FACULTY OF PSYCHOLOGY AND EDUCATIONAL SCIENCES
RESEARCH UNIT OF BRAIN AND COGNITION
LABORATORY OF EXPERIMENTAL PSYCHOLOGY

Prägnanz in visual perception and aesthetic appreciation

Interactions between stimulus, person, and context

Eline Van Geert

Doctoral thesis offered to obtain
the degree of Doctor of Psychology (PhD)

Supervisor: Prof. Dr. Johan Wagemans

2023

FACULTY OF PSYCHOLOGY AND
EDUCATIONAL SCIENCES

LABORATORY OF EXPERIMENTAL PSYCHOLOGY
RESEARCH UNIT OF BRAIN AND COGNITION

Prägnanz in visual perception and aesthetic appreciation

Interactions between stimulus, person, and context

Eline Van Geert

Doctoral thesis offered to obtain the degree of
Doctor of Psychology (PhD)

Supervisor: Prof. Dr. Johan Wagemans

2023

Cover illustration and illustrations at the start of each part were created by Joris Snaet.
This thesis was created using the Quarto publishing system. Both a pdf version and an interactive HTML version are available online (https://elinevg.github.io/pragnanz). Online documentation appending some of the projects discussed in this dissertation can be found on the Open Science Framework (https://doi.org/10.17605/osf.io/7rsyf).

Summary

Things look as they do, not only because of the visual input an individual receives, but also because of the way in which the viewer organizes the input in a specific context (cf. Chapter 1). The law of *Prägnanz* states that psycho-physical organization will always be as 'good' as possible given the prevailing circumstances. What does it mean for an organization to be 'good', however, and how do viewers clarify the incoming visual stimuli to achieve the best possible percept?

A 'good' psycho-physical organization will contain at least some form of unity or regularity, and will become better when it is autonomous, complete, simple of structure, element rich, expressive, and/or meaningful (cf. Chapter 2). Prägnanz thus not only deals with purely figural aspects, but also with how purely and compellingly the phenomenal structure embodies an essence. Moreover, it comprises aspects related to both order and complexity.

To achieve the best or clearest overall organization possible, human perceivers use *internal representations of good Gestalts as reference points* to clarify the visual input. Whereas sometimes these reference points lead to increased *sensitivity to change*, they can also act as robust magnets, decreasing perceptual distance between neighboring stimuli. In a first empirical study (cf. Chapter 3), we show how the existence of strong internal reference points can explain differences in categorization, discrimination, and similarity judgments across a series of figures gradually transforming from one recognizable shape to another.

If no strong pre-existing reference points are available that are similar enough to the incoming visual stimuli, the *immediate context* can play a more extensive role in disambiguating the visual input. In a Registered Report using multistable dot lattices as stimuli (cf. Chapter 4), we confirmed differences in how individuals combine previous input and experience with current input in their perception, and showed stability of these individual differences over one to two weeks' time. Furthermore, we developed an efficient Bayesian observer model that can predict both the *attractive and repulsive* immediate context effects found in this task (cf. Chapter 5).

How will the internal reference points or the immediate context be used to clarify the incoming visual stimuli and arrive at a clearer, more prägnant percept? In every perceptual event, both simplification and complication will occur, albeit to a different extent depending on the strength of the internal and external conditions (related to the viewer and the visual input, resp.). Whereas unnecessary details, differentiating the visual input from the point of comparison, will be leveled or removed, important characteristics distinguishing the input from the reference will be emphasized. As we can conclude from our research, however, *which* characteristics will be either superfluous or essential, and as a consequence which characteristics will be either *simplified or complicated*, will partially depend on the context in which the stimulus is presented (cf. Chapter 6).

Order and complexity are not only important aspects of Prägnanz, they also contribute to our experience of aesthetic appreciation. To improve reproducibility of research assessing the relation between order, complexity, and aesthetic appreciation in the visual modality, we developed the Order & Complexity Toolbox for Aesthetics (OCTA; cf. Chapter 7), a Python package and online application focused on creating multi-element displays varying on different order and complexity dimensions (e.g., shape, color, size of elements). Although primarily focused on aesthetics, OCTA can also be used in research on perceptual organization. Some first empirical studies using OCTA stimuli (not part of this dissertation) teach us that whereas order is almost never disfavored, the appreciation of complexity is more context-dependent. This relates closely to the core of Prägnanz: as some form of regularity is always required, how much intricacy a viewer can handle differs between contexts and individuals.

In sum (cf. Chapter 8), to clarify and aesthetically evaluate visual stimuli, tendencies are at work that are both antagonistic and complementary: although they tend to decrease each other's influence, they also work together towards a better perceptual organization. What the optimal balance of both tendencies entails exactly will depend on the input the individual receives, the individual in question, the context in which the individual receives the input, as well as their interactions.

Samenvatting

Dingen zien er niet alleen uit zoals ze eruitzien vanwege de visuele input die een individu ontvangt, maar ook vanwege de manier waarop de kijker deze input in een specifieke context organiseert (cf. Hoofdstuk 1). De wet van *Prägnanz* stelt dat psycho-fysische organisatie altijd zo 'goed' als mogelijk zal zijn gegeven de heersende omstandigheden. Wat betekent het echter voor een organisatie om 'goed' te zijn, en hoe verduidelijken kijkers de binnenkomende visuele prikkels om tot de best mogelijke waarneming ervan te komen?

Een 'goede' psycho-fysische organisatie zal op z'n minst een vorm van eenheid of regelmaat bevatten, en wordt beter wanneer ze daarnaast ook autonoom, volledig, eenvoudig van structuur, rijk aan elementen, expressief en/of betekenisvol is (cf. Hoofdstuk 2). Prägnanz gaat dus niet alleen over puur vormelijke aspecten, maar ook over hoe puur en overtuigend de waargenomen structuur een essentie belichaamt. Bovendien omvat Prägnanz zowel aspecten gerelateerd aan orde als aspecten gerelateerd aan complexiteit.

Om tot de beste of duidelijkste overkoepelende organisatie mogelijk te komen, gebruiken menselijke waarnemers *interne voorstellingen van goede Gestalten als referentiepunten*, om zo de binnenkomende visuele prikkels te verduidelijken. Alhoewel deze referentiepunten soms leiden tot een verhoogde *sensitiviteit voor verandering*, kunnen ze ook als robuuste magneten functioneren door het waargenomen verschil tussen naburige stimuli te verkleinen. In een eerste empirische studie (cf. Hoofdstuk 3) tonen we hoe het bestaan van sterke interne referentiepunten verschillen kan verklaren in categorisatie-, discriminatie-, en similariteitsbeoordelingen over een reeks van figuren die gradueel van een herkenbare naar een andere herkenbare figuur transformeren.

Als er geen sterke reeds bestaande referentiepunten beschikbaar zijn die voldoende lijken op de binnenkomende visuele prikkels, dan kan de *onmiddellijke context* een belangrijkere rol gaan spelen in het disambigueren van de visuele input. In een Gepreregistreerd Rapport dat gebruik maakt van multistabiele stippenrasters als stimuli (cf. Hoofdstuk 4), bevestigden we verschillen in hoe individuen voorafgaande input en ervaring in hun waarneming combineren met de huidige input. Verder toonden we aan dat deze individuele verschillen stabiel bleven over één tot twee weken tijd. Daarnaast ontwikkelden we een efficiënte Bayesiaanse waarnemersmodel dat zowel de *aantrekkende* als *afstotende* effecten van de onmiddellijke context kon voorspellen die we in deze taak vonden (cf. Hoofdstuk 5).

Hoe gaan we de interne referentiepunten of onmiddellijke context dan gebruiken om de binnenkomende visuele prikkels te verduidelijken en zo een duidelijker, meer prägnante waarneming te bekomen? In elke perceptuele handeling zullen zowel simplificatie als complicatie voorkomen, al zij het in verschillende mate afhankelijk van de sterkte van de interne en externe omstandigheden (resp. gerelateerd aan de kijker en de visuele input). Terwijl onnodige details waarin de visuele input verschilt van het vergelijkingspunt afgezwakt of verwijderd zullen worden, zullen belangrijke kenmerken die de visuele input onderscheiden van de referentie net benadrukt of uitvergroot worden. Zoals we echter kunnen besluiten uit ons onderzoek: *welke* kenmerken overbodig of net essentieel zullen zijn, en als gevolg daarvan welke kenmerken *gesimplificeerd of gecompliceerd* zullen worden, zal gedeeltelijk afhangen van de context waarin de stimulus wordt aangeboden (cf. Hoofdstuk 6).

Orde en complexiteit zijn niet alleen belangrijke aspecten van Prägnanz, ze dragen ook bij aan het ervaren van esthetische appreciatie. Om de reproduceerbaarheid te verbeteren van onderzoek dat de relatie tussen orde, complexiteit, en esthetische appreciatie bestudeert in de visuele modaliteit, hebben we de Orde & Complexiteits-Toolbox voor Esthetiek (OCTA; cf. Hoofdstuk 7) ontwikkeld, een Pythonpakket en online applicatie die focussen op de creatie van multi-elementdisplays die variëren op verschillende orde- en complexiteitsdimensies (bijvoorbeeld vorm, kleur, grootte van de elementen). Hoewel de tool voornamelijk gecreëerd is voor toepassing binnen de esthetiek, kan OCTA ook gebruikt worden in onderzoek over perceptuele organisatie. Een aantal eerste empirische studies die OCTA stimuli gebruiken (die geen deel uitmaken van dit doctoraat) leren ons dat hoewel orde bijna nooit wordt afgekeurd, de appreciatie van complexiteit meer contextafhankelijk is. Dit sluit sterk aan bij de kern van Prägnanz: waar enige vorm van regelmaat een algemene vereiste is, verschilt de mate van complexiteit die een kijker aankan tussen contexten en individuen.

Over het algemeen (cf. Hoofdstuk 8) kunnen we dus stellen dat er, om de visuele input te verduidelijken en esthetisch the evalueren, tendensen aan het werk zijn die zowel antagonistisch als complementair zijn: hoewel ze elkaars invloed verminderen, werken ze samen toe naar een betere perceptuele organisatie. Waar het optimale evenwichtspunt ligt zal afhankelijk zijn van de input die het individu binnenkrijgt, het individu in kwestie, de context waarin het individu deze input ontvangt, alsook hun interacties.

Dankwoord - Thanks

"If I have seen further, it is by standing on the shoulders of giants."
– Isaac Newton (1676)

De voorbije vijfenhalf jaar heb ik aan mijn doctoraat gebouwd, met als centrale vraag waarom we de dingen waarnemen zoals we ze waarnemen. Daarbij gaat het niet enkel om het puur visuele waarnemen, maar ook over waarnemen of perceptie in de bredere zin, over hoe we naar de werkelijkheid kijken, die vorm geven, organiseren, en interpreteren. Zoals ook blijkt uit de titel van mijn doctoraat zie ik daarbij een belangrijke rol voor interacties tussen aspecten van de binnenkomende stimulatie, de persoon (mezelf als waarnemer en onderzoeker), en de context (mijn persoonlijke omgeving zowel als mijn onderzoeksomgeving). Zonder die specifieke persoonlijke en academische context had dit doctoraat er ongetwijfeld anders uitgezien, of was het er misschien zelfs niet geweest. Ik heb het gevoel dat het werken aan dit doctoraat me verder heeft gebracht, zowel in mijn kijk op de menselijke psychologie als in mijn kijk op en omgang met de wetenschappelijke wereld. En als ik werkelijk verder heb kunnen kijken, dan is dat omdat ik op de vele schouders mocht staan die mij hebben ondersteund. Zoals Isaac Newton (1676) het zei: "If I have seen further, it is by standing on the shoulders of giants." Hieronder mijn woorden van dank voor die vele unieke en prachtige mensen die mij hebben ondersteund.

Johan - Bedankt om van bij de start in mij te geloven en mij de tijd en ruimte te bieden om mijn eigen weg in de onderzoekswereld te banen. Onder jouw begeleiding heb ik me onder andere kunnen bekwamen in theoretische topics en in open en reproduceerbare onderzoekspraktijken, twee van mijn persoonlijke stokpaardjes. Voor de onvoorwaardelijke steun die je mij hebt geboden en voor je onwrikbare geloof in mijn kwaliteiten en groeimogelijkheden ben ik je voor altijd dankbaar. Het is dat geloof en de daarbijhorende aanmoediging en steun die ik nodig had om verder te komen dan ik zelf mogelijk achtte. Zonder de onnoemelijk vele kansen die je mij hebt geboden en nog steeds biedt, stond ik niet waar ik nu sta.

Pieter - Je stimuleerde mijn drive om effectief iets te willen veranderen aan de status quo in academia, door zelf het goede voorbeeld te geven en mij aan te moedigen hetzelfde te doen. Dankzij jou zette ik bijvoorbeeld mee mijn schouders onder de universiteitsbrede ReproducibiliTea journal club die we opstartten aan KU Leuven. Dat je wat meer relaxed en down to earth bent is iets wat ik heel erg in je bewonder, en waar ik nog veel van kan leren. Op een bepaald moment zei je dat de studie die ik voor ogen had "een doctoraat op zich" was, en dat advies om niet te veel tegelijk te willen doen is er één waar ik vaak aan terugdenk.

Prof. Dr. Wolf Vanpaemel, Prof. Dr. Floris de Lange, Prof. Dr. Vebjørn Ekroll, Dr. Aenne Brielmann - Thank you for being in my doctoral examination committee, for taking the time to read my dissertation, and for providing helpful and constructive feedback. I also want to thank all members of my mid-term advisory committee for their questions and suggestions.

Julia, Steven, Nori, Rong, Jonas, Yaniv, Elisabeth, and all my other past, present, or future collaborators - Thank you for your feedback on my work, for exchanging ideas with me, for collaborating with me in the past or present, and/or having agreed to collaborate with me in the near future. I am looking forward to bring our shared projects to a success.

Agna - Mama van het GestaltReVision labo, bedankt om al die jaren onze steun en toeverlaat te zijn. Je hebt me niet alleen administratief en praktisch ondersteund, maar was ook begaan met het meer persoonlijke en emotionele welzijn van elk van ons. Bedankt om al die jaren een baken van vertrouwen te zijn voor alle onderzoekers binnen het GestaltReVision team.

Christophe - Onze IT guy, je hebt niet alleen mijn eerste experiment geprogrammeerd, door je enthousiasme voor Python en je motiverende en interactieve lessenreeks heb ik zelf ook de Pythonmicrobe te pakken gekregen. Verder zou de Order & Complexity Toolbox for Aesthetics er zonder jou ook helemaal anders uitgezien hebben! Bedankt voor al je wetenschappelijke bijdrages aan mijn werk, alsook voor je niet-wetenschappelijke bijdrages aan de hartelijke sfeer op het werk.

Kirsten - Bedankt om me in het laatste jaar van mijn doctoraat praktisch en administratief te ondersteunen. Hoe vaak ik ook langskwam met vragen, je was telkens bereid om me verder te helpen waar mogelijk.

Jaana, Nathalie, en Lisa - Ik ben samen met jullie aan dit PhD avontuur begonnen. We hebben het alle vier op heel andere manieren aangepakt, en dat ging voor geen enkel van ons zonder slag of stoot. Jullie hebben me doen inzien hoe belangrijk een goede work-life balance is, en hoe kwetsbaar we als doctoraatsstudenten zijn voor allerlei vormen van druk (druk die we onszelf opleggen, maar ook druk van buitenaf). Ik heb er zelfs een song over geschreven! Bedankt voor de fijne start samen, en voor de inzichten die jullie me hebben gebracht.

Liesse, Tina, Daniel, and Astrid - Thank you for conducting your master's internship or thesis under my supervision. You gave me the opportunity to grow as a supervisor and learn from my mistakes (e.g., supervising three interns at the same time)! Tina and Liesse, I am very grateful for your crucial scientific contributions to my work. You pushed my work further than I could have imagined.

Agnes, Maja, Aline, Sander, and all other members of the Friends of Theory journal club - Thank you very much for bringing up all those interesting papers, for exchanging your ideas and thoughts on these theoretical topics, for reminding me of the importance of good theory in psychological science, and for sustaining my interest in theoretical topics in psychology.

Olivia, Natalie, Aline, and all other organizers of the ReprodubiliTea journal clubs at KU Leuven - Thank you for the time and effort you dedicated to the organization of the journal club, and for all your other efforts to make science more open and reproducible.

Lore, Charlotte, Ben, Claudia, Thiago, Sander, Miguel, Jan and Ans, Anne-Sofie, Chris, Elisabeth, Eleftheria, Laurie-Anne, Nofar, Boris, Nicole, Derya, Yaniv, Lisa, Gonzalo, Maarten, Tjerkje, Linda, Joan, and all other members of the GestaltReVision lab that I had the pleasure knowing - Thank you for being such wonderful labmates. Each of you contributed a unique part to what the lab has become, and has inspired me in different ways. Elisabeth en Eleftheria, jullie motivatie en inzet om aan open en reproduceerbaar onderzoek te doen geeft mij hoop voor de toekomst. Bedankt ook voor jullie luisterend oor wanneer ik daar nood aan had. Chris, thank you for proofreading some of the articles part of my PhD and for improving their readability!

Nathan, Hanne, Brenda, Anne-Merel, Brendan, Marleen, Sofie, and everyone else I forgot to mention here - Thank you for contributing to the pleasant broader research environment of Brain & Cognition. Hanne, bedankt voor je focus op het belang van goede wetenschap. Ik kijk naar je op als onderzoeker en als mens.

Many others who contributed to my doctoral path - In addition to the people I have mentioned above, there are many others I would like to thank. Graag wil ik Dhr. Joris Snaet bedanken voor het maken van de prachtige cartoons voor mijn doctoraatsproefschrift. To everyone who participated in my research, thank you very much! Ook aan alle Nederlandstalige deelnemers aan mijn onderzoek, heel erg bedankt! Thank you to the organizers and participants of all the workshops, summer schools, and winter schools that I attended. Participating in these events has been an important part of my development as an open and reproducible researcher. Thank you also to those reviewers who constructively pushed me further and helped me improve my manuscripts. I would also like to thank the Research Foundation - Flanders (FWO) for the financial support they provided me for four and a half years. Furthermore, I am grateful to Max Wertheimer, Kurt Koffka, Wolfgang Köhler, Wolfgang Metzger, Rudolf Arnheim, Edwin Rausch, Angelika Hüppe, and many others for their foundational and inspiring work on Prägnanz. On a more practical side, I want to thank the developers of the R, Python, and Javascript software and packages. Without them, my PhD would have looked completely different.

Naast de mensen die ik in het kader van mijn doctoraat leerde kennen, zijn er ook heel wat mensen uit mijn persoonlijke omgeving die ik graag wil bedanken.

Alle mensen die mee aan mijn weg als onderzoeker en mens bouwden zelfs voor ik mijn doctoraat startte - De discipline om hard te werken heb ik op het Sint-Vincentiusinstituut in Dendermonde gekweekt. Bedankt aan alle leerkrachten op mijn middelbare school die in mij geloofden en zo bijgedragen hebben aan de persoon die ik vandaag ben. Ook tijdens mijn bachelor- en masteropleiding psychologie heb ik veel kansen gekregen om me verder te ontwikkelen. Hier wil ik graag het onderzoeksteam o.l.v. Prof. Kuppens bedanken voor mijn leerrijke student-onderzoekerschap, Prof. Francis Tuerlinckx voor het begeleiden van het European Seminar on Quantitative Psychology en voor de mogelijkheid om aan het project dat ik daar startte verder te werken als jobstudent, Prof. Wolf Vanpaemel voor de introductie in het belang van reproduceerbaarheid van onderzoek, en Prof. Vera Hoorens voor de vele mogelijkheden en ondersteuning die ik kreeg tijdens mijn onderzoeksstage (o.a., het uitwerken van mijn eerste preregistratie). A big thank you also to the organizers of the Junior Researcher Programme (JRP) and all inspiring people I met during this experience, either at the summer school, during the project, or during the internship.

Kelsey, Stien, Paulien, Sebastiaan, Frederick, Pengqian, and all other Theory & Research students I got to know during my bachelor and master program - Thank you for all fun moments we spent together, either during studying or in our free time.

Jolien, Hanne, Laura, Stéphanie, Maxime, Anaïs, Eline, Sofie, Annika, en Sherin - Gelukkig zijn er ook nog andere dingen in het leven dan studies of werk. Bedankt voor alle fijne gesprekken, weekendjes, uitstappen of activiteiten die ik samen met jullie mocht beleven!

Familie - Nonkel Pieter, bedankt om me telkens het beste te wensen en me aan te moedigen! Dominiek en Claudine, bedankt om mij zo'n warme welkom te geven in de familie D'hooge.

Meter en peter - Jullie kunnen er spijtig genoeg niet meer bijzijn vandaag. Ik ben heel dankbaar dat ik jullie toch nog zo lang heb mogen kennen. Beiden geloofden jullie heel sterk in mijn kunnen, en ik ben er zeker van dat jullie heel trots op me geweest zouden zijn.

Mama en papa - Meer dan iedereen wil ik jullie bedanken voor jullie onvoorwaardelijke steun, het aanhoren van al mijn twijfels en vragen, en het oneindig helpen waar jullie konden. Jullie hebben een immense rol gespeeld in mijn ontwikkeling als mens en als wetenschapper. Ik denk bijvoorbeeld aan de vele musea en wetenschapsexpo's waar jullie me al van jongsaf aan mee naartoe namen. Daarnaast hebben jullie me ook steeds met heel jullie hart ondersteund om mijn eigen interesses te volgen. Was dat Grieks-Latijn? Dan was dat Grieks-Latijn. Was dat psychologie? Dan was dat psychologie. Woorden schieten tekort om aan te geven wat jullie nog steeds voor mij betekenen.

Joachim - Bedankt om zo'n liefdevolle en ondersteunende partner te zijn. Dankzij jou heb ik geleerd om te genieten van de kleine dingen in het leven, en je staat me bij op alle kleine en grote momenten. Verder heb je als geen ander ervaren hoe stressvol de laatste maanden voor het indienen van mijn doctoraat zijn geweest. Bedankt voor je geduld wanneer ik niet veel tijd had.

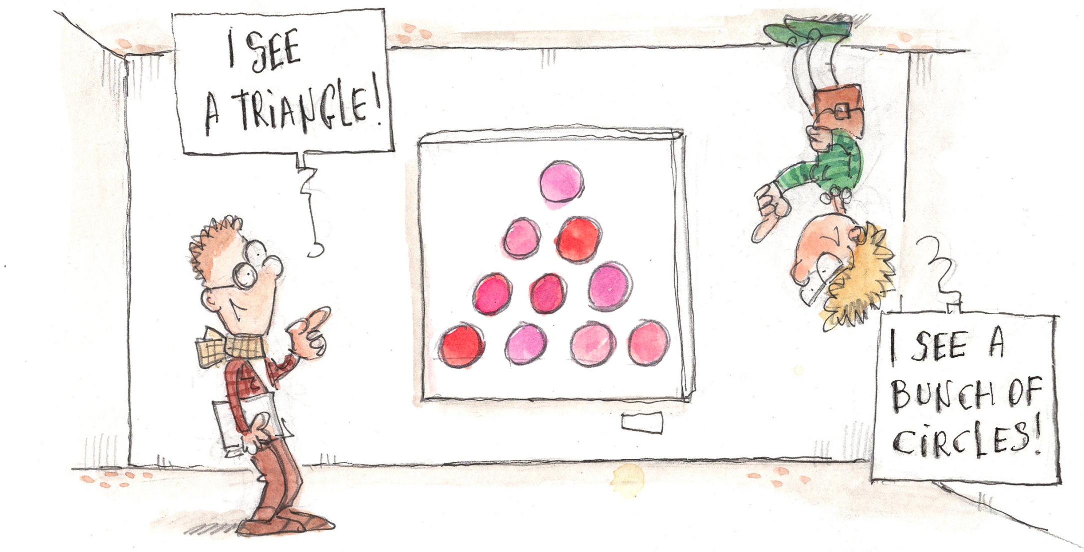
Ik ben me er dus heel bewust van dat ik hier vandaag sta met heel wat privileges: een stimulerende thuisomgeving, een liefhebbende partner, een vormende school, een rots van een promotor, en een motiverende labomgeving. Dankjewel om me elk op jullie manier zo goed te ondersteunen! En ondanks die vele steun die ik van elk van jullie heb gekregen, vind ik dat ik vandaag toch ook mezelf mag bedanken. Dank je, Eline, voor je inzet en doorzettingsvermogen tijdens het hele doctoraatstraject.

Table of contents

Prägnanz in visual perception and aesthetic appreciation

Part I

Introduction



© Joris Snaet

How we perceive things depends on how we *organize* the incoming visual stimulation in a specific context.

Chapter 1
General introduction

As humans, our visual abilities are amongst the most important tools we possess to enable immediate interaction with the world around us. Without visual perception, many everyday activities like preparing breakfast, navigating the streets, and reading, become much more exhausting or even impossible.

Although vision is sometimes presented as fully determined by the physical input it receives (i.e., light signals falling on the receptor cells in the retina), this is not the full story of how our perceptual experiences come about. An evident illustration is the existence of multistable figures like the duck-rabbit figure or the Necker cube (see Figure 1.1). Although the physical input stays exactly the same, each of these figures is perceived differently across individuals, contexts, and time. Visual perception is an active construction process, not only for obviously multistable figures. Even in so called clear-cut displays inter- and intra-individual differences exist based on, e.g., perceptual skills, viewing mode, expertise, vigilance level, and temporal or spatial context. Koffka (1935) already acknowledged that things look as they do, not because things are what they are (veridicality), nor because of the proximal stimuli (i.e., the physical stimulations of the sensory receptors) are what they are, but "because of the field organization to which the proximal stimulus distribution gives rise" (p. 98). In other words, to understand why and how we visually perceive the world in the way we do, we need to study the laws of psycho-physical[1] organization (Koffka, 1935).

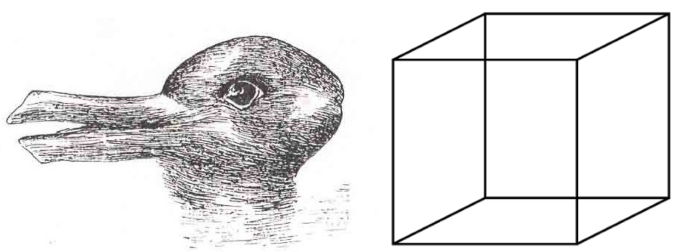


Figure 1.1: The duck-rabbit figure by Jastrow (1899) and the Necker cube (Necker, 1832).

