4.4. Interestingly, Nordfang et al. (2013) found salience to multiply with the task-relevance based influences from the potential category memberships[3].

[3]To be completely frank, I only became aware of this work after I had tried to capture salience in TVA's attentional weight, w, which is the more general attentional weight. This approach is reported in Section 6.

$$w_x = \kappa_x \sum_{j \in R} \eta(x, j)\pi_j \tag{4.4}$$

TVA-TOJ

TVA is a theory that aims to describe visual attention in general. Although it is usually used with the whole and partial report experimental design (e.g., Nordfang et al., 2013), it is also applied to model other experimental paradigms like visual search (Logan, 1996), attentional dwell time (Petersen, Kyllingsbæk, & Bundesen, 2013), and more recently the TOJ (Tünnermann, Petersen, & Scharlau, 2015).

The TOJ design is particularly interesting because it features a simple decision task comparable to the design by Nothdurft (2000) and a model that is less complex than the model by Logan (1996) and even less complex than basic

TVA[4]. It features two events that have to be judged according to their temporal order. These events are usually called probe, p, and reference, r, and usually refer to the onset of stimuli (although other events are possible: for offsets, see, e.g., Vingilis-Jaremko, Ferber, & Pratt, 2008). Both events are separated by the stimulus onset asynchrony (SOA). The closer the SOA is to 0, the more uncertainty in the judgment is to be expected, which is indicated by the performance being close to guessing probability. However, if attention is directed either goal-directed or stimulus-driven to the probe stimulus, this point of maximum uncertainty shifts. This is interpreted as a shift of subjective simultaneity. The point of subjective simultaneity (PSS) is shifted so that the reference has to be shown objectively earlier to create the impression of subjective simultaneity. This phenomenon could be caused either by a speed-up of the attended probe, a slow-down of the unattended reference, or both. If TVA is adopted as a model, TVA provides a clear prediction about the cause: It is both influences, speed-up of attended and slow-down of unattended stimulus, because the attentional weight is shifted, resulting in a redistribution of speed. This has indeed been confirmed by research independent of TVA (Weiß & Scharlau, 2011).

[4]Because it is simpler, it does not allow to measure all TVA parameters. Thus, simplicity does come at a cost.

The PSS is, however, not immediately obvious but estimated from the data. To estimate the PSS, a sigmoid ("S"-shaped) function is assumed and fitted to the data. See Wichmann and Hill (2001) for a non-Bayesian approach and Kuss, Jäkel, and Wichmann (2005) for a Bayesian approach to this estimation. As explained in both articles, there are multiple functions that can be used to estimate the PSS. A reason why multiple functions can be used is that all of them do a reasonably good job at describing the data. However, they are not derived from a psychological theory — that is, all functions are descriptive models, not logical models. This comes at the cost that their parameters, apart from the PSS itself, may be difficult to interpret.

Another approach is to derive a function from theory that is focused on the explanation of TOJ: Tünnermann et al. (2015) use TVA for an explanation of TOJ: A specific formal TOJ model is derived from the more general formal description of attention as proposed by TVA. The central idea is that the function that describes the data collected from the judgments is a consequence of the visual processing rates of both probe stimulus, p, and reference stimulus, r. By shifting attentional weight from one stimulus to the other, the overall fixed visual processing capacity, C, is distributed differently, causing the different processing speeds. While plausible, when expressed verbally, it is not obvious how to apply TVA in such a way. Tünnermann et al. (2015) described the function for judging "probe first", P_p, based on the processing rate for probe, v_p, reference, v_r, and Δt (corresponding to SOA) as in Equation 4.5. It holds that $\Delta t = \text{SOA} + t_0^p - t_0^r$, where t_0^p and t_0^r are the maximum ineffective exposure durations for the two stimuli.

$$P_p(v_p, v_r, \Delta t) = \begin{cases} 1 - e^{v_p|\Delta t|} + e^{v_p|\Delta t|}\left(\frac{v_p}{v_p+v_r}\right), & \text{for } \Delta t < 0 \\ e^{v_r|\Delta t|}\left(\frac{v_p}{v_p+v_r}\right), & \text{for } \Delta t \geq 0 \end{cases} \tag{4.5}$$

The term $1-e^{-v_p|\Delta t|}$ represents the probability that probe stimulus, p, is encoded before reference stimulus, r, and begins the race to VSTM. The term $e^{v_p|\Delta t|}$ represents the probability that probe stimulus, p, is not encoded before the reference stimulus, r, starts its race. Then, the probability of encoding probe stimulus, p, is first given by Luce's choice axiom $\frac{v_p}{v_p+v_r} = \int_0^\infty v_p e^{-v_p t} \cdot e^{-v_r t} \mathrm{d}t$.

This formal model allows applying TVA to TOJ-based designs featuring different stimulus material. This approach has been described in detail in Tünnermann, Krüger, and Scharlau (2017) for different TOJ designs.

Summing up, TVA provides a formal theory in which a few parameters and fixed processes yield a good fit for data of multiple experimental designs involving attention. TVA can be used to provide explicit theoretic models like the TVA-based TOJ model as described by Bailer-Jones (2009) and thus may provide good conditions for an explicit and tight link between observed data and theoretical explanation.

Chapter 5

Bayesian inference

> How often have I said to you that when you have eliminated the impossible, whatever remains, however improbable, must be the truth?
>
> (Doyle, 1890, p. 111)

Bayesian statistics are used in all the articles included in the present dissertation. This chapter introduces key concepts of Bayesian statistics and also examines the reasons for using it. Bayesian statistics is actually a subcategory of Bayesian inference. Bayesian inference as a term does not give away any property of what is meant by the term; there is another term that captures the whole idea beautifully: *inverse probability*[1]. Bayesian inference — or inverse probability — describes how you can reason from effect to cause, if you know how the cause evokes the effect.

[1]a term that can be tracked back even further than 1800 to Laplace (Stigler et al., 1986)

In the first edition of his introductory book, Kruschke (2010, Chapter 4.3, p. 63) clearly formulates three things that can be done using Bayesian statistics: estimate parameters, compare models, and predict new data. I found this presentation helpful to understand that a particular Bayesian analysis can pursue either goal. In the second edition (Kruschke, 2014, Chapter 2.1, p. 16), Bayesian analysis is presented more generally as the reallocation of credibility across possibilities much like the famous Sherlock Holmes quote at the beginning of this chapter.

In the accessible and highly recommendable introduction by Kruschke (2014), coin tosses are used as an example for hundreds of pages to illustrate the principles of Bayesian inference (see Chapter 5.3 for a complete example).

Assume, as Kruschke (2014) does, for example, that we want to estimate the bias of a coin. For this estimation, we need the *likelihood function*. The likelihood function expresses how the bias causes the results of coin tosses. This direction from cause to effect will later be reverted in the estimation. The likelihood function expresses our modeling assumptions in a straightforward way. For the estimation of the bias from observations, the observations — also data or evidence — is needed as the second component[2]. In this example, the evidence is a sequence of heads and tails yielded from tossing the coin. The third and final component to a Bayesian inference is the *prior*. The prior is an initial distribution of credibility of parameter values — it is a probability distribution. For example, you may assume that most coins are fair and that the extreme biases are less likely than small or moderate biases. Thus, the prior is symmetrical around the parameter value of an unbiased coin. Choosing a good prior is a topic on its own (see, e.g., Vanpaemel, 2011). Importantly, if you have the three components, likelihood function, evidence, and prior, Bayesian inference yields a *posterior*. Like the prior, the posterior is a probability distribution. The posterior is an updated version of the prior and corresponds to a rational reallocation of credibility after observing the data. In the example, one may observe more tails than heads, leading to a shift of the prior towards the respective bias.

[2]With the two components, the analyst can already perform maximum likelihood estimation which is almost a Bayesian inference: The maximum likelihood estimation is performed by the mathematical equivalent of asking which parameter value is the most likely to have produced the outcome.

Reading results from Bayesian statistics is more intuitive than reading frequentist results (e.g., Oakes, 1986, Chapter 3). A prominent difference is that parameter estimations are reported not as point estimated, i.e., the mean is 4.2 but rather as probability distributions. These distributions do not only show the most likely value at its highest point but also quantify the uncertainty in the estimation: The more uncertain, the more the distribution is spread out. Two central concepts in reporting parameter estimations are the *maximum a posteriori probability* (MAP) and the *highest density interval* (HDI) or credibility interval. The MAP corresponds to the highest point in the probability distribution and the HDI usually corresponds to the interval in which 95 % of parameter values are concentrated.

Bayesian statistics can be imagined as using an inference machine: In goes a model (determining the likelihood), a prior, and evidence; out comes the posterior. The posterior integrates prior belief with evidence according to the model. This inference works according to the Bayes theorem. If we call the data, D, the model, M, and the parameters we are interested in θ, one could write (Kruschke, 2013, p. 48) :

$$\underbrace{p(\theta|D,M)}_{\text{posterior}} = \underbrace{p(D|\theta,M)}_{\text{likelihood}}\underbrace{p(\theta|M)}_{\text{prior}} / \underbrace{p(D|M)}_{\text{evidence}} \tag{5.1}$$

where the evidence is

$$p(D|M) = \int d\theta p(D|\theta,M)p(\theta|M). \tag{5.2}$$

For a thorough introduction into how to do Bayesian statistics, see Kruschke (2013). Usually, when introducing this formal relationship, the model M is omitted except for when models ought to be compared. The model, M, however, determines the likelihood function and as there are neither right nor wrong models, it is valuable to stress that each inference depends on the used model.

In contrast to typical frequentist methods, the integral in Equation 5.2 is not guaranteed to have a solution and thus may be impossible to compute in closed-form. Even if it has a solution, it can be very hard to find depending on the used model. This is also the reason why the application of Bayesian statistics has been extremely hard for models that are interesting in actual empirical research. However, progress in algorithmic approximation allows to simply approximate this integral even for complex models on an average personal computer (for a historical overview, see, e.g., Leonard, 2014; Fienberg, 2006). The algorithms used to compute the posterior are guaranteed to converge to the right solution. However, how can the user know if the so-called parallel chains have converged? Usually, the approximation is executed multiple times in parallel. If all solutions are similar, the algorithms likely reached convergence.

The models used in Bayesian statistics have in common that they all can be represented qualitatively or structurally by directed acyclic graphs. Besides this structure, there is a second quantitative component; a conditional probability distribution that determines how a node in the graph depends on its predecessors. The notation of the models used in Bayesian estimations varies across literature. The notation in this work is taken from Lee and Wagenmakers (2014).

Given the evidence, e.g., from a behavioral experiment, one may ask which of two theoretically plausible models is better supported by the data. Interestingly, this can be solved by parameter estimation, that is, in the exact way that has already been described in this section. See Kruschke (2013) for a detailed explanation of how to use this approach to draw inference analogously to the t-Test. Another approach is to compute the Bayes Factor (Rouder et al., 2009; Wagenmakers et al., 2010). The Bayes factor quantifies the support of one model over the other. The Bayes Factor can, however, be difficult to compute for arbitrary models. An alternative presents itself in the deviance information criterion (Spiegelhalter, Best, Carlin, & Van Der Linde, 2002) and widely applicable information criterion (Watanabe, 2010). These information criteria can be computed for each model and evidence individually and compared afterward.

Predicting new data can be done in many ways. For example, one could start with merely a prior and a model to compute which data values would follow if the prior is adequate. In this work, however, the prediction will always be made in the form of *posterior predictive* distributions. Posterior predictive distributions provide an estimate of possible unobserved values based on the values that were observed. This, for example, allows estimating the response behavior for unobserved experimental conditions like additional SOAs in the TOJ design.

A recent review of the usage of Bayesian methods in psychology is provided by Van de Schoot, Winter, Ryan, Zondervan-Zwijnenburg, and Depaoli (2017). Bayesian inference was introduced in Section 3 to bridge the gap between statistical inference and scientific inference (see also, Krüger et al., 2018) and Van de Schoot et al. (2017)

review its use in the published psychological literature. Around half of the reviewed publications use Bayesian inference to cope with statistical regression problems; the other half uses Bayesian inference as part of a theory or in a computational model. Together, these two applications account for more than 85 % of the reviewed articles.

The articles included in this dissertation fall under the regression-based category because Bayesian inference is not part of the theoretical explanation but used as a statistical tool. Van de Schoot et al. (2017) summarize the reasons for choosing Bayesian statistical inference: With the use of complex models, researchers report that they are practically forced to use Bayesian methods because other solutions do not exist or are comparably difficult to implement. Further reasons include, but are not limited to, the quality of parameter estimations, the fact that Bayesian methods are used when assumptions of NHST alternatives are violated, greater modeling flexibility, handling of missing data, the possibility of including prior knowledge into the analysis, available tools for model selection and better performance for small sample sizes. This multitude of reasons shows that there are many practical differences between Bayesian inference is frequentist inference.

For the present work, the *ease of modeling* was an early reason for adopting Bayesian inference. It made the implementation of the formulae by Tünnermann et al. (2015) comparably easy. Also, further in my research, different theoretical models could be formalized easily into graphical models (Krüger, Tünnermann, & Scharlau, 2017). Directly connected to this point are the different tools for *model selection*: Once the model has been specified, different information criteria, the Bayes factor, or a parameter estimation can be used to compare different models. Because the models are complex technical problems in one of the comparison techniques can be overcome by resorting to one of the alternatives.

One example for the ease of modeling is *hierarchical models*: Bayesian Hierarchical Graphical models allow to model the sources of uncertainty according to the experimental design. For example, in the literature, TVA parameters are often computed per participant by means of maximum likelihood estimations. The resulting parameters are then used in NHST. However, the parameter estimates are associated with different degrees of certainty. These degrees of certainty are included in the Bayesian Posterior. Modeling this as a hierarchical model involves straightforward assuming the parameters of the posterior distribution are drawn for a shared population distribution. Thus, it is possible that the source of uncertainty between trials of an individual participant is a different source of uncertainty than the one causing inter-individual differences.

Another major reason for choosing Bayesian methods is that their results are more in line with the questions that we are interested in (Dienes, 2011). Rather than providing the probability of data under assumed parameter values, Bayesian inference provides the probability parameters under fixed data. This allows for estimations of the relative probability of competing theories given some observations.

The ability to predict new data for unobserved cases was not among the primary decision criteria. However, being able to predict data with an implemented model and theoretical assumptions, made the models connect to the data in yet another way and offers a plausibility check: Does the developed model reproduce all characteristics of the data (for the investigation of whether a plateau should be assumed in TOJ modeling, see Tünnermann & Scharlau, 2018b).

Bayesian statistics are overall advantageous for working with models; however, there are limitations. Although it is in principle possible to build highly complex models, these models have to be handled by an appropriate inference algorithm. These algorithms work according to the Markov chain Monte Carlo (MCMC) technique implemented in software packages like jags (Plummer, 2003) () or pymc3 (Salvatier, Wiecki, & Fonnesbeck, 2016). Consequentially, the analyst needs to be a programmer with knowledge of the estimation process in order to choose a package and implement and run the model successfully.

Another counterargument is that the models and priors may differ between researchers causing even more freedom for the researcher that may be problematic for reproducibility (Simmons, Nelson, & Simonsohn, 2011). Thus, "objectivity"[3] may not be warranted in the sense as it is often suggested by frequentist toolbox methods. This may, however, be advantageous because it requires statistical thinking as called for by Gigerenzer (2018). The advantages of "subjectivity" in data analysis are discussed by Rouder, Morey, and Wagenmakers (2016).

[3] As described by Hacking (2001), "subjective" and "objective" are terms that can easily distract from fruitful discussion of inductive logic and lead to empty polemics.

If the advantages of hierarchical Bayesian statistics have to be expressed by a single argument, it would be the ease of modeling the data according to intellectual judgments about the domain.

Subjectivity versus objectivity?

It has been argued that Bayesian statistics it is more subjective than NHST (e.g., Simmons et al., 2011). Subjectivity is clearly a problem that science has to deal with. Thus, increasing subjectivity may naively sound so undesirable that it may cause a reluctance to learn the Bayesian perspective. In their article on subjectivity in data analysis, Rouder, Morey, and Wagenmakers (2016) succinctly stated that Bayesian methods require "an overhaul of the relations between models, data, and evidence." (p. 1). Further, they argue that an increase in subjectivity is desirable in scientific data analysis. Consequentially, their argument seems to support the superficial counterargument that Bayesian methods are more subjective. While examining their argument, the reader will notice that Rouder, Morey, and Wagenmakers (2016) propose to transfer a part of the verbal scientific discourse into data analysis by spelling out one's hypothesis by modeling and exposing it to criticism by putting forth the respective idea in a more falsifiable way. So, it could also be argued that it increases the overall objectivity in science.

This and similar arguments rose the suspicion in me that "subjective" and "objective" cause more obfuscation than clarification in debates involving Bayesian inference. This obfuscation is caused by the difficulty to actually assess what is objective. As scientists, we should be aware of these epistemological difficulties all too well. The problem with this type of "argument" has been described most vividly by Hacking (2001):

> How often have you heard this sort of conversation?
>
> James: That's just your subjective opinion.
>
> Mary: Nonsense, it is an objective fact.
>
> How often have *you* talked just like that?
>
> Don't get into that rut, in philosophy or in the rest of your life. *James* and *Mary* are not arguing, they are just slinging mud at each other. (p. 131)

He further explains that both Bayesian and frequentist conceptualizations of probability can be called objective but for other reasons: Whereas frequentist thinking is about a state of the world, Bayesian reasoning is objective because of a logical relationship between evidence and proposition. Certainly, it is possible to take the stance that science is about the world, and thus only the frequentist approach provides the right type of objective inference. However, science is to a similar degree about propositions that ought to explain and predict the world. Bayesian reasoning provides an objective link between propositions and evidence. Thus, a case for the need for Bayesian inference can be made as well. It is important to note that this is not a practical debate but depends on one's stance towards the philosophy of science.

Popper's falsification is the basis for scientific inference in psychology (Dienes, 2008). However, practically, psychologists rather provide evidence for their own often vague theories by rejecting a theoretically uninteresting null-hypothesis. Even introductory works on the philosophy of science show what makes falsification problematic. In fact, after the first half of his excellent book, Chalmers (2013) explains in Chapter 7 that many theories would not have come into existence if falsification were actually followed as a paradigm historically. The lack of strong theories (Muthukrishna & Henrich, 2019) and overarching theoretical frameworks (Fitzgerald &

Whitaker, 2010b) may partially be connected to this stance towards the philosophy of science.

What does the philosophy of science have to do with Bayesian statistics? As Dienes (2011) noted, there are two definitions for rationality: One is to justify one's beliefs sufficiently; The other one is to have subjected one's belief to critical scrutiny. Because of psychology's foundation from the philosophy of science, it may be difficult to appreciate the rationality and objectivity that comes with the justification of beliefs. Thus, I argue, that Bayesian statistics are not more subjective but make the justification part of rationality — a part that is usually done vaguely and verbally in psychology — part of the statistical analysis. Consequentially, it is not a question of subjectivity but what inference the scientist wants to make. Do they focus on a state of the world or on evidence for a proposition?

Chapter 6

A quantification of visual salience

> Statistical inference techniques are good for what they were developed for, mostly making decisions about the probable success of agriculture, industrial, and drug interventions, but they are not especially appropriate to scientific inference which, in the final analysis, is trying to model what is going on, not merely to decide if one variable affects another.
>
> R. Duncan Luce (Taagepera, 2008, p. v)

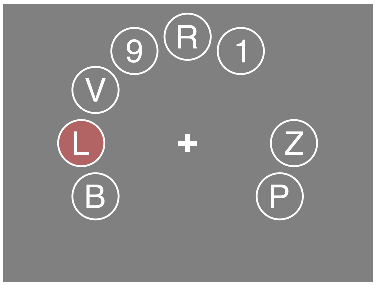
In this chapter, I use the concepts from the previous chapters to propose a measure of salience. This approach combines multi-element displays, temporal-order judgment (TOJ), and the Bundesen's Theory of Visual Attention (TVA) to provide a theoretically justified measure for visual salience that can capture the different aspects of salience including that salience arises from different contrast dimensions, that it follows a time course, and that goal-directed influences are related to it. In this pursuit, there will be different terms that should be clear from the start of this chapter: Salience *quantification* means any form of quantitative representation of the impact of physical contrasts on attention. By salience *measure*, I mean a quantification of salience as well as the means to actually estimate its value from data. Salience *manipulation* means a change in the physical contrast between a stimulus and its surrounding.

Rationale

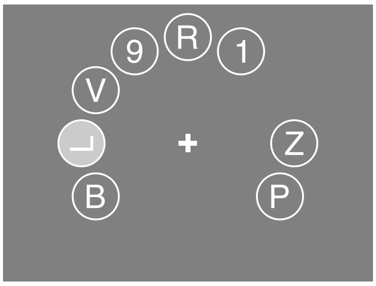
Recalling that TVA was extended by a parameter representing salience called κ, the reader may wonder why not to use the same experimental setup as the one developed by Nordfang et al. (2013). After all, the authors were able to estimate a value describing how strongly a non-task-relevant physical contrast affected visual attention. This is exactly what I mean by salience measure. However, Nordfang et al. (2013) used only a single salience manipulation. Namely, one stimulus was presented against a red background, whereas the other letter's backgrounds were gray. Even this single and minor manipulation of salience had to be tested carefully in pre-tests to ensure that it does not affect the sensory evidence of the stimulus for being a particular letter.

For manipulations of other contrasts dimensions like rotation or luminance, it is rather obvious that they will affect the evidence of the stimulus in the letter report designs usually used with TVA. For example, the evidence for the stimulus "L", x_L, for being the particular letter L that is the sensory evidence, $\eta(x_L, L)$: Luminance contrast will change the visibility of the letter whereas the rotation deviates from the canonical orientation of letters such that it is much harder to recognize. This problem also illustrates the point why the stimulus x_L and its identity as the letter L are distinguished in TVA: A letter "L" in bad handwriting may be considerably more difficult to identify. Thus, even if the letter "L" is highly task-relevant it may not receive much attention because it is simply unclear whether it belongs to the relevant category. Consequentially, salience manipulations must not affect the sensory evidence because it would be unclear whether a change in attentional weight must be attributed to sensory evidence or visual salience. Whereas a color salience manipulation can be chosen so that it does not affect sensory evidence in this design, it may be difficult to compare its salience value to other colors for which it is not guaranteed that they do not affect sensory evidence. For other contrast dimensions like luminance and orientation contrast, it is clear that the more contrast, the more the sensory evidence is affected and thus, the salience measure is confounded by the sensory evidence.

Figure 6.1 shows a sketch of the original displays used by Nordfang et al. (2013) and a (hypothetical) version with luminance and orientation contrast manipulation to illustrate the described problems. Consequentially, this work is a theoretical foundation, but does not provide a broadly applicable salience measure.



(a) Stimulus material by Nordfang et al. (2013).



(b) Stimulus material with hypothetical orientation and luminance contrast.

Figure 6.1: The salience measure by Nordfang et al. (2013) allows only to measure salience caused by contrast that does not affect the sensory evidence for the stimulus identity.

To come up with a potential solution, it is helpful to examine what happens according to TVA if a salience manipulation is added to the partial report design. TVA cannot describe attention if the visual stimuli change dynamically over different episodes[1]. Instead, TVA assumes a stimulus material that is presented at a certain point in time and stays constant until it is masked. Depending on the time of masking, the visual system is more or less able to process the presented stimuli. If the presentation duration is long, the visual short-term memory (VSTM) can store up to four or five of the presented stimulus identities. The encoding process is modeled rather simply because the chance of success is modeled by a hazard rate, as shown in Equation 4.1 — as mentioned in Chapter 4, models for evidence accumulation in TVA exist but are far more complex. In this example, the stimulus "L" is associated with a processing speed to be encoded as an instance of the letter L, $v(x_L, L)$ according to TVA's rate equation, Equation 4.2.

Please note that it is not impossible to encode x_L as something other than its true identity. For example, the stimulus x_L could erroneously be encoded as the letter "I". Thus, one stimulus may be associated with multiple processing speeds, in the example, x_L may either be correctly encoded according to rate $v(x_L, L)$ or confused with the letter "I" according to usually a much smaller rate $v(x_L, I)$. If a stimulus does not have a clear membership in one of the task-relevant categories and such confusions are possible, TVA becomes very difficult to use. This problem is the reason why overlearned stimuli such as letters and digits are used with TVA. Also, possible confusions are the reason why some letters are excluded from the stimulus material. The point of this short explanation is to justify that, in the fol-

[1] Although Schneider (2013) proposes how TVA may be extended to cover different episodes, TVA's formalism has been extended to attentional dwell time (Petersen et al., 2013) but does not cover temporal episodes in general.

lowing, it is — as in the usual TVA experiments — assumed that confusion is negligible (for an example how confusions can be modeled and may even explain phenomena, see Tünnermann & Scharlau, 2018a). This is important because it implies that a stimulus, for example, x_L, is only associated with precisely one processing rate for encoding it correctly which is written as $v_L = v(x_L, L)$ for the sake of brevity.

What determines the processing rate,v_L, of a stimulus x_L? The processing rate determines whether or not a stimulus arrives in VSTM for later recollection. Each stimulus is associated with precisely one processing rate such that this process may be envisioned as a race like it was presented in Chapter 4. However, the processing rate depends on the attentional weights, w, because it is assumed that the overall processing capacity, the sum of all processing rates, stays constant. Thus, what happens is a redistribution of the fixed overall processing capacity according to the attentional weights. The attentional weights are determined based on the weight equation, Equation 4.3. The central point is that the actual encoding of the stimulus and the computation of the attentional weights can be understood as two waves of processing (for an illustration, see Tünnermann et al., 2015). In the first wave, the attentional weights are determined; in the second wave, the stimuli race towards the VSTM. The second wave is not problematic for salience measurement at all. However, an experimental design for quantifying salience in terms of TVA's parameters would have to ensure that the first wave of processing is only affected by manipulating salience.

Summing up, there are two waves of processing of which only the first causes problems for a salience measurement in the designs typically used with TVA. A solution would have to make sure that salience manipulations do not affect the recognition of a stimulus.

Typical salience experiments work with cluttered multi-element displays. The justification for using this kind of stimulus material stems from the assumed underlying processing (as in, e.g., Li, 2002; Koch & Ullman, 1985). Such a display is shown in Figure 6.2. The figure illustrates how a difference in color hue, luminance, and orientation can

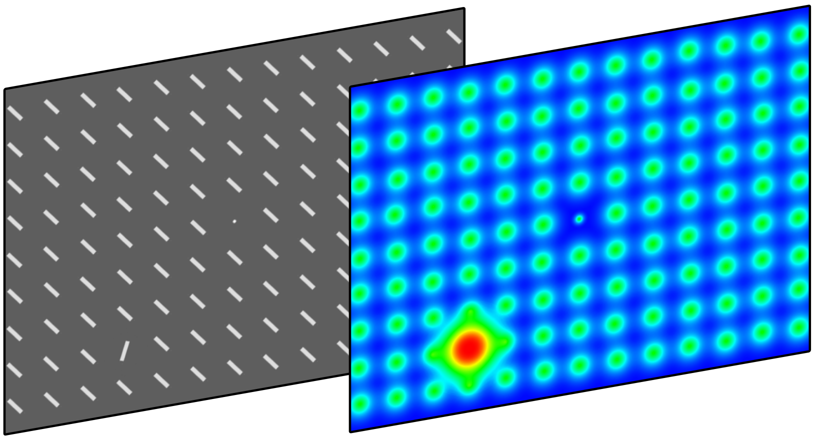


Figure 6.2: Sketch of a multi-element display used to produce a particular physical contrast that can direct attention and the hypothetical stimulus-driven distribution of attention after TVA's first wave of processing.

cause a bar to stand out (upper right corner) whereas an identical bar (lower left corner) is much harder to spot. Showing such a stimulus display will cause a distribution of attention. For example, this has been used by Donk and Soesman (2010) in a design similar to Posner cueing. A display of bars was shown to the participants. Four fixed positions were used in a second step to "probe" attention. In this step, the original bars were masked by star composed of multiple bars. One of these stars was rotating and was a probe for which the participant had to indicate its location by a corresponding button press — top left, top right, bottom left, bottom right. Responses were fast when the probe was shown at a location previously occupied by a salient element and slow when the location was not previously occupied by a salient bar.

Figure 6.2 sketches how salience displays are able to direct attention as described by the first wave of processing in TVA. However, to probe the strength of salience at the location of the unique element, there has to be some sort of task such that observable behavior allows estimating the unobservable strength of salience. So, firstly, this task must have an interpretation in the TVA framework. Secondly, the task will influence attention by the task-relevance of certain stimuli. TVA is very specific about these influences and how they contribute to the attentional weight.

Different tasks have been used in the past for salience measurement: In probe detection (e.g., Donk & Soesman, 2010) a TVA-based interpretation is difficult. Although encoding the probe as the category "probe" may exactly be what a TVA processing rate v_{probe} should stand for, the actual response time is not modeled by TVA. Speeded responses include a motor component that TVA does not model. Also, TVA does not predict the actual time needed for an encoding process but the probability that it happens at a particular point in time. I do not want to discuss in-depth whether such a probe detection experiment can be modeled by TVA because there are developments in these directions (Blurton et al., 2016, e.g.,). However, the resulting model would be far from simple and include a response time, which is not part of the TVA formulation used in usual TVA-based research.

Also, search-based designs, which are common in studies of salience, have been interpreted in terms of TVA (Logan, 1996). This TVA-based search model is, however, quite complex and more difficult to relate to salience compared to the work by Nordfang et al. (2013), Nordfang, Staugaard, and Bundesen (2017).

Accuracy-based designs in attention research are more promising candidates for a TVA interpretation. Attended stimuli are perceived earlier than unattended stimuli. This conjecture has a long history in psychology (see, e.g., Titchener, 1908, p. 251) and has more recently been backed up by a wide range of experimental findings subsumed under the term *prior entry*, as introduced in Chapter 3. This is fitting to TVA in so far as more attentional weight, w, should increase the rate of processing, v, such that a stimulus finishes its race towards VSTM quicker. This would mean that the attended stimulus has a processing speed advantage on average when compared to an otherwise similar but unattended stimulus.

TOJ studies have been used to research salience but applied modeling assumptions different from the TVA-based TOJ model introduced in Section 4. The TOJ has been used for salience measurement with the multi-element displays (Donk & Soesman, 2011). However, it is not obvious how TOJ responses can be used to infer the strength of salience.

Donk and Soesman (2011), for example, used an ANOVA analysis and conceptualized their two conditions and five SOAs as a 2×5 within-design. It was assumed that the strength of salience is different in both conditions, and indeed the ANOVA revealed a main effect of experimental condition. This is, however, not a model that expresses a particular theoretical idea of what salience changes do to the data. So, it is a descriptive model in Taagepera's (2008) terms.

Descriptive models lack an explicit link to the theory regarding the phenomenon but may be useful nonetheless. However, using an ANOVA to analyse TOJ data raises questions concerning the quality of the provided description: ANOVA assumes, roughly speaking, that the difference between levels of SOA is roughly the same and has the same variance within each level. However, a large SOA will result in an easy decision; The corresponding data cannot vary beyond always choosing the right order for the observed events. If the case becomes more difficult because of SOAs closer to 0, a different pattern of variance is to be expected logically. Also, the model assumptions underlying the ANOVA analysis do not allow to conclude whether there was no attentional difference in one condition and an attentional difference in the other or an attentional difference in both conditions. In the latter case, a significant main effect of the experimental condition means simply that salience did not have equal effects on both conditions, yet two different effects of salience are possible which is not discussed by the authors.

A way to improve the analysis of TOJs and move closer to measure the impact of the salience manipulation quantitatively is to determine the SOA at which both stimuli would have been subjectively perceived to appear simultaneously. For example, if — hypothetically — one and a half time as much attention is directed towards the probe stimulus, p compared to the reference stimulus, r, the reference r has to be shown 20 ms earlier to create the subjective impression of both stimuli to appear simultaneous. This difference in SOA that would cause subjective simultaneity is called *point of subjective simultaneity* (PSS). TOJ designs applying the method of constant stimuli cannot include this SOA in advance. However, if a relationship between SOA and responses is assumed in the form of a mathematical function, the data can be used to estimate said function. Once, this function has been estimated, the PSS becomes known. The size of PSS can be used to approximate the influence of attention. The actual value for the PSS estimation depends, however, on the assumed function linking SOA to responsible behavior. This function is usually chosen so that its shape resembles the expected data patterns. Naturally, there exists an indefinite number of functions that can resemble a particular pattern. Moreover, the PSS is not the only degree of freedom because the data show that the slope of the underlying function changes. This parameter is called *difference limen* (DL) and is much harder to interpret (for details on functions and their estimation, see Kuss et al., 2005; Wichmann & Hill, 2001). In terms of Taagepera (2008), these functions are descriptive because they are chosen depending on their ability to describe the data but do not necessarily correspond to a theory on the cognitive processes causing the TOJ[2].

[2]One may argue that the PSS is a connection to theory, thus if the model provides a respective parameter it should be called a logical model. However, the actual shape of the function is not derived from theory. Thus, I call the models descriptive.

Tünnermann et al. (2015) provide a mathematical model of how TVA's second wave of processing can cause prior entry in the TOJ design. The mathematical model's details are described in Section 4. Here, it is important that the TOJ can be interpreted as this second wave of processing according to Bundesen's TVA. What remains open is whether a salience manipulation can be applied during the first wave of processing to quantify the causes change in the theoretical attention parameters of TVA.

For a salience measure, a TVA interpretation of a TOJ design looks quite promising. This combination would allow using multi-element displays to manipulate salience while this design would also allow inferring a formal attention parameter expressing particularly the effect of attention on visual selection. In contrast to earlier studies (Donk & Soesman, 2011), the result would neither be an effect size related to the dependent measure nor a quantitative measure like the PSS whose relation to salience is not explicitly spelled out but the amount of attention according to TVA's selection mechanism. The TVA-TOJ formal model would link the observed data from the prior-entry phenomenon to a formal theory of attention and thus provide a psychological explanation of how a particular salience value is connected to a particular pattern of responses (Krüger et al., 2018). However, being so explicit about how to interpret TOJs comes at a cost: The overall processing capacity should remain constant. Otherwise the attentional weight is surely interpretable, but we would have to admit that the salience manipulations has effects beyond a re-weighting of attention in TVA. Also, according to the weight equation by (Nordfang et al., 2013), the sensory evidence and pertinence have to be kept constant in the TOJ design. Whether his approach is feasible is explored in the first article of this cumulative dissertation.

Summary of Article 1

The first published article, henceforth called Article 1, investigates whether the parameter, w_p, of the TVA-TOJ model can actually be used to measure salience in multi-element displays. All in all, three experiments have been conducted to test different experimental manipulations. A fourth experiment was conducted to test the successful experimental manipulation with another salience dimension. The difference between the first three experiments lies in how to conduct the TOJ. The three designs use onsets, offsets, and flicker events, all of which had to be judged according to their temporal order. Each experiment comprises two experimental conditions: In the neutral condition probe and reference have the same orientation as the other elements in the display; in the salient condition, the probe's orientation differs by 90°. In both conditions, the experiments yield an estimation of two parameters for the population: Overall processing capacity, C, and attentional weight, w_p, for the probe stimulus p(the attentional weight for the reference, r, is necessary $w_r = 1 - w_p$ because only two task-relevant stimuli exist in the design). As explained in the previous section, the overall processing capacity, C, is expected to be unaffected by a salience manipulation.

In the neutral condition, the attentional advantage of the probe, w_p, should be $.5$ because this corresponds to no attentional advantage for either stimulus. The experimental condition comprises a reference stimulus that, again, is orientated as the other elements of the display. The probe stimulus has a maximal orientation contrast of 90° should show a $w_p > .5$ because there should be an attentional advantage for the probe stimulus if the salience measurement is possible like it was derived from theory.

The anticipation of an exact value (for C and w_p in the neutral condition) is an example for Taagepera's (2008) critique of the social science: He argues that physics usually expects a particular value and is interested in deviations whereas social science usually sets up a no-difference hypothesis and is interested in any deviation. This is a notable difference from typical psychological experiments because here we assume particular values based on the logic of the underlying theory as Taagepera proposes.

Experiment 1

The first experiment tested whether the onset of probe and reference is a suitable event for the TOJ to measure the attentional advantage of the probe stimulus over the reference stimulus. To this end, a multi-element display is presented first. After an initial interval, the probe and reference are shown with an abrupt onset.

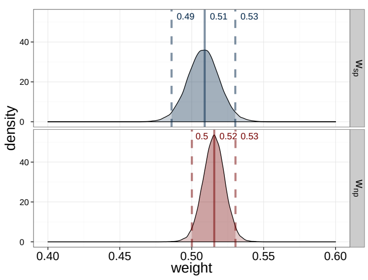


Figure 6.3: Results for the attentional weight estimation of Experiment 1 of Article 1. The salient probe's attentional weight, w_{sp}, does not differ from the weight in the neutral condition where the probe was not salient.

The results show that the attentional weight of the salient probe stimulus, w_sp, in the experimental condition is not different from the attentional weight of the nonsalient probe, w_sp, in the neutral condition. This is indicated by largely overlapping HDIs. Thus, the hypothesis that such an onset-based TOJ allows measuring salience as the parameter w was falsified. So, the conclusion is that the experimental manipulation does not work with onsets.

It is likely that the gaps were salient because they deviated from the pattern in the rest of the display. Li (2002) gives a neuronal explanation of why the absence of a stimulus can be salient: The homogeneous stimuli in the adjacent receptive fields cause an expectancy for the receptive field processing the blank spot. Although the blank spot does not show anything in particular, its deviation from its surrounding makes the receptive field deviate from its expected behavior. According to Li (2002) this is how salience arises. The reason why there is no difference in the attentional weight in the experimental condition is not that there was no salience but rather that there was the same salience for both elements during TVA's first wave of processing.

Interestingly, the Bayesian Analysis reveals a posterior distribution that was shifted slightly but distinctively towards a higher weight for the probe stimulus even in the neutral condition where both stimuli were indistinguishable. Because Bayesian Inference yield a distribution rather than a point estimate, it is clearly visible that the theoretically expected value is barely on the edge of this distribution and that there is clearly more evidence for an attentional weight $w_p > .5$.

In retrospect, it became evident that this shift occurred because of the temporal expectation of the probe stimulus. The probe stimulus' onset always happened after a fixed interval, whereas the reference stimulus' onset varied according to the SOA. This fits well with the experiments on temporal expectation by Vangkilde et al. (2012): The authors found that temporal expectation can indeed increase the visual processing speed.

It speaks, however, for the Bayesian Analysis that it does reveal that something unexpected is going on. If, on the contrary, a simple test would have been made whether the expected value of $.5$ is still a credible parameter value, such a test may or may not have been positive. In Bayesian parameter estimation, however, we see that a distribution that should theoretically be centered on $.5$ is obviously shifted away from this point. Furthermore, this quantitative insight into the credible parameter distribution allows comparing the nuisance influence to the size of salience's influence in Experiment 3.

It is clearly a weakness in the experimental design that no random jitter was included from the beginning of the trial until the TOJ. This design is problematic because the probe's onset will occur at an unforeseeable point in time, the reference's onset, however, will always occur at exactly the same point in time from the start of the trial. So, there is a higher expectancy for the point in time when the reference's onset occurs.

Experiment 2

Experiment 2 is highly similar to Experiment 2. Instead of an onset, we tested the offset of a stimulus as the event to judge with the TOJ. The rationale is that if the gaps were likely equally salient during TVA's first wave of processing and the salience of the stimulus cannot change the attentional weight prior to the processing of its own onset. Consequentially, showing both stimuli from the beginning of the trial onward should distribute attention according to their physical contrasts such that the salient stimulus is more attended than the nonsalient when the task-relevant event occurs.

Results show that indeed, attention was affected; however, against our expectations, the offset of the salient stimulus was processed not faster but slower than the reference stimulus. This can be deduced from the attentional weight smaller than $.5$, as shown in Figure 6.4. Because I did not pursue the causes of this results may further, I do not have a better interpretation than at the time of writing Article 1 for this unexpected results: It can be explained by the difficulty to get rid of a representation of a salient element in VSTM which may take longer than getting rid of a nonsalient representation and thus realizing whether the respective stimulus has vanished. However, in order to be sure about this explanation, further research has to be conducted.

Again, it is concluded that offsets — like onsets — do not cause data in line with the hypothesized first wave of processing where the salient stimulus gains more attention than the reference stimulus and the second wave of processing in which said advantage leads to faster processing of the salient stimulus.

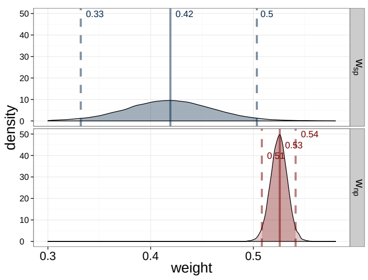


Figure 6.4: Results for the attentional weight estimation of Experiment 2 of Article 1. A Difference between the salient probe's attentional weight, w_{sp}, (experimental condition) and the nonsalient probe's attentional weight, w_{np}, (neutral condition) is clearly visible. However, the salient stimulus turns out to have less attentional weight associated with it. The opposite finding was expected.

Experiment 3

The third experiment is the core contribution of Article 1: It shows that when a flicker, a brief offset followed by an immediate onset, is an event that seems suitable to probe the distribution of attentional weights in a multi-element display. Results show that w_{sp} is high for the salient probe in the experimental condition and clearly distinct from the lower w_{np} for the nonsalient probe in the neutral condition, as shown in Figure 6.5.

Additionally, I find these results encouraging for the FRIM model underlying TVA. It assumes that the same resources are merely redistributed by attention. Thus, the overall processing rate should not change. Exactly these results were yielded by all the experiments of Article 1. In all of these experiments, the other free parameter besides the attentional weight, w, the overall processing capacity, C, stayed constant (for figures see the original article). This result is also in line with TOJ literature that proposes that attention neither causes only a speedup of the attended stimulus nor a slowdown of the unattended stimulus but both effects simultaneous (Weiß, Hilkenmeier, & Scharlau, 2013).

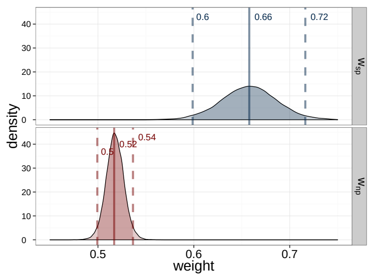


Figure 6.5: Results for the attentional weight estimation of Experiment 3 of Article 1. The difference between the salient probe's attentional weight, w_{sp}, (experimental condition) and the nonsalient probe's attentional weight, w_{np}, (neutral condition) is clearly visible and occurs in the expected direction.

Experiment 4

Experiment 4 extends the procedure to another salience dimension, luminance contrast. Also, instead of the neutral and the salience conditions, two salience conditions are introduced. These two conditions entailing low and high salience ought to show whether differences in salience can be distinguished by the modeling and empirical approach. The neutral condition is omitted because it is theoretically expected for the attentional weight for the probe stimulus w_p to be at the $.5$ level. This hypothesis has also been supported empirically by Experiment 3. The results show that luminance contrast can be measured both for the low and high salience conditions. The low salience condition shows a smaller w_p, attentional weight of the probe stimulus, than the high salience condition. The high luminance contrast has an attentional weight comparable to the high orientation contrast of $90°$.

A critique that could be voiced here and also for Article 2 is that the detection of the offset and onset may not be completely independent of the stimulus luminance contrast to the background. For example, imagine a probe stimulus that is barely visible because it is drawn in dark gray on a black screen. If the stimulus itself is close to the detection threshold, the detection of the offset and onset is affected. However, in the present experiment, the low salience and high salience stimulus are both clearly visible. To improve with regard to a possible confound and also to enable the experimenter to test motion salience, I have developed an improved stimulus material that swaps an internal pattern instead of a complete offset and onset. This is explained in detail in Section 6.6.

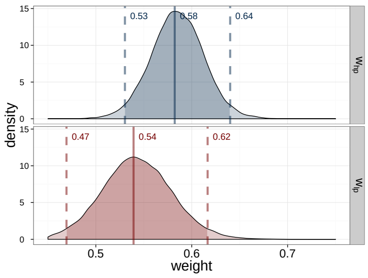


Figure 6.6: Results for the attentional weight estimation of Experiment 4 of Article 1. The attentional weights ob the high luminance contrast, w_{hp}, and the low luminance contrast, w_{lp}, are both higher than the neutral value of $.5$ but as expected w_{lp} is larger than $.5$ but smaller than w_{hp}.

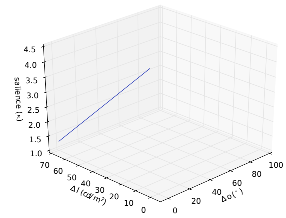
Conclusion

All in all, Article 1 can be seen as a proof of concept. It shows that the common TOJ designs (onset and offset of stimuli) are not suited for measuring the attentional weight of a salient stimulus. Yet, with an adequate event, a flicker, experimental results are in line with theoretical reasoning. Furthermore, different salience dimensions can be manipulated, which has exemplarily been shown for orientation and luminance contrast. For a salience measurement based on TVA, two questions have to be answered: Can the salience of less extreme differences be measured with sufficient precision and can this value be related to a specific salience value, a common currency of salience?

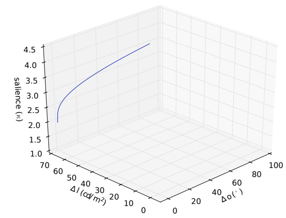
Summary of Article 2

Article 2 of this cumulative dissertation has the goals to find a common currency of salience first and then to use this common currency to answer the question whether two salience dimensions interact. The salience currency is based on the procedure developed in Experiment 3 of Article 1 and the salience parameter κ that was introduced to TVA Nordfang et al. (2013). In Article 1, the salience measure was the general measure of *relative* attention in TVA, the attentional weights. These weights, however, can only be determined in relation to each other. That means, if you multiply all attentional weights by a certain factor, they will yield the exact same outcome mathematically as before. However, Nordfang et al. (2013) simply assumed that a nonsalient stimulus X to have a $\kappa_x = 1$. Thus, a nonsalient reference stimulus in the TOJ should allow measuring an exact quantitative representation for the strength of salience of the salient probe stimulus in the TOJ.

Furthermore, it should be answered whether this measurement of salience is sufficiently precise to distinguish the salience of different levels of physical contrast. And if so, how salience arises independently of physical contrast. After establishing how salience arises within a single dimension of physical contrast, it is checked how two such dimensions combine. To pursue these goals, Article 2 starts with adapting Experiment 3 of Article 1.



(a) Expected lower bound: maximum



(b) Expected upper bound: addition

Figure 6.7: Increasing physical contrast in equally sized steps will increase salience. However, it is not clear how to describe this increase. Possible findings include the linear increase. That is a fixed increase in contrast always increases salience by a fixed amount. Alternatively, it is likely that the salience increase diminishes if the contrast is already high. This relationship can be modeled by a logarithmic or power function.

Experiment 1

Adapting the Experiment 3 of Article 1, this experiment introduces more experimental salience conditions. The goal of Experiment 1 is to understand how salience rises when physical contrast is increased. For example, a linear increase is a possibility. Also, a logarithmic-looking increase is plausible that would mean the larger the already existing contrast, the smaller the gain of adding a fixed-size step. These two alternatives are sketched in Figure 6.7.

Diminishing returns from a fixed contrast increase can be modeled by a logarithmic function or Steven's power law (1957). It is important to understand here that I do not look for the "law" of salience but rather want to understand how it works. These two approaches are contrasted by Cummins (2000). This context, this distinction means that whatever function is more appropriate is used as a model but not proclaimed as the true law of salience. The three experimental conditions comprise 30°, 60°, and 90° of orientation contrast respectively to allow us to see how salience increases with physical contrast.

In the modeling part prior to Experiment 1 in Article 2, it is explained how the relative attentional weight w_p and the absolute amount of salience κ are connected. Namely, the salience value of the nonsalient probe is set to 1 as it is designed to be nonsalient (a salience value of 1 instead of 0 is used as a neutral value because salience and other attentional influences multiply according to the CORE equation Nordfang et al., 2013). This allows to compute κ_p, the salience value of the probe stimulus p to be $\kappa_p = \frac{w_p}{1+w_p}$. So, the reader may think of κ_p as the theoretically more appropriate parameter name for the salience value that, however, does not carry more information than the w parameter.

From the results in Figure 6.8, it is apparent that the 60° and 90° orientation contrasts do not differ much. This does, however, not mean that the method does not work as intended. To the contrary, it is not well studied how salience rises from linearly increasing steps of contrast. In fact, psychophysics shows that physical quantities and the perception of their intensity are often related by a logarithmic (Fechner, Howes, & Boring, 1966) or according to a power function (Stevens, 1957). If the relation between orientation contrast and the amount of salience follows a logarithmic function or a power function with an appropriate exponent, it would be expected that the influence of equally sized steps of contrast recedes. In any case, a linear increase of salience is not supported by these results.

The analysis of Experiment 1 shows that the power law is an appropriate model. The appropriateness of this function stems from two lines of reasoning. First, it can be argued for from a theoretical position as a generalization of Fechner's log function. Secondly, it has to fit better to the data than the alternatives. These empirically tested alternatives are a logarithmic relation between orientation contrast and amount of salience and to simply assume independent κ values for each condition (this is a model that corresponds to no systematic connection between the values). A model comparison revealed that the power-law model fits better to the data if the model complexity is also taken into account. However, more than three experimental conditions may have been helpful to determine the exact shape of the

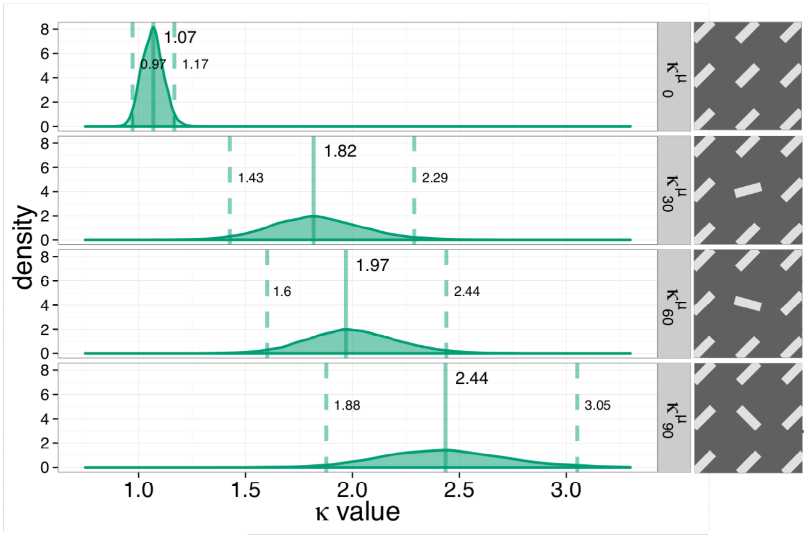


Figure 6.8: Results of the salience estimation of Experiment 1 of Article 2. These results show that a linear increase is highly unlikely. Although the data suggests an increase even for the step from 60° to 90° of contrast, the clearest difference is from 0° to 30° of contrast because the 95 % HDIs (dashed lines) are not overlapping that means a highly credible difference.

increase of salience. Thus, Experiment 2 was conducted with more experimental conditions.