Koffka (1935) presented **the law of *Prägnanz* as the main principle to guide this research on psycho-physical organization**. The law of Prägnanz states that psycho-physical organization will always be as 'good' as the prevailing conditions allow: we will always organize the input in the best way possible given the prevailing circumstances. However, (a) what does it mean for an organization to be *'good'* or *'prägnant'*, (b) *how* do we clarify the input to arrive at a better organization, and (c) what are the *prevailing conditions* to take into account? Notwithstanding

[1] I call this 'psycho-physical' organization because the organization is influenced both by physical stimulation of the sensory receptors (i.e., the proximal stimuli) and psychological factors depending on the perceiving organism in question. This should not be confused with Fechner's notion of psychophysics, which concerns the quantitative mapping of physical stimuli onto psychological entities (e.g., between stimulus intensity and sensation strength).

the abundant reference to Prägnanz in many journal article introductions and discussion sections since its emergence, modern-day mentions of Prägnanz often lack these further clarifications. In addition, the original Gestalt psychological context in which Prägnanz originated got lost, and the interpretation of Prägnanz has changed together with the shifting theoretical context (i.e., increasing focus on information processing theories).

The overall aim of this dissertation is to **arrive at a more fine-grained understanding of Prägnanz** (i.e., "goodness" of organization) and its added value for current theories and research on human visual perception and aesthetic appreciation. Thoroughly understanding what Prägnanz and the law of Prägnanz mean and *can* mean for current perception science first of all requires a thorough investigation of the existing literature on the Prägnanz concept. This includes studying the original sources introducing and extending the concept as well as other sources clarifying later interpretations, proposed quantifications, and critiques of the Prägnanz concept. Given the lack of digitalization and translation for many of the original German sources concerning Prägnanz, gathering and digesting this literature was not a trivial task. From this literature review (cf. Chapter 2 as well as the introduction below), it will become clear that the law of *Prägnanz* was posited *as a general overarching principle*. The goal was to stimulate concrete further investigation into specific organizational principles falling under this law, as well as into the ways in which these principles interact under diverse circumstances, depending on the stimulus, person, and context in question.

Taking Prägnanz as a generative framework for further research is exactly what the other chapters of this dissertation exemplify. The studies do not directly test the overarching principle, but take it as a starting point to explore the principles governing psycho-physical organization in more concrete cases.

In what follows, I give a sneak preview of what is to come. First, I discuss the most important lessons learned from reviewing the existing Prägnanz literature. Then, a brief introduction to each part of the dissertation follows. Afterwards, I highlight how each part connects to the overall Prägnanz framework.

1.1 Main aspects regarding Prägnanz

As mentioned, the *law of Prägnanz* states that psychological organization will always be as 'good' as possible given the prevailing conditions. But what does this imply concretely? To make the Prägnanz law a useful statement, it needs to be specified further (a) what a *'good'* psychological organization entails, (b) *how* the Prägnanz tendency can be realized, and (c) *which conditions* need to be taken into account (cf. also Chapter 2). Although the Gestalt school did provide answers to these questions, more recent references to Prägnanz often lack these clarifications.

1.1.1 What is a 'good' organization?

Prägnanz concerns the goodness or simplicity of an *experienced* overall organization. Importantly, Prägnanz thus is a property of percepts — not stimuli. Hence, what a 'good' organization entails, will depend not only on the incoming visual stimulation, but also on the perceiving organism and the context in which the input is encountered. Moreover, as Prägnanz focuses on the goodness of the experienced *overall organization*, it is neither to be equated with the simplicity of the elements part of the overall organization, nor with the simplicity of the process to form the overall organization.

To be experienced as 'good' or 'prägnant', a psychological organization needs to be different from a simple sum of its sensory elements: it has to be a *Gestalt* (Koffka, 1935; Smith, 1988; Wertheimer, 1922). A Gestalt is defined here as an experienced ensemble of elements that mutually support and determine one another (Ash, 1995; Köhler, 1920; Sundqvist, 2003). Therefore, to be experienced as a Gestalt, *at least some form of unity or regularity* should be experienced in the organization (Rausch, 1966). A Gestalt can be differentiated from a 'complex', in which the sensory elements are experienced as completely independent of each other — as a pure and-summation (Koffka, 1935; Smith, 1988).

The experienced unity or regularity can be due to purely figural or structural aspects, but can also be based on a match between structure and meaning (i.e., how purely and compellingly the structure represents an 'essence' or 'way of being'; Metzger, 1941), or on how strongly the experienced organization interacts with already existing knowledge within the perceiver (i.e., how meaningful the perceiver experiences the organization to be). For example, the artist Constantin Brâncuși represented the concept of a bird by bringing it back to its essence: one feather in this case (cf. Figure 1.2). Furthermore, even on the figural level, an organization can be considered good on diverse grounds, related to both order (i.e., the structure and organization of elements[2]) and complexity (i.e., the quantity and variety of elements).

Figure 1.2: *Bird in Space* by Constantin Brâncuși (1928). Part of the collection of the Museum of Modern Art in New York. Presented here for illustrative purposes under the Fair Use principle. © Succession Brancusi - All rights reserved (ARS) 2018.

Rausch (1966) summarized these factors influencing the goodness of an experienced organization in seven Prägnanz aspects. The presence of at least some form of (1) lawfulness or regularity is posited as the first and only necessary Prägnanz aspect. In addition, the goodness of an organization can increase when the organization is perceived as (2) autonomous rather than derived, (3) complete rather than disrupted, (4) simple of structure rather than complicated of structure, (5) element rich rather than element poor, (6) expressive rather than expressionless,

[2]Although I refer to 'elements' here, I mean this in a broader sense than only the sensory elements part of the organization. The organization can, for example, be associated with multiple congruent meanings (cf. Chapter 2), which can also contribute to the order and complexity in the organization.

and (7) meaningful rather than meaningless. Prägnanz or goodness of organization is thus a multifaceted concept: an experienced organization can be 'good' or 'better' for many different reasons.

Psychological organizations that excel in their Prägnanz can influence our experience of new incoming information: they can serve as a reference to which the input is internally compared. In this context, the term Prägnanz steps [*Prägnanzstufen*] is used to refer to prägnant forms that serve as reference regions on a dimension (e.g., the right angle in the realm of all possible angles).

1.1.2 How do we clarify the input to arrive at a better organization?

The tendency towards Prägnanz of a Gestalt indicates a tendency present in *every* process of psychological organization to come to the best, most clear-cut (i.e., most prägnant) organization given the prevailing conditions (Arnheim, 1975; Koffka, 1935; Köhler, 1920; Metzger, 1941). But *how* do we clarify the incoming stimulation to arrive at a better organization? In all cases, the incoming stimulation is compared to a 'reference'. This reference can be an internal representation of a good Gestalt (i.e., a Prägnanz step) or a reference present in the immediate spatial or temporal context. For example, when perceiving a chihuahua, the height of this chihuahua may be compared to the height of our internal representation of a prototypical dog. On the other hand, the height of the chihuahua may be compared to a locally present reference, for example, to the height of a fence present next to the dog, or to another even smaller chihuahua seen just before. Rather than a binary distinction between reference points and non-reference points, a gradual Prägnanz function applies to each variable dimension (e.g., height), with some regions showing higher Prägnanz than others. Furthermore, the course of this Prägnanz function per domain and dimension can differ between individuals and contexts. For example, the owner of a chihuahua may have different reference regions for dog height than the owner of a shepherd.

Gestalt psychology posited two main ways in which we can use this reference to clarify the incoming stimulation and come to a better Gestalt. On the one hand, we can remove or downsize unimportant details (i.e., simplification, leveling, assimilation, attraction) to make the incoming stimulation more similar to our reference (Arnheim, 1986; Koffka, 1935; Metzger, 1941). On the other hand, we can add, emphasize, or intensify characteristic features of the current stimulation (i.e., complication, sharpening, articulation, repulsion) to make the incoming stimulation stand out more clearly in comparison to the reference (Arnheim, 1986; Koffka, 1935; Metzger, 1941). Both of these tendencies can contribute to the emergence of a 'better', clearer, simpler Gestalt. Koffka (1935) referred to these tendencies as 'minimum simplicity' (i.e., the simplicity of uniformity) and 'maximum simplicity' (i.e., the simplicity of perfect articulation). Arnheim (1986) referred to them as 'tension-reducing' (i.e., decreasing the difference from the reference) and 'tension-enhancing' tendencies (i.e., increasing the difference from the reference and thereby making the currently experienced organization more unique). Koffka (1935) presented minimum and maximum simplicity as strict alternatives, and when viewed from a one-dimensional standpoint they certainly are: one either perceives the chihuahua's height as more similar to the height of a prototypical dog than is actually the case (i.e., simplification), or one exaggerates the smallness of the chihuahua even more than in reality (i.e., complication). I agree with Arnheim (1986), however, that simplification and complication are concurrently present in every perceptual event. In any real-life situation, the incoming stimulation is inherently multidimensional. In our example, the chihuahua is not only evaluated on its height, but maybe also based on its color, cuteness, etc. Therefore, in any multidimensional situation, simplification and complication — although antagonistic in the one-dimensional case — can be complementary and work together to increase the goodness of the experienced organization (cf. also Arnheim, 1986; Hubbell, 1940; Köhler, 1951/1993).

As a sidenote: Whereas one could equate simplification with literally 'removing' features and complication with 'adding' features, this does not have to be the case. For example, a square missing one of its four sides may be simplified by adding a sideline, resulting in a complete square. On the other hand, also removing a part of an organization can complicate an organization, e.g., removing a sideline from a full square.

Reference points can thus serve a double function. On the one hand, a reference point can serve as a magnet: assimilation to the reference may occur, especially when the incoming stimulation is rather weak. This tendency makes perception more robust: perceived organizations will be closer together than the stimuli from which they originate. On the other hand, a reference point can serve as an anchor: they can increase sensitivity in their vicinity (i.e., increase the ability to notice small deviations from the reference). In that sense, reference points support both robustness and sensitivity in visual experience. Under weak stimulus conditions (i.e., under low visibility) or when a specific difference from the reference is deemed unimportant, attraction to the reference will dominate. Under clear stimulus conditions and when a specific difference is experienced as significant, reference points will increase discrimination sensitivity in those regions where deviations matter most (i.e., close to the reference). The stimulating effect of reference points on discrimination sensitivity also allows for the formation of new reference levels in between existing ones when this becomes behaviorally or functionally useful. Which of the features of the organization will be treated as characteristic and which as unimportant will depend on the individual and the context in which the organization is perceived. That is, to determine which features will be part of the experienced organization's *essence*, it will matter in comparison to which *reference* the organization is perceived. Importantly, this dependence of the preferred tendency (i.e., simplification or complication) on the prevailing conditions does not make the Prägnanz tendency an empty statement, it does not imply that 'anything is possible'. What it does imply is that individual and context need to be taken into the equation in further research on Prägnanz tendencies, to determine which tendency will occur for which feature dimension under which concrete conditions.

In addition to the distinction between simplification and complication tendencies, also primary and secondary Prägnanz tendencies can be distinguished (Hüppe, 1984). The *primary Prägnanz tendency* considers the relation between the incoming stimulation and the phenomenally experienced organization resulting from the stimulation. More specifically, the primary Prägnanz tendency leads to deviations from the incoming stimulation to the perceived organization that are not directly noticeable by the observer. In other words, every phenomenal experience already deviates from the stimulus in the direction of Prägnanz. The *secondary Prägnanz tendency* operates on the level of phenomenal experience. It concerns the tendency to evaluate a perceived organization based on its experienced closeness to a prägnant form (i.e., a point of comparison, not necessarily phenomenally present). This secondary Prägnanz tendency may sometimes reach awareness. Both Prägnanz tendencies cannot be seen as completely independent, however: To be able to make statements about secondary Prägnanz, an organized perceptual field (influenced by primary Prägnanz) is preassumed. Importantly, I do not think of these primary and secondary Prägnanz tendencies as successive processes. On the contrary, I believe that according to traditional Gestalt theory, there is not first an organization of the perceptual field and only then more high-level cognitive identification or classification, but rather one complex dynamic process of Gestalt formation (see also Kruse, 1986).

The Prägnanz tendency, i.e., the tendency towards the best psychological organization possible given the prevailing conditions, can thus not only lead to clearer *percepts* (i.e., primary Prägnanz tendency), but it can also play a role in how we — cognitively, emotionally, or aesthetically — *evaluate* (and communicate) our psychologically experienced organization (i.e., secondary Prägnanz tendency). In my dissertation, I focus on human *visual* perception and how it may influence several cognitive and aesthetic evaluations of those perceived organizations.

As perceptual processing of the incoming information is necessary to be able to aesthetically evaluate a percept, the close relation between perception and aesthetics cannot be neglected. von Ehrenfels (1922) equated beauty and Gestalt height, which he defined as the product of unity (of the whole) and multiplicity (of the parts; von Ehrenfels, 1916). Moreover, Koffka (1940) called perception artistic and both Metzger (1941) and Arnheim (1975) noticed the presence of simplification and complication tendencies in artistic practice. Although most authors agree that there is a close relationship between aesthetics and Gestalt, there are two differing views on the exact relation. On the one hand, as von Ehrenfels (1916, 1922) proposed, aesthetic appreciation may be based on the absolute goodness level of the experienced organization (i.e., Prägnanz *height*). On the other hand, aesthetic appreciation may arise together with a consciously experienced *increase* in Prägnanz (i.e., the strength of the experienced Prägnanz *tendency*), and this does not necessarily relate to the absolute Prägnanz height. This second view relates closely to some other existing accounts of aesthetic appreciation, including the predictive processing accounts of Van de Cruys & Wagemans (2011) and Chetverikov & Kristjánsson (2016) as well as the focus on pleasure by insights into Gestalt proposed by Muth and Carbon (Muth et al., 2013; Muth & Carbon, 2013, 2016). Both views may act complementarily as well, however, and future research can investigate the relation between aesthetics and Prägnanz in more detail.

Although this dissertation focuses on human visual perception and aesthetic appreciation, the tendency towards prägnant overall organizations is meant to be a very *general* tendency. The Prägnanz tendency was expected to be present in all forms of psychological organization (e.g., Koffka, 1935; Köhler, 1920; Metzger, 1936/2006), for all types of stimulation (i.e., simple or complex), and for all types of species. In this context, it is important to stress that the Prägnanz *tendency*, present in every psychological organizational process, will not always result in a 'good' Gestalt in the absolute sense (Arnheim, 1987; Koffka, 1935; Metzger, 1941). The tendency towards the most prägnant Gestalt should thus always be seen as relative to the prevailing conditions.

1.1.3 What are the prevailing conditions to take into account?

As psychological organization takes place in an organism, it is constrained by the conditions outlined by the organism. Hence, in determining the goodness of an organization, not only the stimulus constellation will play a role, but also how the stimulus interacts with an individual in a specific spatial and temporal context. When it concerns psycho-physical processes like human perceptual organization, there are both external and internal conditions to consider (Koffka, 1935). *External conditions* are created by the proximal stimuli, i.e., the excitations within the receptor organs to which the light rays coming from the physical object give rise. Under weak external conditions (e.g., because of brief exposure time, low contrast, small size), Prägnanz tendencies will get more room to play a role and can even lead to tangible dislocations and distortions compared to the external stimulation (Koffka, 1935). The factors mentioned under weak external conditions relate to *low visibility*. It is important, however, to explicitly distinguish this overall uncertainty from any uncertainty or ambiguity present in the *features* of the stimulus (e.g., uncertainty in directions present in a random dot kinematogram, ambiguity in orientations present in a multistable dot lattice). *Internal conditions* are related to the structure and/or state of the organism's nervous system. Within the internal conditions, more permanent ones (related to the structure of the nervous system, influenced by both inheritance and previous experience) and more temporary ones (related to, e.g., vigilance, fatigue, needs, attitudes, interests, attentions; Koffka, 1935) can be distinguished.

1.1.4 The many faces of 'Prägnanz'[3]

Based on the extensive Gestalt literature on the topic, I conclude that Prägnanz has many faces. For instance, Prägnanz has been used to refer to both (a) a tendency present in every organizational process and (b) the goodness level of an experienced organization resulting from this Prägnanz tendency. Moreover, the term 'Prägnanz steps' is used to refer to reference regions with high Prägnanz when a single perceptual dimension is varied. In addition, Prägnanz is also regularly referred to in the context of aesthetic appreciation and artistic practice.

Furthermore, also within each of these use contexts, the multifacetedness of Prägnanz prevails. As a property of an experienced organization, Metzger (1941) stressed how both figural order and the pure, compelling embodiment of an essence are essential to understand the full meaning of Prägnanz, and Rausch (1966) posited seven general factors contributing to Prägnanz. As a tendency in every organizational process, Prägnanz can show as an attractive and/or a repulsive tendency relative to an internal or a local reference. Furthermore, the Prägnanz tendency can be primary (i.e., an unconscious difference between stimulus and percept) or secondary (i.e., a sometimes conscious evaluation of the closeness of an experienced organization to a reference). Furthermore, the seven Prägnanz aspects posited by Rausch (1966) can also point to the existence of different Prägnanz tendencies.

Although some may view this multifacetedness of Prägnanz as problematic, and leave Prägnanz and Gestalt theory behind because of this experienced 'vagueness', it is important to clarify the original goal with which Prägnanz was posited. The Prägnanz principle was never meant as a magical one-fits-all solution, and should therefore not be seen solely as an outcome of concrete research results, but also as a device for making new discoveries (Wertheimer, 1924/1999). By using Gestalt theory and the Prägnanz principle as a generative framework for future research, and studying more specific principles of organization and their interaction in concrete cases (Rausch, 1966), we can come to a better understanding of the principles underlying psychological organization.

Hence, the multifacetedness of Prägnanz, and its dependence on input, person, and context, does not imply that 'anything is possible'. Prägnanz provides a framework that highlights potential universality and diversity (i.e., constraints on generality) in psychological organization: different stimuli can lead to similar perceived organizations within an individual, while the same stimulation can be perceived differently depending on individual and context.

[3] The title of this section was inspired by Arnheim's (1986) article *"The two faces of Gestalt psychology"*.

The consideration of these interactions and dependencies does not render *specific* organizational principles following from the Prägnanz framework and under prespecified conditions untestable. And when it concerns Prägnanz as the overall framework, its validity should follow from observing these specific organizational principles under diverse circumstances. Only by organizing all these concrete observations in a coherent whole, we can come to the knowledge of a system (Koffka, 1935; cf. also Chapter 8). Hence, only *because of* its multifacetedness does Prägnanz provide a valuable generative framework for current cognitive science. In my dissertation, I attempt to bring this perspective on Prägnanz into practice. Each of the themes and studies introduced below relates to this overall framework, and in this introduction I aim to make the connection clear.

1.2 Preview on empirical work

1.2.1 Robustness and sensitivity

The part on robustness and sensitivity deals with how we can use *internal representations of good Gestalts* (i.e., Prägnanz steps) as a reference for comparison to arrive at a clearer percept. More specifically, these internal representations can support perceptual stability — related to robust categorization — while they can also support sensitivity to perceptual change — related to discrimination. In the empirical study reported in Chapter 3, we were interested in investigating the *category boundary effect* in perceptual discrimination and similarity judgments. The category boundary effect entails that even when keeping the physical difference between stimuli the same, differences between stimuli belonging to the same category are perceived as smaller than differences between stimuli belonging to different categories. Typically, this is tested and interpreted as an effect of the category boundary: performance on trials with between-category stimulus pairs is compared to performance on trials with within-category stimulus pairs (while equating the size of the difference between the stimuli in both types of pairs). However, we propose that the existence of *reference points* (i.e., exemplars that serve as a point of comparison) can explain the occurrence of the category boundary effect.

We asked participants to perform categorization, discrimination, and similarity judgments for different sets of morph figures (cf. Figure 1.3). Some of these sets morphed between clearly recognizable shapes (e.g., a car and a tortoise), other sets morphed between non-recognizable abstract shapes. Although we replicated the overall category boundary effect for both discrimination and similarity, we show that what mattered for actually predicting discrimination performance and similarity judgments was not the type of stimulus pair (i.e., within-category or between-category), but (a) the strength of the reference points and (b) the difference from the reference points for each of the individual stimuli in the pair.

1.2.2 Hysteresis and adaptation

The part on hysteresis and adaptation deals with how we can use the *immediate temporal context* — including both perceptual history and stimulus history — to clarify our perception of the incoming stimulation. More specifically, earlier research has found an attractive effect of the previous percept on the current percept (i.e., *hysteresis*: we more often perceive what we have perceived before) and a repulsive effect of the previous stimulus evidence on the current percept (i.e., *adaptation*: we less often perceive the organization for which there was most evidence

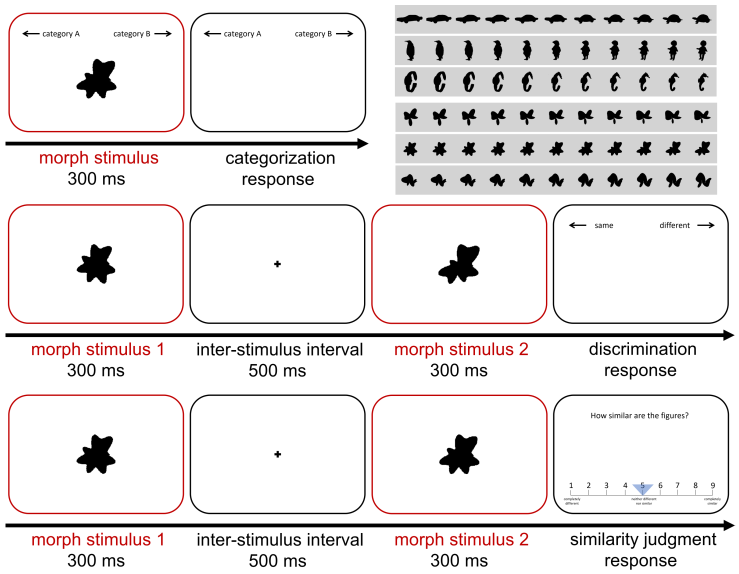


Figure 1.3: Schematic overview of categorization, discrimination, and similarity judgment tasks used in Chapter 3. The upper right shows the recognizable and non-recognizable morph series used in the experiments. *Note.* For reasons of visibility, the shown trial components in this figure have black shapes on a white background. The actual experiment had black shapes on a grey background.

in the previous stimulus; e.g., Gepshtein & Kubovy, 2005; Schwiedrzik et al., 2014). In these earlier studies, these effects were often studied at the group level, averaged across participants. In Chapter 4, we investigate potential individual differences in the size and direction of these immediate temporal context effects, in a multistable dot lattice paradigm (cf. Figure 1.4). Furthermore, we test the temporal stability of these individual differences. Although almost everyone showed an attractive effect of the previous percept and a repulsive effect of the previous stimulus evidence, not every single participant did. Furthermore, participants differed consistently in the extent to which they used previous input and experience in combination with current input to shape their perception, and these individual differences showed stable across one to two weeks' time. In Chapter 5, we develop an efficient Bayesian observer model and discuss how it can explain the co-occurrence of attractive and repulsive temporal context effects in this multistable dot lattice paradigm. An efficient Bayesian observer model differs from traditional Bayesian inference in that it assumes variable encoding precision of feature values in line with their frequency of occurrence, and as a consequence can predict biases away from the peak of the prior. Efficient encoding and likelihood repulsion on the stimulus level could explain the repulsive effect of the previous stimulus evidence, while perceptual prior attraction could explain the attractive effect of the previous percept.

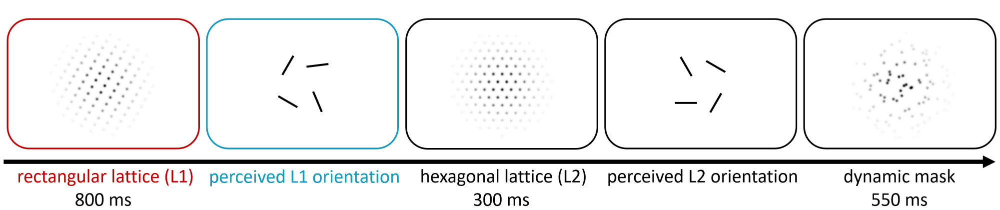


Figure 1.4: Schematic overview of the multistable dot lattice paradigm used in Chapter 4. *Note.* For reasons of visibility, the shown trial components in this figure have black dots on a white background. The actual experiment had white dots on a grey background.

1.2.3 Simplification and complication

The part on simplification and complication deals with how we can use the *immediate spatial context* to clarify our perception of the incoming stimulation. Gestalt theory posited two ways to make an organization more clear: either by removing unimportant details and making the experienced organization more similar to a reference (i.e., simplification), or by emphasizing characteristic features and making the organization stand out more from the reference (i.e., complication; Arnheim, 1986; Koffka, 1935). But when will a feature be simplified and when will it be complicated? In Chapter 6, we investigated whether the importance of a feature for discrimination among alternatives influences which organizational tendency will occur. Participants were presented with four figures composed of simple geometrical shapes simultaneously. One of these figures was indicated as the target figure, and participants were asked to reconstruct this target figure in such a way that another participant would be able to recognize it among the alternatives. Importance of a feature for discrimination was operationalized in three different ways (cf. Figure 1.5). Firstly, the four figures differed either qualitatively or only quantitatively (i.e., far or close context). Secondly, in close contexts, figures varied on two feature dimensions, and for one feature the range of variability across the alternatives was larger than for the other feature. Thirdly, in the case of a smaller variability range, the target figure was either at the extreme of the range or had an in-between value. The results indicated that each of these manipulations influenced the probability with which a feature was simplified or complicated. More specifically, complication occurred more often for the feature with an extreme value, for the feature exhibiting more variability, and for the features of figures presented in the close context, than for the feature with a non-extreme value, exhibiting less variability, or in the far context (i.e., more complication for the examples in the left columns than in the right columns of Figure 1.5). In other words, the immediate spatial context in which a percept is formed could influence which features were experienced as characteristic or rather unnecessary.