As expected, the analysis revealed additionally that the overall processing capacity, C, stays constant in the experiment. This parameter is another free parameter in the model but should theoretically not be changed by the manipulation.

Experiment 2

Because the power-law model of how salience increases with physical contrast was tested because the data suggested it, it can be ascribed to the context of discovery rather than the context of justification (Reichenbach & Richardson, 1938). Thus, Experiment 2 was conducted to justify what was found in Experiment 1. A step size of 15° was chosen so that the shape of the underlying function may become more obvious. These additional steps resulted in seven conditions. The results are shown in Figure 6.9. Again the three models, logarithmic increase (one parameter), power-law (two parameters) and independent κ-values (seven parameters) were tested. The salience increase suggested by the three models is depicted in Figure 6.10.

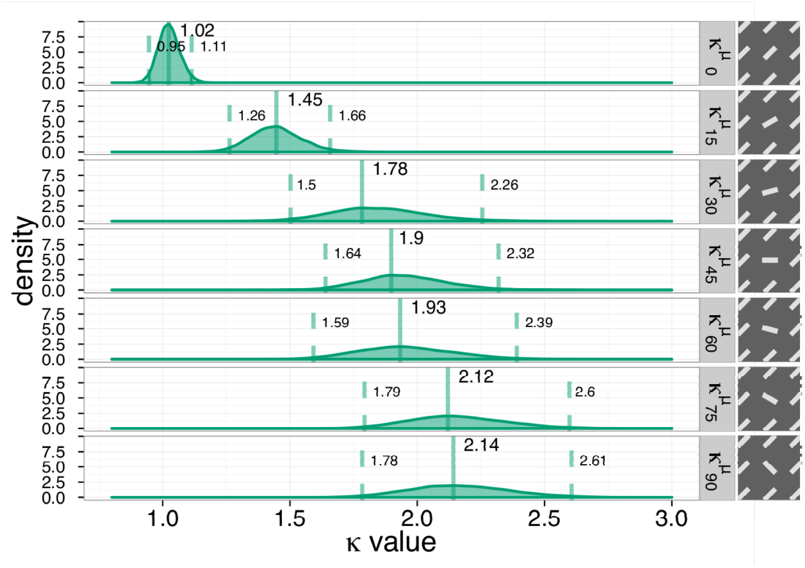


Figure 6.9: Results of the salience estimation of Experiment 2 of Article 2. In contrast to the results of Experiment 1, these estimations allow a much better impression of the diminishing salience returns for a fixed step contrast increase. Further analyses revealed that a power function is an adequate description.

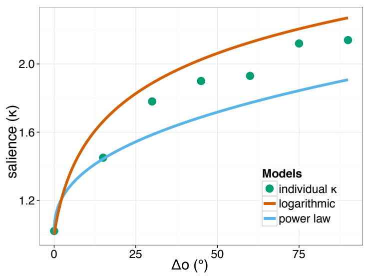


Figure 6.10: This figure shows how strongly the three models differ with respect to the estimated salience value. Please note that the dots are no raw data but are punctual salience estimations assuming that the increase of salience is not connected to the increase in other conditions at all. Also, note that the models have different amounts of free parameters: The individual estimation has seven free parameters, the power function two and the logarithmic model one parameter.

This study, again, yielded that salience does clearly not rise linearly with linear steps of physical contrast increase. Visually, the increase in salience looks like a logarithmic increase. Roughly speaking, this means the more salience, the more contrast is needed to increase the salience of the same amount. Both logarithmic and power-law-based functions have been tested again to describe this relationship. Like in Experiment 1, the power-law model turns out the be an adequate model. In the model comparison, not only fit to the data was taken into account but also the number of free parameters. Again, the analysis revealed that the overall processing capacity, C, stays constant.

Thus, the first contribution of this Article 2 to the scientific discourse is that the increase of salience can adequately be modeled by a power law within at least the orientation dimensions. Previously, the logarithmic model has been used by Huang and Pashler (2005); however, only a few data points led to this hypothesis.

Experiment 3

It stands to reason that the power-law model may generalize to other salience dimensions. Thus, luminance contrast was again tested as another easy to manipulate salience dimension in Experiment 3 of Article 2.

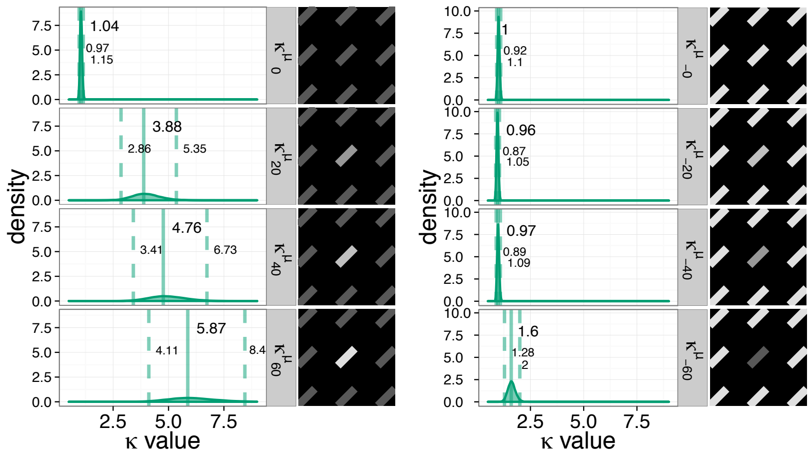
The general idea of this experiment was to test different levels of luminance contrast analogously to Experiment 1 and Experiment 2 of Article 2. However, I did not only test luminance salient stimuli that were brighter than their surrounding but was interested in what happened when the element that stands out is actually darker than the surrounding elements. As currently modeled, only the difference in physical contrast of unique stimulus and surrounding homogenous stimuli is used to determine the physical contrast. The intention for these additional conditions was — maybe a bit naively — to test whether the power-law model can also describe the increase in κ in these conditions. In retrospect, I think this experiment may have been stronger if the additional "dark amongst bright" were omitted because I did not completely reverse the situation because I keep the background color constant so that the unique salience stimulus appeared less bright than the surrounding elements but brighter than the background color. In this contrast, much more contrast between surrounding elements and the unique element is necessary to yield a measurable effect on attention because the unique stimulus is not a luminance outlier in the presented display but in between the luminance values of surrounding elements and background screen.

In congruence with this post-hoc reasoning, the power-law described the increase in κ well if the unique element was indeed the most luminant in the whole visual field. So, the model was appropriate in this case. However, the four conditions in which the unique stimulus' luminance value was between those of surrounding and background elements, salience was nearly not affected for the same steps of contrast increase as in the first four conditions. These four conditions serve to show the limits of the simplistic model that salience arises merely because of the physical contrasts between foreground and background. Although these conditions can easily be criticized, I would like to stress that the model and experimental procedure allow looking into how strongly attention is actually affected by a particular element when compared to an inconspicuous element of the display so that it becomes obvious how inadequate the simplistic difference in luminance contrast between homogenous elements and the unique element is in this case. Figure 6.11 shows the estimated κ value and may help the reader to imagine the conditions with the help of the small sketches on the right side.

To summarize Experiment 3, it shows first that the power-law model is applicable to particular luminance contrasts, but not to all. Thus, Experiment 3 reveals the limitation of the model that the luminance of the background screen is not independent of the actual salience value determined by the procedure. Because all background screens have a particular luminance, it has to be kept constant in the experiments but must be remembered to restrict the generality of the determined salience values. In further research, a better formal description of the contrast between attended element and surrounding may be found.

Experiment 4

Besides establishing a model for how salience rises within a contrast dimension, Article 2 has the second goal to show how two dimensions interact. Here different ideas have been proposed in the past. All of them agree that if two types of salient contrasts are combined, the result is more salient than the most salient individual contrast but less salient than the individual contrasts added up. However, the degree of this interaction seems to vary depending on the two types of contrast being combined. These previous works can be criticized in thus far as they worked with an explicit model of how salience rises within a dimension and I have shown in the previous experiments that salience does not rise linearly with contrast. Secondly, only very few salience estimations are used in the literature to come up with estimations of the degree of interaction. However, if



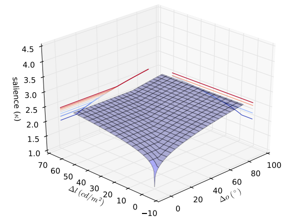
(a) Unique element successively more luminant than homogeneous elements.

(b) Unique element successively less luminant than homogeneous elements.

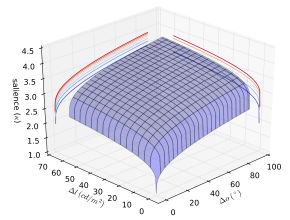
Figure 6.11: Salience estimation of Experiment 3 of Article 3. Whereas a unique element that has the most luminant surface in the display causes salience describable be the power function model, the same contrast levels do not cause a comparable effect if the unique element is less luminant than the homogeneous elements but more luminant than the background. These results show the limitation of the simplistic model that only takes into account differences between homogeneous and unique elements.

maximum, addition, or as proposed a sort of discounted addition were the true systems according to which combined salience is computed, this should lead to particular patterns of salience estimation in a factorial design of for example four times four contrast levels. In Figure 6.12 different ways of computing overall salience have been sketched.

Experiment 4 corresponds exactly to the idea that it should be possible to determine the amount of discount or interaction between the two contrast dimension in creating a combined effect on attention. To this end, four levels of orientation contrast and four levels of luminance contrasts have been used in a factorial design. It should be possible to estimate the degree of interaction by a model that states the idea of a discounted addition formally. If you think about the sketches in Figure 6.12, this means that the addition model is taken and a fixed factor is added for all additions. This means that if a non-zero contrast from one dimension is added to another non-zero contrast from another dimension, individual salience is added up but with a discount of say 30 % for example. Whether this discount is 10 %, 20 %



(a) Expected lower bound: maximum



(b) Expected upper bound: addition

Figure 6.12: The literature suggests that combining two physical contrasts from different contrast dimensions causes a salience that is at least as strong as the salience value of the most salient of the two contrasts (maximum) and never more salient that the individually determined salience values added up (addition). These two bound on the expectation of salience values are depicted in the figures.

or 80 % could be estimated by the model based on how the salience arises within each dimension.

To illustrate the results of this experiment, Figure 6.13 shows the independent κ estimations as green dots as well as the fitted model. As in Experiment 1 and 2 of Article 2, both dots and plane represent models. The underlying logic is that independently estimating the salience has much more degrees of freedom but may not be an economical description because it squanders degrees of freedom where they may add little to the model's ability to describe the data concisely. On the other hand, assuming a relation between these salience values reduces the number of free parameters but limits how the data can be fitted. Yet an assumed function may turn out as the better model if it can still describe the data well while reducing the number of free parameters. The results show that the discount that best explains the data under the stated model assumption is 0. That means salience from the individual contrasts is simply added up. Whereas this analysis can itself be counted as model comparison by parameter estimation (Kruschke, 2013) (meaning a parameter that a model entailing a discount of 0% is equal to the model not having this particular parameter.) we conducted another model comparison by formulating the competing model that simply does not assume any discount. The no-discount model came out slightly on top again.

The interpretation of these results is that it is better to assume a simple linear addition than a fixed discount factor. However, this analysis does not rule out complex discounting mechanisms. The proposed modeling and empirical approach, however, allows testing different discounting mechanisms as soon as they are proposed. For the time being, however, a fixed percentual discount as previously proposed seems worse than a simple addition for describing the data of a multi-contrast-level factorial design.

Summing up this chapter, I have introduced, tested, and tackled research questions by combining an empirical and modeling approach to salience measurement. This approach makes use of multi-element displays and the prior-entry phenomenon that is interpreted with an overarching

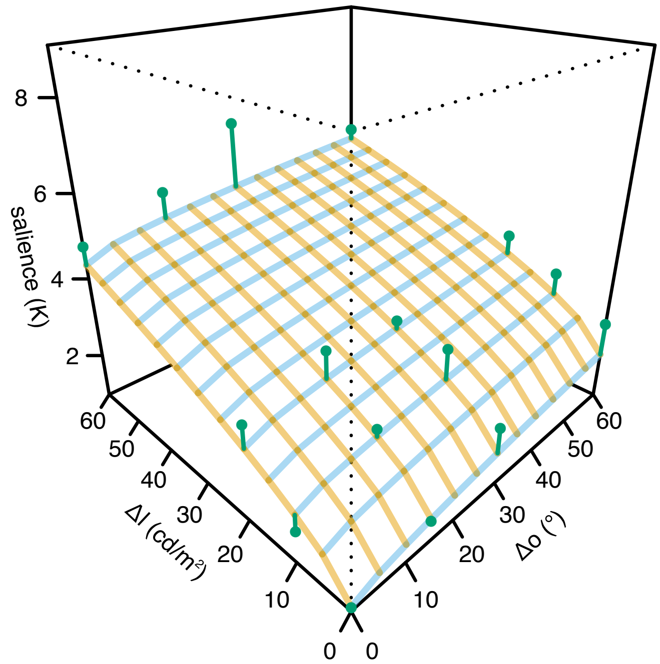


Figure 6.13: Results of the salience modeling of Experiment 4. Green dots show the salience estimation for each condition individually whereas the plane shows how salience is modeled by an increase according to a power function within each dimension and an addition across both dimensions.

formal theory of attention. As such, the resulting quantitative measure of visual salience is yet another way of quantifying visual salience. What is different with this measure is, however, that there is a tight link between data and theory provided by an explicitly stated logical and statistical model. The proposed salience measure, κ, has a specific function within the TVA framework, and additionally, the experimental procedure allows to manipulate and compare many physical contrasts that ought to affect salience. Also, it explicitly treats task-relevance of unique stimuli in salience experiments: Because the two positions are known in advance and are always the same, the locations have the same task-relevance. Thus, it can be argued that whatever salience adds on top is because of its effect on attention. In the following chapter, the proposed combination of experimental design and model is tested for different situations and research questions.

Chapter 7

Quantifying salience-related phenomena

Beyond the foundations of TVA-based salience measurement, many more questions can be addressed by using the proposed method. In this dissertation, Article 1 and Article 2 are the foundation of a new quantification of salience. That is why they and the ideas contained therein were presented in the separate previous chapter. Once established, many salience-related phenomena can be studied with a TVA- based interpretation of TOJ data. Such further research is presented in the present chapter. These smaller studies have not been written up as a full-sized paper. There are various reasons for this: Often, the research question is rather small, or results do not match expectations from previous research so that a further replication is reasonable before communicating the results to a broader audience. An exception is the work on the time course of salience because it has been developed to such a degree that it can be published as an article in its own right.

Time course of visual salience (Article 4)

The strength of salience follows a time course. Article 4 extends the applicability of salience measure to this phenomenon. We expected that the salience parameter, κ, varies over the duration of the first second after the onset of the salience display. Roughly speaking, the literature shows that salience rises in strength up to 150 ms and declines afterward.

Experiment 1 of Article 4 was designed to replicate the results that were expected from the literature. Although the hypothesized time course was statistically shown for the salience parameter κ, this time course was much less pronounced than the time course of the overall processing capacity C: In contrast to our expectations, Experiment 1 revealed a distinct reduction in overall visual processing capacity that was supposed to be similar in all conditions because it describes the overall visual processing capacity, the general ability to solve the TOJ task.

To rule out that the TVA-based model is simply a bad model, and its parameters thus do not adequately describe the data, a model comparison was conducted: Compared to the common logistic function, the TVA-based model provides a slightly better fit for the data while gaining a precise theoretical meaning for its parameters compared to the merely descriptive logistic function. Thus, I do not argue that the TVA-based model provides the better fit in general — although this may be the case, it is difficult to show it convincingly — but that the TVA-based model does not provide an unreasonably worse fit compared to common psychometric functions.

Experiment 2 justifies the conclusion from the Experiment 1 by replicating the result. In order to be comparable to previously published studies, a single aspect of the design has been changed: Instead of a fully randomized design, a blocked design was used — similar to Donk and Soesman (2011). As in Experiment 1, the data observed in the 50 ms and 100 ms intervals is much better explained by a change in visual processing capacity and salience than a change exclusively in salience. Processing capacity reaches stable values in the range expected from previous experiments for the 200 ms, 400 ms, and 800 ms conditions. Roughly speaking, this means that the participants' processing speed in the TOJ was severely and gradually reduced in the first two conditions. This is roughly equivalent to a change in the so-called difference limen parameter of comparable descriptive psychometric functions, e.g., as used by Matthias et al. (2010) to analyze the effect of alertness on the TOJ performance. A change in attention, in contrast, shifts the processing speed from the unattended to the attended stimulus (Weiß et al., 2013; Tünnermann et al., 2015).

Furthermore, the paper argues — independently of whether the reader agrees on the usefulness of the TVA-based model[1] — that a model-based approach is needed to disentangle attention and other effects on selection. A model of visual selection must explicate the consequence of a parameter change in attention as opposed to the general ability to solve the task (alertness or strength sensory evidence). The discussion closes with the observation that modeling helps to refine the concepts in psychology.

[1] A model based on the parameters PSS and DL would of course also reveal a difference in DL, so the result does not completely depend on TVA, but TVA offers a cognitive model for explaining the parameters in terms of the visual selection process.

In-depth comparison of salience dimensions

In Experiment 4 of Article 2 (See Section 6), I have shown how luminance and orientation contrasts interact; however, other dimensions may behave differently. Nothdurft (2000) has already shown this in his comprehensive study. His methods were based on perception. Interestingly, no other author has — to the best of my knowledge — provided a comprehensive comparison of the possible interaction of different salience dimensions. Similarly to Nothdurft's study, a TVA-based analysis of color, motion, orientation, and luminance contrasts in TOJs is possible and may confirm or contradict the estimations by Nothdurft.

To test for possible interactions, an improved design had to be implemented because the flicker creates a big change in luminance contrast. The flicker is also problematic for moving stimuli because a stimulus moving quickly might appear as "jumping" whereas a slow-moving stimulus flickers as intended. Thus, stimulus material was developed that consists of small bars with an internal pattern of light and dark areas. The overall luminance of the stimulus

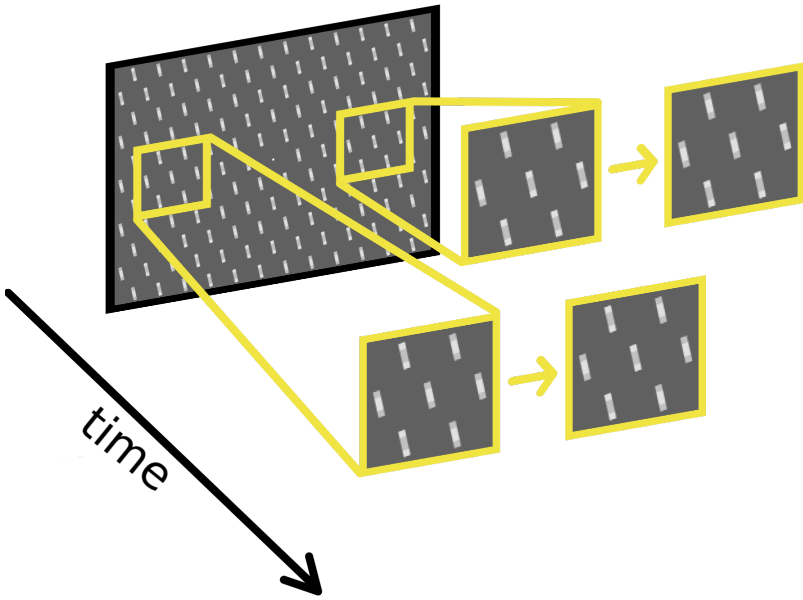


Figure 7.1: Sketch of TOJ task with bars flipping patterns.

could still be manipulated experimentally. The pattern could be flipped and hence creates a noticeable change for the TOJ without changing the overall luminance. A set of experiments was conducted to evaluate this stimulus material.

Figure 7.2: Newly developed stimulus material with internal pattern that can be flipped.

Flipping patterns instead of flickering stimuli

To test the new stimulus material, multiple pre-tests were conducted with different patterns and small numbers of participants. All pre-tests showed a reduced overall processing capacity, C. This finding is very robust and in line with the theory because the pattern flip provides less visual evidence for an event when compared to a complete offset and onset of the stimulus. This, in turn, makes it more difficult to judge correctly. Yet, a proper replication was needed as this explanation was conceived after seeing the data. The central Experiment 3 of Paper 1 was replicated with the new stimulus material as a conceptual replication of the original result. The used stimulus material is depicted in Figure 7.1. In this sketch, a TOJ salience-display is depicted — probe is nonsalient in this example. The TOJ is implemented by the flipping of the internal pattern, which is also enlarged in the sketch. Additionally, a large version of the two states of the stimulus is shown in Figure 7.2.

There were 31 participants in this experiment.

Results show — as expected from the pre-studies — that overall processing capacity, C, is around 23 Hz. As already mentioned, this result is expected because there is less visual evidence for the event to be judged in TOJ. Yet, the main finding that processing capacity is not affected by the experimental manipulation could be replicated successfully. More precisely, this means that an orientation difference of 0° leads to an attentional weight of $.5$. Increasing the orientation contrast in incremental steps of 30°, yields a steep initial rise in attentional weight and a diminishing increase for further steps as the power-function model suggests. Yet the absolute values of the w-parameter are slightly reduced in comparison to the original experiment. There is a sound theoretical explanation for the difference. The TOJ task is harder with the flipping patterns; this is evident from the reduced slope of the function, the corresponding low overall processing capacity, C, and of course, the reduced physical evidence for a change. Thus, it is likely that participants focus their attention more strongly on the a-priori-known relevant locations. Please note that there was always a mix of task-relevance and stimulus-driven attention involved in the design; the task-relevance-based attention was, however, kept constant by the a-priori-known relevant locations. Also, note that the explanation is a post-hoc explanation which is not backed by an additional replication. No additional replication was conducted because the finding of the original Article 1 was successfully replicated and numerically identical estimates were not expected in the first place.

To conclude, the new stimulus material works well enough to replicate the result of Experiment 3 of Article 1. Even if the manipulation's effectiveness is reduced because of the demanding task, salience can be measured yielding the same conclusions as in Article 1.

Comparison of red-green and blue-yellow color axes

The color axes red-green and blue-yellow of the CIE L*a*b color space were compared in a 4×4 factorial design similar to that of Article 2. Although the CIE L*a*b color space is designed in such a way that distances in the color space correspond to visually perceived color differences, it has not yet been tested whether a fixed distance yields the same advantage for attention on both dimensions. Also, their interactions are unknown and have been investigated by the same model as in Article 2.

The actual color differences in the implementation is based on the color range of a color-calibrated (X-Rite ColorMunki Display) CRT-monitor (Iiyama Vision Master Pro 512). The monitor was able to display color values in the range of -50 to 50 for the a and the b parameter in the CIE L*a*b color system. Thus, the maximum difference was 100 for each parameter that corresponds to 1.0 for the maximum difference in the analysis, 0.5 and 0.25 stand for a difference of 50 and 25, respectively. The used luminance value, L parameter, was 70.

The Bayesian hierarchical models and analyses were the same as described in Article 2: One for the independent estimation of salience and one for the power-function model. Overall, there were 25 participants.

Results show that both dimensions scale differently. This is evident from the hyper-parameter plot in Figure 7.3 and the respective two parameters for the power function, the k, and the n parameter. These parameters link the yellow-blue and the red-green contrast in CIE L*a*b to the salience parameter κ. Although difficult to spot from the numbers alone, the 3d plot in Figure 7.4 shows that yellow-blue contrast ($\Delta_{\text{yellow-blue}}$ has an influence on salience (κ) that is twice as large as red-green contrast ($\Delta_{\text{red-green}}$). The DIC is 5588 (this value can be compared to other models' DICs for the same data to find the best model. This procedure also accounts for model complexity, for further details, see Article 2). If the salience, κ, is estimated for each condition independently, i.e., without assuming a relation between them in the model, MAPs for salience are obtained as marked by the green dots in Figure 7.4. The DIC for this model is 5891. Thus, assuming the power-function relation implies a better model than the individual salience estimation.

Both dimensions are independent in terms of a simple discount or bonus for a situation in which both dimensions are combined. This result is the interpretation of the parameter a that is centered on 0. As expected from Section 7, the overall processing capacity is reduced in comparison to the experiments in Article 2.

What do these results mean? First of all, the color was less effective in influencing salience in comparison to the experiments on luminance and orientation contrast in Article 2. Although the CIE L*a*b color space was designed so that distances in this space are proportional to the perceived differences, the same distances on the yellow-blue axis are nearly twice as salient (as estimated for the whole population). Both dimensions are at least not affected by a simple discount or bonus for their combination. Thus, they are either independent or interact in a more complex manner. Apart from this difficult-to-interpret situation, the experiment shows that properties that are a priori similarly salient can be investigated in detail trough the modeling and empirical approach presented in this work[2].

[2] An experiment of similar design was also conducted for color (yellow-blue contrast) and motion contrast. This experiment, however, was limited by several factors (it had to be shorted in order to comply with course credit regulations. Also, the levels of experimental motion-contrast manipulation were limited by the possibly salient collisions of elements). Because of these limitations, the experiment revealed little more than that motion contrast is salient as expected from the literature. This experiment is mentioned here to disclose all lines of experimental research even if unfinished.

Origin of orientation salience

Orientation salience has been implemented in many models; however, what does actually cause the strength of salience? In an experiment based on just-notable differences in orientation, Foster and Ward (1991) found that the absolute orientation of stimuli is relevant in addition to the difference in orientation. For example, showing a background of vertical lines and a singleton that is rotated $60°$ clockwise makes the difference easier to spot then when singleton and background orientations are swapped. This leads to the assumption that a particular orientation contrast does not suffice to predict the effect on salience. There are two explanations why this may be the case: The first one is provided by Foster and Ward (1991) who argue that the

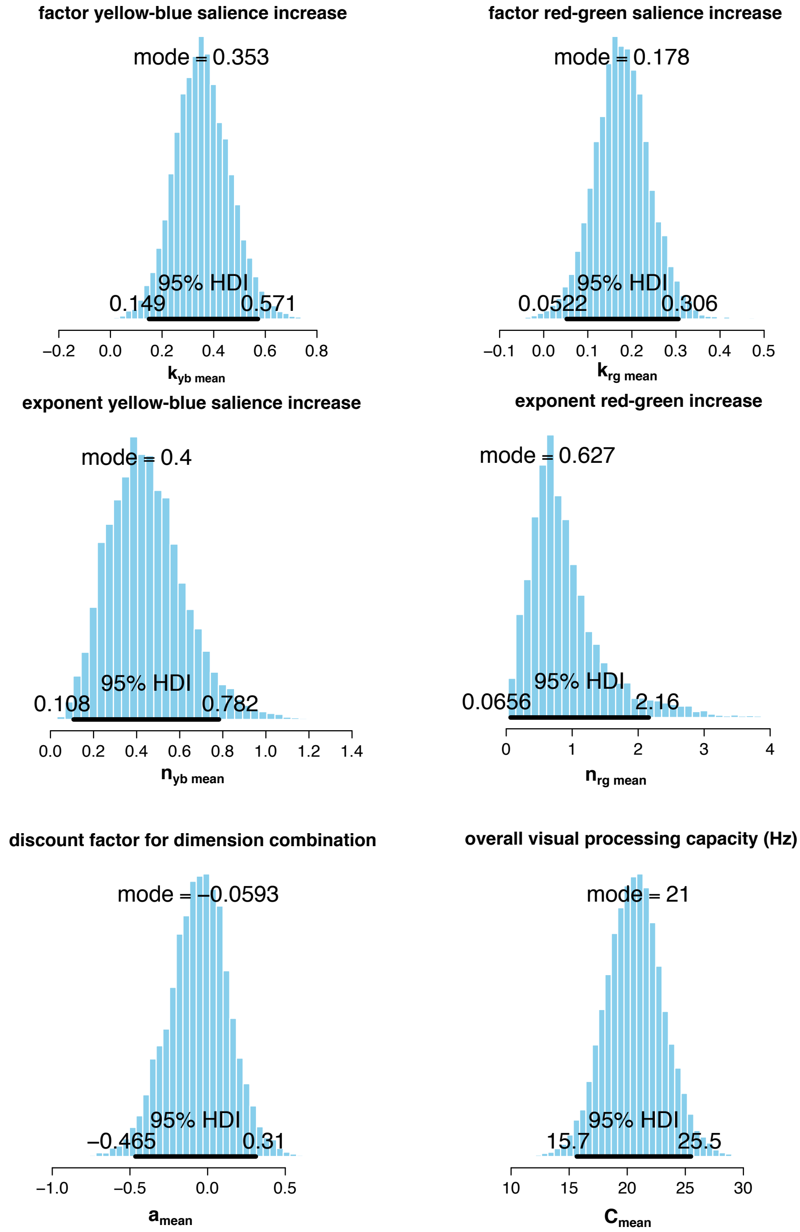


Figure 7.3: Hyper-Parameter estimation of the experiment comparing red-green and- yellow-blue contrast.

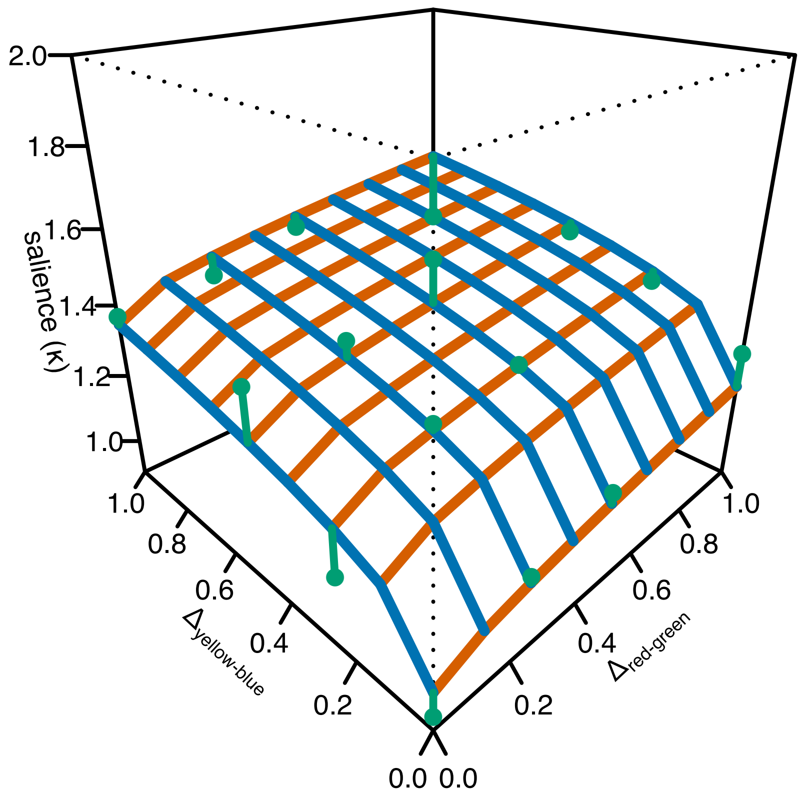


Figure 7.4: Salience as dependent on red-green and- yellow-blue contrast based on MAPs. The green dots shows the saliences estimation if each conditions attentional advantage is modelled as a individual dependent variable.

visual system responds to some orientations more strongly than others. The V1-based salience model by Li (2002) implicitly gives another salience-based explanation: The vertical lines were positioned on a grid; thus, they created a long virtual line that was interrupted between the elements. A deviation from this virtual line caused a much stronger response based on the neuronal architecture in Li's model than the background elements that were rotated by 60° and thus did not form a virtual line.

It is important to mention that it has never been tested whether the absolute orientations are important additionally to the orientation contrast to predict the effect on attention. However, Wolfe (2007) assumes that the finding by Foster and Ward (1991) should probably be used to model orientation salience. We tested this hypothesis.

Results show that absolute orientation matters. This is evident from the salience parameter, κ, estimation for the vertical background lines in comparison to the diagonal, –60°, lines as shown in Figure 7.7. This image shows the vertical background conditions. The first four distributions show the salience, TVA's κ-parameter, for an increase of orientation contrast on a vertically oriented elements (levels: 0°, 15°,30°,60° of orientation contrast; see Figure 7.5); The next four distributions show the same increase of contrast on a diagonal background (–60°, see Figure 7.6). When comparing both groups, it becomes evident that, the salience is stronger if the contrasts are presented on the diagonally orientated background. So, the highest orientation contrast of 60° on the vertically oriented elements (fourth distribution from top of Figure 7.7) is roughly as strong as the 30° contrast on the diagonally oriented background (seventh distribution from top of Figure 7.7). The overall processing capacity, C, shown in Figure 7.8, stays constant as theoretically expected although it is a free model parameter. Thus, the hypothesis that the salience of orientation does not only depend on orientation contrasts but also on the absolute orientation of the background is clearly favored over the hypothesis derived from Li's (2002) work. Li (2002) would have predicted that the deviation from virtual lines causes the salience — an influence not controlled by Foster and Ward (1991).

Although one hypothesis is clearly favored by the results, the results are difficult to publish without replication because based on Foster and Ward (1991) a stronger salience was expected from the contrasts presented on the vertically oriented background. Thus, the quantitative results — in contrast to the qualitative assessment that salience of orientation contrast also depends on the absolute orientations involved — are neither in line with the work by Li (2002) nor the work by Foster and Ward (1991).

Salience, task relevance and location

The work on conjunctions of salience and task relevance and their effect on locations specific attention was made possible by a DAAD grant. With this grant, I was able to visit the Center for Visual Cognition in Copenhagen for three months. Amongst the researchers, I met there

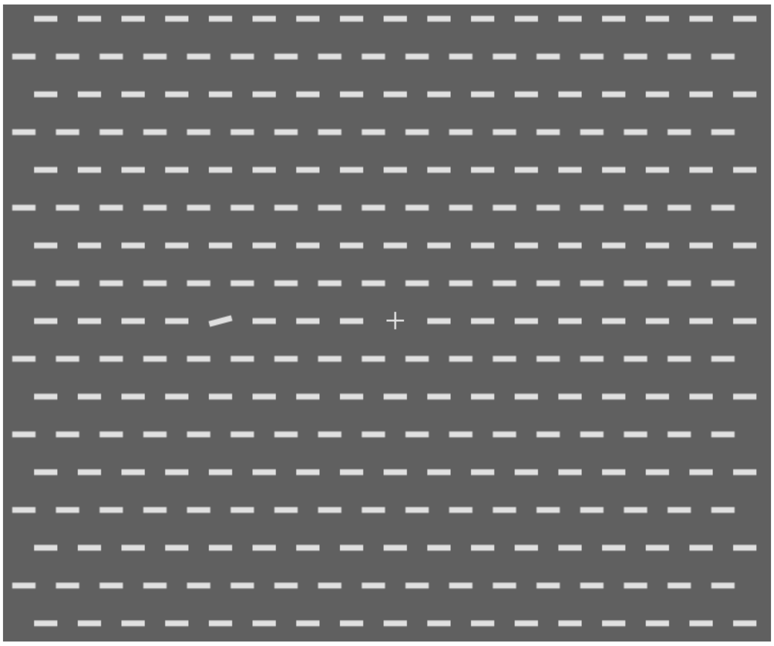


Figure 7.5: Background orientation 0° example

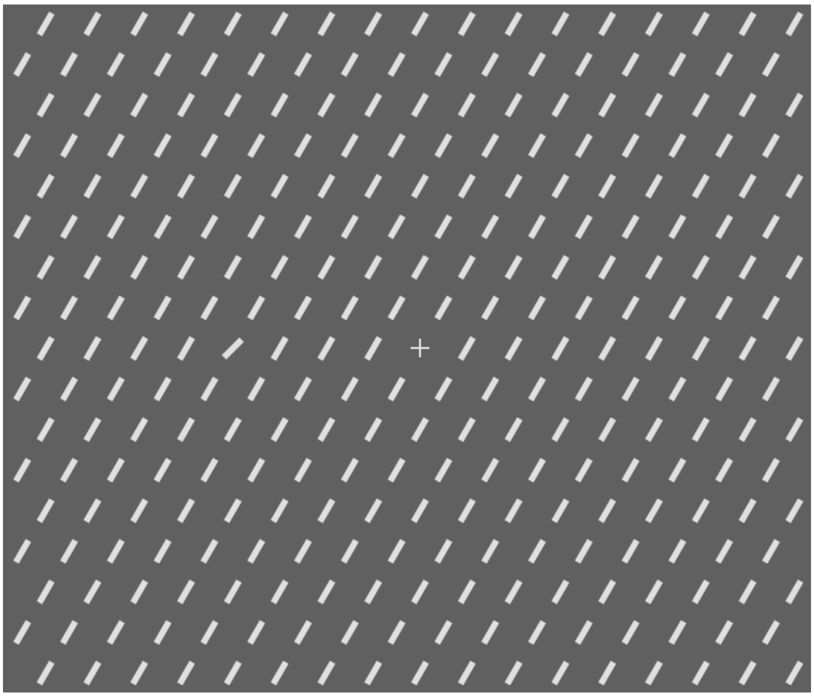


Figure 7.6: Background orientation –60° example

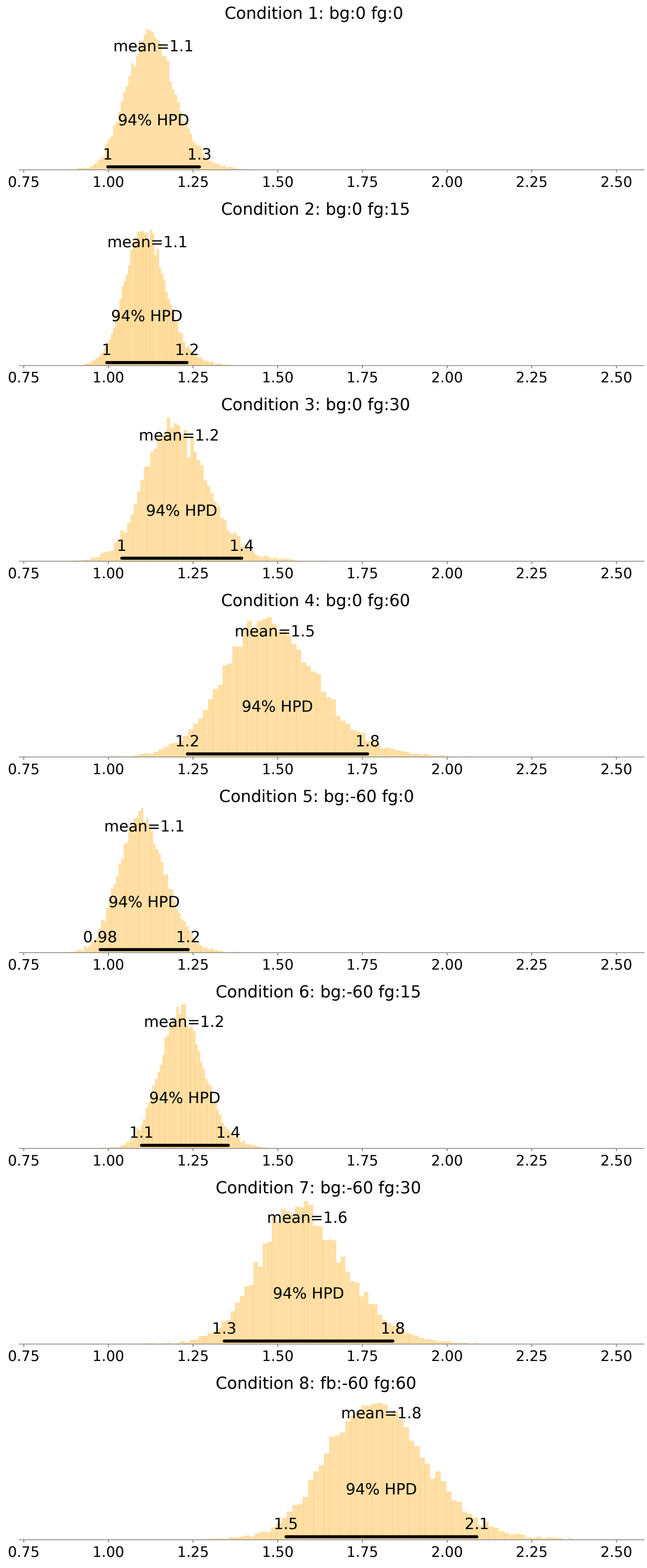


Figure 7.7: Salience parameter, κ, rate per condition.

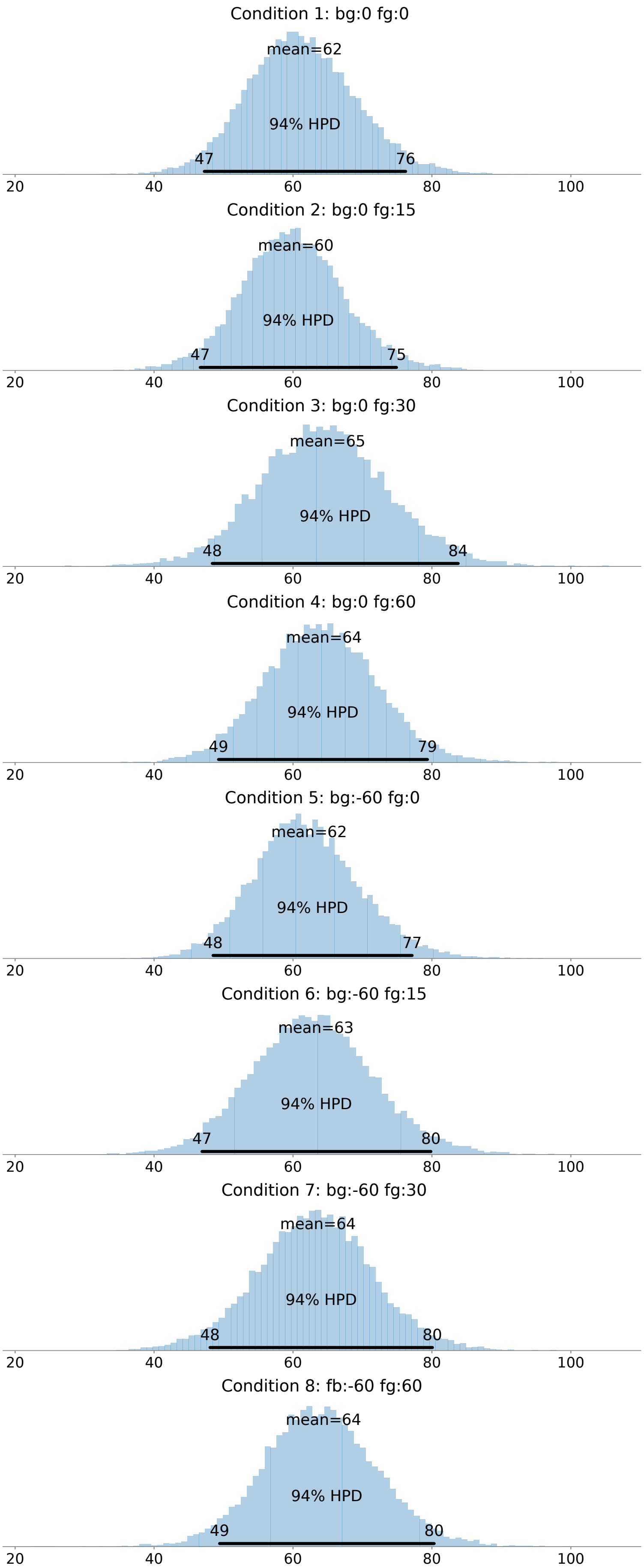


Figure 7.8: Overall visual processing, C, rate per condition.

were Claus Bundesen, Axel Larsen, Thomas Habekost, and Søren Kyllingsbæk. All are deeply involved in TVA modeling and related experiments. The closest contact I was with Anders Petersen and Signe Vangkilde. For their support and warm welcome, I am very grateful.

During my time there, I generated a few ideas for TVA-based research[3]. The most promising research idea I came up with was based on a recent publication by Nordfang et al. (2017). The central idea of this publication is that location is special when compared to feature-based attention. Originally, this has not been part of TVA. This publication addresses this missing link by drawing on many years of observing location-specific preferences of individual participants and refining the empirical finding to a new weight equation that generalizes the salience-specific findings by Nordfang et al. (2013). Whereas this extension was conceived from letter-report designs, I wanted to check the new equation with the TVA-based model of TOJ.

[3] I am particularly grateful for all researchers that took the time and discussed these ideas with me because these discussions helped me understand that an epistemically interesting question still needs an interested audience. I cannot remember who formulated it this straight, but one response to an idea was "This is indeed an open research question given the literature, but I doubt that anyone outside this room would be interested in its answer."

Check a new TVA weight equation in a different experimental paradigm requires theoretical and empirical work. In particular, hypotheses for the new experimental setup have to be derived from the equations, and also the manipulations must affect exactly the constructs the TVA parameters are supposed to represent.

To explain the reasoning behind the experiments on location-specific attention in TVA, it is apt to start with the weight equations. The original weight equation by Bundesen (1990) as shown previously in Equation 4.3 and for ease of comparison also repeated here as Equation 7.1. This equation interpreted by understanding the attentional weight[4] as the result of multiplying η and π that stand for the sensoric evidence and pertinence, respectively. Both, η and π, depend on j in the equation, meaning that a stimulus, x, provides a sensory evidence for being a particular thing, a thing from the category j. Also, categories differ in relevance; this is represented by the respective pertinence, π_j, for category j. The sum of this equation means that all relevant categories from the participants' long-term memory have to be considered. In general, this seems quite difficult to handle because of the potentially large num-

[4] In TVA stimuli race in parallel to be represented in short-term memory. The attentional weight represents an attention-based distribution of processing resources. Stimuli receiving more resources — have a higher weight — possess an advantage in the race to short-term memory.

ber of categories in the participants' long-term memory. This problem is addressed by clever experimental design in which only a few categories matter such that all others can be assumed to have a pertinence of 0, such that the evidence is multiplied by 0 and hence not relevant for the attentional weight.

$$w_x = \sum_{j \in R} \eta(x, j)\pi_j. \tag{7.1}$$

The TVA extension by Nordfang et al. (2013) is comparably simple as shown in Equation 7.2. The only new parameter is κ that has already been introduced as the salience parameter. So interpreting the equation yields that salience must work multiplicatively with feature-based influences on attention.

$$w_x = \kappa_x \sum_{j \in R} \eta(x, j)\pi_j. \tag{7.2}$$

Another more recent TVA extension by Nordfang et al. (2017) is more complex but also more general by assuming that all location-specific attentional influences work separately from feature-based attention. This idea is expressed formally by Equation 7.3.

$$w_x = \sum_{\text{spatial locations } l} \eta(x, l)\pi_l \sum_{\text{nonspatial features } j} \eta(x, j)\pi_j. \tag{7.3}$$

There is a certain beauty in the symmetry of this formulation: The spatial location, l, as well as the nonspatial features, j, work similar in principle but are summed up individually, and afterward, they are integrated by multiplication.

The two competing models' predictions should be testable with the TOJ design. Both equations have been developed using data from the partial report design. However, the models predictions should generalize to other instances of visual selection and recognition. Equation 7.2 and Equation 7.3 make different prediction for the selection of combinations of location specific and non-location-specific features affecting attention. These combinations can be produced in TOJ. Endogenous and exogenous attention manipulations are common in TOJ designs (e.g., Shore et al., 2001) and allow for a manipulation of different location specific and non-location-specific features.

To come up with two competing hypotheses to test whether Equation 7.2 or Equation 7.3 describes the distribution of attention in TOJ more accurately, a connection has to be made with the attentional manipulations in TOJ. In prior-entry research, manipulation of location-specific attention has been done with endogenous and exogenous cues as reviewed by Shore et al. (2001). Both manipulations affect spatial locations. When examining both equations, it becomes evident that according to Equation 7.2, salience always has a multiplicative impact whereas in Equation 7.3 salience and other location-specific influences should add up in the left sum of the equation. Assuming that the task-relevance of the flicker is always the same, a combination of endogenous, task-relevance of location, and exogenous, salience, attention produces quite different numerical attentional weight because one is based on adding the other equation assumes a multiplication. These thinking yields two mutually exclusive hypotheses:

H_+ The attentional weight of endogenous and exogenous location cues in TOJ add up.

$H_\times$ The attentional weight of endogenous and exogenous location cues in TOJ multiply.

There is, however, a problem in checking how the attentional weights behave. Attentional weights in TVA are only assessable up to a constant. That is, only the ratio can be estimated. For two stimuli, a ration of 0.5 means that their actual weight values may be 1 and 2 or similarly possible 21 and 42.

Before solving this problem, it has to be checked whether endogenous attention manipulations work in the TVA-based TOJ model as expected — when presented individually or in conjunction with salience.

To address the problems systematically, I first report the pre-studies that were conducted to check if endogenous cues and the TVA-based TOJ analysis work as theoretically predicted (the cue affects only the weight parameter, not the overall processing capacity).

Pre-studys

The pre-studies were conducted to develop the actual experiments. There are two difference between the pre-studies and the actual experiments. First, there was no planned sample size. The small and not pre-planned sample size is related to the second difference. Second, the hypotheses for the pre-studies were note scientifically new but derived from general assertions. It is because of these differences that I refrain from a lengthy individual explanation of each pre-study. Instead, I sketch the development of the experiments with its main findings.

To make the abstract description of pre-studies a bit more tangible and illustrate why they are reported at all, I start with an example. In addition to the salience manipulation, I needed a location-specific task-relevance manipulation. Such endogenous attention manipulations are well known. For example, Posner (1980) achieved an endogenous attention manipulation by providing a cue with 80% validity. Consequentially, I designed multiple potential locations for the TOJ, I cued one of them with 80% validity. Four individuals generated data in sessions lasting for over one hour. TVA's attentional weight and overall capacity distributions were strongly overlapping for the valid cue and nonvalid cue condition.

Let me digress for two paragraphs from the description of the pre-studies to analyze what I was doing from an epistemically perspective. Clearly, the limited sample does not suffice for a convincing claim that endogenous cueing does or does not work for TOJs. However, for my particular realization of this cueing procedure four healthy participants did not show a visible shift in the posterior distributions of both, the w and the C parameter. Here, the difference between Bayesian and frequentist thinking becomes evident again: Clearly, four observed subjects means more information for a decision regarding the design than zero observed subjects. The Bayesian approach to statistics now considers the data as fixed. So, the fixed data (information) is used to estimate the model parameters which are treated as not-fixed random variables. In the frequentist logic, the data (events) is randomly drawn from an infinity collective, i.e. the data is treated as a random variable. The model parameters, e.g., a mean difference of 0, is treated as fixed parameter in this logic. Thus, for the frequentist thinking we would have had to define the collective before conducting the pre-study otherwise, it did not just have a low statistical power but cannot be analysed at all — defining the collective post-hoc after seeing the data is a sure way of of "fooling" oneself as Feynman (1974) puts it (for great explanation of the argument — even it the article goes much further, see Gelman & Loken, 2013). Consequentially, I argue that all pre-studies with out pre-planned sample sizes are actually not necessarily bad practice but serve our intuitive Bayesian thinking about the strength of evidence (for arguments that Bayesian logic may or may not be more in line with the questions psychologists pose, see Dienes, 2011).