1.2.4 Order and complexity

The part on order and complexity deals with how the factors that contribute to increased clarity (i.e., Prägnanz) of our percepts can also contribute to aesthetic appreciation. As mentioned above under “What is a ‘good’ organization?”, both order and complexity can increase the clarity of an experienced organization. Earlier research has often found relations between order, complexity, and aesthetic appreciation (for an overview, see Van Geert & Wagemans, 2020), but the exact type and direction of this relation has remained unclear. Many studies have investigated the influences of order and complexity separately, have focused on specific types of order, ignored the multidimensionality of order and complexity, and/or did not manipulate order and complexity in a standardized manner. In Chapter 7, we present the Order & Complexity Toolbox for Aesthetics (OCTA), a Python toolbox that provides a free and easy way to create reproducible multi-element displays including both order and complexity manipulations. Furthermore, OCTA is also available as a point-and-click Shiny application (cf. Figure 1.6), which allows researchers without any programming experience to also use the toolbox and construct reproducible stimuli for their research.

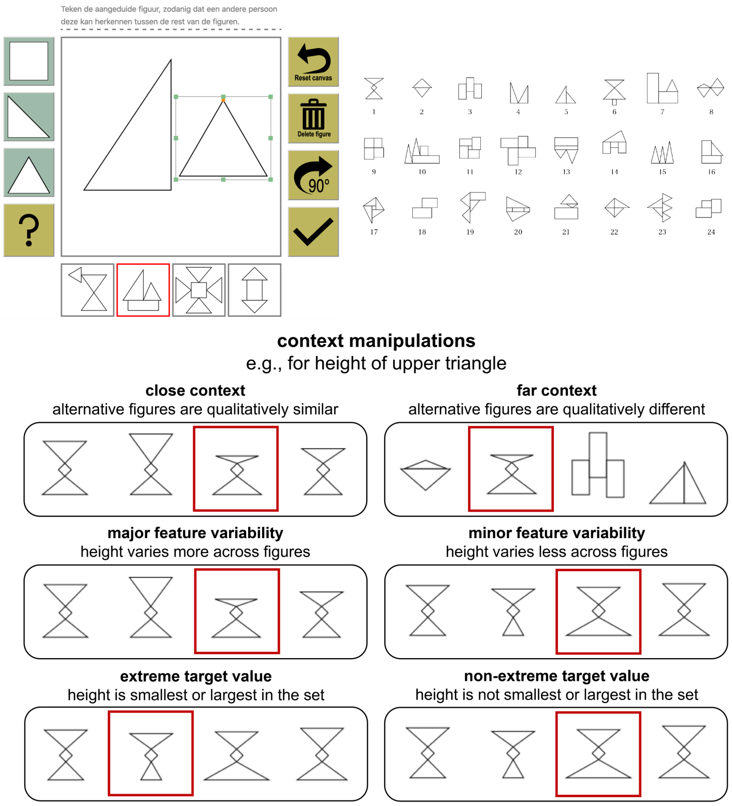


Figure 1.5: Overview of the drawing task, stimuli, and context manipulations used in Chapter 6.

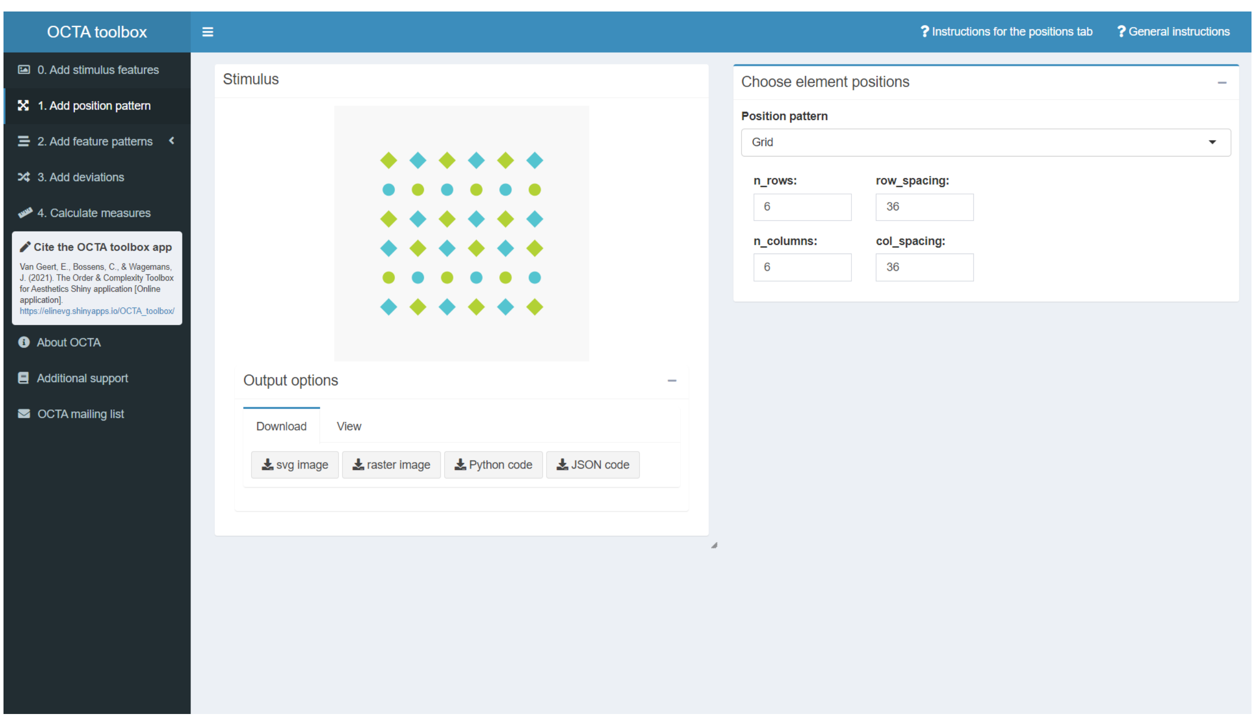


Figure 1.6: The OCTA Shiny application. Click the link to try out the application.

1.3 How does it all connect?

1.3.1 Robustness and sensitivity

The part on robustness and sensitivity focuses on how *internal representations of good Gestalts* (i.e., Prägnanz steps) are used *as reference points* to clarify the visual input. On the one hand, prägnant Gestalts can act as anchors, increasing sensitivity to change and discrimination performance around them. On the other hand, prägnant Gestalts can act as robust magnets, decreasing perceptual distance and worsening discrimination (Köhler, 1920; Stadler et al., 1979; Wertheimer, 1923). Relatedly, it is more difficult to transform prägnant figures into non-prägnant ones than the other way around (Goldmeier, 1937, 1982): also in that sense, prägnant Gestalts are more robust. As Wertheimer (1923) described it, shapes close to a prägnant step appear as [perceptually] different from but [categorically] related to the prägnant Gestalt: “as a somewhat ‘poorer’ version of it” (p. 318). In other words, although shapes close to the prägnant form can be *perceptually* discriminated from it, they are still *categorized* in relation to this prägnant form: the prägnant form serves as a reference point.

This double role of internal reference points in the Gestalt literature is congruent with the conclusion drawn by Quinn (2000), who discussed task context as a moderating factor for the role of perceptual reference points on discrimination sensitivity. On the one hand, increased sensitivity around reference points (above threshold) is present in tasks involving direct perceptual comparisons with plenty of perceptual evidence present and minimal memory demands. On the other hand, decreased sensitivity around reference points occurs for tasks in which a currently available stimulus was compared to stimuli stored in memory (i.e., for which there was limited perceptual input). Put differently, repulsive effects leading to increased sensitivity around reference points may be more related to ‘low-level’ perceptual processing, while attractive effects that decrease sensitivity and increase robustness may arise from ‘higher-level’ perceptual or cognitive processing (Quinn, 2000).

In Chapter 3, we propose that the existence of internal reference points (i.e., Prägnanz steps that serve as a point of comparison) and an attractive tendency towards these Prägnanz steps can explain the occurrence of the category boundary effect. Furthermore, we explain how the attractive tendency towards Prägnanz steps can also explain directional asymmetries in within-category pairs, i.e., worse discrimination when comparing a stimulus further away from the reference point to a stimulus closer to the reference point than the other way around.

In addition to the expected influence of Prägnanz steps on discrimination and similarity, we also expected an effect on categorization. As Prägnanz steps have been associated with more robust categorization — less prägnant Gestalts change their category membership more easily (Wertheimer, 1923) — we expected increasing categorization performance based the closer the stimulus was to an internal reference point.

The chapter thus focuses on investigating the relevance of internal reference points (i.e., Prägnanz steps) and the perceptual distance of the current stimulus from those internal reference points as determining factors for what the best possible perceptual and cognitive organization will be. Given that participants provide judgments that go further than only reporting their percept, both primary and secondary Prägnanz may be involved.

1.3.2 Hysteresis and adaptation

How do participants clarify the input when no clear internal reference points are present? In the section above, on possible Prägnanz tendencies, I discussed that sometimes, for example when no internal reference is available, a local reference may be used. This local reference is present in the immediate temporal or spatial context of the stimulus. In this part on hysteresis and adaptation, I focus on how *immediate temporal context* can be used as a (local) reference to clarify the visual input.

In the dot lattice study reported in Chapter 4, the current visual input was always ambiguous, and was also presented for only a short duration (i.e., weak stimulus conditions). The clarity of the *previous stimulus* varied across trials, however, and this stimulus was presented for a longer duration. Using an existing multistable dot lattice paradigm (Gepshtein & Kubovy, 2005; Schwiedrzik et al., 2014), we investigated individual differences in hysteresis and adaptation. The main questions of this study included whether (a) individual differences exist in the size of these temporal context effects, and if so, (b) whether everyone shows these effects in the expected direction, and (c) whether these individual differences are stable across one to two weeks' time. In addition to these main interests, we investigated much more, all of which will be discussed in Chapter 4. Of particular interest to this general introduction may be that we also investigated the relation between the strength of long-term absolute orientation biases (i.e., strength of Prägnanz steps for orientation) and that of the attractive and repulsive short-term temporal context effects. Also, a control task was included to investigate whether the attractive temporal context effect was based on the previous percept and/or on the previous response/decision. We conducted this study as a Registered Report, which means that we first specified our methods and analysis plan and got in-principle-acceptance from the journal before we actually collected and analyzed the data.

Chapter 5 describes a model that can predict co-occurring attractive and repulsive temporal context effects using efficient encoding and Bayesian decoding. Moreover, the model gives an idea about how the attractive effect of the previous percept and the repulsive effect of the previous stimulus evidence may arise. Efficient Bayesian observer models (Mao & Stocker, 2022; Ni & Stocker, 2023; Wei & Stocker, 2015; Wei & Stocker, 2017) assume variable encoding precision of feature values in line with their (long- and/or short-term) frequency of occurrence: those feature values that occur more frequently are more accurately represented. In addition, frequency of occurrence also influences prior expectations for a feature value to occur. For example, cardinal orientations are more common than oblique ones, and therefore humans expect more cardinal orientations and represent them more accurately compared to oblique ones.

In contrast to other Bayesian observer models, efficient ones (a) can predict perceptual biases away from the peak of the prior (Wei & Stocker, 2015) and (b) take the dissimilarity between stimulus space and sensory space into account, leading to differential predictions for the influence of internal sensory noise (i.e., reduced *visual* strength as in Huang, 2015; Huang, 2022; e.g., less visible orientation because of brief presentation duration) and external stimulus noise (i.e., reduced *featural* strength as in Huang, 2015; Huang, 2022; e.g., high variance in the orientations present in the stimulus).

Several aspects of efficient Bayesian observer models relate closely to the Prägnanz framework discussed above. Firstly, efficient encoding puts a limit on the organism's capacity to process the incoming stimulation. This assumption of a capacity limit can be categorized as part of the internal conditions discussed by Koffka (1935), and was referred to by Metzger (1941) as 'comprehension capacity' [*Fassungsvermögen*], influencing whether an organization would be experienced as unclear and chaotic or as rich and including multiple regularities. Furthermore, Koenderink et al. (2018) described this capacity limit as a 'structural complexity bottleneck'. When the structural complexity of a stimulus exceeds the level of structural complexity that the organism can process, the stimulus will no longer be recognized as a "picture", but rather as "featureless" or "just noise" (Koenderink et al., 2018). This capacity limit may be subject to contextual and individual differences, and thereby influence which Prägnanz tendencies will occur.

Secondly, the distinction between internal sensory noise and external stimulus noise may remind you of the distinction I made when relating Koffka's (1935) weak external conditions to low visibility but not uncertainty present in the features of the stimulus. Koffka's weak external conditions (i.e., low visibility) thus relate to the idea of high internal sensory noise in the efficient Bayesian observer model.

1.3.3 Simplification and complication

In this theme, I focus on how the Prägnanz tendency can play out by concurrent simplification, i.e., removing or weakening distracting, unnecessary details, and complication, i.e., intensifying characteristic features of the visual input (Arnheim, 1986). Which features are characteristic or rather unnecessary, however, and as a consequence, which features will more often be sharpened or leveled, respectively?

In Chapter 6, we research how the *immediate spatial context* can play a role in which features of the visual input are simplified or complicated to arrive at the best possible psychological organization given the conditions. We expected that in line with one of Metzger's (1941) definitions of prägnant Gestalts (i.e., good Gestalts as those structures that most purely and compellingly represent an essence), the *essence* of a Gestalt may be context-dependent, and this could influence whether simplification or complication of a feature leads to the best organization in the specific context. More specifically, we hypothesized that the importance of a feature for discrimination within a specific (task) context would influence which organizational tendency would occur.

In the previous themes, the studies focused on situations in which the current stimulus was shown very briefly. In other words, the 'external conditions' were rather weak (Koffka, 1935). In this study, external conditions were rather strong: the stimulus and the surrounding context were shown for an unlimited amount of time. Furthermore, participants were instructed to draw a target figure in such a way that the another participant would be able to recognize it among the alternatives. The study may thus have targeted a rather conscious deviation from the target figure, to enhance its communicative value — and thus also its Prägnanz — despite the strong stimulus factors. These — potentially explicit — simplification and complication tendencies may be seen as rather similar to some of the conscious Prägnanz tendencies in artistic practice. Especially cartoon artists directly apply these tendencies: they exaggerate characteristic features (i.e., complication) as well as cleanse the stimulus from other, distracting details (i.e., simplification). This also relates to Hüppe's (1984) description of the secondary Prägnanz tendency as a (potentially conscious) evaluation of the closeness of an experienced organization to a reference. As indicated, however, even in situations in which conscious deviations may be at play, primary Prägnanz tendencies can have an (unconscious) underlying influence too.

1.3.4 Order and complexity

The goodness of an experienced overall organization may increase with increasing order — think of the order-related, tension-reducing aspects listed by Rausch (1966), i.e., lawfulness, autonomy, integrity, and simplicity of structure — but can also increase with increasing intricacy or complexity — think of Rausch's (1966) Prägnanz aspects of element richness, expressiveness, meaningfulness. Order and complexity are not only important contributors to a better, more prägnant organization, they have also often been related to our experience of aesthetic appreciation (Van Geert & Wagemans, 2020). Furthermore, order- and complexity-stimulating Prägnanz *tendencies* are as present in artistic practice as they are in perceptual organization. Metzger (1941) for example indicated that true artists will go beyond their models in the direction of Prägnanz. To do so, they can both intensify characteristic features (i.e., complication) or weaken unimportant details (i.e., simplification). But how do Prägnanz and aesthetic appreciation relate precisely?

This part on order and complexity discusses two potential perspectives on the relation between good psychological organization and *aesthetic appreciation*. On the one hand, von Ehrenfels (1922) called beauty nothing else than Gestalt height, which he earlier defined as the product of unity (of the whole) and multiplicity (of the parts; von Ehrenfels, 1916). This view focuses on the importance of the *absolute Prägnanz level* of a perceived organization for its aesthetic appreciation.

On the other hand, aesthetic appreciation could also be connected to the *relative increase in Prägnanz* that the perceiver experiences (i.e., the strength of the experienced Prägnanz *tendency*). This second view makes experiencing at least some lawfulness or regularity a requirement for both Prägnanz and aesthetic appreciation: some level of order always needs to be presented for an organization to be 'good' and to be appreciated. For complexity, however, the relation with Prägnanz and appreciation would be slightly different. If it is mainly the relative *increase* in Prägnanz that matters for aesthetic appreciation, complexity can play a larger role in aesthetics than in perceptual organization. Complexity-related factors (i.e., element richness, expressiveness, or meaningfulness) can increase the Prägnanz level of a psychological organization if the complexity stays within the capacity limits of the organism and does not diminish perceived order. In aesthetic appreciation, however, more perceived complexity means more room for increases in Prägnanz. A perceiver's expectation to be able to handle the complexity (the individual's perceived capacity limit) will still put a cap on the maximum level of complexity that can be appreciated. However, given that in aesthetics vagueness has not the same negative life-related consequences as in perceptual organization, in appreciation it is less crucial that the final experienced organization is maximally ordered. If only the potential to increase Prägnanz matters (and less the absolute Prägnanz level), and the perceiving organism has a relatively high capacity limit, complexity may play a bigger role in aesthetic appreciation than in perception. As indicated above, this second view relates closely to the predictive processing accounts of, for example, Van de Cruys & Wagemans (2011) and Chetverikov & Kristjánsson (2016) as well as the focus on pleasure by insights into Gestalt proposed by Muth and Carbon (Muth et al., 2013; Muth & Carbon, 2013, 2016).

Both views — appreciation as Prägnanz *height* and appreciation as experienced strength of the Prägnanz *tendency* — may act in complementary ways as well, however. This combination of views is in line with the pleasure-interest model of aesthetic liking (PIA model) proposed by Graf and Landwehr (2015, 2017). The PIA model suggests two separate pathways towards aesthetic liking: one resulting from immediate automatic processing of the visual input (resulting in the experience of *pleasure*; related to processing fluency), and a second one resulting from perceiver-driven controlled processing (resulting in the experience of *interest*). Translated to the order, complexity, and Prägnanz terminology (cf. also Van Geert & Wagemans, 2020), we suggest that interest (often for rather complex, element rich stimuli) only occurs on the condition that the perceiver is able to psychologically organize this complexity and experience enough order. In contrast, pleasure may result immediately (for rather simple, meager stimuli), as less Prägnanz tendencies are necessary to organize the input.

In our view, a minimum level of unity or order may be a prerequisite for aesthetic appreciation as it is for Prägnanz, but aesthetic appreciation is expected to arise together with a conscious *increase* in Prägnanz, and therefore also complexity may play an important role. Given that capacity limits may depend on individual and context, strong individual differences could be expected when it comes to the appreciation of complexity. Furthermore, in case both Prägnanz height and the strength of the experienced Prägnanz tendency matter for aesthetic appreciation (in line with the PIA model), the importance of both relations may differ between individuals and contexts.

In an empirical study using images of neatly organized compositions (Van Geert & Wagemans, 2021), which was part of my master's thesis, I found that soothingness and fascination differ in their relations with complexity, but both relate positively to order and aesthetic appreciation. Furthermore, individual participants' correlations of perceived soothingness and fascination with perceived complexity were much more variable than the correlations with perceived order. In a follow-up on this study, we compared native Chinese-speaking and native Dutch-speaking participants, and found the positive relation between appreciation and order to be cross-culturally consistent, but the relation between appreciation and complexity showed to be cross-culturally diverse (Van Geert et al., 2021).

In these above mentioned studies I conducted on the relation between order, complexity, and appreciation, stimulus materials were collected from the internet, and consequently were not parametrically controlled. Other existing research mainly focused on the separate influence of order and complexity on aesthetic appreciation, focused on rather specific types of order (i.e., balance or symmetry), and/or ignored the multidimensionality of order and complexity. I believe that progress has also been hindered by the lack of an easy way to create reproducible and expandible stimulus sets, including both order and complexity manipulations on multiple stimulus dimensions (e.g., color, shape, size, orientation).

Chapter 7 presents the Order & Complexity Toolbox for Aesthetics, a Python toolbox and online point-and-click application we created to enhance the *reproducibility of stimuli used in research on the connection between order, complexity, and appreciation* in the visual modality. This chapter focuses on how Prägnanz, and more specifically order and complexity-related aspects of Prägnanz, can play a role beyond perceptual organization, in forming aesthetic evaluations of visual stimuli.

1.3.5 General discussion

To perceptually clarify the incoming stimulation, both *attractive* and *repulsive* tendencies can concurrently be at play. While the attractive tendencies will make the perceived organization more similar to the reference, the repulsive tendencies will increase the perceived difference between the currently experienced organization and the reference. The reference can be based on an *internal representation of a good Gestalt*, the *immediate temporal context*, and/or the *immediate spatial context* surrounding the incoming stimulation. These attractive tension-reducing and repulsive tension-increasing tendencies — robustness and sensitivity (cf. Chapter 3), hysteresis and adaptation (cf. Chapters 4 and Chapter 5), simplification and complication (cf. Chapter 6), and order and complexity (cf. Chapter 7) — are antagonistic to some extent, but also work together (i.e., complement each other) to arrive at the best possible psychological organization given the prevailing conditions (i.e., the Prägnanz principle), to optimize a *balance* between what we already know and the new input we receive. Importantly, the effects of Prägnanz are not limited to perceptual organization, but have consequences in many other psychological processes, including aesthetic appreciation. For a longer overview of the general conclusions that can be drawn from this dissertation, as well as an outlook to future directions, I refer to the General discussion (cf. Chapter 8).

Part II

What is Prägnanz?

How 'good' an organization is does not only depend on figural aspects, but also on how *meaningful* a specific organization is to us.

Chapter 2
Prägnanz in visual perception

How do we perceptually and cognitively organize incoming stimulation? A century ago, Gestalt psychologists posited the law of Prägnanz: psychological organization will always be as 'good' as possible given the prevailing conditions. To make the Prägnanz law a useful statement, it needs to be specified further (a) what a *'good'* psychological organization entails, (b) *how* the Prägnanz tendency can be realized, and (c) *which conditions* need to be taken into account. Although the Gestalt school did provide answers to these questions, modern-day mentions of *Prägnanz* or good Gestalt often lack these clarifications.

The concept of *Prägnanz* has been (mis)understood in many different ways, and by looking back on the rich history of the concept, we will attempt to present a more fine-grained view and promote a renewed understanding of the central role of Prägnanz in visual perception and beyond. We review Gestalt psychology's answers to the questions listed above, and also discuss the four main uses of the Prägnanz concept in more detail: (a) a Prägnanz tendency in each organizational process, (b) Prägnanz as a property of a Gestalt, (c) Prägnanz steps as internal reference points, and (d) Prägnanz in relation to aesthetic appreciation.

As a key takeaway, Prägnanz is a multifaceted Gestalt psychological concept indicating the "goodness" of an experienced organization. Both the removal of unnecessary details and the emphasis on characteristic features of the overall organization compared to a reference organization can contribute to the emergence of a 'better' Gestalt. The stimulus constellation is not the only factor in determining the goodness of an organization, also the stimulus' interaction with an individual in a specific spatial and temporal context plays a role.

Taking the ideas on Prägnanz as a generative framework and keeping the original Gestalt psychological context in mind, future research on perceptual organization can improve our understanding of the principles underlying psychological organization by further specifying how different organizational principles interact in concrete situations.

Van Geert, E., & Wagemans, J. (under review). Prägnanz in visual perception. Preprint available from
https://doi.org/10.31234/osf.io/pxdg4

Preprint Figshare

Prägnanz in visual perception

How do we perceptually and cognitively organize incoming stimuli? A century ago, Gestalt psychologists posited the law of Prägnanz: psychological organization will always be as 'good' as possible given the prevailing conditions. Although very commonly referred to in journal article introduction and discussion sections on perceptual organization, further clarification of (a) what a *'good'* psychological organization entails, (b) *how* the Prägnanz tendency can be realized, and (c) *which conditions* need to be taken into account, is often lacking in these modern-day references to Prägnanz. In addition, the Prägnanz concept has been (mis)understood in many different ways (cf. Figure 2.1). For example, in more recent years, *Prägnanz* has often been equated with element simplicity and the minimum principle: It was assumed that a visual *stimulus* is prägnant when it consists of *few elements*. The roots of the concept suggest a different, much richer and more complex interpretation, however. Rather than discussing a property of *stimuli*, it concerns a property of phenomenal experience, including the observer and the context as crucial interacting components. Moreover, not simplicity of the elements is central to the concept of *Prägnanz*, but simplicity of the whole, complete configuration as perceived by the observer, i.e., *simplicity of the Gestalt*. Part of the narrowing and misconceptions concerning Prägnanz may be due to many of the original sources – written in German – not being translated in English. Furthermore, if they were translated, this often happened only partially, or rather recently (e.g., Metzger, 1936/2006; Wertheimer et al., 2012; cf. also Wagemans, Elder, et al., 2012; Wagemans, 2015).

By looking back on the history of Prägnanz, we attempt to present a more fine-grained view and promote a renewed understanding of the central role of Prägnanz in visual perception and beyond. We review Gestalt psychology's answers to the questions listed above (cf. Figure 2.2), and discuss the four main uses of the Prägnanz concept in more detail: (a) a Prägnanz tendency in each organizational process, (b) Prägnanz as a property of a Gestalt, (c) Prägnanz steps as internal reference points, and (d) Prägnanz in relation to aesthetic appreciation (cf. Figure 2.3). Furthermore, we counter common critiques and reject alternative conceptualizations of Prägnanz by pointing back to the original intentions of Gestalt psychologists when positing the Prägnanz principle. Importantly, the Prägnanz principle was not meant as a magical one-fits-all solution, and it should be seen not only as an outcome of concrete research results but also as a device to stimulate further research (Wertheimer, 1924/1999): by using Gestalt theory and the Prägnanz principle as a generative framework for future research, and by studying more specific principles of organization and their interaction in concrete cases (Rausch, 1966), we can come to a better understanding of the principles underlying psychological organization.

2.1 Prägnanz in all its facets

At the end of the nineteenth and the beginning of the twentieth century, the widespread view on perception (and science in general) was elementaristic: researchers believed that a perceptual experience was reducible to its elementary sensations. This is, in fact, still the dominant view in most areas of experimental psychology and cognitive (neuro)science (which also demonstrates why a careful review and analysis of the alternative view is still relevant today). Wertheimer (1922, 1924/1999) and the Berlin Gestalt school questioned the immediate givenness

General remarks concerning Prägnanz

Prägnanz does not concern stimuli, but percepts

- Rather than discussing a property of stimuli, it concerns a property of phenomenal experience, including the observer and the context as crucial interacting components.
- Prägnanz can only apply to stimuli by approximation, i.e., by making abstraction of any internal conditions influencing Prägnanz height.

Prägnanz does not concern element simplicity, nor simplicity of the organizational process, but concerns simplicity of the resulting overall organization (i.e., Gestalt)

- Simplicity of a Gestalt can be distinguished from simplicity of the individual components: simple stimuli do not necessarily produce simple perceptual groupings (Wertheimer, 1923, 2012).
- Unity or regularity is not constrained to uniformity or homogeneity, but can also entail other types of structuring (Rausch, 1966).
- Both regularity and complexity (incl. element *richness*) can contribute to the Prägnanz of a Gestalt (i.e., they both make the overall organization simpler, more clear-cut, and more unambiguous; Rausch, 1966).

The Prägnanz tendency does not always lead to highly prägnant Gestalts in the absolute sense, only to the most prägnant Gestalt possible *given the prevailing external and internal conditions* (Arnheim, 1987; Koffka, 1935; Metzger, 1941)

- External forces are created within the receptor organs by the proximal stimuli and will constrain the Prägnanz tendency.
- Internal forces are related to the human nervous system and will draw the percept towards the most prägnant organization possible.

Prägnanz does not only concern human visual perceptual organization, but is meant as a very general concept

- The Prägnanz tendency is expected to occur in all sensory modalities as well as in other, more cognitive areas of psychological organization (Köhler, 1920; Metzger, 1936/2006), including memory, productive thinking, and problem solving (Kanizsa, 1975; Metzger, 1941; Sorge, 1940; Wertheimer, 1959; Wulf, 1922).
- The Prägnanz tendency is not limited to the perception of very simple proximal stimuli or the use of simple organizational principles like symmetry, but also occurs under more complex conditions (Arnheim, 1986; Köhler, 1920).
- The Prägnanz tendency is expected to occur in all different species, and differences between them are expected to be due to differences in the relative strength of different Gestalt principles rather than to differences in the nature of these principles (Metzger, 1941).

Prägnanz steps do not indicate discrete, stepwise but continuous, gradual changes

- A variable dimension as having a Prägnanz function across its domain: some regions have higher Prägnanz than others (Rausch, 1966), and it clearly concerns a gradual concept. Although we call them Prägnanz 'steps', these do not refer to a stepwise function (Rausch, 1952).

Prägnanz steps are not always the main factor determining organization

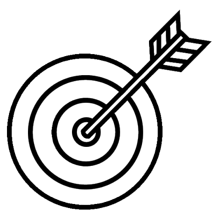
- Even when Prägnanz steps are present on the perceptual dimension in question, a concurrently present local reference may be used, e.g., when inspecting a clock we can use the small hand as comparison for the big hand (Rausch, 1966).
- If a stimulus is presented as part of an ordered series, the factor of objective set or setting [Einstellung] can come into play, and Prägnanz steps will no longer be the only factor determining the resulting organization (Wertheimer, 1923).
- Strong external forces (e.g., because of long presentation duration, high intensity, large size) will diminish the influence that internal forces (including the presence of Prägnanz steps) can have on the end Gestalt (Koffka, 1935).

Prägnanz is not an outcome, but a device

- The Prägnanz principle does not specify what will happen exactly in each and every specific situation (Arnheim, 1987; Rausch, 1952), but creates a general framework and stimulates further investigation concerning specific organizational principles falling under this law and how these different principles interact (Wertheimer, 1924).

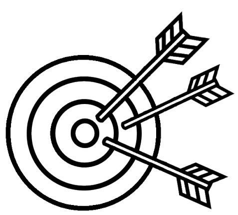
Figure 2.1: Important general remarks concerning the Prägnanz concept. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.21977504.

Main clarifications concerning Prägnanz



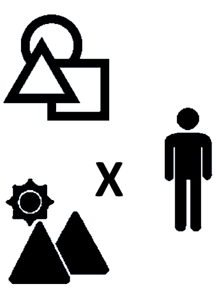
What does a *’good’* organization entail?

- an experienced organization is ’good’
  (1) when it contains **at least some form of unity or regularity**
- given all other things being equal, an organization increases in ’goodness’ when it is experienced
  (2) as **autonomous** rather than derived,
  (3) as **complete** rather than distorted,
  (4) as having a **simple structure** rather than a complicated structure,
  (5) as **element rich** rather than meager,
  (6) as **expressive** rather than expressionless, and
  (7) as **meaningful** rather than meaningless
- Prägnanz thus comprises both order- and complexity-related aspects, and is not only based on purely figural criteria, but also concerns how purely and compellingly the organization represents an *essence*



How can the Prägnanz tendency be realized?

- by **downplaying or removing unessential details** (i.e., tension-reducing tendencies, simplification, leveling, uniformity, minimum simplicity)
- by **intensifying or adding characteristic features** (i.e., tension-enhancing tendencies, complication, sharpening, perfect articulation, maximum simplicity)
- these tendencies are **present in every organizational process** and can be seen as **antagonistic but complementary**: both contribute to the goodness of the overall organization



Which conditions need to be taken into account?

- **external conditions**: created in the receptor organs by the proximal stimuli (i.e., the excitations to which the light rays coming from the physical object give rise)
- **internal conditions**: related to the human nervous system, including *more permanent conditions* (related to the structure of the nervous system, influenced by both inheritance and previous experience) and *more temporary conditions* (related to, e.g., vigilance, fatigue, needs, attitudes, interests, attentions)
- in case of weak external conditions (e.g., short exposure time, low intensity, small size), there is more room for the Prägnanz tendency to play a role

Figure 2.2: Main clarifications concerning Prägnanz. *Note.* The icons in this Figure were adapted from the following CC BY licensed icons from the Noun Project: target by Support Designs, head by Maxim Kulikov, shapes by Andrejs Kirma, and landscape by Baboon designs. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.21977522.

of independent elementary sensations and instead posited the primacy of the whole and direct influences of the whole on the perception of the elements: our experience of every individual element is, from the start, influenced by our organization of the whole. For example, in a melody, a particular tone will be experienced differently depending on which role this tone has in the melody (e.g., leading tone vs. tonic; Wertheimer, 1922, 1924/1999). Uncovering the principles guiding the spontaneous self-organization of the phenomenal field – without dissolving the perceptual experience by taking an item-per-item approach – is a key task for Gestalt psychology, and will lead to a better understanding of groupings and divisions in perceptual experience (Ellis, 1938; Wertheimer, 1922, 1923; Wertheimer et al., 2012).

Gestalt psychology thus takes it as its core task to reveal the principles that govern spontaneous self-organization in our phenomenal experience, including visual perception. Koffka (1935) designates the *Prägnanz principle* (i.e., the tendency towards the best possible overall organization given the prevailing conditions) as the most important principle to guide research on perceptual organization. This tendency towards Prägnanz of Gestalts was seen as the general principle overarching more specific principles of organization (e.g., grouping by proximity, similarity, good continuation). Under weak stimulus conditions, this tendency gets more room to play a role and can even lead to tangible dislocations and distortions compared to the external stimulation. Both the removal or softening of unnecessary details (i.e., simplification, leveling) and the addition or emphasis on characteristic features of the Gestalt organization (i.e., complication, sharpening) can take place and lead to a 'better' overall organization.

In addition to the Prägnanz tendency present in every organizational process, the term *Prägnanz* is also used *as a property* to characterize the organization or Gestalt *resulting* from this organizational process. To be 'prägnant', a psychological organization needs to be different from a simple sum of its elements: it has to be a Gestalt. In contrast to von Ehrenfels and the Austrian Gestalt psychological school, who defined Gestalt as a quality or characteric of an ensemble of items (Smith, 1988), Wertheimer (1922) and the other members of the Berlin Gestalt school defined a Gestalt as an ensemble of items that mutually support and determine one another (Sundqvist, 2003). Following the Berlin school, we could thus say that something is a Gestalt when our phenomenal experience is different from the experience of a pure sum of sensory elements — different from a pure and-summation (Koffka, 1935; Smith, 1988). This criterion for a phenomenal experience to contain at least some form of unity or regularity also comes back as the first and only necessary criterion in Rausch's (1966) specification of Prägnanz (cf. the section on "Rausch's (1966) Prägnanz aspects" below). In addition, Prägnanz can increase when the organization is perceived as autonomous rather than derived, complete rather than disrupted, simple of structure rather than complicated of structure, element rich rather than element poor, expressive rather than expressionless, and meaningful rather than meaningless. A psychological organization can thus be 'good' for several reasons, not only based on purely figural criteria. Moreover, Prägnanz is increased not only by increasing unity or regularity of the whole, overall organization, but also by increasing intricacy of its underlying components, of its relation between structure and meaning, and of its interaction with already existing knowledge structures in the organism.

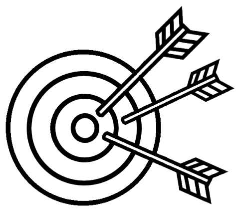
Psychological organizations that excel in their *Prägnanz* can influence our phenomenal experience of new incoming information: they serve as a reference to which the input is internally compared. The term *Prägnanz steps*

[*Prägnanzstufen*] is used in this context to refer to prägnant forms that serve *as reference regions* on a univariate dimension (e.g., the right angle in the realm of all possible angles). These Prägnanz steps serve a double function: on the one hand, assimilation to these Prägnanz steps may occur (especially when the external stimulus factors are weak); on the other hand, these reference points can increase sensitivity in their vicinity (to increase the ability to notice small deviations from the Prägnanz step). In that sense, Prägnanz steps support both robustness and sensitivity in visual experience: under weak stimulus conditions (i.e., under uncertainty) or when a specific difference is deemed unimportant, their stimulating effect to adhere to the best possible organization will dominate; under clear stimulus conditions and when a specific difference is significant, their influence on discrimination sensitivity in those regions where deviations matter most (i.e., close to the Prägnanz steps) will dominate. Their stimulating effect on discrimination sensitivity also allows for the formation of new reference levels in between existing ones when this becomes behaviorally or functionally useful. Rather than a binary distinction between reference points and non-reference points, a gradual Prägnanz function applies to each variable dimension, with some regions showing higher Prägnanz than others (e.g., the right angle as a higher Prägnanz step than the Prägnanz steps of sharp and obtuse angles, Rausch, 1966). The course of this Prägnanz function per domain and dimension can differ between individuals and contexts. For example, experience may elicit the formation of more and narrower Prägnanz steps on a dimension (Rausch, 1966; Wertheimer, 1923): individuals who deal with angles frequently (e.g., designers, architects) may have additional Prägnanz steps around 45° angles (besides 0° and 90° angles), and their Prägnanz steps may be narrower (e.g., only ranging from 89° to 91° around 90° angles instead of ranging from 87° to 93°).

When we psychologically organize incoming stimulation, we can *not only* describe or classify the experienced organization in a purely structural or semantic sense, but we can also evaluate our aesthetic experience of this organization. Since perceptual processing of the incoming information is necessary to be able to aesthetically evaluate our percept, the close relation between perception and *aesthetics* cannot be neglected. von Ehrenfels (1916, 1922) called beauty nothing else than Gestalt height, which he defined as the product of unity (of the whole) and multiplicity (of the parts). Unity-in-variety is also a major principle in design (e.g., Post et al., 2016). Moreover, Koffka (1940) called perception artistic and both Metzger (1941) and Arnheim (1975) noticed the presence of simplification and complication tendencies in artistic practice. In our view, aesthetic appreciation may arise together with a conscious increase in Prägnanz (i.e., the strength of the experienced Prägnanz tendency). This view also relates closely to other accounts of aesthetic appreciation, including the predictive processing accounts of Van de Cruys & Wagemans (2011) and Chetverikov & Kristjánsson (2016) as well as the focus on pleasure by insights into Gestalt proposed by Muth and Carbon (Muth et al., 2013; Muth & Carbon, 2013, 2016). On the other hand, aesthetic appreciation could also be based on the absolute level of Prägnanz experienced, and this does not necessarily relate to the strength of the Prägnanz *tendency*. Nevertheless, both views may act in complementary ways as well.

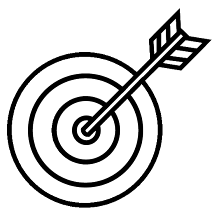
In what follows, we will discuss the four main uses of the Prägnanz concept mentioned above (cf. Figure 2.3) in more detail: (a) a Prägnanz tendency in each organizational process, (b) Prägnanz as a property of a Gestalt, (c) Prägnanz steps as internal reference points, and (d) Prägnanz in relation to aesthetic appreciation. Although these

Four main uses of Prägnanz



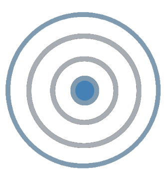
Prägnanz as a principle, tendency, process

- tendency towards the best possible overall experienced organization (i.e., Gestalt) given the prevailing external and internal conditions
- general principle overarching more specific principles of organization
- manifests itself in downplaying of unnecessary details (i.e., simplification, attraction) and emphasizing of characteristic features (i.e., complication, repulsion)
- how far tendency can go and how tendency manifests itself depends on the prevailing conditions (e.g., external stimulus domain, stimulus strength, energy level of the perceiver, temporal and spatial context)



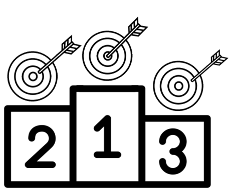
Prägnanz as a property, characteristic

- goodness level of the organization resulting from the Prägnanz tendency
- present when the experienced organization is different from a simple sum of its sensory elements, i.e., when the experienced organization contains at least some form of unity or regularity
- higher when the organization is experienced as autonomous, complete, simple of structure, element rich, expressive, and / or meaningful
- comprises both order- and complexity-related aspects, and is not only based on purely figural criteria
- also referred to as good Gestalt, Gestalt strength, Gestalt height, Prägnanz height, or Prägnanz dimensionality



Prägnanz steps, ranges as an internal reference points

- those regions on a (univariate) perceptual dimension that contain highly prägnant organizations, and thereby can serve as a reference for comparison
- serve a double function: can attract percepts to level unimportant details, but can also repel percepts to emphasize significant differences (and to allow for the formation of new Prägnanz steps when functionally relevant)
- by doing so, support both robustness and sensitivity in visual experience
- rather a gradually varying Prägnanz function across the dimension than sudden, discrete steps; function dependent on stimulus, person, and context



Prägnanz in relation to appreciation and artistic practice

- comparing several experienced organizations, or comparing one experienced organization to an internal reference (i.e., a multidimensional Prägnanz function)
- opinion 1: appreciation is related to an organization's Prägnanz *height*
- opinion 2: appreciation is related to the strength of the Prägnanz *tendency*
- artists exemplify multiple ways of balancing simplification and complication
- artists create stimuli in order for them to be prägnant or to allow for strong Prägnanz tendencies

Figure 2.3: Four main uses of Prägnanz. *Note.* The icons in this Figure were adapted from the following CC BY licensed icons from the Noun Project: target by Support Designs and podium by Prettycons. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.21977540.

uses overlap in many ways, we distinguish them here as they all focus on a different facet of the overall concept. A full understanding of Prägnanz encompasses all of these facets and their interrelations in a compelling whole, however.

2.1.1 A Prägnanz tendency present in each organizing process

When we view Prägnanz as a tendency present in each organizing process, we speak of the **Prägnanz law**, the **Prägnanz principle**, or the **Prägnanz tendency** (cf. Figure 7.6). Etymologically, Prägnanz derives from the German verb 'prägen' (i.e., to mint a coin) and the Latin verb 'premere' (i.e., to press a point), and therefore Prägnanz refers to being sharply grasping, unambiguous, clear, or distinct (Arnheim, 1975; Wagemans, 2018)[1]. The Prägnanz tendency thus concerns a tendency, present in all forms of psychological organization, to evolve in the direction of a more clear-cut overall organization (i.e., a better Gestalt). This is equivalent to evolving in the direction of minimal structural energy to arrive at a stable organization. The first written mention of Prägnanz as a tendency comes from Wertheimer (Schumann, 1914), who described his success in ascertaining, under several Gestalt laws of a general nature, *"ein Gesetz der Tendenz zum Zustandekommen einfacher Gestaltung (Gesetz 'zur Prägnanz der Gestalt')"* (i.e., a law of the tendency to come towards a simple Gestalt or a law towards Prägnanz of a Gestalt, Schumann, 1914, p. 149). Importantly, simplicity of the Gestalt can be distinguished from simplicity of the individual components: simple stimuli do not necessarily produce simple perceptual groupings (Wertheimer, 1923; Wertheimer et al., 2012).

Wertheimer (1923) proposed the law of good Gestalt as an overarching law of which the other Gestalt laws are special cases: When forming a percept and a grouping, the "best" overall organization, the simplest "whole" will win, and each specific law (e.g., proximity, similarity, good continuation; cf. also Wagemans, Elder, et al., 2012; Wagemans, 2018; Wertheimer, 1923) gives an indication of what the "best" grouping will be. For example, the law of similarity indicates that there is a tendency for uniformly colored parts to group together. Following the law of similarity, the "best", most simple overall organization will thus be the one with uniformly colored components (e.g., red parts grouping together, blue parts grouping together, Wertheimer, 1923), but this may oppose the best organization according to another Gestalt principle (e.g., the law of proximity).

What happens if several Gestalt principles are concurrently at play (e.g., Figure 2.5)? If several Gestalt principles are working in the same direction, this will lead to stronger inner cohesion and sharper segmentation (Metzger, 1941), in other words, to a better, more prägnant Gestalt. Conflicting Gestalt factors will yield one of five possible states: (1) one of the principles is stronger and wins; (2) the result is ambiguous and there is switching between two possible organizations; (3) the result is unclear, chaotic; (4) a richer organization forms in which both Gestalt factors play a role; or (5) one of the Gestalt factors wins but the end Gestalt is slightly changed based on the other Gestalt factor (Metzger, 1941). Inter- and intra-individual differences will play an important role in determining whether (3) or (4) will be the case (e.g., comprehension capacity [*Fassungsvermögen*], expertise, energy level, mood, Metzger, 1941). Importantly, the law of Prägnanz does not specify what will happen exactly in each and every specific situation (Arnheim, 1987; Rausch, 1952), but it creates a general framework and stimulates further investigation

[1] In contrast to what many assume, Prägnanz is etymologically not related to the English word pregnancy (derived from the Latin word 'praegnans,' meaning rich in potential content, Arnheim, 1975).

2

Glossary

Prägnanz good Gestalt	indicates the goodness level of an experienced psychological organization. To be prägnant, the organization needs to contain at least some form of unity or regularity (i.e., needs to be different from a pure sum of its sensory elements). Goodness increases when the organization is experienced as autonomous, complete, simple of structure, element rich, expressive, and/or meaningful.
Gestalt height Gestalt strength	can be used as synonyms for Prägnanz or good Gestalt. They refer to the extent to which the parts of the experienced organization are mutually dependent, form a whole. Strong or high Gestalten are those Gestalten in which the mutual dependence among the parts is so great that there is no room for any displacement or changes without an influence on all other parts of the whole (Ash, 1995; Köhler, 1920). For example, a rose is a higher Gestalt than a heap of sand (von Ehrenfels, 1916, 1932/1937).
Prägnanz law Prägnanz tendency Prägnanz principle	concerns a tendency towards the best possible experienced organization (i.e., Gestalt) given the prevailing external and internal conditions. This tendency can manifest itself through simplification as well as complication.
simplification leveling minimum simplicity	indicates tension-reducing tendencies that downplay or remove unessential details when psychologically organizing the incoming stimulus in the given context and organism.
complication sharpening maximum simplicity	indicate tension-enhancing tendencies that emphasize or add characteristic features when psychologically organizing the incoming stimulus in the given context and organism.
external conditions	are created within the receptor organs by the proximal stimuli (i.e., the excitations to which the light rays coming from the physical object give rise). They constrain the Prägnanz tendency.
internal conditions	are related to the human nervous system, and include more permanent conditions (related to the structure of the nervous system, influenced by both inheritance and previous experience) as well as more temporary conditions (related to, e.g., vigilance, fatigue, needs, attitudes, interests, attentions). The stronger they are, the stronger the Prägnanz tendency.
primary Prägnanz	concerns the relation between stimulus and perceived organization and leads to deviations from the stimulus to the percept that are not directly noticeable by the observer. Put differently, every perceptual organization already deviates from the stimulus in the direction of Prägnanz.
secondary Prägnanz	concerns the tendency to evaluate a phenomenon based on its experienced closeness to a prägnant form (i.e., an internal point of comparison, not necessarily phenomenally present)
Prägnanz steps Prägnanz function	indicate(s) those regions on a (univariate) perceptual dimension that contain highly prägnant organizations, and thereby can serve as a reference for comparison. Rather than a binary distinction between reference points and non-reference points, a gradual Prägnanz function applies to each variable dimension, with some regions showing higher Prägnanz than others.
Prägnanz height Prägnanz dimensionality	can be used as a synonym for Prägnanz or good Gestalt. Rausch (1952) used these terms to refer to the Prägnanz function when multiple variables are concurrently at play.

Figure 2.4: Glossary of Prägnanz-related terminology. Retrieved from doi.org/10.6084/m9.figshare.21977546.

concerning specific organizational principles falling under this law and how these different principles interact (Wertheimer, 1924/1999).

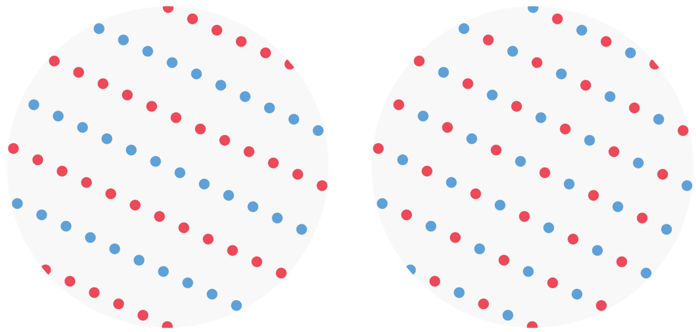


Figure 2.5: Example of a dot lattice in which proximity grouping and color similarity grouping are either congruent or incongruent. These lattices were created using the OCTA toolbox (Van Geert, Bossens, et al., 2022). Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.21977558.

Köhler (1920) focused on the equivalence between Wertheimer's tendency towards Prägnanz of a Gestalt and the physical tendency towards minimal structural energy (attained when in a stable, stationary state). In processes ending in a stable state, there is a tendency in the direction of minimal structural energy, i.e., a tendency to achieve minimal structural energy in the end state or resulting organization given what is possible under the prevailing conditions (Köhler, 1920). As only the final structure or organization – and not the corresponding energy level – is available in phenomenal experience, we cannot determine whether a perceived organization corresponds to the minimal energy level based on phenomenology alone. One could only infer whether the Prägnanz principle of minimal structural energy is realized in the nervous system for a specific perceived organization if a simple relationship existed between the energy level and the perceived organization (Köhler, 1920; cf. also Pepperell, 2018).

The best-known classical description of the law of *Prägnanz* is probably the one by Koffka (1935) in his English book on the *"Principles of Gestalt Psychology"*: "psychological organization will always be as 'good' as the prevailing conditions allow" (p. 110). In this definition, the term 'good' is undefined, but entails properties such as regularity, symmetry, simplicity, and others (Koffka, 1935, p. 110). Koffka (1935) presents the law in the context of his discussion on finding the true solution to the question "Why do things look as they do?". False solutions mentioned include "because things are what they are" (veridicality; cf. Box 1) and "because the proximal stimuli [i.e., the excitations to which the light rays coming from the physical object give rise] are what they are". Things look as they do "because of the field organization to which the proximal stimulus distribution gives rise. [...] It means that we have to study the laws of organization" (Koffka, 1935, p. 98). The law of *Prägnanz* is mentioned as the main principle to guide research on psychophysical[2] (i.e., perceptual) organization (Koffka, 1935, p. 110).

In sum, the tendency towards Prägnanz of a Gestalt indicates a tendency present in every process of psychological organization to come to the organization that — when taking into account the given conditions — has minimal

[2]Koffka referred to this type of organization as 'psycho-physical' as the formation of these organizations is influenced by both internal (i.e., psychological) and external (i.e., physical) factors. Koffka's notion should not be confused with Fechner's notion of psychophysics, which concerns the quantitative mapping of physical stimuli onto psychological entities (e.g., between stimulus intensity and sensation strength).

structural energy and is the most clear-cut and simple. Although it concerns a maximal tendency in the direction of high Prägnanz, it does not mean that every organizational process will lead to a simple, clear-cut Gestalt in the absolute sense (Arnheim, 1987; Koffka, 1935; Metzger, 1941). The tendency towards the most prägnant Gestalt should thus always be seen as relative to the prevailing conditions. In addition to a further specification of what a 'good', 'clear-cut', or 'simple' organization entails (cf. Prägnanz as a property), this formulation of the Prägnanz tendency requires clarification of two additional elements: (a) *how* the Prägnanz tendency can be realized; and (b) *which* prevailing *conditions* need to be taken into account.

2.1.1.1 *How* can the Prägnanz tendency be realized?

First, *how* can the tendency towards prägnant Gestalts be realized? In psychological organization either *as much* or *as little* will happen as the prevailing conditions permit (Koffka, 1935). Whereas **minimum simplicity** indicates the simplicity of uniformity, **maximum simplicity** indicates the simplicity of perfect articulation (Koffka, 1935, cf. Figure 2.6). Simplification and complication do not have to be perceived as necessary, mutually-exclusive alternatives, however (Köhler, 1951/1993). To achieve Prägnanz of a perceived whole, some components may need to develop in different directions (Köhler, 1951/1993). Arnheim (1986) viewed simplification and complication as antagonistic but complementary tendencies present in every perceptual event. Whereas tension-reducing tendencies (i.e., simplification, leveling, minimum simplicity) remove unessential details, tension-enhancing tendencies (i.e., complication, sharpening, accentuation, pointing, articulation, maximum simplicity) intensify characteristic features of a Gestalt structure (Arnheim, 1986; Metzger, 1941)[3]. In this way, both simplification and complication can contribute to the Prägnanz of a Gestalt (i.e., they both make the overall organization simpler, more clear-cut, and more unambiguous). Recent work (Prasad & Bainbridge, 2022) concerning the visual Mandela effect in memory is consistent with this idea: some images from popular iconography elicit consistent, specific false memories in the direction of a better Gestalt, regardless of whether it concerns downplaying specific details (i.e., simplification; e.g., a golden instead of a silver leg for C-3PO from the Star Wars franchise) or intensifying characteristic features (i.e., complication; e.g., a black-tipped instead of an almost completely yellow tail for Pikachu from the Pokémon franchise).