Following the Bayesian understanding of the pre-study — that was never intended to make a strong public claim — I made decisions regarding the design. In this first attempt, the cued location of probe and reference were located on opposite locations on an imaginary circle around the fixation mark. A central cue, a bar, was used to cue to positions. For example, upper right and lower left, middle left and middle right, or lower right and upper left location. The locations were on the opposite site of the circle such that the distance between probe and reference was equal in all conditions. Based on the pre-study, I decided that the evidence for a successful endogenous attention manipulation caused by the cue was too small to conduct a full study using this design. Note that I neither know whether this manipulations does not work in general or because of my realization or sample. So, simply put, I used the statistics for my own decision process rather than to support a general public claim.

Coming back to the results of the pre-studies, a second pre-study showed that a simple rephrasing of the TOJ task can direct more attention to one of the two TOJ locations:

A central cue, small arrow, indicates which of the two stimuli has to be be judged. Of course, logically the TOJ task remains the same. One cannot judge one of the stimuli without the other. Yet, the phrasing of the task puts sufficient emphasis on the pointed out location such that a strong shift in the attentional weight, w, becomes visible. A further pre-study revealed that this does work comparably even if the other noncued stimulus location is known in advance. This is also the reason why, I report this phenomenon to be caused by the phrasing and not by the prior knowledge of one location while the other location is not yet known to the participant — an explanation that I anticipated to be adequate originally.

A further results of the planning and pre-study phase was that I realized that a possible interaction of endogenous and exogenous attentional manipulation is actually not so easy to determine because TVA's attentional weight can only be determined up to a factor. For example, an attentional weight for stimulus x of $w_x = 3$ does in itself not mean anything because it depends on the other attentional weights. If there was only one further stimulus y, and it had the weight $w_y = 2$ than the actual information contained in the values is that x has a relative advantages of $\frac{3}{2}$. Their individual values can however be as well $w_x = 6$ and $w_y = 4$, or any value combination for w_x and w_y such that it holds that $\frac{w_x}{w_y} = \frac{3}{2}$. The relevance of this fact is that attentional weight cannot be compared easily in a factorial design even though the values can statistically be determined easily. To avoid this problem, the processing speeds of the individual stimuli will be used for analysis. Table 7.1 shows the conditions and comparisons for the intended analyses. There have to be four conditions. They result from a factorial combination of endogenous cue, a small central arrow that indicates which stimulus' flicker hast to be rated, and exogenous cue, low and high salience, either congruent on the endogenously cued stimulus on incongruent on the noncued stimulus.

The basic analysis idea for this design is that a salience increase is much higher if it interacts multiplicatively with another attention manipulation ($H_\times$) than if such a manip-

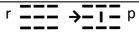
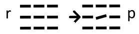
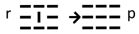
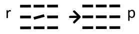
Condition	Sketch	v_p	v_r	Speed-up salience	Speed-up difference
1		$v_{p_{90p}}$	$v_{r_{90p}}$	$v_{\Delta p} = v_{p_{90p}} - v_{p_{15p}}$	$v_{\Delta} = v_{\Delta p} - v_{\Delta r}$
2		$v_{p_{15p}}$	$v_{r_{15p}}$		
3		$v_{p_{90r}}$	$v_{r_{90r}}$	$v_{\Delta r} = v_{r_{90r}} - v_{r_{15r}}$	
4		$v_{p_{15r}}$	$v_{r_{15r}}$		

Table 7.1: The four conditions of pre-study Experiment 2 and the respective processing rates of the probe and reference stimulus.

ulation is merely additive (H_+). To check these hypotheses, the processing speed increase from low to high salience is computed between Condition 1 and 2, as well as Condition 3 and 4, as shown in column "Speed-up salience" of Table 7.1. The column "Speed-up difference" than shows how both rate differences may be compared by subtraction.

The data from the pre-study implementing the design presented in Table 7.1 did not show a difference caused by the different levels of the salience manipulation. Although this was not expected — previously orientation differences increases caused clearly observable changes in attentional weight, I decided to conduct a full experiment because odd results may have been caused by the limited sample and the design was far enough developed to provide a potentially interesting result — in the sense that this combination of manipulations and analysis has not yet been conducted.

Summing up, the work in Copenhagen resulted in a research question, experimental design, and analysis to answer whether a recent TVA extension provides a better explanation of the TOJ than the earlier model.

Full Experiments

After the initial idea and pre-studies developed in Copenhagen, I refined the idea further and conducted two Experiments. These Experiments aimed at specifically discerning whether the weight equation by Nordfang et al. (2013) or Nordfang et al. (2017) is more fitting in the TVA-TOJ setting with two factors of spatial attention. Two hypotheses have been derived during the introduction to these sections:

H_+ The attentional weight of endogenous and exogenous location cues in TOJ adds up.

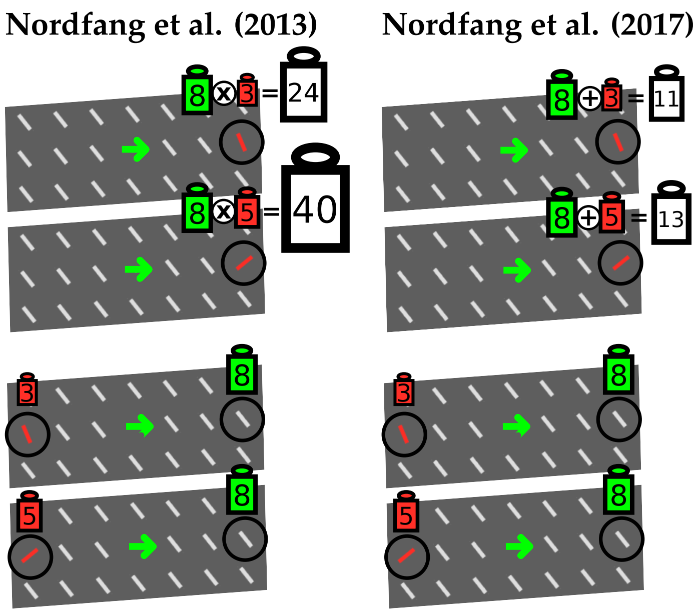


Figure 7.9: Sketch of possible attentional weights and their interaction.

$H_\times$ The attentional weight of endogenous and exogenous location cues in TOJ multiply.

The design is equal to the design of the pre-study of Table 7.1. However, because the pre-study experiments have suggested that salience combined with the endogenous cue does not change the processing rates much, a stronger salience cue has been used. The salience cue consists of a 90° orientation contrast and a color contrast (where the salient element is red and the surrounding gray).

Again, in this design, checking the hypotheses H_+ and $H_\times$ is difficult because the absolute attentional weight cannot be estimated by TVA. A comparison of processing rates between conditions in which both cues coincide at the same location an conditions in which they appear at separate locations can be made.

A hypothetical distribution of attentional weight according to both hypotheses is sketched in Figure 7.9. Whereas, according to Nordfang et al. (2013), salience has to multiply with any other attentional factor, the newer formulation suggests that different influences of spatial attention add up similarly as different feature-based attentional influences add up in the original TVA weight equation (Nordfang et al., 2017). The sketch illustrates the reasoning behind the four conditions in the experiment. To help to explain the reasoning behind the conditions, the term congruent is helpful: congruent conditions are conditions in which both attentional manipulations are applied to the same stimulus, whereas incongruent conditions are conditions in which one stimulus is particularly task-relevant while the other stimulus is salient.

If indeed the influences multiply, then the difference between the congruent conditions should be much higher in terms of attentional weight than the difference between the incongruent conditions. Table 7.1 showed how the differences in processing rates caused by an increase in salience could be used to test the hypotheses: The effect of an increase in salience can be observed in the congruent and in the incongruent conditions. If salience and endogenous cue affect the rates additively (H_+), a difference in rates will be observed comparable to the difference between low and high salience in the incongruent condition. If both cues work multiplicatively ($H_\times$), however, a much larger difference in rates should be observable between the low and high salience processing rates for congruent conditions are compared to those of incongruent conditions.

Experiments

Two experiments were conducted that enabled a rate comparison, as explained in the previous paragraph. Both used a 2×2 design with low and high salience either in a congruent or incongruent condition, i.e., at the same or different location as an endogenous cue.

The endogenous cue was implemented by instructing the participants to judge whether the stimulus at the indicated location flicked first or second. This manipulation had been tested successfully in the described series of pre-studies.

Experiment 1 used an orientation and color contrast. The color of the salient stimulus was always red. The experimental factor was whether it was its orientation salience was low or high. There were 29 participants.

Experiment 2 used motion contrast and orientation contrast. To better see possible differences caused by the factor salience, motion and orientation were increased from the level low to the level of high salience. There were, again,

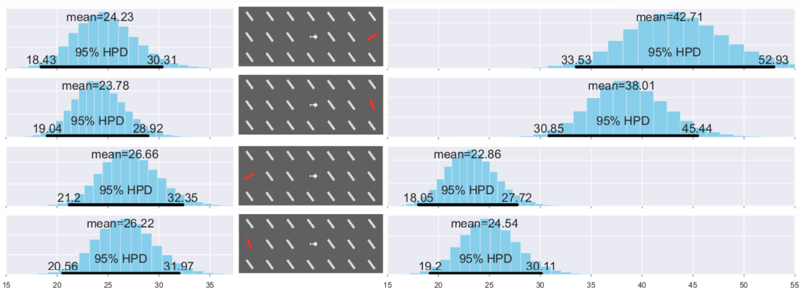


Figure 7.10: Estimated distribution of the speed-up difference of Experiment 3.

29 participants.

Results and interpretation

The the estimated mean processing speeds are showen in Figure 7.10 for the first experiment and in Figure 7.11 for the second experiment.

By comparing the high salience with the low salience condition, it becomes obvious that there is no clear main effect of the salience manipulation. This is a problem for the planned comparisons because they assume that each manipulation has a clear main effect — merely the interaction was assumed to be unknown.

The absence of a clear redistribution caused by salience indicates that whatever process combines endogenous and exogenous cues is more complicated than initially assumed.

So, comparing the introduced hypotheses by subtracting processing speeds is not possible. A preliminary conclusion from this line of research is that although TVA models many details, its applicability even if circumstances are slightly changed has to be evaluated carefully. Still this detailed description can be seen as an advantage because a deviation from expected behavior may not become obvious unless explicitly stated.

Discussion

This section marks a loose end but also symbolizes a learning experience. First, I will discuss the loose end aspect of this series of experiments. Afterward, I will draw a conclusion for working with TVA and formal modeling in

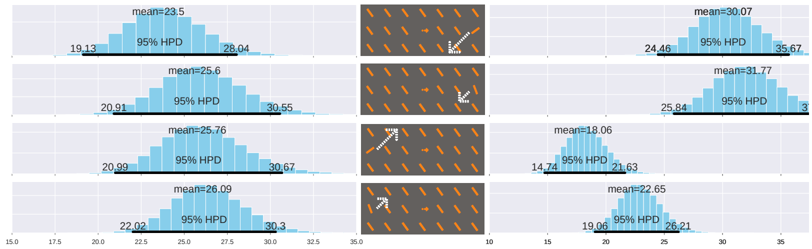


Figure 7.11: Estimated distribution of the speed-up difference of Experiment 3.

general. The reason for the loose end characteristic of this series of experiments is that the results of both experiments show that the initial theory-driven assumption that only the attentional weights change throughout the experimental conditions is wrong. The attentional weights change indeed. If forced to give an answer to the initial research hypothesis, it is possible to say that weights add up according to the 2017 formulation of the weight equation rather than the multiplicative interaction required by the 2013 formulation. This is justified by the comparable difference in attentional weight from between the congruent and incongruent conditions: In the congruent condition, the weight difference is not larger than the difference in the incongruent conditions. However, this answer does only work superficially and is not justified after closer examination: It is obvious that the overall capacity changes form the congruent to the incongruent cases. Succinctly put, the difference between congruent and incongruent cases is thus no mere redistribution of the same resources. Such a redistribution was, however, initially assumed (and tested positively in pre-studies). Thus, it was initially assumed that both central cue and salience are supposed to affect attentional weight exclusively. Whereas the result to the contrary sparks an interesting research question in itself (How do the congruent and incongruent conditions affect visual processing speed), it makes a comparison of attentional weights difficult: Attentional weights in TVA are always relative because when all of them are multiplied by a certain factor, for example, α, the observable result stays exactly the same. In so far, the sketch in Figure 7.9 maybe misleading the used weight should not be 8, 5, and 3 but 8α, 5α, and 3α instead. If the overall capacity stays the same, it is reasonable to assume the same frame of reference because the resources of the same process are redistributed. If the process changes, however, there may be the weights 8α, 5α, and 3α in the situation A, whereas 8β, 5β, and 3β are the weight in situation B. They will look indistinguishable, yet β may be quite difficult from α. There may be different arguments about the meaning of the comparisons between the experimental conditions in the described experiments.

A possible explanation for the observed results may be that attention helps to resolve ambiguity in the TOJ design: The TOJ design creates situations that are difficult to resolve correctly because of the limited evidence that short SOAs provide for the actual order. It may be possible that attention is then used as a prior in Bayesian statistics: If there is only a little evidence, the prior becomes more and more decisive for how the posterior looks. In terms of the Bayesian Brain perspective on cognition, the Brain would simply use another readily available signal to disambiguate a situation in which evidence is too little for reaching a decision boundary. If this perspective was adequate than it is easy to see that one attention-increasing manipulation on both stimuli creates much more ambiguity than two attention-increasing manipulations on one stimulus, thus the task becomes easier in the latter case, which would be associated with higher overall processing capacity, C, that was indeed found. This hypothesis could be easily tested by testing whether C stays constant in an experiment with either one salient stimulus or both stimuli salient. Whereas we conducted a comparable experiment as a pre-study in the lab once and did not see this behavior, it may be particularly hard to disambiguate if the attentional cues are not from the same type. A TOJ experiment where there was a goal-directed and a stimulus-driven at the same time was indeed not conducted within our lab. If only this set up makes disambiguation difficult, it could be counted as evidence for disjoint representations of salience and relevance such that incongruent cues within these representations (e.g., multiple orientation contrast manipulations) are easily resolved whereas incongruencies between salience and relevance cannot be used conjointly to disambiguate in the TOJ design.

For the time being, it is, however, rigorous to not draw speculative conclusions from these experiments. This inconclusiveness is why this section marks a loose end. However, it also marks a learning situation.

In hindsight, I would not again derive such specific conclusions in advance and base a whole experiment on such specific predictions. On the other hand, it is also not advisable to apply a method of analysis that not so sensitive for the theoretically different concepts. For example, computing the point of subjective simultaneity would surely yield a result, yet the change in the difference limen would probably be simply ignored. In so far, the experiments made the point that it is good to be so specific that one can be wrong. In particular, when understood as biased competition, goal-directed and stimulus-driven attentional manipulations do not simply change the attentional weights when modeled as independent fixed-capacity races. Without a formal model, this mismatch between assumptions and data patterns would not have become evident.

Chapter 8

Conclusion

> Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful.
>
> George E. P. Box, (Box & Draper, 1987, p. 74)

This work started with a simple question: "If a physical contrast attracts attention and is consequentially salient, how strongly does it affect attention?". The cover image illustrates this point by showing two green horizontal bars one of which — upper right corner — is clearly salient whereas the other — lower left corner — is clearly non-salient. In between both extremes lies a continuum of potential contrasts with varying degrees of salience. So, how can we measure and compare salience from different contrasts? [1].

[1] Actually the image is part of the animation that varies the salience of two elements by changing the background elements as shown in Figure 8.1

As a solution, I propose to use a model of visual selection based on TVA to assign numbers to the contrasts' salience based on the strength of their impact on attention. The theoretical foundation has been developed by Nordfang et al. (2013), but their suggestions were very difficult to apply because prestudies had to ensure that salience manipulations do not affect other parameters such as sensory evidence. Thus, I used a TVA-based interpretation of the TOJ task developed by Tünnermann et al. (2015) to combine both with cluttered multi-element displays. I presented three articles (Article 1, 2, and 4) developing and applying this modeling and empirical approach together with further experiments that support the broad applicability. The method presented allows salience estimation, model comparisons

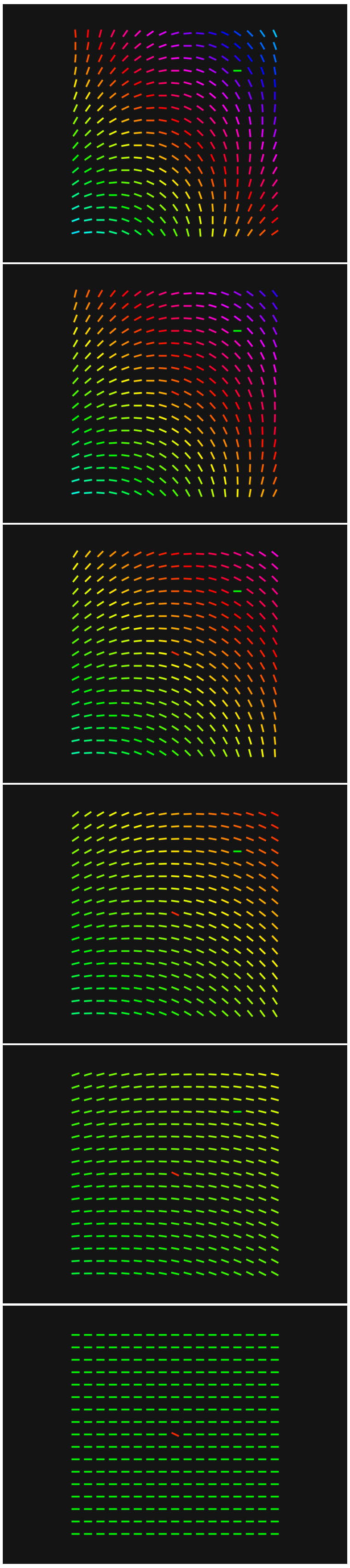


Figure 8.1: An animation illustrating different degrees of salience by systematically changing the surrounding of three fixed stimuli (lower left, upper right corner, and center).

(to test alternative explanations) and the prediction of new data (e.g., for checking the fit or simulating yet unobserved conditions). Furthermore, the proposed method can be used with many physical contrast and timing manipulations. Also, the task is simple enough to be embedded in further tasks or complex environments. These results are limited by the acceptance of the auxiliary assumptions for modeling (e.g., no evidence accumulation is used in this simple TVA model). However, because the assumptions are incorporated in the formal model, a critic may pin criticism on specific parts and even compute a model comparison once their own idea has been incorporated into a competing model. As such, I do not present *the* quantification of visual salience but *a* quantification, a way how to scientifically deal with the strength of salience so that theory and data are well connected by a model that is arguably not wrong enough to not be useful.

Before I draw a more elaborate conclusion, let me argue why a psychologist seeking to understand attention should be interested in the estimation of salience at all. There are two parts to this answer: First, theoretically we know a lot about attention. Second, we do not explicitly use this knowledge for our analyses. So, instead of using the theory to specify what we expect to find in the data, we set up null hypotheses that we are not actually interested in. If such a null hypothesis is rejected, we use this as evidence for an alternative that supposedly matches previous results without stating this alternative model explicitly. Results obtained by this method are often overestimated with respect to their inferential power (Oakes, 1986; Gigerenzer, 2018; Cohen, 1990; Cumming, 2013). This overestimation of hypothesis tests may lead to the illusion that other statistical tools are not needed and may even be harmful as they seemingly introduce less “objective” practices. Rouder, Morey, Verhagen, et al. (2016) call this the “free lunch” illusion and argue for the value of modeling the hypotheses of individual studies to test the scientifically interesting explanations rather than theoretically uninteresting null models. Explicitly modeling alternatives does not mean to become a proponent of Bayesianism in the epistemic sense;

the method works fine with the hypothetico-deductive scientific method (Gelman & Shalizi, 2013).

I find it very important to remark that using NHST techniques is not a bad decision in general. However, it is most suitable when little is known about the process that causes the data (Little, 2006; Efron, 2005). Thus, for attention it is possible and advisable to explicitly state the model that is supposed to be true according to the researchers' expectations from the literature. The lack of (formal) theorizing and modeling can severely limit empirical research as it becomes difficult to judge which results are in line with previous works and which are surprising. To a degree, it explains a part of the reproducibility crisis discussed in Section 2 (Muthukrishna & Henrich, 2019).

For visual attention, formal frameworks exist that already helped accumulating knowledge. Most notably similarity-choice based theories — TVA is one of those — have been reviewed to provide the best explanations when compared to other formal approaches to attention(Logan, 2004). Although modeling has been arguably neglected in the past, a recent rise and interest in modeling has occurred that is inconspicuous to a degree that Rodgers (2010) calls it a quiet revolution. Taagepera (2008), for instance, distinguishes between descriptive models — those that fit the data — and logic models — those that reflect the theoretical judgments about the observed quantities. He argues that incorporating theory into models makes social science and psychology more scientific. In fact, this fits Popper's opinion that tests have to be severe to be useful.

The advantages of parameter estimation as a statistical technique have been advocated by Cumming (2013). Additionally, Taagepera (2008) argues that it is even better to estimate parameters that are of theoretical import, and Kruschke continues that Bayesian methods are even better for estimation (Kruschke & Liddell, 2018), while, amongst others, Rouder and Lu (2005) argues that hierarchical models are particularly apt for nonlinear models of cognition. Yet, the connection between scientific reasoning and statistical reasoning is rarely spelled out explicitly.

Bailer-Jones (2009) presents a historical review and a consequential philosophical position that models link data and theory in the hypothetico-deductive scientific method. A core idea is that models are applied hierarchically: Not only the data has to be modeled to enable statistics but also the experimental setting has to be interpreted with a theory to deal explicitly with auxiliary hypotheses needed to make the theory applicable. Applying theories is, according to Bailer-Jones (2009), never straightforward because they are by design maximally abstract to be as general as possible. This property comes at the cost that an explicit connection to a situation has to be grounded by additional assumptions. If these assumptions are made implicitly, it is very difficult to identify what exactly has been falsified when a finding is apparently not in line with a theory. It is a faulty theory or an auxiliary assumption that has been shown to be wrong. A side effect of this method is that it produces explanations that link theoretical ideas with observable data not by appealing to scientific laws[2] but apt simplifications — models, a thought that is spelled out in Article 3 and inspired by Bailer-Jones' (2009) work on scientific models.

[2] Although the Hempel-Oppenheimer model is initially appealing, it is very hard to defend epistemically.

So what do these ideas in modeling have to do with understanding attention and estimating a salience parameter? Attention is a form of selection. Stimuli compete for representation and are thus not attended to because of a property they have (as, e.g., done by James, 1890) but by comparing their properties to those of other stimuli in the visual field. Attention gates the access to limited resources like representation for conscious perception. Thus, already small differences may result in a different pattern of selected stimuli. To understand whether a salience manipulation does lead to a result in line with the literature, it is important to make auxiliary assumptions explicit e.g. in a formal model. This also allows backtracking if a result apparently refutes the theory: Does it really refute the theory or an assumption made particularly for the present experimental setup?

Summing up, we know a lot about attention and need to explicate this knowledge to see whether results are in line with this body of research. Especially, when much is already known, NHST may create an illusion of free lunch, i.e., that significant results can be obtained without referring explicitly to previous theory in the statistical procedure (Rouder, Morey, Verhagen, et al., 2016). Explicitly appealing to the theory of attention is important to see whether salience manipulations actually affect attention and selection via attention or by other mechanisms. If, on the other hand, salience always affects selection via attention, then this explicit quantity may be used to check for a common currency of salience as hypothesized by e.g., Treue (2003).

Explicitly incorporating theory is both an advantage and a limitation of the presented approach. If a sequential process model with inhibition of return is assumed, as in many computational salience models, the actual numbers estimated for salience in TVA'a κ would be bereft of their original meaning — a meaning that is defined by TVA's equations that assume parallel processing. This is not to say that the κ estimates are necessarily quantitatively different from estimates based on such a different model.

Only the future will show whether the TVA-based salience measure is useful in research. We already undertook a couple of experiments to test how TVA can be applied in situations with less experimental control than in those presented in this dissertation and the results promise that precise estimation of attentional parameters like salience is possible in a more applied context. Beyond a potential practical usefulness, it is my hope that this work among many others shows that psychology has strong, if not necessarily universal, theories about cognition.

Appendix

Erweiterte deutsche Zusammenfassung

Visuelle Auffälligkeit bestimmt, wie wir die Welt wahrnehmen. Sie hat einen direkten Einfluss auf die visuelle Aufmerksamkeit. Während diese Tatsache bereits seit langem bekannt ist, ist es umso verwunderlicher, dass eine genaue quantitative Bestimmung der visuellen Auffälligkeit bis heute noch nicht gelungen ist, obwohl immer wieder die Annahme geäußert wird, dass sich verschiedene Faktoren zu einer gemeinsamen Aufmerksamkeitswährung verrechnen lassen (Treue, 2003). Ziel dieser Arbeit ist es, einen theoretisch gerechtfertigten quantitativen Wert für die visuelle Auffälligkeit zu bestimmen. Die Schwierigkeit dabei besteht darin, dass die visuelle Auffälligkeit zwar von physikalischen Kontrasten hervorgerufen wird, die für sich genommen gut zu messen sind, wie auffällig diese Kontraste jedoch für den Betrachter sind, nicht direkt beobachtbar ist. Ebenfalls ist die Aufmerksamkeitsausrichtung im Raum nicht direkt zu beobachten.