Which features will be treated as unessential and which as characteristic? The features that may be noticed refer to differences from a reference used (i.e., a local reference, or an internal reference, cf. 'Prägnanz steps'). Which of these features will be treated as characteristic will depend on the individual and the context in which the organization is perceived (e.g., Van Geert, Frérart, et al., 2022).

Importantly, whereas one could equate simplification with literally 'removing' features and complication with 'adding' features, this does not have to be the case. For example, a square missing one of its four sides may be

[3] Note that the terms we mention as approximately equivalent here have previously been used to indicate slightly different tendencies. For example, complication and simplification were defined as adding and removing parts, respectively (e.g., Fehrer, 1935; Hubbell, 1940). Sharpening, accentuation, and pointing were defined as emphasizing or exaggerating particularities of a figure, making the characteristics of a figure more differentiated (Fehrer, 1935; Hubbell, 1940; Wulf, 1922). The term leveling referred to downplaying particularities of a figure, making the characteristics of a figure less differentiated (Fehrer, 1935; Hubbell, 1940; Wulf, 1922). In addition, Wulf (1922) defined normalizing as increasingly the resemblance to a familiar structure.

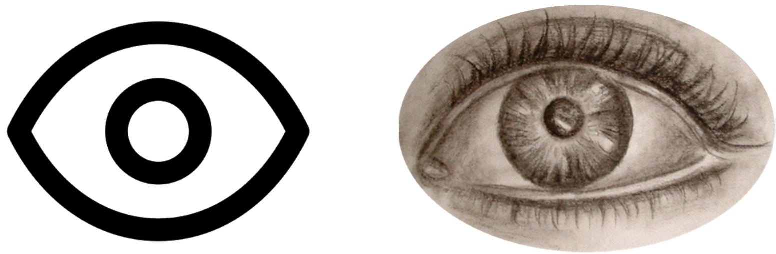


Figure 2.6: Example of minimum simplicity (i.e., the simplicity of uniformity) and maximum simplicity (i.e., the simplicity of perfect articulation. *Note.* In this Figure, the CC BY licensed eye icon by Shiva from the Noun Project and picture of an eye drawing by Lucky Lynda were used. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.21977564.

simplified by adding a sideline, resulting in a complete square. On the other hand, also removing a part of an organization can complicate an organization, e.g., removing a sideline from a full square.

Inspired by Kanizsa (1979)'s distinction between primary perceptual processes (i.e., autochthonous forces leading to the organization of the perceptual field) and secondary perceptual processes (e.g., identification, classification), Hüppe (1984) distinguished primary and secondary Prägnanz (cf. Figure 2.7). The **primary Prägnanz** tendency considers the relation between stimulus and phenomenon and leads to deviations from the stimulus to the percept that are not directly noticeable by the observer. Put differently, every phenomenal experience already deviates from the stimulus in the direction of Prägnanz. For example, a parallellogram may be perceived as more rectangular than it actually is (cf. left side of Figure 2.7). The **secondary Prägnanz** tendency operates on the phenomenal level and concerns the tendency to evaluate a phenomenon based on its experienced closeness to a prägnant form (i.e., an internal point of comparison, not necessarily phenomenally present). For example, a perceiver may cognitively evaluate the parallellogram as 'almost a rectangle' (cf. right side of Figure 2.7). Importantly, both Prägnanz tendencies cannot be seen as completely independent: To be able to make statements about secondary Prägnanz, an organized perceptual field (influenced by primary Prägnanz) is preassumed.[4]

2.1.1.2 *Which conditions* influence the course of the Prägnanz tendency?

Second, what do the prevailing conditions entail? As psychological organization takes place in an organism, it is constrained by the conditions outlined by the organism. When it concerns psychophysical processes like human perceptual organization, there are both external and internal conditions to consider (Koffka, 1935). **External conditions** are created within the receptor organs by the proximal stimuli. The *proximal* stimuli, i.e., the excitations to which the light rays coming from the physical object give rise, are in their turn influenced by the *distal* stimuli (i.e., the physical objects), the nature of the light source, the position of the viewer in relation to the distal stimuli and the light source, etc. (Koffka, 1935). **Internal conditions** are related to the structure and state of the human nervous system. Within the internal conditions, more permanent ones (related to the structure of the nervous system,

[4]Although we do see the value of this distinction between Prägnanz tendencies, we do not think of these Prägnanz tendencies as *successive* processes in the way Kanizsa (1979) described primary and secondary processes. We believe that according to traditional Gestalt theory, there is not first an organization of the perceptual field and only then more high-level cognitive identification or classification, but rather one complex dynamic process of Gestalt formation (see also Kruse, 1986).

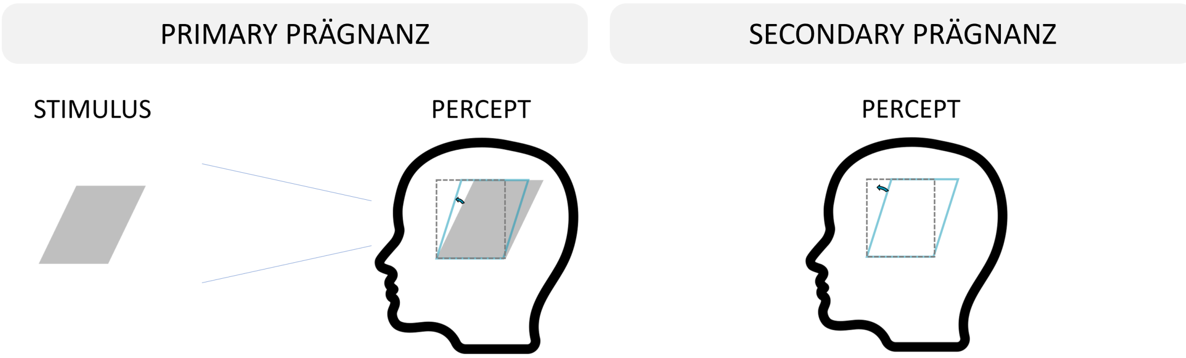


Figure 2.7: Visual representation of primary and secondary Prägnanz tendencies. The parallellogram represents the incoming stimulus. The shape of the figure is not always perceived veridically. The perceiving individual has a reference distribution of the most prägnant geometric shapes (i.e., Prägnanz steps). In the case of this parallellogram and individual, the rectangle is the reference figure. When perceptually organizing the incoming figure, the perceived figure already deviates from the stimulus (i.e., primary Prägnanz tendency). In this case, the parallellogram is perceived as more rectangular than it actually is (i.e., primary simplification), but under different conditions the parallellogram may be perceived as less rectangular than it actually is (i.e., primary complication). For the perceiver, there is no direct way to be aware of this first deviation. Secondly, the perceiver can also consciously evaluate a perceived organization in relation to a Prägnanz step. In this example, the individual evaluates the shape as 'almost a rectangle' (i.e., secondary simplification). Under different conditions, the individual may evaluate the shape as more different from a rectangle than it actually is (i.e., secondary complication). Closeness of the current percept to the internal reference point is directly observable by the perceiver. *Note.* In this Figure, the CC BY licensed head icon by Maxim Kulikov from the Noun Project was used. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.21977597.

influenced by both inheritance and previous experience[5] can be distinguished from more temporary ones (related to, e.g., vigilance, fatigue, needs, attitudes, interests, attentions, Koffka, 1935). One internal condition influencing the tendency towards minimum versus maximum simplicity that Koffka (1935) discussed was the level of vigilance or the energy level of the organism: low activity levels would lead to uniformity and minimum simplicity (i.e., simplification), whereas high activity levels would lead to good articulation (i.e., maximum simplicity, complication, Koffka, 1935).

The internal and external conditions will serve as two separate organizing forces in perceptual organization. While the internal forces of organization will draw the percept towards the most prägnant organization possible, the external forces will constrain the Prägnanz tendency (Koffka, 1935). When both internal and external forces act in the same direction, very stable organizations should result. In contrast, conflicting internal and external forces will yield a less stable organization (Koffka, 1935).

Although internal forces of organization are present even under conditions of strong external forces (Koffka, 1935), weak external forces (e.g., because of short exposure time, low intensity, small size) will give more room to the internal forces to alter the end Gestalt, producing considerable dislocations which lead to a more stable end state (e.g., Figure 2.8). These internal forces can even lead to the addition of new lines if that leads to a better end result (Koffka, 1935). Take note that Koffka also related the laws of organization to the simplicity of the resulting Gestalt, not the simplicity of the process: "the process of organization depends upon the properties of its result" (Koffka, 1935, p. 151).

Although these internal and external conditions influence which Prägnanz tendencies will occur, this conditional dependence does not imply randomness or arbitrariness: it is not the case that 'anything is possible' (Wertheimer,

[5]This also includes, for example, the effects of expertise (Koffka, 1935; Metzger, 1941, 1954, 1936/2006).

1923; Wertheimer et al., 2012). What it does imply is that stimulus, individual, and context need to be taken into the equation to determine which Prägnanz tendencies will occur under which concrete conditions. Specifically because of these dependencies, the Gestalt psychologists described their Gestalt principles as *ceteris paribus principles*: a Gestalt principle is supposed to hold only within the constraints of the prevailing (internal and external) conditions (Wagemans, 2018).

2.1.1.3 How general is the Prägnanz tendency?

Although this paper focuses on Prägnanz in visual perception, the tendency towards prägnant Gestalts is not at all limited to visual perceptual organization, but was proposed as a *general* tendency present in all forms of psychological organization (Koffka, 1935; e.g., Köhler, 1920; Metzger, 1936/2006). This generality can be interpreted in three different ways. Firstly, the tendency towards prägnant Gestalts is expected to occur and/or has been studied not only in visual perception, but *in all sensory modalities*, as well as *in other, more cognitive areas* of psychological organization (Köhler, 1920; Metzger, 1936/2006), including memory, productive thinking, and problem solving (Kanizsa, 1975; Metzger, 1941; Sorge, 1940; Wertheimer, 1959; Wulf, 1922). Secondly, the Prägnanz tendency is not limited to the perception of very simple proximal stimuli or the use of simple organizational principles like symmetry, but *also occurs under more complex conditions* (Arnheim, 1986; Köhler, 1920). Thirdly, the Prägnanz tendency is expected to occur *in all different species*, and differences between the species are expected to be due to differences in the relative strength of different Gestalt principles rather than to differences in the nature of these principles (Metzger, 1941).

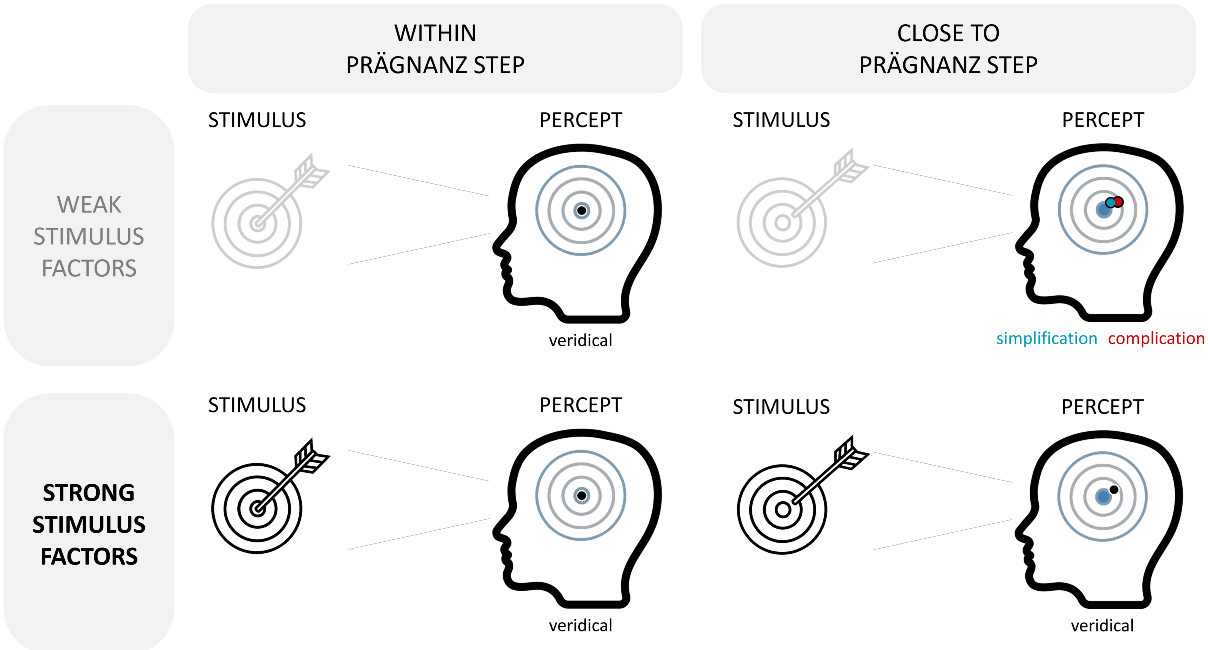


Figure 2.8: Visual representation of internal and external conditions influencing the Prägnanz tendency. The dartboard represents the incoming stimulus. The location of the arrow hitting the board is not always perceived veridically. The perceiving individual has a reference distribution of the most prägnant locations (i.e., Prägnanz steps). In the case of the dartboard and this individual, the bullseye is the most prägnant region. The circular lines indicated on the board also have some Prägnanz, and the outer circle has a bit more Prägnanz, but less than the bullseye region. When perceiving the location of the arrow on the board, the perceiving individual will take the visual input into account, but also his/her own reference distribution. The tendency towards a better overall organization (i.e., the Prägnanz tendency) will have more room to influence the percept when the external stimulus factors are weak (e.g., because of diminished contrast, short presentation duration, small size). In case the stimulus value (i.e., in this case the location of the arrow on the board) falls within a highly prägnant region, no strong tendency will occur. When the value of the stimulus is close to a prägnant region but falls outside, the Prägnanz tendency will be the largest. Under some circumstances, simplification may occur, pulling the stimulus value closer to the prägnant region (i.e., the bullseye in this case). On the other hand, complication may occur, more clearly differentiating the experienced value from the prägnant region. Which of those two tendencies will occur will depend on the stimulus, person, and context. *Note.* This figure is clearly a simplification of the situation and shows four extreme cases. In any real-life situation, (a) many other stimulus-, person-, and context-related factors are at play (e.g., luminance sensitivity, energy level, stimulus history), (b) stimuli and percepts are multidimensional, which could lead to simplification on one perceptual dimension and complication on another, and (c) the transition between veridicality, simplification, and complication is gradual rather than strictly defined, percepts can be more or less simplified or complicated compared to the actual stimulus value. Furthermore, stimuli that fall exactly within our Prägnanz steps on all important dimensions are very rare compared to stimuli that fall outside of our Prägnanz steps for at least one dimension. The icons in this Figure were adapted from the following CC BY licensed icons from the Noun Project: target by Support Designs and head by Maxim Kulikov. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.21977612.

Box 1: Veridicality and Prägnanz

What is perception aimed at? Or as (Koffka, 1935, p. 75) famously phrased it, "Why do things look as they do"? Some researchers believe that perception is aimed at achieving a sufficiently veridical representation of the physical world — and this is actually still the mainstream view in current vision science (as noted by Koenderink, 2014, 2015). For example, researchers who approach perception as Bayesian inference often assume a direct link between the physical and the phenomenal world. This belief in veridicality also comes back in the idea that "the senses cannot lie", and in the separation of illusory vs. non-illusory perception (Koffka, 1935; Rogers, 2014). However, if perception is always co-determined by the organizational processes present in the organism, what differentiates 'illusions' from other percepts (Rogers, 2014)?

The Gestalt theorists did not view veridicality as a direct aim of perception. Instead, the seeming correspondence between experienced and physical world results from the presence of similar natural, autochthonous principles guiding spontaneous self-organization in both the experienced and the physical world (Bischof, 1966; Hüppe, 1984). Because the course of this spontaneous self-organization depends on the properties of the organism in which the organizational process takes place, it is impossible to predict phenomenal experience from the stimulus conditions alone (Hüppe, 1984; Koffka, 1935). Hence, knowledge about the physical world alone is not enough to accurately predict perception. No physical unit is necessary nor sufficient for the appearance of a perceptual unit (Koffka, 1935). Also in current cognitive science, the main distinction made is between 'stimulus' and 'response', whereas the necessary step of organizing the stimulus into a 'percept' is often overlooked (e.g., in the literature on serial dependence, Bosch et al., 2020; Pascucci et al., 2023, 2019; Sadil et al., 2021; Sheehan & Serences, 2023; cf. also Van Geert, Moors, et al., 2022).

Disagreements about whether perception is aimed at veridicality have also played an important role in the 'simplicity' vs. 'likelihood' debate (e.g., Chater, 1996; Leeuwenberg & Boselie, 1988; Pomerantz & Kubovy, 1986; van der Helm, 2000). In this debate, the 'simplicity' principle — which states that we will see the simplest possible interpretation of the sensory pattern — is contrasted with the 'likelihood' principle — which states that we will perceive the most likely interpretation that fits the sensory pattern. The simplicity or minimum principle has traditionally been related to the law of Prägnanz (Wagemans, Feldman, et al., 2012), but — as will become clear from this review — the original conception of the Prägnanz principle was much broader than only a focus on minimum simplicity (cf. also Luccio, 2019).

Similarly to Gestalt psychology, also some more recent authors have explicitly opposed the idea of veridical perception. For example, Hoffman's (Hoffman, 2009; Hoffman et al., 2015) interface theory of perception stresses the importance of evolution and natural selection in shaping perception. Perception is in that sense aimed at fitness, not truth. Similarly, Koenderink (2014, 2015, 2019) describes perception as an idiosyncratic user interface, which is — through evolutionary tendencies — optimized for utility, not veridicality.

2.1.2 Prägnanz as a phenomenal property of the resulting organization

In the former part we referred to Prägnanz as a tendency present in every process of psychological organization. However, Prägnanz can also be used to refer to a property of the Gestalt *resulting* from this organizational process. Importantly, **Prägnanz** or **good Gestalt** as a property should be seen as inherently related to the foregoing descriptions of Gestalt and the tendency towards prägnant, simple Gestalts. Nevertheless, it is not the case that every organizational process results in a 'good' Gestalt in the absolute sense: it will only result in the best Gestalt possible given the prevailing internal and external conditions. In what follows, we will shed light on the diverse aspects of the Prägnanz concept that have been brought forward to clarify what a 'good' or 'simple' Gestalt entails.

A first way to characterize Prägnanz as a property is by defining it as **Gestalt height or strength** (Köhler, 1920; von Ehrenfels, 1916, 1932/1937): Strong Gestalts are those Gestalts in which the mutual dependence among the parts is so great that there is no room for any displacement or changes without an influence on all other parts of the whole (Ash, 1995; Köhler, 1920). For example, a rose is a higher Gestalt than a heap of sand (von Ehrenfels, 1916, 1932/1937). The emphasis on strong interdependence between the parts as the defining feature for strong Gestalts may remind you of the original meaning of the Gestalt concept as proposed by the Berlin Gestalt school (i.e., an ensemble of items that mutually support and determine one another; Sundqvist, 2003). von Ehrenfels (1932/1937) presented a clear criterion to identify higher Gestalts: "Higher Gestalts are those in which the product of unity of the whole and manifoldness of the parts [*das Produkt von Einheitlichkeit des Ganzen und Mannigfaltigkeit der Teile*] is greater". Consequently, when one keeps the degree of unity constant, those Gestalts that embrace a greater multiplicity of the parts will be better. Equivalently, for a fixed degree of multiplicity, those Gestalts that more strongly unify this multiplicity will be better (von Ehrenfels, 1916; translated in Smith, 1988). The importance of unity or regularity is also part of Metzger's (1941) description of Prägnanz: prägnant Gestalts show an outstanding [*ausgezeichnete*] and consequently persistent order (from a purely figural, i.e., non-semantic, perspective). Also more recent literature on how regularity and non-accidental properties increase degree-of-objecthood (e.g., Biederman, 1987; Feldman, 2003; Kubilius et al., 2014, 2017; Strother & Kubovy, 2012; Wagemans, 1992) relates to this characterization of Prägnanz as Gestalt strength.

A second aspect of Prägnanz that Metzger (1941) put forward, next to the aspect of strong figural unity, relates to the etymological meaning of Prägnanz (i.e., being sharply grasping, unambiguous, clear, or distinct, Arnheim, 1975). Metzger (1941) discussed three types of properties of a whole: (a) their structure [*Struktur oder Gefüge*], e.g., straight, round, angular, or symmetrical; (b) their whole quality or texture [*Ganzqualität oder -beschaffenheit*], which is material-related, e.g., transparent, rough, or shiny; and (c) their essence [*Wesen*], e.g., friendly, female, peaceful, or proud.[6] For each essence [*Wesen*], to the extent that it shows itself in structures [*Gefügen*], there is *a completely specified structure in which the essence is most pure and compelling* (Metzger, 1941, p. 62). This structure is called "ausgezeichnet" or "prägnant".[7]

For Metzger, both figural order and the pure, compelling embodiment of an essence are essential to understand the full meaning of Prägnanz, and often — if not always — go together (Metzger, 1941).

[6] Although this distinction between properties may remind readers of Koffka (1940)'s primary, secondary, and tertiary qualities, the classifications are not identical.

[7] It must be clear that there is not a random relation between essence and structure: where the structure is, the essence is too (Metzger, 1941). The perfection of a poem, for example, lies in the fact that its meaning or essence becomes clear already in the sounds, not just in the words (Metzger, 1941). Another example could be the expressive fonts and typesetting used by Paul van Ostaijen, where the way of positioning the words and the fonts used already embody the meaning of the text. These mutual relationships between whole properties are not bidirectional, however: not for every possible structure [*Gefüge*] and also not for every possible material whole quality [*Beschaffenheit*] there is a special essence or being [*Wesen*] (Metzger, 1941).

2.1.2.1 Call for qualititative and quantitative refinement and disambiguation of Prägnanz

Two common critiques on Prägnanz as a phenomenal property are the lack of a sufficiently precise qualitative definition and the lack of a quantitative measure (Metzger, 1966).

2.1.2.1.1 The lack of a sufficiently precise qualitative definition of Prägnanz

The lack of a sufficiently precise qualitative definition of Prägnanz has often led researchers to point out the ambiguity of the concept as well as to suggest narrower concepts to replace Prägnanz.

Petermann (1931), for example, argued that Prägnanz can only be the start of a formalization as it is much too unclear and undefined. He viewed the Prägnanz tendency as a danger for further progress, as it can serve as a magical solution. Wellek (1959) indicated the dual meaning of Prägnanz as related to clarity and simplicity [***rein figuraler Prägnanz***] as well as meaningfulness and expressiveness [***Sinnprägnanz***]. Although Wellek (1959) believed both aspects of Prägnanz can sometimes covary, he argued that this is predominantly not the case. Therefore, he posits that a distinction between these two meanings is necessary and Prägnanz as a whole is an ambiguous, and hence useless, concept. Kanizsa & Luccio (1986) also pointed to the ambiguity of the Prägnanz concept and distinguished between Prägnanz as excellence, uniqueness, “outstandingness” [***Ausgezeichnetheit***] and as simplicity and stability [***Einfachheit und Stabilität***].

Although these critiques may seem valid at first sight, they may be nuanced or viewed as less destructive based on the foregoing discussion of the origins of the concept. It is true that the Prägnanz tendency can hinder further progress when used as a magical solution (as mentioned by Petermann, 1931), but this was already clear from Wertheimer’s focus on the Prägnanz principle as a general framework to stimulate further investigation into more concrete organizational principles falling under this law and their interactions (Arnheim, 1987; Luchins & Luchins, 1998; Rausch, 1952; Wertheimer, 1924/1999). As Wellek (1959) indicates, Prägnanz is indeed a multifaceted term, and further specifying different aspects of Prägnanz and their relative importance under different conditions should be a continued research endeavor. Although this specification is far from finished, Rausch (1966) undertook a significant effort to clarify Prägnanz as a concept (cf. the section on “Rausch’s (1966) Prägnanz aspects” below). In the light of its Gestalt psychological context, the seeming ambiguity in the meaning of Prägnanz that Kanizsa & Luccio (1986) indicate (uniqueness vs. simplicity and stability) is maybe less of an ambiguity than it first seems: because organizational processes resulting in a stationary state will tend towards that organization which has minimal energy requirements, the Prägnanz tendency distinguishes stable, simple end Gestalts from other possible states and makes them unique (see also Zimmer, 1991). Gestalt psychologists have deliberately kept these various possibilities open, as part of their view of Prägnanz as a broad, multifaceted concept — a general principle that can take many forms. By distinguishing between diverse facets of the Prägnanz concept, the so-called ambiguity or vagueness can be avoided, and empirical research can investigate under which conditions the different manifestations occur. Conceptual refinement is thus a necessary condition for empirical progress, which is an important motivation for this review detailing the nuances in the original meanings of Prägnanz.

2.1.2.1.2 The lack of a quantitative measure for Prägnanz

Some researchers aimed to replace the qualitative Prägnanz concept with a narrower, more easily quantifiable concept. For example, Hochberg (1968) suggested that an objective definition of 'simplicity' is needed if we want to be able to predict how an ambiguous image will be perceived. Also, he compared the lack of a quantitative measure for simplicity with the lack of a quantitative way to determine likelihood (i.e., how do we know that what we perceive is the most likely interpretation?; cf. also the 'simplicity'-'likelihood' debate, e.g., Chater, 1996; Leeuwenberg & Boselie, 1988; Pomerantz & Kubovy, 1986; van der Helm, 2000).

Goldmeier (1937, 1972) reduced Prägnanz to the concept of **singularity**: singular (i.e., unique) qualities are those qualities that are very sensitive to change, and are contrasted with qualities that have a range character and consequently are insensitive to change. Yet Arnheim (1987) called the concept of singularity misleading: things can be unique for many reasons that are totally unrelated to Prägnanz. Singularity is only a secondary consequence of the purity of prägnant forms (Arnheim, 1987).

In the light of information theoretical approaches, Prägnanz has been equated with high **internal redundancy** (i.e., low information content; Attneave, 1954; Hochberg & McAlister, 1953)[8]. Hochberg & McAlister (1953) proposed to search for parallels of the qualitatively and subjectively formulated Gestalt principles of perceptual organization by analyzing the objective properties of the stimulus constellation. As an objective definition of perceptual "goodness" they posited the frequency of occurrence of, or the relative time span devoted to, each perceptual organization that a stimulus may elicit. They hypothesized that the organization requiring the least information to be specified, will be the most likely to be perceived. To approximate figural goodness in this way, it was deemed important (a) to empirically determine the dimensions on which information needs to be scored, and (b) to demonstrate a correlation between the frequency with which different organizations of a stimulus constellation occur and the calculated information scores for a large set of stimuli. Attneave (1954) operationalized Prägnanz as internal redundancy: the most prägnant figure will be the one with the highest degree of internal redundancy[9]. In his view, the different grouping principles specified in Gestalt psychology all stimulate redundancy.