Der Einfluss von visueller Auffälligkeit auf die visuelle Aufmerksamkeit kann auch mit dem Begriff Salienz beschrieben werden. Bisherige Methoden, die den Einfluss von Salienz auf die visuelle Aufmerksamkeit quantitativ erfassen, sind sehr vielfältig und in ihren Ergebnissen teilweise widersprüchlich, wenn es um den Einfluss von Kombinationen von individuell salienten Kontrasten handelt. Es sich wichtig, darauf hinzuweisen, dass Salienz unterschiedlich verstanden werden kann. So kann es sich um eine Eigenschaft handeln, die die Aufmerksamkeitsausrichtung betrifft (saliente Reize werden stärker beachtet) oder die die Wahrnehmung betrifft (saliente Reize erscheinen uns als auffällig). Für ein Beispiel, in dem es um Wahrnehumg geht, siehe Kerzel u. a. (2011a). Insbesondere bei algorithmsichen Salienzmodellen ist oft unklar, welches Konstrukt sie konkret modellieren (What do salience models predict). Diese Modelle zeigen jedoch auch die Interdisziplinarität des Themas: Auch künstliche Handelnde, etwa Roboter, können von Aufmerksamkeitsmechanismen profitieren, und es hat sich um die Algorithmische Modellierung eine eigene Subdisziplin gebildet (für eine Übersicht siehe Frintrop). Während einige Ideen aus diesem Bereich wertvoll sind[3], konzentriert sich diese Arbeit auf die psychologische Literatur.

[3]Ein Beispiel ist das auf informationstheoretischen Annahmen basierende Modell von Bruce und Tsotsos (2009), das grob gesprochen die Stärke von Salienz darin begründet sieht, wie gut sich ein Bildinhalt durch seine Umgebung vorhersagen lässt. In dem Gedankenexperiment in Abschnitt 3 kann man erproben, dass dieses Konzept sich gut mit dem Alltagsphänomen Salienz deckt.

Für diese erweiterte Zusammenfassung kann man die relevantesten Befunde zu Salienz wie folgend zusammenfassen. Wie bereits erwähnt ist vorstellbar, dass Salienz zu einer gemeinsamen Aufmerksamkeitswährung beiträgt, sich also qualitativ verschiedene Einflüsse auf Aufmerksamkeit in einer internen quantitativen Repräsentation zusammenfassen (Treue). Zentral ist dabei, Salienz nicht mit Aufgabenrelevanz zu konfundieren, was leicht bei Suchaufgaben passiert, in denen die stimulusgetriebenen Einflüssen von Kontrast damit konfundiert sind, dass genau diese das Zielelement kennzeichnen. Dies muss keinesfalls immer problematisch ein, schränkt möglicherweise jedoch die Generalisierbarkeit ein. Für eine aktuelle Taxonomie von Aufmerksamkeit siehe Chun u. a. (2011)[4]. Im Kontrast zu frühen Ideen über Auffälligkeit etwa bei James (1890) sind einzelne Eigenschaften nicht etwa per se auffällig, sondern weil sie von ihrer Umgebung in entscheidenden Merkmalen unterschiedlich sind. Dies wurde von Duncan und Humphreys (1989) gezeigt, indem einerseits die Ähnlichkeit von auffälligem Reiz und Hintergrund variiert wurde, andererseits aber auch die Homogenität des Hintergrunds. Diese Experimente zeigen klar, dass Salienz durch den Kontrast zur Umgebung besteht.

[4]Für mich ist es immer noch überraschend, wie stark aktuelle Taxonomien noch immer der von James (1890) ähneln.

Ebenfalls eine zentrale Eigenschaft von Salienz ist, dass sie zumindest was die Beeinflussung der Aufmerksamkeit angeht, zeitlich nicht konstant ist. Reviews zeigen, dass stimulusgetriebene Aufmerksamkeit, zu der auch Salienz zählt, nur in bestimmten Zeitfenstern nach Onset stark auf die Aufmerksamkeit einwirken. Für Salienz wurde dies unter anderem von Donk und van Zoest (2008), Donk und Soesman (2010) gezeigt.

Vielleicht ist es offen sichtlich für die Leserin oder den Leser, aber es gibt natürlich mehrere Arten von Kontrast, die nicht alle zur Salienz beitragen Wolfe und Horowitz, 2017; Wolfe und Horowitz, 2004. Daher werden in der vorliegenden Arbeit nur solche Kontraste verwendet, deren Einwirkung auf die Aufmerksamkeit als gesichert gelten kann.

Die vorliegende Doktorarbeit beschäftigt sich mit einem weiteren Verfahren um Salienz zu messen. Jedoch unterscheidet sich das Herangehen von früheren Arbeiten, indem diesem neu erarbeitete Salienzmaß ein Modell zugrunde liegt. Der Vorteil des Modells ist, dass die Verbindung der beiden latenten Variablen, Salienz und visuelle Aufmerksamkeit, im Bezug auf die visuelle Selektion explizit formuliert wird.

Das verwendete Modell basiert auf Bundesens Theorie der visuellen Aufmerksamkeit (TVA; Bundesen, 1990), die mittlerweile seit 30 Jahren existiert und kontinuierlich erweitert wird. Bundesens Theorie zeichnet sich dadurch aus, dass sich verschiedene theoretisch relevante Parameter der visuellen Aufmerksamkeit mit hoher Präzision individuell bestimmen lassen. Dies macht die Theorie zum einen für die Grundlagenforschung, zum anderen aber auch für angewandte Forschung im Bereich der klinischen und der Neuropsychologie interessant. Zu den Parametern der Theorie zählen die visuelle Verarbeitungsgeschwindigkeit, v, Aufmerksamkeitsgewichte, w, sowie weitere Parameter, die für diese Arbeit von geringerer Bedeutung sind.

Der zentrale Erfolg dieser Arbeit besteht darin, die Forschung zu visueller Salienz mit der Theorie von Bundesen zu verbinden, in dem ein Experimentaldesign und Modell entwickelt wurde, welches die Displays aus der Salienzforschung mit einem theoriebasierten Modell zusammenbringt. In der Salienzforschung werden üblicherweise Displays verwendet die sehr viele Elemente beinhalten (multielement displays), dazu gibt es mindestens ein Element, das eine besondere, eine saliente, Eigenschaft aufweist. Derartige Salienzmanipualationen lassen sich nur mit erheblichem

Aufwand mit den typischen von der TVA modellierten Designs einsetzen (Nordfang u. a., 2013). In der vorliegenden Arbeit wurde für das Experimentaldesign auf das zeitliche Reihenfolgeurteil (temporal-order judgment; TOJ) zurückgegriffen. Eine Möglichkeit, dieses Design und TVA zu verbinden, liefert Tünnermann u. a. (2015). Aufbauend auf dieser Arbeit werden Multi-element displays mit TOJ kombiniert sowie ein spezielles TVA-basiertes Modell entwickelt, in dem die Salienzstärke einen eigenen Parameter, κ, erhält, der die angestrebte Quantifizierung der Salienz darstellt.

Bundesens Theorie der visuellen Aufmerksamkeit

Bundesens Theorie der visuellen Aufmerksamkeit ist eine Theorie des Wiedererkennens und Auswählens von visuellen Reizen. Historisch betrachtet ist die Theorie eine Kombination von Luces Auswahlregel (Luce, 1959) mit den sogenannten Race Modus (Shibuya & Bundesen, 1988). Dies kann man sich vereinfacht als ein Pferderennen vorstellen, bei dem alle visuellen Stimuli wie Pferde in einem Wettrennen gegeneinander antreten: Nur die Schnellsten der visuellen Reize werden im begrenzten visuellen Kurzzeitgedächtnis encodiert (entspricht dem Race-Modell). Andere Eigenschaften der formalen Theorie von Bundesen passen nicht sehr gut zu der Analogie des Pferderennens: Zum Beispiel wird angenommen, dass die gesamte visuelle Verarbeitungskapazität für eine Aufgabe immer gleich ist. Jedoch wird diese Kapazität durch Aufmerksamkeit auf die unterschiedlichen Reize im visuellen Feld verteilt. Dabei kommt Luces Auswahlregel zum Einsatz, die besagt, dass einerseits die Ähnlichkeit des Reizes mit eine Kategorie beachtet werden muss, andererseits auch die Aufgabenrelevanz von verschiedenen Kategorien. So kann ein Reiz zu einer wichtigen Kategorie gehören und sollte daher auffällig sein. Gleichzeitig kann derselbe Reiz aber nur schwer als zu dieser Kategorie gehörig zu erkennen sein. Beispielsweise kann man sich vorstellen, dass der Buchstabe "O" mit der Ziffer "0" verwechselt wird, sodass, wenn Buchstaben berichtet werden sollen, der Buchstabe "O" weniger Aufmerksamkeit auf sich zieht als andere Buchstaben, weil er für eine "0" gehalten werden kann.

Formal basiert Bundesens Theorie auf mehreren Gleichungen, die ungefähr das formal ausdrücken, was die vorherige Analogie ausdrückt. Bei diesen Gleichungen handelt es sich um die Ratengleichung und um die Gewichtsgleichung.

$$v(x,i) = \eta(x,i)\beta_i \frac{w_x}{\sum_{z \in S} w_z} \tag{8.1}$$

Die Ratengleichung, Gleichung 8.1, gibt an, mit welcher Rate ein bestimmter Reiz als eine bestimmte Kategorie encodiert wird. Während dies auch das Modellieren von Verwechslungen erlaubt, dass z.B. ein q als p encodiert wird (siehe Tünnermann cues fate), werden relevante Verwechslungen durch das Experimentaldesign minimiert. Vereinfacht kann man die Gleichung also so verstehen, dass sie angibt, mit welcher Rate ein Reiz ins VSTM encodiert wird. Es handelt sich um eine Rate, da die TVA die tatsächliche Ankunftszeit im VSTM durch eine Hazardrate in einem fixed-capacity independent-race Modell (FIRM; Shibuya & Bundesen, 1988) modelliert.

$$w_x = \sum_{j \in R} \eta(x,j)\pi_j \tag{8.2}$$

Die Gewichtsgleichung, Gleichung 8.2, gibt das Aufmerksamkeitsgewicht für einen Reiz x an. Sie basiert auf mehreren Parametern, die mit der Aufgabenrelevanz und der sensorischen Evidenz, die vom Reiz ausgeht, zusammenhängen. Für eine genauere Beschreibung, siehe Abschnitt 4.

In den vergangenen drei Jahrzehnten wurde diese Theorie beständig weiterentwickelt: Einerseits erläutert die NTVA (Bundesen u. a., 2005, 2011) die neuronalen Grundlagen der Theorie und dient zur Herleitung empirisch testbarer Vorhersagen (Li, Kozyrev, Kyllingsbæk u. a., 2016), andererseits gibt es immer wieder Überlegungen, den Gültigkeitsbereich der Theorie auszuweiten, zum Beispiel in Bezug auf zeitliche Episoden (Schneider, Anders) oder das Modellieren von Reaktionszeiten (Blurton). Darüber hinaus wird die TVA und die durch sie mögliche Aufmerksamkeitsparameterschätzung in vielen Arbeiten verwendet. Dabei handelt es sich zum Beispiel um Grundlangenforschung (z.B. Vangkilde), klinische Forschung (z. B., Finke u. a., 2005) und Neuro-Biopsychologie (Li, Kozyrev, Kyllingsbæk u. a., 2016, z. B., wie bereits erwähnt,).

Von besonderer Relevanz für diese Doktorarbeit sind die Veröffentlichungen von Nordfang u. a. (2013), Nordfang u. a. (2017), die sich mit der Ausweitung der Theorie auf die Auswirkung von physikalischen Kontrasten intensiv beschäftigt hat. Während die Basisgleichungen von Bundesen (1990) erfordern, dass nur Eigenschaften Aufmerksamkeit auf sich ziehen können, die eine Relevanz ungleich 0 haben, so vermögen doch auch irrelevante Reize auch Aufmerksamkeit auf sich zu lenken (z.B., Huang & Pashler, 2005). Die erste Arbeit, Nordfang u. a. (2013), bildet eine Grundlage um derartige Phänomene in der TVA zu beschreiben, weil sie den theoretischen Parameter κ einführt und eine Abwandlung der Gewichtsgleichung vorstellt, bei der der Parameter eine zentrale Rolle spielt. Jedoch zeigt die Untersuchung auch, wie kompliziert das Verwenden einer Salienzmanipulation mit dem partiellen Report ist. Bereits eine Salienzmanipulation mit einer Stufe erforderte eine eigene Vorstudie, da die Salienz sich auf die sensorische Evidenz und folglich auf die Aufgabenrelevanz des Reizes auswirken könnte. Aufbauend auf dieser Arbeit haben Nordfang u. a. (2017) die Frage gestellt, ob sich die TVA–Gewichtsgleichung für räumliche Aufmerksamkeit (statt der typischen feature-basierten Aufmerksamkeit) generalisieren lässt. Das Ergebnis ist, dass räumliche Aufmerksamkeit ebenfalls nach der Auswahlregel von Luce verteilt wird. Beide Faktoren multiplizieren sich im neu entwickelten Modell.

Zeitliche Reihenfolgeurteile

Wie bereits im vorherigen Absatz beschrieben, ist der Einsatz von Salienzmanipulationen im partiellen Report-Design schwierig. Dabei kommt die Vorarbeit von Tünnermann et al. (2015) zum Tragen: Tünnermann hat bereits die TVA–

Gleichungen benutzt, um das zeitliche Reihenfolgeurteil (temporal-order judgment; TOJ) (temporal-order judgment, TOJ; für eine Übersicht siehe Shore u. a., 2001; Spence & Parise, 2010) zu modellieren.

Das zeitliche Reihenfolgeurteil-Paradigma besteht aus zwei Ereignissen. Oft sind die Ereignisse Onsets von Reizen, welche sich im Rahmen des Artikels 1 für die TVA-basierte Salienzmessung jedoch nicht als angemessen herausgestellt haben. Diese Ereignisse werden als Probe und Reference bezeichnet, um sie unterscheidbar zu machen. Dabei ist Probe das Ereignis, bei dem potentiell eine Aufmerksamkeitsmanipulation stattfindet, während Reference dem Namen entsprechend einen Referenzpunkt bietet. Die Schwierigkeit im Bewerten der zeitlichen Reihenfolge von Probe und Reference besteht darin, dass sie von einem zeitlichen Intervall getrennt sind (SOA), sodass das eine Ereignis kurz vor dem anderen eintreten kann oder umgekehrt. Die Urteile der Versuchspersonen stellen eine genauigkeitsbasierte abhängige Variable da.

Das TOJ lässt sich mit der TVA modellieren. Das Modell lässt sich so erläutern, dass die zwei Ereignisse, die im zeitlichen Reihenfolgeurteil zu beurteilen sind, durch ein Wettrennen der beiden modelliert werden. Hierbei ist zentral dass die TVA zwei Verarbeitungswellen unterscheidet (z.B., Tünnermann u. a., 2015). Die erste Welle bestimmt die Aufmerksamkeitsgewichte der einzelnen Reize, während in der zweiten Welle die Reize in das visuelle Kurzzeitgedächtnis encodiert werden. Tünnermanns Modell befasst sich zunächst mit der zweiten Welle. Formal betrachtet können die Ratengleichungen für Probe und Reference — diese Namen dienen als Bezeichner für die zwei zu beurteilenden Ereignisse — so umgeformt werden, dass sich eine psychometrische Funktion ergibt — nicht unähnlich zu jenen psychometrischen Funktionen, die bereits häufiger für die Analyse von TOJs verwendet werden. Im Unterschied zu diesen Funktionen ist die auf der TVA basierende Funktion jedoch aus einer Theorie hergeleitet und bietet somit die Möglichkeit Parameter mit exakter theoretischer Bedeutung zu schätzen.

Für den mit den TOJ vertrauten Leser lohnt sich ein Vergleich mit den typischen psychometrischen Funktion (z.B., Wichmann & Hill, 2001; Kuss u. a., 2005), die in diesem Bereich eingesetzt werden. Diese Funktionen zeichnen sich durch 2 Parameter aus, den Punkt der subjektiven Gleichzeitigkeit (point of subjective simultaneity; PSS) und das differnce limen (DL), das umgangssprachlich die Steilheit der Steigung der Funktion bestimmt. Ohne eine unterliegende formale Theorie wird dabei der PSS als der Einfluss der Aufmerksamkeit verstanden und das DL als ein Maß für die Aufgabenschwierigkeit oder Genauigkeit des Urteils. Diese Art der beschreibenden Modellierung wird von Taagepera (2008) auch deskriptives Modell genannt, da es ein kompakte Beschreibung der Daten darstellt, nicht aber die Logik deren Entstehung berücksichtigt.

Das TVA-basierte Modell verfügt ebenfalls Über zwei Parameter. Diese sind wie bereits erwähnt aus der Theorie hergeleitet und beschreiben die Verarbeitungsrate des einen sowie des anderen Reizes. Taagepera (2008) nennt diese Art von Modell ein logisches Modell, da es die Daten auf der Basis von Überlegungen zu deren theoretische Ursprung modelliert. Die beiden Verarbeitungsraten lassen sich in die Kapazität und das Aufmerksamkeitsgewicht umrechnen, sodass ähnlich wie bei den klassischen psychometrischen Funktionen zwei Parameter getrennt existieren. Während der eine Parameter den Einfluss von Aufmerksamkeit beschreibt und der andere Parameter die Kapazität die zur korrekten Lösung der Aufgabe zur Verfügung steht.

Zeitliche Reihenfolgeurteile können ebenfalls mit Multi-Element Displays Kombiniert werden (Donk & Soesman, 2011) Diese Arbeiten gehen jedoch nicht modellbasiert vor, nicht einmal mit einem der klassischen psychometrischen Modelle.

Bayessche Statistik

Die Bayessche Statistik lässt sich am besten im Kontrast zu der frequentierte Statistik einführen. Den beiden Formen von Statistik liegt dabei ein unterschiedliches Wahrscheinlichkeitskonzept zugrunde. Dies bedeutet, dass es sich bei beiden Wahrscheinlichkeitskonzepten um Wahrscheinlichkeiten im mathematischen Sinne handelt, sprich sie den mathematischen Anforderungen an einer Wahrscheinlichkeit genüge tun, jedoch in der Welt etwas anderes bedeuten. Dies lässt sich am besten mit einem konkreten Beispiel erläutern: Überlegen sie, wo sie ihr Auto abgestellt haben oder wo Sie es abstellen würden. In der frequentistischen Denkweise gibt es nun nur 2 Möglichkeiten für die Position ihres Autos: Entweder das Auto steht noch an dem von Ihnen ausgewählten Ort oder es steht an einem anderen Ort. In der Bayesianischen Denkweise können Sie im Gegensatz dazu eine Wahrscheinlichkeit dafür angeben, dass die Proposition "Mein Auto steht an Ort X." wahr ist. Dieser Unterschied ist darin begründet, dass in der Bayesschen Statistik für jede Proposition eine Wahrscheinlichkeit per Definition angegeben werden kann, während in der frequentistischen Denkweise ein Kollektiv gebildet werden muss, für das dann eine Wahrscheinlichkeit berechnet werden kann. Diese Unterschiede können, obwohl sie erst einmal oberflächlich scheinen, tiefgreifende Folgen haben. Interessanterweise führen die epistemisch unterschiedlichen Ansätze für viele konkrete Statistikproblem zum selben Ergebnis (855 t-tests paper). Eine gute Einführung in die induktive Logik und somit den Erkenntnistheoretischenteil bietet Hecking an. Eine gute praktische Einführung findet man Kruschke (2014) und einen Vergleich beider Methoden für die Psychologische Forschung bei Dienes (2011).

Für die vorliegende Arbeit ist verkürzt gesagt relevant, dass die Bayesische Statistik empfohlen wird, wenn man bereits einige Modellanahmen hat. Während die frequentistische Statistik empfohlen wird in Fällen, in denen es wenig bis kein Vorwissen ober die Wirkmechanismen gibt (Little, 2006; Efron, 2005). Neben dieser Empfehlung bietet die Base Statistik viele praktische Vorteile, wenn man die Lücke zwischen Theorie und Daten mittels einem Modell schließen möchte. Diese Überlegungen habe ich auf Basis von Bailer-Jones (2009) Buch über Modelle in der Wissenschaft in Artikel 3 verschriftlicht.

Um die Ergebnisse dieser Arbeit zu verstehen, müssen einige grundlegende Ideen der Bayessche Statistik eingeführt werden. So ergibt eine Parameterschätzung immer eine Verteilung statt eines einzigen konkreten Wertes. Diese Verteilung spiegelt die Unsicherheit über diesen Wert wieder: Ist sie verhältnismäßig breit, besteht große Unsicherheit; Ist sie verhältnismäßig schmal, besteht geringe Unsicherheit über den gesuchten Wert. Diese Information fasst das highest-density interval (HDI) zusammen. Das HDI beschreibt das Intervall, in dem sich 95% der wahrscheinlichen Parameter befinden. Da die Bayessche Statistik für alle möglichen Propositionen eine Wahrscheinlichkeit angeben kann, steht und fällt die Sinnhaftigkeit einer Parameterschätzung mit der Sinnhaftigkeit des verwendeten Modells. Die Qualität der Modelle die hier vorgestellt und gerechnet werden ist zum einen durch deren theoretische Herleitung abgesichert, zum anderen aber auch durch deren empirische Überprüfung, zum Beispiel durch Vergleiche mit anderen ebenfalls theoretisch hergeleiteten Modellen oder Vergleich von vorhergesagten und beobachteten Daten.

Kumulative Artikel der Dissertation

Dieses kumulative Dissertation besteht insgesamt aus 4 einzelnen Artikeln sowie eine Reihe von Experimenten, die bisher in keinem Artikel zusammengefasst wurden. Während sich drei der Artikel ganz direkt mit der Quantifizierung von visueller Salienz beschäftigen, verfolgt der letzte Artikel ein anderes Ziel. Von einer Vogelperspektive aus wird das Arbeiten mit Modellen betrachtet, insbesondere wie Modelldaten und Theorie zu einer quantitativen Erklärung verbinden. Dieser Artikel nimmt aus mehreren Gründen eine Sonderstellung ein: Einerseits hat er, wie bereits erwähnt, einen anderen Blickwinkel, andererseits wurde dieser Artikel auch in Kooperation mit anderen Autoren geschrieben, sodass er als ein Zusatz zu dieser Promotion anzusehen ist, auch wenn bei diesem gemeinsamen Schreiben eine klare Trennung von Aufgaben erfolgt ist.

Der erste Artikel befasst sich mit dem Versuch, zeitliche Reihenfolgeurteile, Salienz und die TVA Modellierung von Tünnermann zusammenzubringen. Der Artikel zeigt ein Experimentaldesign, mit dem dies gelingt. Der zweite Artikel untersucht genauer, wie physische Kontraste und Aufmerksamkeitsparameter quantitativ zusammenhängen. Auch wird ein Salienzmaß auf Basis von (Nordfang u. a., 2013) eingeführt, welches erlaubt einzele saliente Eigenschaften mit Kombinationen von salienten Eigenschaften zu vergleichen. Dieser Artikel zeigt, dass eine quantitative Schätzung von Salienz innerhalb einer Art von Kontrast sowie übergreifend für eine Kombination mehrerer Arten von Kontrast möglich ist. Der dritte Artikel zum Thema Salienzmessung beschäftigt sich mit einer besonderen Eigenschaft der Salienz, nämlich der, dass sie zeitlich nicht stabil ist. Stattdessen gibt es eine Phase im Bereich von 50 bis 200 Millisekunden, in der Salienz eine starke Wirkung auf die Aufmerksamkeit hat. Diese Variation der Salienzstärke über die Zeit wurde ebenfalls durch das vorgestellte Modell ermittelt. Dieser Artikel liefert ein überraschendes Ergebnis, indem deutlich wird, dass die visuelle Verarbeitungskapazität stärker variiert als die Aufmerksamkeit und somit der Verdacht nahe liegt, dass die frühe Aufmerksamkeitsmessung im Bereich unter 150 ms möglicherweise mit einem anderen kognitiven Prozess konfundiert ist.

Artikel 1

Artikel 1 dokumentiert den Versuch die TVA-basierte Modellierung von TOJ mit Multi-Element-Salienzdisplays zu kombinieren. Dazu wird zunächst auf die Theorie eingegangen, die bereits auf den vorangegangenen Seiten dieser Zusammenfassung erörtert wurde. Kernstück der Arbeit sind die vier Experimente. Das Ergebnis ist einerseits, dass gezeigt wird, welche Formen des TOJs für diese Kombination Aufmerksamkeit messbar machen und welche nicht, andererseits wird bereits durch die vierstufige Salienzmanipulation klar, dass das physischer Kontrast und Aufmerksamkeitsgewicht nicht linear zusammenhängen.

Experiment 1 basiert darauf, dass das TOJ über einen Onset realisiert wird. Konkret wurde ein Multi-Element Display eingeblendet, viele kleine Balken auf einen imaginären Gitter, die die selbe Orientierung hatten. Rechts und links des Fixationskreuzes blieben zwei Lücken im Gitter.