Garner (1974) proposed subset size to be the critical aspect of redundancy in relation to pattern goodness: good patterns are part of small subsets and this relates to high internal redundancy. This subset size can be compared to the size of the total set, which contains all stimuli that can be produced given a specific set of dimensions and levels (Garner, 1974). Whereas in some cases the actual subset size may be known, an inferred subset can be produced for any single stimulus. Garner (1974) describes evidence for the beneficial effects of pattern goodness as indicated by inferred subset size on several information processing tasks, including perceptual discrimination, recognition memory, reproduction memory, and verbal encoding of stimulus patterns (Attneave, 1955; Checkosky & Whitlock, 1973; Clement, 1964; Clement & Varnadoe, 1967; Glanzer & Clark, 1963; Pomerantz, 1977; Pomerantz & Garner, 1973). One such finding is that good patterns are encoded more rapidly (Garner, 1974; Pomerantz, 1977). Often

[8] As will become clear, this is in contrast with Rausch's (1966) fifth Prägnanz aspect concerning element richness as contributing to a better overall organization or Gestalt.

[9] This leaves the complexity or multiplicity aspects as indicated by Rausch (1966) — discussed under the section on "Rausch's (1966) Prägnanz aspects" — untouched, however, and exclusively focuses on unity and simplicity as characteristic for Prägnanz.

sets of dot patterns are used in these tasks, with the number of reflections and 90° rotations that lead to different patterns than the given one determining subset size.

According to Palmer (1982), good figures are those that have greater transformational invariance (i.e., contain more symmetries). More specifically, Palmer (1991) extends Garner's (1974) idea of reflection and rotation subsets and proposes a figure to be good when there are more local and global transformations (e.g., rotation, reflection) that leave the figure unchanged.

In structural information theory (Leeuwenberg & van der Helm, 2012), another approach to quantifying Prägnanz and the Prägnanz principle, the stimulus organization that we will perceive is expected to be the one containing the least structural information (i.e., simplicity or descriptive minimum principle). Three basic types of regularity are distinguished: iteration (e.g., AAAA), symmetry (e.g., ABBA), and alternation (e.g., ABCB). The stimulus organization can involve hierarchically organized levels, of which the highest hierarchical level (i.e., the 'superstructure') will determine the perceived unity (Leeuwenberg & van der Helm, 1991). Structural information theory is mainly meant for intra-stimulus comparisons of different possible organizations of the same stimulus (Leeuwenberg & van der Helm, 2012; van der Helm, 2017; van Lier et al., 1994): for each stimulus, the most structurally simple interpretation will be perceived. When different stimuli are compared, the simplest descriptions for each of the stimuli are used and figural goodness is used as a criterion (Leeuwenberg & van der Helm, 2012). In this context, figural goodness is explicitly distinguished from simplicity and is defined as the detectability of (or weight of evidence for) a regularity in a stimulus.

Importantly, a reduction of the Prägnanz concept to internal redundancy or inferred subset size completely focuses on stimulus-related aspects and leaves out any influences of the observer on Prägnanz (see also Koenderink et al., 2018). Hochberg (2003) later admitted that this purely stimulus-based approach was too simplistic: for example, attention and meaningfulness or familiarity will also influence the organization that we perceive (e.g., Peterson & Gibson, 1994). Metzger (1975) pointed to the specificity of information theory-based research's results when treating Prägnanz as a negative entropy or redundancy measure, and strongly doubted the generalizability of these as measures for Prägnanz.

Whereas structural information theory focuses on how stimuli are perceived rather than on aspects of the stimuli themselves, the theory also leaves out any influences of the observer on what the most prägnant organization will be. Hence, this theory also does not provide a satisfying formalization of Prägnanz.

In many of these quantifications of Prägnanz, the focus has been on 'simplicity' rather than the more meaning- or complexity-related aspects of the Prägnanz concept (Hüppe, 1984; Koenderink et al., 2018; Luccio, 2019; cf. the section on "Rausch's (1966) Prägnanz aspects" below). For example, the complexity of the individual components in the stimulus constellation is not taken into account. This is possibly a consequence of the type of stimulus materials used in most studies: in dot lattices for example, element richness, expressiveness, and meaningfulness do not play a role (Hüppe, 1984).

Koenderink et al. (2018) investigated whether a quantification of the Prägnanz concept is possible at all, or more specifically, whether there are aspects or interpretations of the concept that can be quantified. They deviate from the tendency to focus on the stimulus structure alone, and also consider the observer as an important determinant of Prägnanz. If a measure for Prägnanz is to be found, it needs to take into account both the structural complexity[10] bottleneck of visual systems (i.e., an upper limit to the complexity that can be processed by the organism) and the relevance to the organism's biological fitness (Koenderink et al., 2018). As an example of Gestalts high in Prägnanz, the releasers or sign stimuli discussed in ethology are mentioned (i.e., stimulus constellations that trigger a fixed behavioral pattern in a particular species). Koenderink et al. (2018) state that in humans, the structural complexity bottleneck might be enough to approximate Prägnanz with some generality. When the structural complexity of a stimulus exceeds the level of structural complexity that the organism can process, the stimulus will no longer be recognized as a "picture", but rather as "featureless" or "just noise" (Koenderink et al., 2018). Within the capacity limits of the visual system, the more structurally complex stimulus will be experienced as more prägnant. This is because more complex patterns (that stay within the capacity limits of the organism) will be experienced as more unique and hence raise the odds of detection, while they are still 'simple enough' to ensure automatic detection (Koenderink et al., 2018). As a concrete example: given that the maximum structural complexity that can be processed by the organism is equal to 4 (M= 4), patterns with a complexity of 4 (N = 4) will be experienced as most prägnant ($P = N/M = 1$). For patterns that exceed the structural complexity bottleneck of the organism ($N > 4$), Prägnanz will drop abruptly} (cf. Figure 2.9).

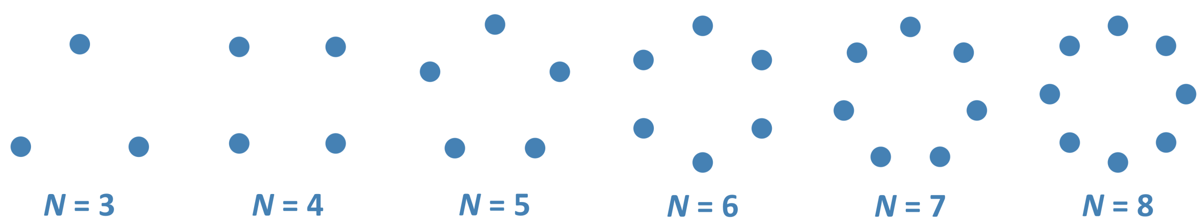


Figure 2.9: Dot patterns suggesting regular polygon shapes. N indicates the structural complexity of the pattern as defined in Koenderink et al. (2018). These figures were created using the OCTA toolbox (Van Geert, Bossens, et al., 2022), with the intention to resemble Figure 7 from Koenderink et al. (2018). Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.21977645.

In general, although several researchers have tried to replace Prägnanz with narrower, more easily quantifiable concepts, the Prägnanz concept – as originally conceived, not as later interpreted – cannot be replaced (Metzger, 1941). Its multifaceted nature certainly needs further specification and study, but this multifacetedness is exactly what is essential to make Prägnanz a viable concept. This also means that any interpretation of Prägnanz as purely figural, not taking into account the individual and context in question, will eventually fail. For example, when visually grouping a set of known objects, essential properties like purpose of use will more often play a determining role than structural (e.g., color, size, shape) or material properties, and complementary objects will preferably be grouped, rather than objects with the same function (Metzger, 1941).

Each of the mentioned quantifications of Prägnanz has led to valuable contributions. Importantly, it is not bad to try to quantify Prägnanz in a specific context for a specific set of stimuli, rather to the contrary. Nevertheless,

[10]Take note that the 'structural complexity' concept of Koenderink et al. (2018) does not refer to Rausch's fourth Prägnanz aspect of the simplicity or complicatedness of structure, but rather to figural complexity in the sense of stimulus entropy or element richness, as in Rausch's (1966) fifth Prägnanz aspect; cf. the section on "Rausch's (1966) Prägnanz aspects" below.

these measures are too preliminary and too stimulus- and context-specific to choose one quantification and thereby replace the overall concept of Prägnanz.

None of the listed critiques counters the essence of the concept (Metzger, 1966), nor do they imply that the scope of Prägnanz's application should be limited (Rausch, 1952). It is only important to recognize that (a) there is a need for further concretisation of Prägnanz, and (b) besides the general principle, more detailed, possibly quantitative, specifications of individual cases are also relevant (Rausch, 1952). Rausch (1966) made a laudable effort to qualitatively clarify different aspects of Prägnanz, and also proposed some early quantitative indicators of Prägnanz.

2.1.2.2 Rausch's (1966) Prägnanz aspects (cf. Figure 2.10)

As Rausch (1966) posits, Prägnanz is a highly complex concept, which makes it necessary to further specify its different aspects. Rausch (1966) distinguished seven Prägnanz aspects, and although these aspects are still complex in themselves (allowing for several different expressions of the same aspect), and the naming of these aspects is somewhat arbitrary (as a term is not always available to capture the full commonality within that aspect), these aspects may help to clarify the diversity of Prägnanz in its original meaning. Rather than only mentioning the overarching Prägnanz concept, Rausch (1966) advises future researchers to also specify which aspect of Prägnanz one refers to. The first four Prägnanz aspects highlight aspects of lawfulness or regularity (i.e., order, unity), the other three focus on aspects of 'fullness', complexity, or multiplicity [*Fülleaspekte*]. Five of the seven specify purely form-related aspects of Prägnanz, whereas the two last aspects are more content-related. Dependent on the phenomenon under consideration, one can use a system of 3, 4, 5, 6, or 7 Prägnanz aspects to evaluate its Prägnanz (Rausch, 1966). In what follows, Rausch's (1966) discussion of Prägnanz aspects is translated and summarized.[11]

2.1.2.2.1 Lawfulness vs. randomness

This first Prägnanz aspect reflects both the clarity of ***unity*** or degree of unity of a complex or Gestalt (or the clarity of the existence of a Gestalt quality) and its ***lawfulness*** or regularity (as opposed to randomness). Importantly, unity cannot only be reached by uniformity or homogeneity, but also by other types of structuring. Uniformity thus only presents a special case. As the experienced lawfulness (i.e., regularity, order) of a complex indicates or co-determines its experienced unity (and lawfulness and unity thus are not completely independent of each other), one can characterize lawfulness (as opposed to randomness) as the decisive factor for this first Prägnanz aspect. Whereas this aspect can be binary in experience (i.e., a complex is either ordered or not), it is also possible to speak of a degree of lawfulness. The other Prägnanz aspects depend on the presence of at least some form of lawfulness or regularity. As a reminder, it is important to take into account that it concerns the perceived lawfulness of a *phenomenal experience*. That means that whereas for one individual, a relative complex stimulus constellation can lead to a phenomenal experience with a random character, another individual or different conditions can yield a lawful phenomenal experience based on the same stimulus.

[11] An earlier English summary of Rausch's (1966) ideas on the Prägnanz concept is available in Smith (1988).

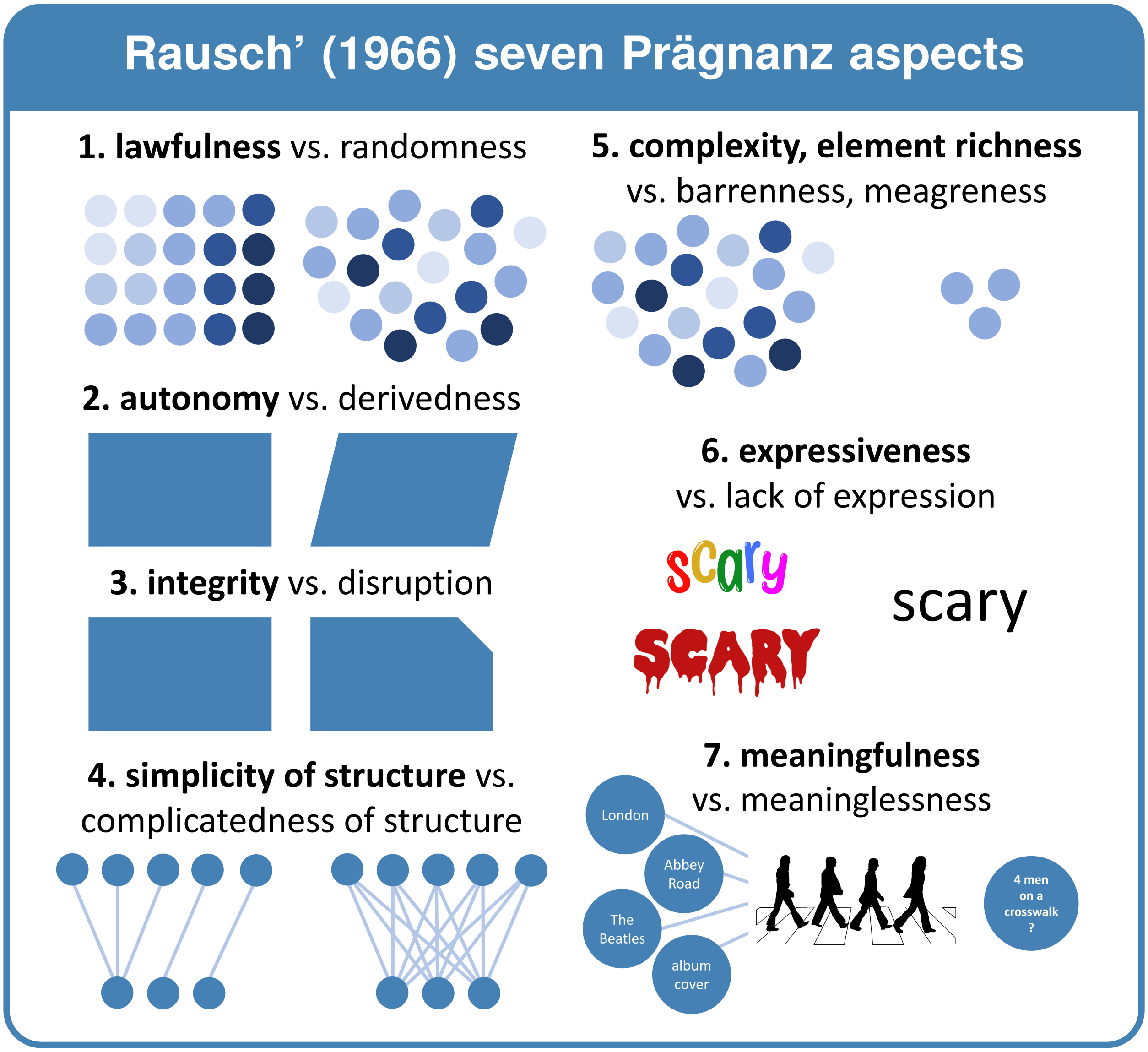


Figure 2.10: Illustration of the seven groups of Prägnanz aspects as defined by Rausch (1966). The Abbey Road icon was adapted from the CC BY licensed Abbey Road icon by Lia Thompson from the Noun Project. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.21977693.

2

2.1.2.2.2 Autonomy vs. derivedness

The second Prägnanz aspect distinguishes phenomena (or phenomenal qualities) that are ***autonomous*** rather than derived, in a binary fashion. For example, one could say that a parallelogram is derived compared to a rectangle (which is seen as autonomous). Similarly, obtuse and sharp angles can be seen as derived from the autonomous right angle. As a phenomenon can be autonomous in one dimension and derived in another, it is important to specify the dimension under evaluation. Derivations can occur in form or shape, but also in position (i.e., location and orientation). The relation between autonomous and derived is asymmetrical and non-reversible: whereas the parallelogram can be seen as derived from the rectangle, the rectangle cannot be seen as derived from the parallelogram. Although Rausch (1966) first posits this Prägnanz aspect to be dichotomous (i.e., a phenomenon is either autonomous or derived), he also acknowledges that a more gradual interpretation of autonomy is possible. For example[12], one could say that a rectangle is derived from a square, and then a rectangle would be more autonomous than a parallelogram, but less autonomous than a square (see also Feldman, 2000; Hendrickx & Wagemans, 1999; Leyton, 1992; Sablé-Meyer et al., 2021; Wagemans et al., 1994).

2.1.2.2.3 Integrity vs. disturbedness

Integrity or completeness rather than disturbedness[13] forms the third (dichotomous) Prägnanz aspect put forward by Rausch (1966). This disturbedness can be related to the stimulus constellation underlying the phenomenon as a whole or more locally, and can manifest in diverse ways: something can be missing, superfluous, or different; something can be 'not yet complete' or 'not complete anymore' (cf. also Spröte et al., 2016). In each of these cases, a complete version indicating how the complex should look serves as a reference. Similar as the relation between autonomous and derived in the second Prägnanz aspect, the relationship between distorted and complete is asymmetric. Although Rausch (1966) first posits this Prägnanz aspect as strictly dichotomous (i.e., a phenomenon is experienced as either complete or disturbed), he later acknowledges the possibility of a more gradual interpretation. One could say that a certain phenomenon or phenomenal quality is more or less disturbed, for example the extent to which an alignment of elements deviates from a straight line[14] (Claessens & Wagemans, 2008; Strother & Kubovy, 2006).

How can one distinguish between derivedness and disturbedness? This may have to do with the distance from the prägnant form. If a parallelogram is 'almost a rectangle', or an angle is 'almost right', it will be perceived as distorted. If the phenomenon is far away from the prägnant form, it may be perceived as derived. The precise distinction between the two is not always clear. In that case, one can talk about a *deviation* from a Prägnanz step (i.e., the complete or autonomous form of the complex, cf. the section on "Prägnanzstufen" below}) to refer to either derivedness or disturbedness.

2.1.2.2.4 Simplicity of structure vs. complicatedness of structure

Although it is easy to confuse this fourth Prägnanz aspect of ***simplicity of structure*** with the first (i.e., lawfulness or regularity), it can be distinguished from it. More specifically, regularity can be simple or complicated. Put differently:

[12] This example is not given in Rausch (1966).
[13] Just like derivations, disruptions can occur in form or shape, but also in position (i.e., location and orientation).
[14] This example is not given in Rausch (1966).

To be prägnant, an organization should at least contain some form of regularity. Within prägnant organizations, there is a tendency towards simple regularities. Simplicity of structure can be seen as a binary concept or as a continuum.

2.1.2.2.5 Complexity (part or element richness) vs. sparseness

Given a fixed level of regularity, a phenomenon will be experienced as more prägnant when it contains a greater number or a greater diversity of components, i.e., when it is more complex (as opposed to sparse). Importantly, ***complexity*** is different from the above mentioned complicatedness: whereas complicatedness relates to the entanglement and intricacy of a regularity or structure, complexity relates to the comprehensiveness and encompassingness of a phenomenon as indicated by the multiplicity and diversity of its parts. This fifth Prägnanz aspect can be used as a dichotomy or as a graded property. As all other Prägnanz aspects, complexity is meant as a phenomenal property, not as a property of the stimulus constellation. Individual and contextual differences can thus occur.

2.1.2.2.6 Expressiveness vs. weakness of expression

A phenomenon is experienced as more prägnant, the more expressive the phenomenon is. With this sixth Prägnanz aspect, Rausch (1966) goes beyond purely structural properties of a phenomenon. ***Expressiveness*** relates to Metzger's (1941) definition of Prägnanz as a pure embodiment of a nature or essence. For this Prägnanz aspect to be usable, the range of variability between different phenomena to be compared should be limited.

2.1.2.2.7 Meaningfulness vs. meaninglessness

The seventh Prägnanz aspect of ***meaningfulness*** deals with being able to connect a phenomenal experience with earlier acquired knowledge. For example, when perceiving an acquaintance, we can connect the face of this person with where he lives, which profession he has, which opinions he holds, etc. The richer the connections with previous knowledge, the more meaningful the experience of the phenomenon. Unlike expressiveness, which is inherently and from the start related to a structure or organization, meaning is added later. Similarly to the sixth Prägnanz aspect (i.e., expressiveness), meaningfulness can in practice only be used as a Prägnanz aspect when the range of variability between different phenomena to be compared is constrained. When Wellek (1959) distinguished purely figural Prägnanz from Sinnprägnanz, the latter referred to both expressiveness and meaningfulness.

2.1.2.3 Implications of Rausch's (1966) Prägnanz aspects for Prägnanz tendencies

As Rausch (1966) indicates, the further specification of Prägnanz in several Prägnanz aspects can also help to distinguish several Prägnanz tendencies. Oftentimes the Prägnanz tendency refers to a general tendency towards lawfulness and unity (i.e., the first Prägnanz aspect). In other cases, the Prägnanz tendency may refer to a tendency towards autonomy, towards completeness, or towards simple structure. Under weak stimulus conditions, it may be the tendency from distorted to complete that is most prominent. The tendency from derived to autonomous may occur in a more conscious manner. Both the tendency towards autonomy and towards completeness are tendencies towards a latent reference (i.e., the autonomous or complete 'original' of the phenomenal experience). Further research can clarify the strength of each of these Prägnanz tendencies under different conditions.

2.1.2.4 Rausch's (1966) quantitative indicators of Prägnanz

Rausch (1966) specified three quantitative characteristics or features (not 'measures') of Prägnanz. Although these quantitative concepts are defined here, they are in need for extension, as are the seven Prägnanz aspects mentioned (Rausch, 1966).

2.1.2.4.1 Density of Prägnanz steps (Prägnanzstufendichte; D)

As a characteristic of an individual, the quantitative indicator of *Prägnanz step density* concerns the degree of differentiation present when distinguishing and using prägnant steps (cf. the section on "Prägnanzstufen" below) in a certain domain. The higher the number of separate, clearly distinguishable ranges on a certain dimension, the higher the density of prägnant steps. It can be interpreted either as a temporary state or as a more permanent trait of an individual.

2.1.2.4.2 Prägnanz strength (Prägnanzstarke; S)

Prägnanz strength takes into account three to seven Prägnanz aspects in a dichotomous fashion, and indicates how many of those Prägnanz aspects are present in a phenomenon. The feature thus has a value between 0 and N, the number of N being the number of Prägnanz aspects taken into account. For example, a rectangle is prägnant on all three first Prägnanz aspects (S = 3), whereas a parallelogram is lawful and complete, but derived (S = 2).

2.1.2.4.3 Autonomy index (Eigenständigkeitsindex; J)

The *autonomy index* Rausch (1966) proposed only takes the second Prägnanz aspect into account, and deals with the question of how many important properties or dimensions of a phenomenon are autonomous or derived. For example, when taking equal width, straightness, and orthogonality into account as properties (cf. Rausch, 1952), a rectangle is autonomous on all three (J = 3), whereas a parallelogram lacks orthogonality and is, in that sense, derived (J = 2).

2.1.3 Prägnanzstufen

Ideas linked to the concept of **Prägnanz steps** [*Prägnanzstufen*] were present already in Wertheimer's (1912) article on numerical thinking in aboriginal people, *"Über das Denken der Naturvölker. I. Zahlen und Zahlgebilde"* (Hüppe, 1984). There, Wertheimer (1912) mentioned *ausgezeichneten Anzahlen* (§10, p. 337–340), literally translated as "out-standing", unique numbers. Some of these unique numbers become apparent from how they are named (e.g., in Andamanese, 10 = orduru = "all"). They are also used for the formation of higher numbers and in the rounding of prices. In addition, numbers close to them are expressed in terms of these unique numbers, which serve as a reference point (e.g., in Ralik Rakater: 8 = "take two away", where 10 is the reference point). These unique numbers serve as a first example of *Prägnanzstufen* (i.e., *Prägnanz* steps or reference regions), a concept elaborated further in Wertheimer's (1923) article *"Untersuchungen zur Lehre von der Gestalt. II."*.

It was only in 1923 that Wertheimer first discussed *Prägnanz* as a Gestalt principle and its relation to Prägnanz steps in more detail, although he conducted the research during his period in Frankfurt (1911-1914, Wertheimer, 1923; translated in Wertheimer et al., 2012). He used series of dots as well as angles of varying degrees to illustrate that in between distinctive regions (in which there is a clear, "winning" grouping or organization available), there are often intermediate series that are "not unequivocal to the same degree, not quite as salient [*prägnant*], 'less definite' in their character, less pronounced, and often more easily seen in terms of one grouping or the other" (Wertheimer, 1923; Wertheimer et al., 2012). The *Prägnanz* principle thus entails that if one varies a component (e.g., the location of a dot in between two other dots) in systematic, physically equidistant steps, the resulting psychological impressions will not be equidistant; the progression will be discontinuous as particular Prägnanz steps (i.e., *Prägnanzstufen*) occur, that each have their own range (of influence).

Between different Prägnanz steps, three situations can occur. *In between* the ranges of influence of different Prägnanz steps, either neutral transition areas can be present (in which percepts are indifferent or meaningless) or ambiguous percepts may occur that fluctuate between two Prägnanz ranges (Rausch, 1966). *Within* the ranges of influence of the Prägnanz steps, less prägnant forms will be experienced as related to the prägnant forms, as somewhat "poorer", incomplete, disturbed versions of them (Rausch, 1966; Wertheimer, 1923; Wertheimer et al., 2012). For example, an angle of 93° may look like a right angle, but not completely (Rausch, 1966). Put differently, less prägnant forms that are close to Prägnanz steps (i.e., in their range of influence) are *evaluated* in relation to these prägnant forms, but can be *perceptually* discriminated from them (cf. also the distinction between primary and secondary Prägnanz tendencies described by Hüppe, 1984).

Prägnanz steps thus serve a double function. On the one hand, assimilation (i.e., attraction, simplification) to the Prägnanz steps may occur, especially when the external conditions are weak (i.e., limited visibility, due to, e.g., brief presentation, low contrast, or small size; Köhler, 1920; Stadler et al., 1979; Wertheimer, 1923; cf. also Van Geert & Wagemans, 2022)[15]. On the other hand, Prägnanz steps can increase sensitivity to change in their vicinity: when viewing conditions are less limited, they can increase the ability to notice small deviations from a Prägnanz step (Goldmeier, 1937, 1982). In that sense, Prägnanz steps can support both robustness and sensitivity in visual experience, via simplification and complication, respectively (cf. the section on "How can the Prägnanz tendency be realized?" above).