An diesen Positionen wurden der Probe- und Referenzreiz eingeblendet. Die Einblendung erfolgte entsprechend eines SOAs. Der Probereiz war potentiell salient. Es gab vier Stufen für die Salienzmanipulation. Die erste Stufe wies keinen Unterschied zu den benachbarten Reizen oder dem Referenzreiz auf. Alle Reize hatten so die selbe Orientierung. In den folgenden drei Stufen wurde der Orientierungskontrast erhöht, indem der Probereiz eine andere, in Stufen extremere, Abweichung von der Orientierung seines Umfeldes hatte.

Durch einen einfachen Onset konnte jedoch keine Aufmerksamkeitsveränderung auf Basis des Orientierungskontrastes gemessen werden. Dies ist wahrscheinlich darin begründet, dass zunächst beide Löcher im Muster, in denen Probe und Referenz eingeblendet werden, gleich salient sind. Im Moment, in dem das Rennen zwischen den beiden Onsets stattfindet, sind die Aufmerksamkeitsgewichte noch auf Basis des vorherigen Bildes mit den zwei freien Stellen verteilt.

Das Experiment hat außerdem eine kleine methodische Schwäche enthüllt: Während der Probereiz zeitlich um den Referenzreiz variierte, wurde der Referenzreiz immer zum selben Zeitpunkt ab Trialstart gezeigt. Entsprechend der Arbeit von Vangkilde u. a. (2012) kann man mit der TVA auch die Aufmerksamkeitseffekte von zeitlicher Erwartung messen. Ein solcher kleiner, aber über alle Experimente konstanter Effekt von zeitlicher Erwartung zeigt sich in dem sehr geringen Aufmerksamkeitsvorteil des Referenzreizes. In späteren Experimenten wurde ein Zufallsintervall zwischen Trialstart und TOJ verwendet, um diese Einflüsse auszuschließen. Dieser kleine Mangel am Design enthüllt jedoch auch eine Stärke der theoriebasierten Analyse, da die Größe des Einflusses quantitativ bestimmt werden kann und im Vergleich zu dem Einfluss von Salienz auf Aufmerksamkeit aus Experiment 3 und 4 deutlich kleiner ausfällt.

Experiment 2 befasste sich mit die Idee, statt des Onsets einen Offset von Reizen für das TOJ zu verwenden, wie es beispielsweise bei Vingilis-Jaremko u. a. (2008) vorkommt. Dies bedeutet, dass zu Beginn des Trials alle Reize, Hintergrundelemente sowie Probe und Referenz, angezeigt wurden. Entsprechend des SOAs wurden dann Probe und Referenz ausgeblendet. Dieses Design führte zu einem deutlich messbaren Unterschied im Aufmerksamkeitsgewicht, jedoch in genau der umgekehrten Richtung, wie erwartet. Der Offset des auffälligen Reizes wurde später und nicht früher als der Offset des nicht auffälligen Reizes wahrgenommen. Eine mögliche Post-hoc-Erklärungen wäre, dass das Auflösen der Bindung von Verarbeitungsressourcen für saliente Reize länger dauert. Jedoch handelt es sich dabei um eine Vermutung, die durch weitere empirische Forschung abgesichert werden müsste. Für das Messen von Salienz ist ein reduziertes Aufmerksamkeitsgewicht durch physischen Kontrast in jedem Fall nicht überzeugend.

Experiment 3 stellt den Versuch dar, das Multi-Element-Salienz-Display möglichst konstant zu halten. Dazu werden weder dauerhaft Elemente ein noch ausgeblendet. Zu den vom SOA bestimmten Zeitpunkten wird stattdessen der entsprechende Reiz kurz aus- und 80 ms später wieder eingeblendet. Bei diesen Ereignissen handelt es sich um klar detektierbare Ereignisse, die aber keine dauerhafte Änderung am Bild erfordern. Die Ergebnisse zeigen, dass dieses Design ein klaren Aufmerksamkeitsgewichtsvorteil für den salienten Reiz zeitigt und dies in Abhängigkeit der Salienzstärke, wie es theoretisch zu erwarten ist, geschieht. Ebenfalls zeigt das Experiment, dass Salienz nicht linear mit physischem Kontrast wächst.

Experiment 4 versucht die Methode auf eine andere Kontrastart anzuwenden. Dazu wurde die Orientierung bei allen Balken konstant gehalten. Salienz wurde durch einen Helligkeitsunterschied zwischen Probe und Umgebung in vergleichbaren vier Stufen erzeugt. Auch hier lassen sich mit dem sogenannten Flicker, kurzer Off- und Onset von Reizen, Aufmerksamkeitsvorteile anhängig vom Kontrast messen.

Artikel 2

Artikel 2 verbessert die Modellierung, sodass nicht nicht der Aufmerksamkeitsvorteil gemessen werden kann, sondern direkt der Salienzwert in Anlehnung an Nordfang u. a. (2013). Im empirischen Teil der Arbeit geht es darum, zu erfassen, wie Salienz in Abhängigkeit von einer Kontrastart, zunächst Orientierungsdifferenz, steigt, wenn dieser Kontrast sukzessive ansteigt. Während Experiment 1 die Ergebnisse aus Artikel 1 Experiment 2 mit dem neuen Modell repliziert, wurde in Experiment 2 zusätzliche Stufen der unabhängigen Variable hinzugefügt, um einen klareren Blick auf deren Verhalten zu ermöglichen. Modellvergleiche zeigen, dass eine Potenzfunktion ein gutes Modell (das Beste unter den getesteten) darstellt, um das Wachstum von Salienz in Abhängigkeit von Kontrast auszudrücken. Die Modellierung ist dabei Stevens Power Law entlehnt und beinhaltet zwei freie Parameter. Zu den alternativen Modellen zählte ein ein-parametriges logarithmisches Modell und ein Modell, welches keinen Zusammenhang zwischen den Experimentalbedingungen annimmt, also einen freien Parameter pro Bedingung besitzt. Die daraus resultierenden Modellvergleiche zeigen, dass es — gegeben die Daten — sinnvoller ist, von einem systematischen Zusammenhang auszugehen als von nicht zusammenhängenden Salienzwerten. Für sich genommen erscheint dieses Ergebnis trivial, jedoch möchte ich damit klar argumentieren, warum genau es vernünftig ist eine Funktion zwischen den physikalischen Kontrasten und dem Salienzwert anzunehmen. Weiterhin konnte gezeigt werden, dass der visuell logarithmisch wirkende Verlauf besser durch eine Potenzfunktion beschrieben werden kann als durch ein simpleres Modell. So kann es nach der Analyse immer noch sein, dass das Potenzfunktionsmodell nur das Beste unter drei schlechten Modellen ist. Jedoch ist es durch die Analysen sichergestellt, dass es sinnvoller ist, als einen auf dem ersten visuellen Eindruck basierenden logarithmischen Zusammenhang oder gar keinen funktionalen Zusammenhang anzunehmen.

Eine weitere Art von Kontrast, der Luminanzkontrast, wurde in Experiment 3 untersucht. Luminanzkontraste können, anders als Orientierungskontrast, durch höhere Intensität beim salienten Reiz (heller als Umgebung) oder durch niedrigere Intensität beim salienten Reiz (dunkler als Umgebung) erzeugt werden. Zumindestens für die durch einen Reiz mit höherer Intensität als Umgebung verursachten Salienzwerte, passt die Potenzfunktion zur Beschreibung des Salienzwachstums. Daher ist es ein Ergebnis dieser Doktorarbeit, dass Salienz innerhalb einer Kontrastart entsprechend einer Potenzfunktion[5] wächst.

Experiment 4 führt beide Modellierungen zusammen, indem die in der Literatur widersprüchlich beantwortete Frage diskutiert wird, wie gut sich Kontrastarten zur Gesamtsalienz ergänzen. Eine Schwierigkeit bei der Literaturrecherche ist, dass es nur wenige Publikationen gibt, die sich für die genaue Stärke von Salienz interessieren — häufiger wird mittels eines Tests die Existenz von Salienz untersucht. Die wenigen Untersuchungen, die Salienzstärke von verschiedenen Kontrastarten kombinieren sind jedoch wenig vergleichbar und insbesondere die umfangreiche Untersuchung von Nothdurft (2000) und das Ergebniss von Koene und Zhaoping (2007) scheinen sich im Bezug darauf zu widersprechen, wie effektiv sich verschiedene Salienzarten kombinieren lassen.

Da bei den sehr unterschiedlichen experimentellen Methoden nicht alle Hilfsannahmen ausformuliert sind, ist es sehr schwer zusagen, in wie weit sich die Ergebnisse stützen oder widersprechen. Diese Situation bringt mich in den Zwiespalt, dass ich einerseits einen ebenfalls das Thema bearbeiten möchte, andererseits eine weitere wiederum neue Methode die Divergenz weiter vergrößert. Jedoch ist es Kerngedanke dieser Doktorarbeit, dass ein aus der Theorie hergeleitetes explizites formales Modell noch nicht alle Widersprüche beseitigt, wohl aber transparent macht, wo diese Widersprüche entstehen können. So kann der kritische Leser, die kritische Leserin direkt auf die Stelle in der Formel zeigen, an der man widersprechen möchte. Ein Beispiel kommt aus der TVA -Forschung selber, in der die ortsspezifische Aufmerksamkeit zusätzlich zu der eigenschaftsbezogenen Aufmerksamkeit modelliert wurde. Noch greifbarer ist es, wenn man sich über Salienz — stimulusgetriebene Aufmerksamkeit — und Aufgabenrelevanz unterhält. Durch die TVA-Formeln kann viel klarer gemacht werden, wo genau sich eine entsprechende Manipulation auswirkt und was genau eine Konstanthaltung zur Folge hat.

[5] Der Begriff *power law* mag einem in den Sinn kommen. Da jedoch Cummins (2000) gut dargelegt hat, welche Tücken der Begriff des Gesetzes für die Spezialwissenschaften mit sich bringt, möchte ich lieber vom einem Modell sprechen.

Um wieder auf die Kombination verschiedener salienter Kontraste zurückzukommen, wurde ein Modell entwickelt, das die in der Literatur verbreitete Idee realisiert, dass es einen pauschalen prozentualen Abzug an Salienz gibt, wenn Kontraste kombiniert werden (Huang & Pashler, 2005; Nothdurft, 2000). Entsprechend des vorher festgestellten Zusammenhangs von Kontrast innerhalb einer Kontrastart, müssten also die beiden entsprechend der Potenzfunktion ermittelten Werte pro Kontrastart addiert werden und ein prozentualer Wert davon abgezogen werden. Genau dies wurde in einem formalen Modell realisiert, in dem die Größe des prozentualen Abzugs als zusätzlicher freier Parameter modelliert wurde.

Dieses Modell wurde eingesetzt, um die Daten aus einem zweifaktoriellen Design — Luminanzkontrast und Orientierungskontrast — mit jeweils vier Stufen— 0%, 25%, 50%, 100% des maximalen Kontrasts—zu analysieren. Die Salienzwerte in Abhängigkeit von den beiden Faktoren lassen sich in einem dreidimensionalen Plot als Ebene darstellen. Dieses Modell hat aufgrund der hierarchischen Struktur bereits eine erhebliche Komplexität, sodass die vom Modell für jede Versuchsperson vorhergesagten Werte mit den tatsächlichen Werten abgeglichen wurden, um zu beurteilen, ob das Modell auch tatsächlich die Daten angemessen abbildet. Die Ergebnisse sind im Anhang von Artikel 2 veröffentlicht und zeigen, dass die Datenmuster durch das Modell nachgebildet werden.

Die modellbasierte Analyse zeigt, dass weder ein pauschaler Abzug noch ein pauschaler Salienzgewinn entsteht, wenn Orientierungskontrast und Luminanzkontrast kombiniert werden. Diese Arbeit gibt damit eine weitere Antwort auf die Frage nach der Kombination von Salienz durch verschiedene Kontrastarten. Durch die explizite Modellierung der in der Literatur vermuteten Beziehung und das aus der Theorie hergeleitete Modell sollen jedoch alle Annahmen transparent gemacht werden sowie eine formaltheoretische Erklärung der Befunde ermöglicht werden.

Artikel 3

Artikel 3 ist im Vergleich zu den anderen drei Publikationen eine Ausnahme, da er sich aus einer Vogelperspektive den Einsatz von Modellen und Analysen in der Psychologie betrachtet. Ein besonderer Fokus liegt dabei auf quantitativen Erklärungen. Kernthese ist dabei, dass eine Hierarchie von Modellen die Kluft zwischen Theorie und beobachtbaren Daten überbrückt. Diese Kluft entsteht, da die Theorien möglichst losgelöst von konkreten Situationen formuliert werden, um ein Maximum an Generalisierbarkeit zu ermöglichen. Angewandt auf konkrete Fälle müssen jedoch Hilfsannahmen gemacht werden und eine Passung zwischen konkreter Situation und theoretischer Konzeptualisierung hergestellt werden.

Die Arbeit beruht zu einem erheblichen Teil auf Bailer-Jones (2009) Arbeit zu Modellen in den Naturwissenschaften. In dieser Arbeit wird das Thema historisch aufgearbeitet. So galten Modelle lange als eine Art unterkomplexe, unvollständige Theorie. Eine Theorie, so die Denkweise, würde ein Modell vollkommen obsolet machen. Besonders die Arbeit von Suppes (1966) zeigt jedoch auf, dass das Anwenden von Theorie auf konkrete Ergebnisse großen Interpretationsspielraum bietet und Objektivität nur durch Offenlegen dieser Interpretationen erreicht werden kann.

Der Artikel vergleicht eine Kombination von Modellierung und Bayesscher Statistik mit Nullhypothesentests einerseits und maschinellem Lernen andererseits (für eine Übersicht zu ML in Psychologie siehe Yarkoni und Westfall (2017). Das Ergebnis ist, dass Modellierung und Bayessche Statistik im Vergleich zu den beiden Alternativen einerseits eine hohe Passung zwischen Daten und modelliertem Verhalten erzielen kann (im Vergleich zu Nullhypothesentests, was sich z.B. in der Voraussage von Daten zeigt); andererseits bleiben die Modelle immer erklärbar und anpassbar — im Vergleich zu festen ML-Methoden, die sozusagen bei Design die Datenanalyse nicht von den Prozessen abhängig machen, die die Daten tatsächlich verursacht haben (Breiman, 2001).

Zusammenfassend kann man sagen, dass der Psychologie durch ML, klassische Nullhypothesentests, Bayesianische Statistik und Modellierungsmethoden viele Analysewerkzeuge zur Verfügung stehen, sodass statistisches Denken (Gigerenzer, 2018) erforderlich ist, um die Methode auszuwählen, die für eine Forschungsfrage den höchsten Erkenntnisgewinn verspricht. Für eine quantitative Erklärung, die eine explizite Verbindung zu den zur Erklärung herangezogenen Theorien hat, ergibt die im Artikel 3 vorgestellte Argumentation, dass die Bayessche Statistik und Modellierung die Methode mit der höchsten Passung zu diesen Zielen ist.

Artikel 4

Artikel 4 beschäftigt sich mit dem zeitlichen Verlauf von Salienz. Artikel 1 und Artikel 2 zeigen, dass ein Modell abgeleitet aus der TVA die Schätzung von Salienzwerten—abhängig von physische Kontrast—mit klarer theoretischer Bedeutung ermöglicht. Ungeklärt ist jedoch, ob ebenfalls der zeitliche Verlauf von Salienz in der selben Weise erfasst werden kann. Daher ist die unabhängige Variable in den beiden Experiment von Artikel 4 die Zeit zwischen Onset des Multi-Element-Displays und des TOJs. Da bisherige Untersuchungen besonders im Bereich von 50 bis 150 ms einen starken Einfluss von Salienz auf. Daher wurden die Intervalle der unabhängigen Variable in folgenden Schritten gewählt: 50, 100, 200, 400 und 800 ms. Die modellbasierte Analyse entspricht der Salienzmessung für unabhängige Bedingungen aus Artikel 2.

Entgegen der Hypothese zeigte sich in Experiment 1, dass die visuelle Verarbeitungskapazität deutlich stärker und in einem klar erkennbaren Muster variiert: In den Bedingungen mit kurzem Intervall, 50 und 100 ms, ist die Verarbeitungsrate jeweils halbiert im Vergleich zum nächst längeren Intervall. Neben entweder dem Aufmerksamkeitsgewicht oder dem Salienzparameter, genannt κ, ist die gesamte Verarbeitungsrate der zweite Parameter, der die Form der benutzten psychometrischen Funktion bestimmt. Bisher wurde dieser Parameter nicht weiter erwähnt, da er theoretisch wie praktisch nicht variieren sollte. In Experiment 1 konnte keine systematische Varianz der Aufmerksamkeit beobachtet werden—es wäre sehr gut möglich gewesen, dass auch beide Parameter einen klaren Verlauf zeigen. Da das Ergebnis so nicht hypothesenkonform ist, haben wir uns für eine Replikation entschieden.

Experiment 2 repliziert die Ergebnisse aus Experiment 1. Der einzige Unterschied ist, dass statt des voll randomisierten Designs ein geblocktes Design gewählt wurde, um mit anderen Studien (Donk & Soesman, 2011, z.B.,) im Bezug auf das Design besser vergleichbar zu sein. Dabei wurde erwartet, dass die Salienz deutlicher zwischen den Bedingungen variiert. Außerdem wurde erwartet, dass die deutliche Kapazitätsveränderung replizierbar ist.

Die Ergebnisse zeigen dabei — wie erwartet — mehr Varianz im Salienzparameter zwischen den Bedingungen und der deutliche Kapazitätsunterschied in den ersten beiden Bedingungen wurde ebenfalls repliziert. Eine genauere Erklärung der Parameter und dazu, wie ihr Zusammenwirken die psychometrische Funktion ergibt ,findet sich im Abschnitt 6.

Diese Ergebnisse zeigen einerseits, dass man mit Intervallen unter 150 ms bei der Salienzmessung vorsichtig sein muss, da möglicherweise auch ein anderer kognitiver Prozess involviert ist, der über das Bestimmen der TVA-Aufmerksamkeitsgewichte in der ersten Verarbeitungswelle hinaus geht. Andererseits zeigt sich, dass das Anwenden eines Modells ermöglicht, Abweichungen von theoretischen Erwartungen sichtbar zumachen und sogar zu quantifizieren.

Schlussfolgerung

Wie bei jedem Ansatz gibt es auch bei dem hier vorgestellten Stärken und Schwächen. Eine Stärke ist, dass viel Vorwissen und Annahmen explizit und teilweise formal verwendet werden. Eine Schwäche ist, dass mit dem Ansatz keine große inhaltliche Frage der kognitiven Psychologie geklärt wird. Stattdessen zeigt die Arbeit, wie konkret mit der Idee einer allgemeinen Aufmerksamkeitswährung für Salienz verfahren werden kann, die schon früher vermutet wurde (Treue, 2003). Insofern bleibt diese Lösung eine mögliche Lösung. Besonders in der kognitiven Psychologie zeichnet sich jedoch ein Trend ab, der von Rodgers (2010) sogar als stille Revolution bezeichnet und gut in die Überlegungen von Taagepera (2008) passt, die dafür plädieren, die Logik der Prozesse, die unsere Daten erzeugen, in unseren Analysen abzubilden. Klassische Methoden können vermitteln, dass dasselbe Ergebnis leichter erzielt werden kann (Rouder, Morey, Verhagen u. a., 2016) oder dass andere Methoden zu einer Subjektivierung von Ergebnissen führen (Rouder, Morey & Wagenmakers, 2016). Jedoch ist es am Beispiel der Physik zu erkennen, dass rein verbale Theorien über einen gewissen Wissensstand hinaus nur noch bedingt helfen. Gerade in der Forschung zur visuellen Aufmerksamkeit gibt es so viele empirische Ergebnisse,dass eine übergeordnete Theoriebildung immer unwahrscheinlicher erscheint. Mangelnde Theorie kann jedoch nicht durch Empirie ausgeglichen werden, da dann ein Maßstab für das fehlt, was als gesichert betrachtet werden sollte und das was aufgrund seiner Neuheit Replikationen bedarf (Muthukrishna & Henrich, 2019). Es ist meine Hoffnung, dass diese Arbeit — mit der vorgestellten Verbindung von Theorie und Daten — in dieser Hinsicht zeitgemäß ist.

Bibliography

Abend, G. (2008). The Meaning of 'Theory'. *Sociological Theory*, *26*(2), 173–199. doi:10.1111/j.1467-9558.2008.00324.x

Acosta, A., Adams (Jr.), R. B., Albohn, D. N., Allard, E. S., Beek, T., Benning, S. D., … Zwaan, R. A. (2016). Registered Replication Report: Strack, Martin, & Stepper (1988). *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *11*(6), 917–928. doi:10.1177/1745691616674458

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., … Wicherts, J. M. (2013). Recommendations for Increasing Replicability in Psychology. *European Journal of Personality*, *27*(2), 108–119. doi:10.1002/per.1919

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in cognitive sciences*, *16*(8), 437–443.

Bailer-Jones, D. M. (2009). *Scientific models in philosophy of science*. Pittsburgh: University of Pittsburgh Press.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature News*, *533*(7604), 452. doi:10.1038/533452a

Bartha, P. (2013). Analogy and analogical reasoning. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy (spring 2019 edition)*. Retrieved from https://plato.stanford.edu/archives/spr2019/entries/reasoning-analogy/

Baumeister, R., Heatherton, T., & Tice, D. (1994). *Losing control: How and why people fail at self-regulation*. San Diego, CA: Academic Press.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of personality and social psychology*, *74*(5), 1252.

Begley, C. G. & Ellis, L. M. (2012). Raise standards for preclinical cancer research. *Nature*, *483*, 531. doi:10.1038/483531a

Betz, T., Kietzmann, T. C., Wilming, N., & König, P. (2010). Investigating task-dependent top-down effects on overt visual attention. *Journal of Vision*, *10*, 15. doi:10.1167/10.3.15

Bieri, P. (2017). *Wie wäre es, gebildet zu sein?* München: Komplett-Media.

Blurton, S., Nielsen, C., Kyllingsbuk, S., & Bundesen, C. (2016). A poisson random walk model for response time and pure accuracy tasks. *Journal of Vision*, *16*(12), 587–587.

Bolacchi, G. (2004). On "social Sciences"and Science. *Behavior and Philosophy*, *32*(2), 465–478. Retrieved December 18, 2019, from www.jstor.org/stable/27759497

Box, G. E. P. (1976). Science and Statistics. *Journal of the American Statistical Association*, *71*, 791–799. doi:10.1080/01621459.1976.10480949

Box, G. & Draper, N. (1987). *Empirical model-building and response surfaces*. Wiley series in probability and mathematical statistics: Applied probability and statistics. Hoboken, NJ: Wiley.

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., … van 't Veer, A. (2014). The Replication Recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, *50*, 217–224. doi:10.1016/j.jesp.2013.10.005

Breiman, L. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231. doi:10.1214/ss/1009213726

Broadbent, D. (1958). *Perception and communication*. London: Pergamon Press.

Bruce, N. D. & Tsotsos, J. K. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of vision*, *9*(3), 5–5.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*, 523–547. doi:10.1037/0033-295X.97.4.523

Bundesen, C. (1998). A computational theory of visual attention. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *353*, 1271–1281. doi:10.1098/rstb.1998.0282

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review*, *112*, 291–328.

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2011). A neural theory of visual attention and short-term memory (NTVA). *Neuropsychologia*. Attention and Short-Term Memory, *49*, 1446–1457. doi:10.1016/j.neuropsychologia.2010.12.006

Bundesen, C., Vangkilde, S., & Petersen, A. (2015). Recent developments in a computational theory of visual attention (TVA). *Vision Research*, *116, Part B*, 210–218. doi:http://dx.doi.org/10.1016/j.visres.2014.11.005

Bundesen, T., C. and Habekost. (2008). *Principles of Visual Attention: Linking Mind and Brain*. Oxford University Press.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. doi:10.1038/nrn3475

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*. Vision Research 50th Anniversary Issue: Part 2, *51*, 1484–1525. doi:10.1016/j.visres.2011.04.012

Carrier, M. (2019). *Wissenschaftstheorie zur einführung*. zur Einführung. Hamburg: Junius Verlag.

Carter, E. C. & McCullough, M. E. (2014). Publication bias and the limited strength model of self-control: Has the evidence for ego depletion been overestimated? *Frontiers in Psychology*, *5*. doi:10.3389/fpsyg.2014.00823

Chalmers, A. F. (2013). *What is this thing called science? (4th ed.)* Berkshire: Open University Press.

Chambers, C. (2017). *The Seven Deadly Sins of Psychology: A Manifesto for Reforming the Culture of Scientific Practice*. Princeton University Press.

Cho, A. (2012). Once Again, Physicists Debunk Faster-Than-Light Neutrinos. Retrieved August 26, 2019, from https://www.sciencemag.org/news/2012/06/once-again-physicists-debunk-faster-light-neutrinos

Christensen, R. (2005). Testing Fisher, Neyman, Pearson, and Bayes. *The American Statistician*, *59*(2), 121–126. doi:10.1198/000313005X20871

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual review of psychology*, *62*, 73–101.

Cohen, J. (1990). Things I Have Learned (So Far). *American Psychologist*, 9.

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

Couffe, C., Mizzi, R., & Michael, G. A. (2016). Salience-based progression of visual attention: Time course. *Psychologie Française*, *61*, 163–175. doi:10.1016/j.psfr.2015.04.003

Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

Cummins, R. (2000). 'how does it work?' versus 'what are the laws?': Two conceptions of psychological explanation. In F. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (Chap. 5, pp. 117–144). MIT Press.

de Saint-Exupéry, A. (1946). *Le petit prince*. Paris: Gallimard.