Furthermore, there is an asymmetric relationship between less prägnant forms and the Prägnanz steps they are related to: "the bad Gestalt looks similar to the out-standing one, but not the other way around" (Metzger, 1941, p. 63; cf. also Goldmeier, 1937, 1982). In addition, a hierarchy of Prägnanz steps can exist: for example, the right angle will be a better Gestalt than the sharp or obtuse angles (following the three first Prägnanz aspects), although the sharp and obtuse angles can also be seen as Prägnanz steps (Rausch, 1966).

Prägnanzstufen or Prägnanz steps are thus exemplars of good Gestalts, as they embody a particular essence purely (Rausch, 1966). The number of Prägnanz steps can increase with experience or time, as new intermediate steps may develop, with the new steps forming as embodiments of special subtypes or subclasses of an essence in areas that are only meaningless intermediate areas for less sensitive individuals (Metzger, 1941; Wertheimer, 1923; Wertheimer et al., 2012). Individuals may not only differ in the number of Prägnanz steps they have on a dimension, but also in the width of their Prägnanz steps: more sensitive individuals may have narrower and more sharply centered Prägnanz steps (Metzger, 1941). These Prägnanz steps were primarily defined on purely

[15]Wertheimer (1923) presented the Prägnanz tendency under limited viewing conditions (i.e., assimilation to the prägnant form in tachistoscopic presentations) as the most extreme evidence for the statement that intermediate steps are psychologically perceived as "poorer" or "imperfect" versions of the Prägnanz steps.

quantitative dimensions, but Wertheimer (1923; 2012) noted that something similar occurs in the purely qualitative domain. For example, the system of Prägnanz steps concerning animal shapes for a zoologist, or concerning colors for a painter, will contain more, but also narrower, Prägnanz steps than the corresponding systems in children (Metzger, 1941). More recently, also the 'geons' (i.e., geometrical ions) that are a crucial part of Biederman's (1987) recognition-by-components (RBC) theory could be seen as reference shapes (i.e., Prägnanz steps) for recognizing object parts. Of course, 3D shapes could be differentiated better than the 2-fold or 3-fold distinctions on the 4 dimensions underlying the 36 components (2x2x3x3) that Biederman proposed as building blocks, but as building blocks for rapid and automatic object recognition, these would be sufficient according to RBC theory.

Rausch (1952) distinguished **Prägnanz height** (or dimensionality) from Prägnanz steps (or the Prägnanz function): Prägnanz height is used when multiple (objective) variables play a role, Prägnanz steps when only one (objective) property is varied. Although we call it Prägnanz 'steps', these do not refer to a stepwise function; we should interpret them as values from an objective variable (Rausch, 1952). In other words, we can look at a variable dimension as having a Prägnanz function across its domain: some regions have higher Prägnanz than others (Rausch, 1966), and it clearly concerns a gradual concept.

For a single objective domain, sometimes Prägnanz steps will be present and sometimes more homogeneous perception will arise. For example, when we speak of the time as 'five minutes before nine', we use nine o'clock as a reference point. When we however use a concurrently present local reference, like the small hand in a clock as comparison for the big hand, Prägnanz steps can be absent (Rausch, 1966). Relatedly, if a stimulus is presented as part of an ordered series, the factor of objective set or setting [*Einstellung*] comes into play, and Prägnanz steps will no longer be the only factor determining the resulting organization (Wertheimer, 1923).

2.1.3.1 Prägnanz steps in the work of Eleanor Rosch

The work of Eleanor Rosch on perceptual and cognitive reference points (e.g., Rosch, 1975) also builds on the idea of prägnant steps. Rosch (1975) viewed stimuli as 'reference points' when other stimuli are seen 'in relation to' them. She indicated that categories are often not clearly delineated, but rather built around prototypes (i.e., clearest cases, best examples). These prototypes exemplify the 'core meaning' of the category (cf. Prägnanz as pure embodiment of an essence). These core meanings around which categories build are in no sense arbitrary, but are given by the human perceptual system: they are more perceptually salient than other exemplars, hence 'natural' prototypes (Rosch, 1973).

We also do not need clear category boundaries to be able to judge the degree of prototypicality of an exemplar, but rather use clear cases as a comparison, an insight she attributed to Wittgenstein (Rosch, 1978). Non-prototype category members trend towards the prototype to a certain extent, and this may lead to systematic asymmetries in, for example, perceived similarity (Rosch, 1975).

Importantly, just as in the Gestalt view, Rosch (1978) does not interpret 'prototypes' as one single value on a dimension, but rather emphasizes the gradualness of prototypicality. In her view, a structure is more prototypical when it has more attributes in common with other members of the category, *and* when it has fewer attributes in common with members of contrasting categories (Rosch, 1978). Which attributes will be perceived is partially dependent on the functional needs of the organism (Rosch, 1978). Degree of protoypicality also correlated with beneficial effects on reaction time, speed of learning, etc. (Rosch, 1978).

Although Rosch (1975) was aware of Wertheimer's work on Prägnanzstufen – she described it as the idea that there are certain 'ideal types' which serve as anchoring points for perception (Rosch, 1975), and her husband was the son of Fritz Heider, a psychologist working in the Gestalt psychological tradition – she never discussed the Gestalt ideas in detail (Bock & Pfeiffer, 1987; Hüppe, 1984).

2.1.4 Aesthetics and Gestalt

When we psychologically organize incoming stimuli, we can *not only* describe or classify the experienced organization in a purely structural or semantic sense, we can also evaluate our aesthetic experience of this organization. Perceptual processing of the input is necessary to be able to aesthetically evaluate our percept, therefore the close relation between perception and aesthetics cannot be neglected.

Goodness of Gestalt has been tied to aesthetic appreciation since the beginnings of Gestalt psychology. von Ehrenfels (1922, p. 50) made the relation clear as follows: "what we call beauty is nothing else than Gestalt height". According to Arnheim (1986, p. 823), Wertheimer spoke of 'good' Gestalts because of the cognitive and aesthetic improvement they bring about. Koffka indicated that violations of organizational principles like good continuation and good shape due to external conditions are felt as violations because they "hurt our sense of beauty" (Koffka, 1935, p. 175). Moreover, those stimulus constellations that are most in agreement with the organizational principles underlying our perception will be judged as most beautiful (Eysenck, 1942).

Aesthetics has a clear relation to the Prägnanz *tendency* as well. Metzger (1941) indicated that true artists will go beyond their models in the direction of Prägnanz, which means that they will come to a structure that more purely and compellingly specifies the essence of content of the artwork. To make the essence clearer, they can use simplification and/or complication strategies. Also in art, these tension-reducing and tension-promoting tendencies are always concurrently present to a certain extent (Arnheim, 1975). In his late works, Piet Mondrian, for example, used rectangularity and primary colors, which served to simplify, but the irregular spacing was a complication, which served to make his work more dynamic (Arnheim, 1975).

The same tendencies — order and complexity, unity and variety, integration and differentiation, simplification and articulation — are thus at play as determinants of aesthetic appreciation and those of good Gestalt or Prägnanz in visual perception (Eysenck, 1942). In perception, we are constrained by the external conditions in our tendency towards Prägnanz, and these external conditions will typically not allow for highly balanced and symmetric percepts. When the external conditions do allow balance and symmetry, however, it will be perceived (Koffka, 1940). In that sense, perception is artistic (Koffka, 1940). Artworks on the other hand are made with balance and symmetry in mind, they are made to serve as as a source of stimulation that results in the perception of a good Gestalt (Koffka, 1940; Smith, 1988).

The artist will thus trigger tendencies in the observer to experience order, but different types of Prägnanz tendencies may be triggered and artists may differ in the type of Prägnanz they aim to maximize (Smith, 1988). Exactly the multidimensionality of Prägnanz makes the work of artists extremely difficult, as it is unclear how different Prägnanz tendencies will interact when concurrently present (Smith, 1988). Knowing about the diverse aspects to Prägnanz might support artists in finding dimensions of aesthetically relevant structure in both their artworks and the reactions to their artworks, without dictating how an artwork should look (Smith, 1988).

In our view, a minimum level of unity or order may be a prerequisite for aesthetic appreciation as it is for Prägnanz, but aesthetic appreciation is expected to arise together with a conscious increase in Prägnanz. This increase may be seen as a comparison between organizations (i.e., one organization is experienced as more prägnant than the other), or as an improvement in the experienced organization keeping the stimulus constant. For example, by extended looking, repeated viewing, and/or expertise, one may notice some kind of higher-order relationship between elements that were initially perceived to be disconnected or in arbitrary positions. The percept then becomes increasingly better organized even though all stimulus elements remain the same. Following this idea, aesthetic appreciation will be higher when a stronger Prägnanz *tendency* is experienced, which can be the result of an increase in any of the mentioned aspects of Prägnanz (either related to order or complexity). It is important to note, however, that increased complexity will only lead to increased appreciation up to the point the observer can still grasp the resulting organization (cf. above).

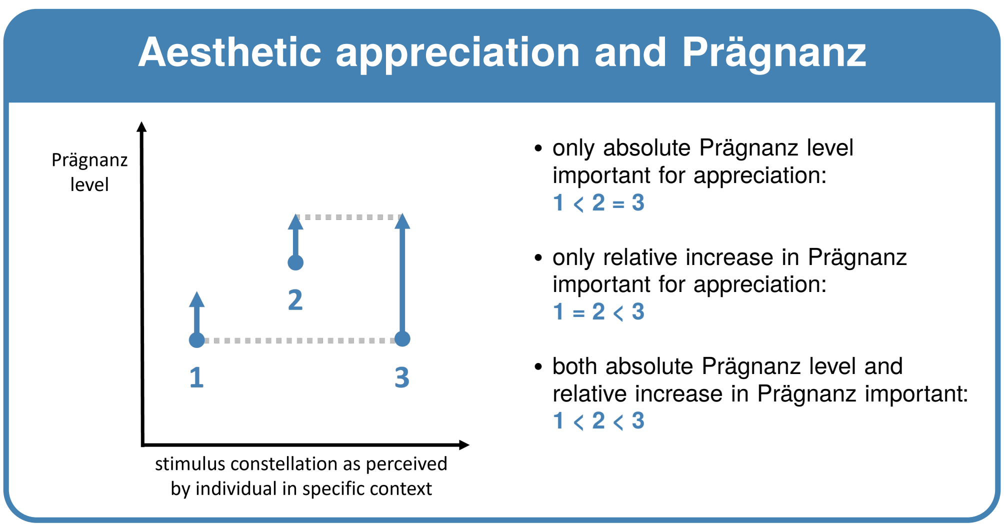


Figure 2.11: Potential relations between aesthetic appreciation and Prägnanz. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.21977762.

This view (see also Van Geert & Wagemans, 2020) is similar to other accounts of aesthetic appreciation, including the predictive processing accounts of Van de Cruys & Wagemans (2011) and Chetverikov & Kristjánsson (2016) as well as the focus on pleasure by insights into Gestalt proposed by Muth and Carbon (Muth et al., 2013; Muth & Carbon, 2013, 2016). An alternative view, more in line with the interpretation by von Ehrenfels (1922) and Eysenck (1942), is that aesthetic appreciation could be based on the absolute level of Prägnanz experienced. The absolute level of Prägnanz (i.e., Prägnanz *height*) does not necessarily relate to the strength of the experienced Prägnanz *tendency* (i.e., relative increase in Prägnanz height; see Figure 2.11). Nevertheless, both views may act complementarily as well. This combination of views is in line with the pleasure-interest model of aesthetic liking proposed by Graf and Landwehr (2015, 2017).

2.2 Discussion and conclusion

By looking back at the history of Prägnanz, we aimed to give a more detailed overview of the different uses and interpretations of the concept than is typically done. We distinguished four main uses: (a) Prägnanz as a tendency present in each process of psychological organization; (b) Prägnanz as a property of the result of such an organizational process; (c) Prägnanz steps as points of comparison when organizing the current stimulus; and (d) Prägnanz in relation to aesthetic appreciation (cf. Figure 2.3).

More specifically, the law of Prägnanz concerns the tendency to achieve the best psychological organization possible given not only the visual input, but also the individual, the context, and their interactions. Importantly, the Prägnanz principle thus concerns experienced organizations — percepts — not stimuli. When the external stimulus factors are weak, the tendency towards Prägnanz can play a larger role. This tendency is realized by comparing the visual input to a (local or internal) reference, using two antagonistic but complementary tendencies. On the one hand unimportant differences are downsized or removed (i.e., leveling, simplification). On the other hand significant, characteristic differences are added or emphasized (i.e., sharpening, complication). Both tendencies contribute to the emergence of a better overall organization. Future research should investigate how these tension-reducing and tension-enhancing tendencies interact under different circumstances and clarify how the reference emerges.

For a psychological organization to be prägnant or 'good', the organization should be perceived as containing at least some form of unity or regularity, and, all other things being equal, Prägnanz increases when the organization is perceived as more autonomous, complete, simple of structure, element rich, expressive, and/or meaningful. Whereas some of these Prägnanz aspects relate to order and unity, other Prägnanz aspects relate to richness and intricacy (i.e., complexity). This highlights that both order and complexity play an important role in the concept of Prägnanz. Given these several aspects to Prägnanz, multiple Prägnanz tendencies can exist, and new research should illuminate which of these tendencies dominate under which conditions as well as how different tendencies interact.

When only one dimension of a stimulus is varied, the values on that dimension that are associated with the most prägnant percepts are called Prägnanz steps [*Prägnanzstufen*]. These Prägnanz steps will serve as reference levels for psychological organization: not only do they serve as a point of comparison for a broad range of percepts, they also make the organism more sensitive to change in their vicinity. The idea of prägnant Gestalts as reference points may however be useful in the multivariate case as well. When multiple stimulus dimensions are varied, the term Prägnanz height or dimensionality is used to indicate the Prägnanz or goodness of a phenomenally experienced organization. In the future, it is worthwhile to explore the diverse consequences of prägnant Gestalts and the interactions between those consequences in more detail and in a more systematic way than was done before, also taking the diverse Prägnanz aspects into account in the choice of the prägnant Gestalts under investigation.

Since the inception of Gestalt psychology, prägnant Gestalts have been proposed to be the percepts that are most aesthetically appreciated, in addition to the ones we tend to perceive when possible. That perception and appreciation are closely related is largely beyond doubt, but the exact relation between them is more difficult to pin down. Whether appreciation is related to the absolute Prägnanz level of a psychological organization, to the relative increase in Prägnanz and thus the strength of the experienced Prägnanz tendency, or to both is subject to further investigation. In addition, the result of these future investigations could depend on the Prägnanz aspects taken into account, so a systematic study of different possible combinations is recommended.

2.2.1 General points of attention when investigating Prägnanz

One of the main goals of Gestalt psychology was to discover the principles governing human perceptual organization as well as the conditions influencing it (Ash, 1995; Koffka, 1935; Wertheimer, 1924/1999). Besides the above mentioned recommendations for future research related to the different uses of Prägnanz, some more general recommendations for future research on Prägnanz can be made.

2.2.1.1 Conduct concrete research with Prägnanz as a guiding principle

Let's use Prägnanz and Gestalt theory more generally as Wertheimer (1924/1999) proposed it: as a framework and device for future, concrete research. Although most contemporary researchers have left the basic ideas of Prägnanz and Gestalt theory behind, we believe that these ideas can serve as important handles for improving our understanding of human psychological organization and visual perceptual organization in particular. As has been pointed out before (Ash, 1995; Smith, 1988; Wagemans, Feldman, et al., 2012), Gestalt theorists' opposition to a simple associationist, elementaristic world view is a point that is still highly relevant in the current research climate. On the other hand, it is wrong to view Prägnanz as a fixed part of a theory to which no theoretical changes can be made, or as a finalized product that should be taken for granted or left aside. Further clarifications and specifications of the workings of the different Prägnanz tendencies in concrete cases are needed to further elaborate the general framework and test specific aspects of it.

2.2.1.2 Respect the richness and multiplicity of Prägnanz as a concept

Let's not simplify Prägnanz to the narrow interpretations it has been given after having been taken out of its original Gestalt theoretical context. Especially, let's consider that percepts and psychological organizations in general originate in organisms, which reveals that "goodness" of organization cannot be determined solely based on stimulus conditions, but needs to take into account the observer in its context. In different contexts, different Prägnanz aspects may become dominant (e.g., Marković & Gvozdenovi, 2001).

2.2.1.3 Investigate potential quantitative indicators of Prägnanz

Let's use the diverse qualitative Prägnanz aspects defined by Rausch (1966) as a starting point for further reflection on different quantitative indicators of Prägnanz and quantitative models of Prägnanz tendencies. Quantifications have been proposed for some of the grouping principles and their interactions (e.g., proximity, similarity, good continuation; Claessens & Wagemans, 2008; Froyen et al., 2015; Jäkel et al., 2016; Kubovy & van den Berg, 2008; Kubovy & Wagemans, 1995; Quinlan & Wilton, 1998). Recently, individual differences in the strength of these grouping principles have also started to be taken into account (Van der Hulst et al., 2022; Van Geert, Moors, et al., 2022). Although each of these quantitative indicators in itself is not enough to replace the overall concept of Prägnanz, specifying the influences and interactions of different Prägnanz tendencies, quantitatively when possible, is the way forward proposed by the Gestalt psychologists themselves (Rausch, 1952; Wertheimer, 1924/1999).

2.2.1.4 Respect the qualitative version of Prägnanz

Let's not overemphasize the importance of a definitive quantitative measure for Prägnanz. Although a quantitative indicator is helpful, not all useful concepts are easy to measure (and not all measures represent useful concepts).[16] So, let's not throw away the baby with the bathwater: Prägnanz is a valuable concept even when it cannot be quantitatively defined as an overarching concept. Koffka (1935) emphasized that although a quantitative formulation of Prägnanz is desirable, it only entails a more *precise* specification of the qualitative formulation, which is not different from it in kind. Rausch (1979/1992) also stressed the importance of paying attention to qualitative research methods and results next to quantitative ones.

2.2.1.5 Try to connect and compare Prägnanz to other perspectives on perception

Let's try to see the connections between the Gestalt theoretical concept of Prägnanz and other perspectives on visual perception. Typically, 'simplicity' and 'likelihood' have been presented as contrasting principles (e.g., Pomerantz & Kubovy, 1986; van der Helm, 2000, cf. also Box 1), while overlap in ideas and predictions following from these ideas seems to have been largely ignored. Given that both Bayesian and Gestalt psychological views posit a connection to regularities in the physical world — be it directly by proposing veridicality or indirectly by proposing parallellism — it is understandable that both views will lead to similar results in many cases (cf. also van der Helm, 2000). Furthermore, Gestalt thought is not in contrast to any influence from learning or previous experience, rather to the contrary. The main difference is that Gestalt theory does not view 'previous experience' as a definitive answer to all questions of human perceptual organization, and it emphasizes the importance of more general principles of organization next to the influences of previous experience. A comparison, confrontation, or synthesis between these diverse views on perception can all advance research and unite the field. As Henle (1987) puts it, some researchers mainly see continuities (i.e., the dromedaries), others mainly see dichotomies (i.e., the camels). There are however many basic issues in psychology that cannot be solved by any of them. We need to overcome dichotomies in another way than by choosing one of the two, simply adding both, or finding a middle ground (Henle, 1987).

2.2.1.6 Look for potential neuroscientific indicators of structural energy

Let's look at the neuroscientific processes underlying the principles of perceptual organization through the lens of Prägnanz. Can we find measures or indicators of structural energy across the brain? According to Köhler (1940), neuroscientific evidence is the only way to find support for the Prägnanz tendency as a tendency towards minimal structural energy. A study by Schurger et al. (2015) found a more stable neural activation pattern for trials in which a stimulus was consciously perceived compared to trials in which the stimulus was not consciously perceived. Although this study is not yet direct evidence for the Prägnanz tendency, it is congruent with Prägnanz as a tendency towards the best, most stable organization.

[16] As William Bruce Cameron (1963) said it: "Not everything that can be counted counts, and not everything that counts can be counted."

2.2.2 Key takeaway

Prägnanz is a multifaceted Gestalt psychological concept indicating the “goodness” of a perceived organization. The stimulus constellation is not the only factor in determining the goodness of an organization, also the stimulus’ interaction with an individual in a specific spatial and temporal context plays a role.

The Prägnanz principle indicates a tendency present in every process of psychological organization to tend towards the most prägnant organization possible. As Prägnanz is a multifaceted concept, several tendencies can be present and their interaction in different stimuli, contexts, and individuals is an important area for further study.

Prägnanz as a concept cannot be reduced to a singular dimension, but entails diverse aspects including those related to order and unity as well as intricacy and complexity. Organizations particularly high in Prägnanz are sometimes used as a reference to which incoming stimuli are compared. In addition, Prägnanz has a close connection to aesthetic appreciation and artistic practice, although the exact relation is subject to further research.

Taking the ideas about Prägnanz as a guiding framework and keeping the original Gestalt psychological context in mind, future concrete research on perceptual organization can retake the path paved by the Gestalt psychologists by further specifying how different organizational principles interact in concrete situations, by respecting the nuanced and multifaceted nature of Prägnanz, and by clarifying which specific aspects of Prägnanz are under investigation.

Bringing different empirical findings together in a broader framework of tendency towards Prägnanz does not mean that it is not worth studying more specific effects, but integrating them in a broader framework will bring us from a lot of scattered pieces of information to knowledge of a system (see also Koffka, 1935).

Part III

Robustness and sensitivity

© Joris Snaet

When the incoming visual stimulation is limited, our own *internal biases* can play a larger role in what we perceive.

Chapter 3

What good is goodness? The effects of reference points on discrimination and categorization of shapes

3

Earlier research reported a category boundary effect on perception: Differences between stimuli belonging to the same category are perceived as smaller than differences between stimuli belonging to different categories even when the physical dissimilarity between the stimuli in the pairs is the same. In this paper, we propose that the existence of reference points (i.e., exemplars that serve as a point of comparison) can explain the occurrence of the category boundary effect as well as the directional asymmetries in within-category pairs. We investigated how reference points influence categorization and discrimination performance, using three different tasks: categorization, successive discrimination, and similarity judgment. We used both recognizable and non-recognizable morph figures as stimuli, assuming that recognizable series have clearer reference points. We replicated the overall category boundary effect for both discrimination and similarity and show the effect's dependence on the strength of the reference points involved. The general category boundary effect is not a proper category boundary effect, however: rather than the type of stimulus pair presented (i.e., within- or between-category) one needs to take into account the distance from the reference points for each of the individual stimuli in the pair to actually predict discrimination performance and similarity judgments. These results provide evidence that reference points on a dimension and their strength have tangible consequences for how we perceive, categorize, and react to stimuli on that dimension. Moreover, our findings remind us of the danger of averaging without looking at underlying data patterns, and of the gains that can be made by seriously exploring consistent variability in extensive data sets.

Van Geert, E., & Wagemans, J. (under review). What good is goodness? The effects of reference points on discrimination and categorization of shapes. Preprint available from: https://doi.org/10.31234/osf.io/6x75c

Preprint | Interactive figures | Materials, data, and code

3

What good is goodness? The effects of reference points on discrimination and categorization of shapes[1]

3.1 Introduction

For many decades, psychologists acknowledge the importance of using psychological rather than physical scaling when assessing similarity between stimuli or differences in reactions to stimuli (Fechner, 1860; Shepard, 1987). Recent work also discusses the importance of using psychological scaling rather than physical scaling in explaining several types of human functioning, performance and behavior (e.g., Schurgin et al., 2020; Sims, 2018). In most of this previous literature, psychological scaling has been used as an explanatory variable, but it is less often considered which factors – besides physical (dis)similarities – influence the formation and dynamic adaptation of these psychological scales. The existence of *reference points* (i.e., exemplars that serve as a point of comparison, in relation to which we perceive other exemplars of that category) may be an important factor in the formation and dynamic adaptation of categories, and consequently, in the formation of psychological scales.

In what follows, we first explain in more detail how categories can influence psychological similarity, with a particular focus on the well-known 'category boundary effect'. We then dive into potential explanations for the existence of this effect and show how the tendency towards reference points can explain the existence of the category boundary effect as well as directional asymmetries that are sometimes found when studying the category boundary effect. At the end of this introduction, we discuss the goals of the current empirical study in more detail.

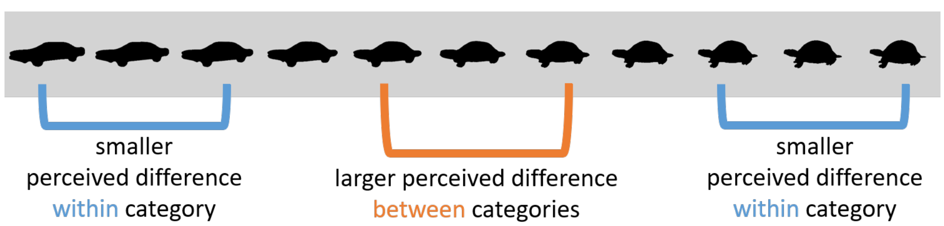


Figure 3.1: Illustration of the category boundary effect (CBE). Keeping physical distance equal, differences between stimuli belonging to the same category are perceived as smaller than differences between stimuli belonging to different categories (i.e., stimulus pairs crossing the category boundary, Harnad, 1987).

3.1.1 Categorization and the category boundary effect

Which factors influence the formation and adaptation of psychological scales? Previous research suggests that the categories an individual employs influence one's perception, a phenomenon referred to as categorical perception (CP; Goldstone & Hendrickson, 2010). One of the most prominent findings related to CP is the category boundary effect (see Figure 3.1)[2]: Keeping physical distance equal, differences between stimuli belonging to the same category are perceived as smaller than differences between stimuli belonging to different categories (i.e., stimulus pairs crossing the category boundary, Harnad, 1987). It is typically assessed by looking for better discrimination performance for physically equidistant pairs of stimuli in which each stimulus belongs to a different category

[1] Garner (1974) posed the question "What good is goodness?" in the second chapter of his book *'The Processing of Information and Structure'*.

[2] Often the category boundary effect is also referred to as categorical perception, but here we define categorical perception more broadly as all possible effects of categorization on perception.

compared to pairs in which both stimuli belong to the same category (Harnad, 1987; Newell & Bulthoff, 2002). This does not mean, however, that within-category differences cannot be noticed at all; the category boundary effect is about a decreased ability rather than a complete inability to discriminate within-category members (Harnad, 2003; see also Shepard, 1987).