Desimone, R. & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, *18*(1), 193–222.

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. London: Palgrave.

Dienes, Z. (2011). Bayesian Versus Orthodox Statistics: Which Side Are You On? *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *6*, 274–290. doi:10.1177/1745691611406920

Dombrowe, I. C., Olivers, C. N. L., & Donk, M. (2010). The time course of color- and luminance-based salience effects. *Frontiers in Psychology*, *1*. doi:10.3389/fpsyg.2010.00189

Donk, M. & Soesman, L. (2010). Salience is only briefly represented: Evidence from probe-detection performance. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 286–302. doi:10.1037/a0017605

Donk, M. & Soesman, L. (2011). Object salience is transiently represented whereas object presence is not: Evidence from temporal order judgment. *Perception*, *40*, 63–73.

Donk, M. & van Zoest, W. (2008). Effects of saliences are short-lived. *Psychological Science*, *19*, 733–739. doi:10.1111/j.1467-9280.2008.02149.x

Doyle, A. (1890). *The sign of four*. London: Penguin Classics.

Duncan, J. & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458. doi:10.1037/0033-295X.96.3.433

Efron, B. (2005). Bayesians, Frequentists, and Scientists. *Journal of the American Statistical Association*, *100*, 1–5. doi:10.1198/016214505000000033

Egeth, H. E. & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual review of psychology*, *48*(1), 269–297.

Einhäuser, W. & König, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience*, *17*, 1089–1097. doi:10.1046/j.1460-9568.2003.02508.x

Engmann, S., Hart, B. M. ', Sieren, T., Onat, S., König, P., & Einhäuser, W. (2009). Saliency on a natural scene background: Effects of color and luminance contrast add linearly. *Attention, Perception, & Psychophysics*, *71*(6), 1337–1352. doi:10.3758/APP.71.6.1337

Fechner, G. T., Howes, D. H., & Boring, E. G. (1966). *Elements of psychophysics*. New York: Holt, Rinehart and Winston.

Feynman, R. P. (1974). Cargo cult science. *Engineering and Science*, *37*(7), 10–13.

Fienberg, S. E. (2006). When did Bayesian inference become "bayesian"? *Bayesian Analysis*, *1*(1), 1–40. doi:10.1214/06-BA101

Finke, K., Bublak, P., Krummenacher, J., Kyllingsbaek, S., Muller, H. J., & Schneider, W. X. (2005). Usability of a theory of visual attention (TVA) for parameter-based measurement of attention I: Evidence from normal subjects. *Journal of the International Neuropsychological Society: JINS*, *11*(7), 832–842.

Fitzgerald, C. J. & Whitaker, M. B. (2010a). Examining the Acceptance of and Resistance to Evolutionary Psychology. *Evolutionary Psychology*, *8*(2), 147470491000800211. doi:10.1177/147470491000800211

Fitzgerald, C. J. & Whitaker, M. B. (2010b). Examining the acceptance of and resistance to evolutionary psychology. *Evolutionary Psychology*, *8*(2), 147470491000800211.

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.

Foster, D. H. & Ward, P. A. (1991). Asymmetries in Oriented-Line Detection Indicate Two Orthogonal Filters in Early Vision. *Proceedings: Biological Sciences*, *243*(1306), 75–81. Retrieved from http://www.jstor.org/stable/76519

Frank, M. C. & Saxe, R. (2012). Teaching Replication. *Perspectives on Psychological Science*, *7*(6), 600–604. doi:10.1177/1745691612460686

Friese, M., Loschelder, D. D., Gieseler, K., Frankenbach, J., & Inzlicht, M. (2018). Is Ego Depletion Real? An Analysis of Arguments. *Personality and Social Psychology Review*, 1088868318762183. doi:10.1177/1088868318762183

Frintrop, S., Rome, E., & Christensen, H. I. (2010). Computational visual attention systems and their cognitive foundations: A survey. *ACM Transactions on Applied Perception*, *7*, 6:1–6:39. doi:10.1145/1658349.1658355

Friston, K. (2012). The history of the future of the bayesian brain. *NeuroImage*, *62*(2), 1230–1233.

Gelman, A. & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University*. Retrieved from http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Gelman, A. & Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 8–38. doi:10.1111/j.2044-8317.2011.02037.x

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*. Statistical Significance, *33*(5), 587–606. doi:10.1016/j.socec.2004.09.033

Gigerenzer, G. (2018). Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science*, *1*(2), 198–218.

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on "Estimating the reproducibility of psychological science". *Science*, *351*(6277), 1037–1037. doi:10.1126/science.aad7243

Giner-Sorolla, R. (2012). Science or Art? How Aesthetic Standards Grease the Way Through the Publication Bottleneck but Undermine Science. *Perspectives on Psychological Science*, *7*(6), 562–571. doi:10.1177/1745691612457576

Godfrey-Smith, P. (2003). *Theory and reality: An introduction to the philosophy of science*. Science and Its Conceptual Foundations series. University of Chicago Press.

Goto, K., Bond, A. B., Burks, M., & Kamil, A. C. (2014). Visual Search and Attention in Blue Jays (Cyanocitta cristata): Associative Cuing and Sequential Priming. *Journal of experimental psychology. Animal learning and cognition*, *40*(2), 185–194. doi:10.1037/xan0000019

Goulson, D. (2000). Why do pollinators visit proportionally fewer flowers in large patches? *Oikos*, *91*(3), 485–492.

Green, C. D. (2015). Why Psychology Isn't Unified, and Probably Never Will Be: *Review of General Psychology*. doi:10.1037/gpr0000051

Habekost, T., Petersen, A., & Vangkilde, S. (2014). Testing attention: Comparing the ant with tva-based assessment. *Behavior research methods*, *46*(1), 81–94.

Hacking, I. (2001). *An introduction to probability and inductive logic*. Cambridge: Cambridge University Press.

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., … Zwienenberg, M. (2016). A Multilab Preregistered Replication of the Ego-Depletion Effect. *Perspectives on Psychological Science*, *11*(4), 546–573. doi:10.1177/1745691616652873

Hagger, M., Wood, C., Stiff, C., & Chatzisarantis, N. (2010). Ego Depletion and the Strength Model of Self-Control: A Meta-Analysis. *Psychological bulletin*, *136*, 495–525. doi:10.1037/a0019486

Haller, H. & Krauss, S. (2002). Misinterpretations of Significance: A Problem Students Share with Their Teachers? *7*(1), 20.

Hanlon, R. (2007). Cephalopod dynamic camouflage. *Current Biology*, *17*(11), R400–R404. doi:10.1016/j.cub.2007.03.034

Hansen, T., Pracejus, L., & Gegenfurtner, K. R. (2009). Color perception in the intermediate periphery of the visual field. *Journal of Vision*, *9*(4), 26–26.

Hofstadter, D. (2007). *I am a strange loop*. New York, NY: Basic Books.

Hofstadter, D. & Dennett, D. (2001). *The mind's i: Fantasies and reflections on self & soul*. Basic Books. Retrieved from https://books.google.de/books?id=SSBJPwAACAAJ

Huang, L. & Pashler, H. (2005). Quantifying object salience by equating distractor effects. *Vision Research*, *45*, 1909–1920. doi:10.1016/j.visres.2005.01.013

Hubel, D. H. & Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *The Journal of Physiology*, *148*, 574–591.

Hutson, M. (2018). Artificial intelligence faces reproducibility crisis. *Science*, *359*(6377), 725–726. doi:10.1126/science.359.6377.725

Itti, L. & Koch, C. (2001a). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*, 194–203. doi:10.1038/35058500

Itti, L. & Koch, C. (2001b). Feature combination strategies for saliency-based visual attention systems. *Journal of Electronic Imaging*, *10*, 161–169. doi:10.1117/1.1333677

James, W. (1890). *The principles of psychology*. New York: Holt.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*, 0956797611430953. doi:10.1177/0956797611430953

Kerzel, D., Schönhammer, J., Burra, N., Born, S., & Souto, D. (2011a). Saliency changes appearance. *PLoS ONE*, *6*, e28292. doi:10.1371/journal.pone.0028292

Kerzel, D., Schönhammer, J., Burra, N., Born, S., & Souto, D. (2011b). Saliency changes appearance. *PloS one*, *6*(12), e28292.

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

Koehler, K., Guo, F., Zhang, S., & Eckstein, M. P. (2014). What do saliency models predict? *Journal of vision*, *14*(3), 14–14.

Koene, A. R. & Zhaoping, L. (2007). Feature-specific interactions in salience from combined feature contrasts: Evidence for a bottom-up saliency map in V1. *Journal of Vision*, *7*, 6. doi:10.1167/7.7.6

Krüger, A. (2020). A quantification of visual salience (cumulative dissertation): Data and scripts. doi:10.17605/OSF.IO/49X37

Krüger, A., Tünnermann, J., Rohlfing, K., & Scharlau, I. (2018). Quantitative explanation as a tight coupling of data, model, and theory. *Archives of Data Science, Series A (Online First)*, *5*(1), A10, 27 S. online. doi:10.5445/KSP/1000087327/10

Krüger, A., Tünnermann, J., & Scharlau, I. (2016). Fast and conspicuous? Quantifying salience with the theory of visual attention. *Advances in Cognitive Psychology*, *12*(1), 20–38. doi:10.5709/acp-0184-1

Krüger, A., Tünnermann, J., & Scharlau, I. (2017). Measuring and modeling salience with the theory of visual attention. *Attention, Perception, & Psychophysics*, 1–22. doi:10.3758/s13414-017-1325-6

Kruschke, J. (2010). *Doing bayesian data analysis: A tutorial introduction with R*. Boston, MA: Academic Press.

Kruschke, J. (2014). *Doing bayesian data analysis: A tutorial with r, jags, and stan*. Boston, MA: Academic Press.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, *142*(2), 573.

Kruschke, J. K. & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206. doi:10.3758/s13423-016-1221-4

Kuhn, T. (1957). *The copernican revolution: Planetary astronomy in the development of western thought*. A Harvard Paperback. Harvard University Press. Retrieved from https://books.google.de/books?id=sWScX%5C_aduGMC

Kuss, M., Jäkel, F., & Wichmann, F. A. (2005). Bayesian inference for psychometric functions. *Journal of Vision*, *5*(5), 8. doi:10.1167/5.5.8

Kyllingsbæk, S., Markussen, B., & Bundesen, C. (2012). Testing a poisson counter model for visual identification of briefly presented, mutually confusable single stimuli in pure accuracy tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 628.

Landeau, L. & Terborgh, J. (1986). Oddity and the 'confusion effect' in predation. *Animal Behaviour*, *34*(5), 1372–1380. doi:10.1016/S0003-3472(86)80208-1

Lee, M. & Wagenmakers, E. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Leonard, T. H. (2014). A personal history of Bayesian statistics. *Wiley Interdisciplinary Reviews: Computational Statistics*, *6*(2), 80–115. doi:10.1002/wics.1293

Li, K., Kozyrev, V., Kyllingsbæk, S., Treue, S., Ditlevsen, S., & Bundesen, C. (2016). Neurons in Primate Visual Cortex Alternate between Responses to Multiple Stimuli in Their Receptive Field. *Frontiers in Computational Neuroscience*, *10*. doi:10.3389/fncom.2016.00141

Li, K., Kozyrev, V., Kyllingsbæk, S., Treue, S., Ditlevsen, S., & Bundesen, C. (2016). Neurons in primate visual cortex alternate between responses to multiple stimuli in their receptive field. *Frontiers in computational neuroscience*, *10*, 141.

Li, Z. (2001). Computational design and nonlinear dynamics of a recurrent network model of the primary visual cortex. *Neural Computation*, *13*, 1749–1780. doi:10.1162/08997660152469332

Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, *6*, 9–16. doi:10.1016/S1364-6613(00)01817-9

Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal Distributions across the Sciences: Keys and CluesOn the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: That is the question. *BioScience*, *51*(5), 341–352. doi:10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2

Little, R. J. (2006). Calibrated Bayes. *The American Statistician*, *60*, 213–223. doi:10.1198/000313006X117837

Logan, G. D. (1996). The code theory of visual attention: An integration of space-based and object-based attention. *Psychological review*, *103*(4), 603.

Logan, G. D. (2004). Cumulative Progress in Formal Theories of Attention. *Annual Review of Psychology*, *55*, 207–234. doi:10.1146/annurev.psych.55.090902.141415

Los, S. A. (2010). Foreperiod and sequential effects: Theory and data. *Attention and time*, *289*(302), 24.

Luce, R. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.

Lyon, A. (2013). Why are Normal Distributions Normal? *The British Journal for the Philosophy of Science*. doi:http://dx.doi.org/10.1093/bjps/axs046

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W. H. Freeman.

Matthias, E., Bublak, P., Müller, H. J., Schneider, W. X., Krummenacher, J., & Finke, K. (2010). The influence of alertness on spatial and nonspatial components of visual attention. *Journal of Experimental Psychology.*

Human Perception and Performance, *36*, 38–56. doi:10.1037/a0017602

Mayo, D. (2016). Don't throw out the error control baby with the bad statistics bathwater: A commentary. *The American Statistician. Retrieved from http://www. tandfonline. com/doi/suppl/10.1080/00031305.2016*, *1154108*.

Miller, J. (2016). Statistical facilitation and the redundant signals effect: What are race and coactivation models? *Attention, Perception, & Psychophysics*, *78*(2), 516–519. doi:10.3758/s13414-015-1017-z

Moore, J. (2015). Pragmatism, mathematical models, and the scientific ideal of prediction and control. *Behavioural Processes*. SI:SQAB 2014, *114*(Supplement C), 2–13. doi:10.1016/j.beproc.2015.01.007

Müller, H. J. & Krummenacher, J. (2006). Locus of dimension weighting: Preattentive or postselective? *Visual Cognition*, *14*(4-8), 490–513.

Muthukrishna, M. & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, *3*. doi:10.1038/s41562-018-0522-1

Nordfang, M., Dyrholm, M., & Bundesen, C. (2013). Identifying bottom-up and top-down components of attentional weight by experimental analysis and computational modeling. *Journal of Experimental Psychology: General*, *142*, 510–535. doi:10.1037/a0029631

Nordfang, M., Staugaard, C., & Bundesen, C. (2017). Attentional weights in vision as products of spatial and nonspatial components. *Psychonomic Bulletin & Review*, 1–9. doi:10.3758/s13423-017-1337-1

Nothdurft, H.-C. (2000). Salience from feature contrast: Additivity across dimensions. *Vision Research*, *40*, 1183–1201. doi:10.1016/S0042-6989(00)00031-6

Oakes, M. W. (1986). *Statistical Inference*. Newton Lower Falls, MA: Epidemiology Resources Incorporated.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). doi:10.1126/science.aac4716

Peng, R. (2015). The reproducibility crisis in science: A statistical counterattack. *Significance*, *12*(3), 30–32. doi:10.1111/j.1740-9713.2015.00827.x. eprint: https://rss.

onlinelibrary.wiley.com/doi/pdf/10.1111/j.1740-9713.2015.00827.x

Petersen, A., Kyllingsbæk, S., & Bundesen, C. (2013). Attentional dwell times for targets and masks. *Journal of vision*, *13*(3), 34–34.

Plummer, M. (2003). {JAGS}: A program for analysis of {bayesian} graphical models using {gibbs} sampling. Proceedings of the 3rd international workshop on distributed statistical computing.

Popper, K. (1959). *The logic of scientific discovery*. Routledge.

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, *32*(1), 3–25.

Poth, C. H. & Schneider, W. X. (2013). Aufmerksamkeit. In A. Stephan & S. Walter (Eds.), *Handbuch kognitionswissenschaft* (pp. 221–230). Weimar: J.B. Metzler.

Reichenbach, H. & Richardson, A. W. (1938). Experience and prediction: An analysis of the foundations and the structure of knowledge.

Rodgers, J. L. (2010). The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *American Psychologist*, *65*, 1–12.

Rouder, J. N., Haaf, J. M., & Aust, F. (2017). From theories to models to predictions: A Bayesian model comparison approach. *PsyArXiv*. doi:10.17605/OSF.IO/JT4TH

Rouder, J. N. & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, *12*(4), 573–604. doi:10.3758/BF03196750

Rouder, J. N., Morey, R. D., Verhagen, J., Province, J. M., & Wagenmakers, E.-J. (2016). Is There a Free Lunch in Inference? *Topics in Cognitive Science*, *8*(3), 520–547. doi:10.1111/tops.12214

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. doi:10.3758/PBR.16.2.225

Rouder, J., Morey, R., & Wagenmakers, E.-J. (2016). The interplay between subjectivity, statistical practice, and psychological science. *Collabra: Psychology*, *2*(1).

Ryan, R. (2013). *The oxford handbook of human motivation*. Oxford Library of Psychology. New York, NY: Oxford University Press.

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in python using PyMC3. *PeerJ Computer Science*, *2*, e55. doi:10.7717/peerj-cs.55

Schneider, W. X. (2013). Selective visual processing across competition episodes: A theory of task-driven visual attention and working memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1628), 20130060.

Schubert, T., Finke, K., Redel, P., Kluckow, S., Müller, H., & Strobach, T. (2015). Video game experience and its influence on visual attention parameters: An investigation using the framework of the Theory of Visual Attention (TVA). *Acta Psychologica*, *157*, 200–214. doi:10.1016/j.actpsy.2015.03.005

Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *Journal of Vision*, *11*, 9. doi:10.1167/11.5.9

Shibuya, H. & Bundesen, C. (1988). Visual selection from multielement displays: Measuring and modeling effects of exposure duration. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 591–600. doi:10.1037/0096-1523.14.4.591

Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human information processing: Ii. perceptual learning, automatic attending and a general theory. *Psychological review*, *84*(2), 127.

Shore, D. I., Spence, C., & Klein, R. M. (2001). Visual Prior Entry. *Psychological Science*, *12*, 205–212. doi:10.1111/1467-9280.00337

Silvis, J. D. & Donk, M. (2014). The effects of saccade-contingent changes on oculomotor capture: Salience is important even beyond the first oculomotor response. *Attention, Perception, & Psychophysics*, *76*, 1803–1814. doi:10.3758/s13414-014-0688-1

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366. doi:10.1177/0956797611417632

Simons, D. J. & Chabris, C. F. (1999). Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events. *Perception*, *28*, 1059–1074. doi:10.1068/p281059

Simonsohn, U. (2014). Posterior-hacking: Selective reporting invalidates bayesian results also. *Available at SSRN 2374040*. doi:10.2139/ssrn.2374040

Song, F., Parekh, S., Hooper, L., Loke, Y. K., Ryder, J., Sutton, A. J., … Harvey, I. (2010). Dissemination and publication of research findings: An updated review of related biases. *Health Technol Assess*, *14*(8), 1–193.

Spence, C. & Parise, C. (2010). Prior-entry: A review. *Consciousness and Cognition*, *19*, 364–379. doi:10.1016/j.concog.2009.12.001

Sperling, G. (1967). Successive approximations to a model for short term memory. *Acta Psychologica*, *27*, 285–292. doi:https://doi.org/10.1016/0001-6918(67)90070-4

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639. doi:10.1111/1467-9868.00353

Stevens, M. & Merilaita, S. (2011). *Animal camouflage: Mechanisms and function*. Cambridge: Cambridge University Press.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*(3), 153–181. doi:10.1037/h0046162

Stigler, S. M. et al. (1986). Laplace's 1774 memoir on inverse probability. *Statistical Science*, *1*(3), 359–363.

Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of personality and social psychology*, *54*(5), 768.

Suppes, P. (1966). Models of Data. In E. Nagel, P. Suppes, & A. Tarski (Eds.), *Studies in Logic and the Foundations of Mathematics* (Vol. 44, pp. 252–261). Logic, Methodology and Philosophy of Science. Elsevier. doi:10.1016/S0049-237X(09)70592-0

Taagepera, R. (2008). *Making Social Sciences More Scientific : The Need for Predictive Models: The Need for Predictive Models*. OUP Oxford.

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, *135*, 77–99. doi:10.1016/j.actpsy.2010.02.006

Theeuwes, J. (2013). Feature-based attention: It is all bottom-up priming. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1628), 20130055.

Tinbergen, L. (1960). The natural control of insects in pinewoods. *Archives neerlandaises de zoologie*, *13*(3), 265–343.

Titchener, E. B. (1908). *Lectures on the elementary psychology of feeling and attention*. New York : The Macmillan Company, 1908.

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

Treue, S. (2003). Visual attention: The where, what, how and why of saliency. *Current Opinion in Neurobiology*, *13*, 428–432. doi:10.1016/S0959-4388(03)00105-3

Tsotsos, J. K. (1990). Analyzing vision at the complexity level. *Behavioral and Brain Sciences*, *13*(03). doi:10.1017/S0140525X00079577

Tudge, L., McSorley, E., Brandt, S. A., & Schubert, T. (2017). Setting things straight: A comparison of measures of saccade trajectory deviation. *Behavior research methods*, *49*(6), 2127–2145.

Tünnermann, J., Petersen, A., & Scharlau, I. (2015). Does attention speed up processing? Decreases and increases of processing rates in visual prior entry. *Journal of Vision*, *15*, 1–1. doi:10.1167/15.3.1

Tünnermann, J. & Scharlau, I. (2018a). Left visual field advantages in temporal-order judgments assessed with the theory of visual attention. In preparation.

Tünnermann, J., Krüger, A., & Scharlau, I. (2017). Measuring attention and visual processing speed by model-based analysis of temporal-order judgments. *JoVE (Journal of Visualized Experiments)*, (119), e54856.

Tünnermann, J. & Scharlau, I. (2018b). Stuck on a plateau? a model-based approach to fundamental issues in visual temporal-order judgments. *Vision*, *2*(3), 29.

Vadillo, M. A., Gold, N., & Osman, M. (2016). The Bitter Truth About Sugar and Willpower: The Limited Evidential Value of the Glucose Model of Ego Depletion. *Psychological Science*, *27*(9), 1207–1214. doi:10.1177/0956797616654911

Van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of bayesian articles in psychology: The last 25 years. *Psychological Methods*, *22*(2), 217.

van Zoest, W., Donk, M., & Van der Stigchel, S. (2012). Stimulus-salience and the time-course of saccade trajectory deviations. *Journal of Vision*, *12*, 16. doi:10.1167/12.8.16

Vangkilde, S., Coull, J. T., & Bundesen, C. (2012). Great expectations: Temporal expectation modulates perceptual processing speed. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 1183–1191. doi:10.1037/a0026343

Vanpaemel, W. (2011). Constructing informative model priors using hierarchical methods. *Journal of Mathematical Psychology*. Special Issue on Hierarchical Bayesian Models, *55*(1), 106–117. doi:10.1016/j.jmp.2010.08.005

Vingilis-Jaremko, L., Ferber, S., & Pratt, J. (2008). Better late than never: How onsets and offsets influence prior entry and exit. *Psychological Research*, *72*(4), 443–450.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189. doi:10.1016/j.cogpsych.2009.12.001

Wagenmakers, E.-J., Verhagen, J., Ly, A., Matzke, D., Steingroever, H., Rouder, J. N., & Morey, R. D. (2017). The need for bayesian hypothesis testing in psychological science. *Psychological science under scrutiny: Recent challenges and proposed solutions*, 123–138.

Watanabe, S. (2010). Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *J. Mach. Learn. Res. 11*, 3571–3594. Retrieved from http://dl.acm.org/citation.cfm?id=1756006.1953045

Weiß, K., Hilkenmeier, F., & Scharlau, I. (2013). Attention and the Speed of Information Processing: Posterior Entry for Unattended Stimuli Instead of Prior Entry for Attended Stimuli. *PLOS ONE*, *8*(1), e54257. doi:10.1371/journal.pone.0054257

Weiß, K. & Scharlau, I. (2011). Simultaneity and temporal order perception: Different sides of the same coin? Evidence from a visual prior-entry study. *The Quarterly Journal of Experimental Psychology*, *64*, 394–416. doi:10.1080/17470218.2010.495783

Westheimer, G. (1998). Lines and gabor functions compared as spatial visual stimuli. *Vision Research*, *38*(4), 487–491.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical Evidence in Experimental Psychology An Empirical Comparison Using 855 t Tests. *Perspectives on Psychological Science*, *6*, 291–298. doi:10.1177/1745691611406923

Wichmann, F. A. & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–1313. doi:10.3758/BF03194544

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*, 419–433.

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202–238. doi:10.3758/BF03200774

Wolfe, J. M. (2007). Guided search 4.0: Current progress with a model of visual search. In W. Gray (Ed.), *Integrated Models of Cognitive Systems* (pp. 99–119). New York: Oxford.

Wolfe, J. M. & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*, 495–501. doi:10.1038/nrn1411

Wolfe, J. M. & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 0058.

Wright, R. & Ward, L. (2008). *Orienting of attention*. New York, NY: Oxford University Press.

Yarkoni, T. & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. doi:10.1177/1745691617693393

Zhaoping, L. (2014). *Understanding vision: Theory, models, and data*. New York, NY: Oxford University Press.

Manuscripts

Original manuscripts

The following pages contain the manuscripts of the articles that are part of this cumulative dissertation. However, the digitally archived version ends on this page respecting the copyrights of the respective academic journals. If you are reading this abridged version, please refer to Table 1 on page ii for the references.

Links to the articles, the primary data, an implemented version of the Bayesian model, and an analysis script including an interactive explanation are also published online (Krüger, 2020).