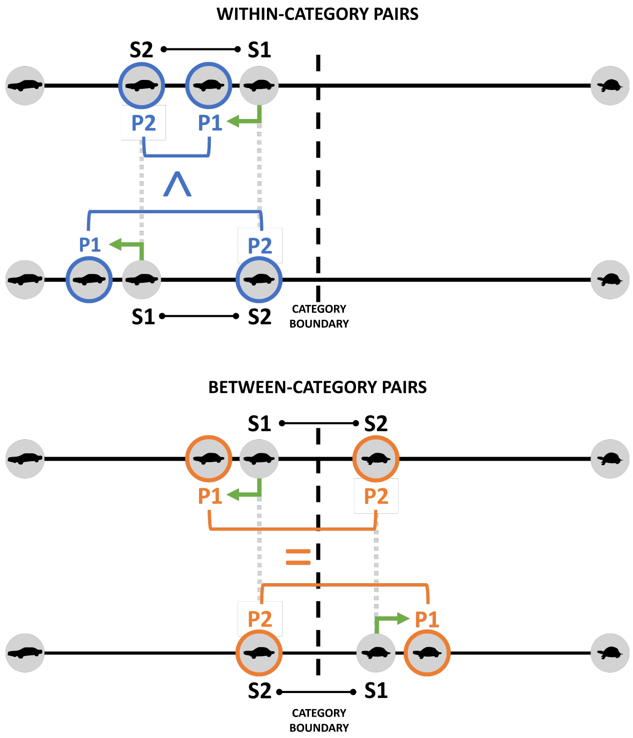


Figure 3.2: Illustration of how asymmetries could arise from the existence of reference points. When comparing stimuli within the same category, the first stimulus (S1) presented would then be drawn (green arrow) towards the reference point, making discrimination more difficult when compared to a second stimulus (S2) closer to the reference point. However, when the first stimulus (S1) presented in the within-category pair is compared to a second stimulus (S2) further away from the reference point, discrimination will become easier rather than more difficult. For between-category pairs however, this tendency towards the reference points will always increase the psychological distance between the stimuli in the pair, leading to the overall observed category boundary effect (i.e., between-category pairs are overall easier to discriminate than physically equidistant within-category pairs). As more time has passed since the presentation of S1, S1 will be more susceptible to memory effects than S2, which was presented more recently and thus is retained more perceptually.

3.1.2 The category boundary effect as a consequence of reference points

Why would discrimination performance be poorer for within-category than for cross-category pairs? Often the category boundary effect has been claimed to be the result of heightened natural discrimination sensitivity around the category boundary (Pastore, 1987; Repp & Liberman, 1987). When discrimination sensitivity was tested at threshold, however, no evidence was found for lower discrimination thresholds at the category boundary (Hanley & Roberson, 2011; Roberson et al., 2007, 2009). The category boundary effect can thus not be explained by heightened natural sensitivity at the boundary, as heightened sensitivity would have resulted in a lower threshold for discrimination at the category boundary compared to other segments of the continuum. Later research showed that within-category discriminability is not *always* poorer than between-category discriminability (Hanley & Roberson, 2011; Roberson et al., 2007): For within-category pairs, an asymmetry in discriminability was found (see Figure 3.2). Discriminability was tested here using an XAB task, in which a two-alternative forced-choice task was presented after a target was shown separately. Only when the target stimulus was closer to the category boundary (and thus more ambiguous) than the distractor, the within-category discriminability was poorer than the discriminability across the category boundary (Hanley & Roberson, 2011).[3] Similar asymmetries have been found, for instance, in similarity judgment (e.g., Op de Beeck et al., 2003a; Panis et al., 2011) and visual search tasks (e.g., Kayaert et al., 2011). In similarity judgment tasks, participants judged two stimuli as more similar when the stimulus further away from the reference point was compared to the stimulus closer to the reference point rather than the other way around (e.g., 99 is more similar to 100 than 100 is to 99, with 100 being the reference point; Op de Beeck et al., 2003a; Panis et al., 2011; Polk et al., 2002; Rosch, 1975; Tversky, 1977). Likewise, performance in a successive discrimination task, assessed by both accuracy and reaction time, was worse when the stimulus further away from the reference point was compared to the stimulus closer to the reference point rather than the opposite. This effect has been given many names (e.g., time-order effect in Patching et al., 2012; directional asymmetries in Polk et al., 2002; asymmetric similarity in Nosofsky, 1991; contraction bias in Ashourian & Loewenstein, 2011). The existence of these asymmetries (i.e., making discrimination more difficult when a stimulus further away from the reference point was compared to a stimulus closer to the reference point) corroborates the idea that heightened natural sensitivity at the boundary is not what explains the category boundary effect.

More importantly, both the asymmetries found and the general category boundary effect may be consequences of the existence of reference points (see Figure 3.2). For the morph series and tasks used in the current study, we define the extremes on the dimension as the logical reference points (i.e., 'clear' exemplars that serve as a point of comparison, in relation to which other exemplars on the dimension are perceived). Rather than a consequence of inter-item similarity, the asymmetries found may arise because of the properties of individual items presented, also referred to as 'stimulus bias' (Nosofsky, 1991). When comparing stimuli within the same category, the first stimulus presented (S1) would then be drawn towards the reference point, making discrimination more difficult when compared to a second stimulus closer to the reference point (S2). However, when the first stimulus presented in the within-category pair is compared to a second stimulus further away from the reference point, discrimination would become easier rather than more difficult. For between-category pairs however, this tendency towards the reference points will always increase the psychological distance between the stimuli in the pair, leading to the overall observed category boundary effect (i.e., between-category pairs are overall easier to discriminate than physically equidistant within-category pairs). As more time has passed since the presentation of S1, S1 will be more susceptible to memory effects than S2, which was presented more recently and thus is retained more perceptually.

The assimilation to prototypes or *perceptual magnet hypothesis* states that reference points shrink the psychological distance around them, which implies decreased sensitivity around reference points (Feldman et al., 2009; Hellström, 2007; Kuhl, 1991; Samuel, 1982). The *perceptual anchor hypothesis* suggests that reference points (i.e., anchors) cause an increase in sensitivity in their vicinity (Acker et al., 1995; Quinn, 2000). Evidence has been found for both of these hypotheses, and task context seems to be a moderating factor (Quinn, 2000): Whereas increased sensitivity around reference points (above threshold) has been found for tasks involving direct perceptual

[3]The opposite asymmetry has been found later on, however, and then it was suggested that the differences in asymmetry may have been due to differences in the choice statistics in the different experiments (Best & Goldstone, 2019).

comparisons with plenty of perceptual evidence present and minimal memory demands, decreased sensitivity around reference points has been reported for tasks in which a currently available stimulus was compared to stimuli stored in memory (i.e., for which there was limited perceptual input). Put differently, anchor effects may be more related to early perceptual processing, whereas magnet effects may arise from later perceptual or cognitive processing (Quinn, 2000).

3.1.2.1 Earlier work on the functions of reference points in the Gestalt psychological literature

This double function of reference points–on the one hand attracting and on the other hand repelling neighboring stimuli–has been suggested much earlier than the 90's of last century. As discussed more extensively elsewhere (Van Geert & Wagemans, 2023), already since the emergence of the concept, "goodness" or *Prägnanz* was related to very low identity tolerance (i.e., very high sensitivity to change, related to the anchor effect) as well as to serving as a reference for a broad range of stimuli (i.e., high robustness against transformation, related to the magnet effect). Wertheimer (1923) indicated that shapes close to a prägnant step appear as [perceptually] different from but [categorically] related to the *prägnant Gestalt*: "as a somewhat 'poorer' version of it" (p. 318). In other words, although shapes close to the prägnant form can be *perceptually* discriminated from it, they are still *categorized* in relation to this prägnant form: the prägnant form serves as a reference point.

In situations in which there is limited perceptual evidence (e.g., in tachistoscopic presentation conditions and tasks involving memory), a tendency towards the prägnant form (i.e., reference point) was observed, with assimilation to the prägnant form (i.e., a tendency towards Prägnanz of the Gestalt, Köhler, 1920; Stadler et al., 1979; Wertheimer, 1923). Consequently, it is more difficult to transform prägnant figures into non-prägnant ones than the other way around (Goldmeier, 1937, 1982). This tendency towards prägnant forms is in line with the asymmetry effects found in the categorical perception literature: If conditions with limited perceptual evidence lead to a tendency to transform less prägnant figures (i.e., figures further away from the reference point) into more prägnant ones (i.e., figures closer to the reference point), this will make stimulus pairs in which the exemplar further away from the reference is compared to an exemplar closer to the reference point to be *perceived as more similar* than when the exemplar closer to the reference point is compared to an exemplar further away from the reference. Similarly, this tendency will make *discrimination* between stimulus pairs in which the less prägnant exemplar is compared to the more prägnant exemplar more difficult than when the more prägnant exemplar is compared to the less prägnant one (see Figure 3.2).

Furthermore, Wertheimer (1923) mentioned that in-between distinctive regions, there are often intermediate series that are "not unequivocal to the same degree, not quite as salient [prägnant], "less definite" in their character, less pronounced, and often more easily seen in terms of one grouping[4] or the other" (Wertheimer, 1923, p. 317; translated by Spillmann, Wertheimer et al., 2012). This description of less prägnant Gestalts indicates that prägnant forms (i.e., reference points) are often related to ease of categorization, as it is easier for less prägnant Gestalts to change their category membership. Empirical work has confirmed that distance from reference points (either at the category center or the extremes) can predict ease of classification (Medin, 1989; Medin & Barsalou, 1987). Relatedly,

[4] Wertheimer probably refers to a cognitive grouping (i.e., categorization) in this sentence.

proximity of a stimulus to the category boundary has been found to increase the difficulty of categorization (Gillebert et al., 2009; Grinband et al., 2006).

Also Goldmeier's (1937, 1982) characterization of Prägnanz as singularity stressed this double functionality of reference points. First and foremost, he described singularity as sensitivity to change: whereas coding accuracy is high in the narrow range of prägnant Gestalts, nonsingular values are coded only approximately. He thus predicted a greater discriminative ability for figures in a small range around the reference points. Secondly, he discussed singularity as a norm or reference: in the near-prägnant zone, feature values of a stimulus are perceived in relation to the reference points on that feature dimension.

Goldmeier's hypothesis relates very closely to the observed category boundary effect and its limits: We can interpret prägnant steps as several categories lying on a dimension, with intermediate steps that are categorized in function of these prägnant steps, but intermediate steps can be perceptually distinguished from prägnant steps when enough perceptual evidence is available. The jumps (i.e., the transitions between a prägnant and an intermediate step) in the dimension relate to the increased discrimination performance around the category boundary, and more importantly to jumps in psychological scaling: "In general, if one varies a component [...], in systematic steps, then the resulting impressions are psychologically not a mass of individually characteristic impressions consisting of evenly balanced matched steps. Rather, particular salient steps [*Prägnanzstufen*] occur, each with its range; the progression shows breaks. Intermediate steps typically appear as related to one of the salient forms [*Prägnanzformen*]." (Wertheimer, 1923, p. 319; translated by Spillmann, Wertheimer et al., 2012).

3.1.2.2 Later work on 'goodness' and reference points influencing perceived similarity and discrimination performance

3.1.2.2.1 'Goodness' influencing encoding and memory generation, similarity influencing comparison

Garner and colleagues built on the earlier Gestalt psychological literature and found support for the effects of pattern 'goodness' on perceptual discrimination (Clement & Varnadoe, 1967; Garner, 1974; Pomerantz, 1977; Pomerantz & Garner, 1973), recognition memory, paired-associates learning, and verbal encoding of stimulus patterns (Garner, 1974). They proposed that discrimination difficulty (operationalized as total discrimination time) depended on three different components or subprocesses: (a) the ease of stimulus encoding of the individual stimuli; (b) the ease of memory generation for the stimulus presented first; and (c) the ease of stimulus comparison of the two alternatives (Garner, 1974). They further argued that the required encoding time and time for memory generation depended on pattern goodness, whereas comparison time depended on stimulus similarity.

Also some more recent observations provide evidence in the direction of a link between goodness and ease of encoding (Irwin & Pachella, 1985). Caddigan et al. (2017) found that the degree to which an image exemplifies its category influences how easily it is detected. Rauschenberger & Yantis (2006) indicate that in visual search tasks, perceptual 'goodness' (i.e., Prägnanz) of the non-target stimuli determines the efficiency of visual search.

3.1.2.2.2 Reference points provide categories with an internal structure

Also Rosch's (1975) work on cognitive reference points built on the Gestalt psychological literature described above. As she indicates, categories are often not clearly delineated, but rather build around prototypes (i.e., clearest cases, best examples), with non-prototype category members tending to the prototype to a certain extent, which may lead to systematic asymmetries in perceived similarity (Tversky, 1977).

3.1.2.2.3 Asymmetric similarity can be explained by differential stimulus bias

Nosofsky (1991), following up on the work by Garner and colleagues (1974), proposed that asymmetric similarity results can alternatively be interpreted as symmetric similarity results together with differential stimulus bias. Whereas similarity relates to a relation between two stimuli, bias relates to individual stimuli. More concretely, if the bias associated with item a is larger than the bias associated with item b, item b will tend to be confused more with item a than item a will be confused with item b (Nosofsky, 1991). Some stimuli are particularly salient in perception and memory, and are more easily encoded than other stimuli. These characteristics relate to individual stimuli rather than pairs of stimuli, and can therefore better be described as stimulus biases than as asymmetric similarities. In other words, there may be prior biases to perceive or remember certain stimuli, independent of the stimulus that is actually presented, and these stimulus biases may consequently influence perceived similarity (Nosofsky, 1991).

3.1.3 The current study

In this study, we investigated how reference points influence categorization and discrimination performance, using three different tasks: (a) a categorization task, (b) a discrimination task, and (c) a similarity judgment task. For the morph series and tasks used, we define the extremes on the dimension as the logical reference points (i.e., 'clear' exemplars that serve as a point of comparison, in relation to which other exemplars on the dimension are perceived). In the categorization task, we assessed whether distance from the reference points influenced ease of classification in both categorization responses and reaction time. In the discrimination and similarity judgment tasks, we investigated whether we could replicate the category boundary effect as well as the expected asymmetries in within-category pairs. To be able to approximately equalize the physical distance of stimuli along a continuum between two category reference points, we used morph images as stimuli (i.e., stimuli created by morphing between exemplars of two different categories, see for example Figure 3.1). As we assumed both the category boundary effect and the directional asymmetries in within-category pairs to be consequences of a tendency towards reference points, the strength of these effects would depend on the strength of the reference points in question. Therefore, we used both recognizable and non-recognizable morph figures as stimuli, as we assumed the recognizable morph series to have better ('clearer') reference points: Whereas recognizable morph series have clear known reference points, morphs between two artificial stimuli can only be learned gradually by building experience with the stimuli. A categorization task was included to evaluate the categorization strength for each of the different stimuli presented. As will become clear from the results, we show that the category boundary effect has little to do with the relation between stimuli in a pair, but is rather a consequence of stimulus bias, as already indicated by Garner (1974) and Nosofsky (1991), for instance. The general 'category boundary effect' only emerged because of averaging across all physically equidistant stimulus pairs without taking each stimulus' distance from the reference points into account. We provide evidence that, in this study, discrimination performance and perceived similarity were - besides the influence of physical (dis)similarities - largely a consequence of categorization strength (related to a tendency towards the reference points) of the individual stimuli in the pair.

3.2 Methods

3.2.1 Transparency and openness

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study. The study design and analyses were not pre-registered. This manuscript was written in RMarkdown using the papaja package (Aust & Barth, 2022) with R code for data analysis integrated into the text. The data, materials, and analysis and manuscript code for the experiment are available on the Open Science Framework (https://doi.org/10.17605/osf.io/ugcd8).

3.2.2 Participants

283 first-year psychology students from KU Leuven participated in the study. The number of participants was dependent on the number of students subscribing and actually showing up to take part in the study in the prespecified time slots for the study planned between November 12 - 16, 2018. Participants were granted one research credit for participation. Of those 283 participants, 250 (88.34%) were female. Age of the participants varied between 17 and 23 years (M_{age} = 18.22 years, SD_{age} = 0.86 years). Age information concerning one participant is missing because of technical issues during data collection. For 272 of the 283 participants (96.11%), Dutch was their mother tongue. The study received ethical approval from the Social and Societal Ethics Committee of the authors' institution (G-2018 06 1266).

3.2.3 Stimuli

Stimuli (see Figure 3.3) were recognizable and non-recognizable morph series containing 11 stimuli each. The recognizable morph series were based on the ones used in Hartendorp et al. (2010) and Burnett & Jellema (2013). The non-recognizable morph series were based on stimuli from Op de Beeck et al. (2003b). We define the "morph level" of the eleven stimuli per series from -5 to 5, with morph level -5 indicating the most left stimulus in the morph series and morph level 5 indicating the most right stimulus in the morph series as presented in Figure 3.3. For the purpose of the study, all recognizable stimuli were converted to greyscale and changed to have a square format. All recognizable and non-recognizable stimuli were made black (0) on a grey (211) background. Images for each morph series were resized based on the mean number of non-grey pixels in that series. Then, images were cropped and squared to the same size for all images. Finally, physical similarity between neighboring stimuli was calculated. For the non-recognizable morph series, the final physical difference between neighboring stimuli in the morph series used was bigger than in the study by Op de Beeck et al. (2003b). This was done to approximately match the physical similarity between neighboring stimuli in the non-recognizable morph series with the physical similarity between neighboring stimuli in the recognizable morph series. Also, the starting amplitude for the non-recognizable stimuli was sometimes changed compared to the one used in Op de Beeck et al. (2003b) to avoid "holes" in the stimuli at the end of the series. The non-recognizable stimuli were generated in Matlab R2018a. Both the recognizable and non-recognizable stimuli were adapted (as described above) using Python 2.7.

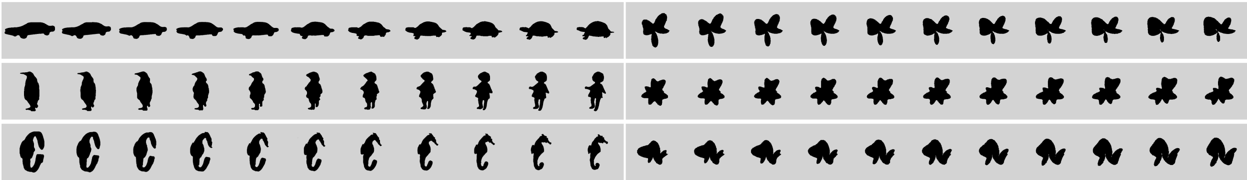


Figure 3.3: Recognizable and non-recognizable morph series used in the experiment. Left: recognizable morph series (car to tortoise, penguin to child, and watch to seahorse). Right: non-recognizable morph series 1, 2, and 3.

3.2.4 Procedure

The experiment was written in Python 2.7 and run on Windows computers with TFT screens of 21.5". Screen luminance and contrast were both set to 65%. Students participated in a room with up to 23 other participants present. After giving informed consent, participants were asked for their participant number, gender, age, and mother tongue. Each participant then completed each main task of the study (categorization, discrimination, and similarity judgment task) twice: once for a recognizable and once for a non-recognizable morph series. Every participant completed one task per morph series: all three tasks were completed for a different recognizable and non-recognizable morph series. The assignment of morph series to tasks was counterbalanced between participants.[5] The order of the three tasks and the order of recognizable versus non-recognizable series was randomized across participants. Before the start of each new task, participants got instruction screens explaining what was expected from them during that task, and they got four example trials with a different morph series (see Figure 3.4). In all tasks, the exact position of the stimuli on the screen was jittered (from -20 to 20 on both x and y axes) to prevent focus on local feature changes only.

Figure 3.4: Non-recognizable morph series used in the example trials.

At the start of the categorization task, participants were shown the four most 'clear' exemplars per category (i.e., the four most left and four most right examples for each morph series of eleven exemplars as shown in Figure 3.3), to give them an idea of the categories. Within each category, the presentation order of the four exemplars was randomized. Each trial in the categorization task consisted of (a) the presentation of a fixation cross (400 ms); (b) the presentation of the stimulus (300 ms); (c) a response screen reminding participants to press the left arrow key for category A and the right arrow key for category B. Which end of the series was labeled as category A was randomized across participants. Each stimulus in the morph series was presented 5 times (55 trials in total), and presentation order was randomized.

Each trial in the discrimination task consisted of (a) the presentation of a fixation cross (400 ms); (b) the presentation of a first stimulus (300 ms); (c) intertrial interval (500 ms); (d) the presentation of a second stimulus (300 ms); and (e) a response screen reminding participants to press the left arrow key for same (different) and the right arrow key for different (same). Which key press was related to same or different was randomized across

[5] To start the experiment, participants typed in a counterbalancing number assigned to their seat in the room (numbers were redistributed after each session). This number determined the assignment of the morph series to the different tasks.

participants. All possible different trials were presented once in each direction, all possible same trials were presented 5 times (165 trials in total), and presentation order was randomized.

Each trial in the similarity judgment task consisted of (a) the presentation of a fixation cross (400 ms); (b) the presentation of a first stimulus (300 ms); (c) intertrial interval (500 ms); (d) the presentation of a second stimulus (300 ms); and (e) a response screen including a 9-point rating scale on which participants indicated how strongly the two figures resembled each other, going from 1 (*very different*) to 9 (*very similar*). All possible different trials were presented once in each direction, all possible same trials were presented twice (132 trials in total), and presentation order was randomized.

At the end of the study, participants were asked to indicate whether they wanted to get a debriefing email and/or an email concerning the results of the study.

3.2.5 Data analysis

We used R [Version 4.0.4; R Core Team (2021)] for all our analyses.[6] Preprocessing included (a) combining datafiles per participant into datafiles for all participants combined; (b) pseudonymizing the data; (c) recoding key press responses in the categorization task when the series was reversed; (d) recoding key press responses in the discrimination task when the same and different buttons were reversed; and (e) adding additional variables to the categorization, discrimination, and similarity judgment data for simplifying further analyses. We excluded categorization data from participants for a particular morph series when their probability to label the stimulus as category B was higher or equal for level 0 compared to level 10, assuming that they used the response labels opposite to how it was instructed. When analyzing categorization response times, we excluded response times below 200 ms and above 3 seconds.

We excluded discrimination data concerning a specific participant and morph series when the mean accuracy for the given participant and morph series was more than two standard deviations below the overall mean accuracy for that morph series. When analyzing discrimination response times, we excluded response times below 200 ms and above 3 seconds.

Similarity ratings were standardized per participant per morph series to decrease the impact of differential scale use across participants.

To investigate our research questions, we fit hierarchical regression models using the brms (Bürkner, 2017) package in R (see below as well as Appendix A for more information). In all these models, we chose to estimate the effects for each of the morph series separately rather than to directly distinguish the groups of recognizable and non-recognizable morph series. This separation allows a more complete investigation of the effects per morph series, and similar results across several morph series of the same type (recognizable vs. non-recognizable) can be treated as replications. With only three morph series per type, it is difficult to generalize the results to the type of series as such (recognizable vs. non-recognizable), and therefore we prefer this small multiples approach.

[6] For an overview of all R packages used, see Appendix A.

3.2.5.1 Categorization responses

In the categorization data, we were interested in whether there was a clear categorization present for all morph series, and whether the categorization strength was stronger for recognizable than for non-recognizable morph series. We fitted a hierarchical Bayesian binomial logistic regression model to the categorization response data with morph level as fixed effect, for each morph series separately, and participant ID within each morph series as random effect for both intercept and slope. To investigate whether categorization strength was stronger for the recognizable than for the non-recognizable morph series, we plot the posterior distributions for the effect of morph level from the model and compare the slope strengths across morph series by plotting contrast distributions for the slopes.

Additionally, we expected categorization response times for exemplars closer to the reference points to be faster than those for more ambiguous exemplars. We therefore fitted a hierarchical Bayesian lognormal regression model to the categorization response time data with morph level and morph level squared as fixed effects, for each morph series separately, and participant ID within each morph series as random effect for both intercept and slopes. To investigate whether the difference in categorization time depending on morph level was larger for the recognizable than for the non-recognizable morph series, we plot the posterior distributions for the quadratic term from the model and compare the slope strengths across morph series by plotting the contrast distributions.

3.2.5.2 Discrimination responses

To investigate the presence of differences in discrimination sensitivity across stimulus pairs, we fitted a hierarchical Bayesian binomial logistic regression model to the discrimination response data with stepsize as fixed effect, for each morph series separately, and with trial stimuli and participant ID within each morph series as random effects for intercept and participant ID within each morph series as random effect for the slope. We plot the posterior predictions from the model for stepsizes 1 to 5 and separately for within- and between-category pairs. To investigate the presence of category boundary effects, we plot the difference in the expectation of the posterior predictive distributions for between-category compared to within-category pairs, per stepsize and morph series. The category boundary was determined as the middle between the two morph levels where the point of subjective equality was crossed, based on the categorization data collected in the study. To investigate whether the effect of stepsize on discrimination sensitivity was stronger for the recognizable than for the non-recognizable morph series, we plot the posterior distributions for the effect of stepsize from the model and compare the slope strengths across morph series by plotting contrast distributions for the slopes. To further investigate other trends in the data, we also plot the posterior predictions from the model for each stimulus pair separately.

To further investigate the presence of an overall category boundary effect (i.e., difference in discrimination sensitivity for equidistant between-category versus within-category pairs), we fitted a hierarchical Bayesian binomial logistic regression model to the discrimination response data for stepsizes 1 to 5 (i.e., the stepsizes which had both within- and between-category pairs) with stepsize as fixed effect, for each morph series and trial type separately, and with trial stimuli and participant ID within each morph series as random effects for intercept and participant ID within each morph series as random effect for the slope. To investigate whether the category boundary effect on discrimination sensitivity was stronger for the recognizable than for the non-recognizable morph series, we plot the posterior distributions for the main effect of trial type as well as for the interaction between stepsize and trial type from the model and compare the slope strengths across morph series by plotting contrast distributions for the slopes.

To investigate the presence of directional asymmetries, we fitted a hierarchical Bayesian binomial logistic regression model to the discrimination response data with stepsize as fixed effect, for each morph series separately, and with ordered trial stimuli and participant ID within each morph series as random effects for intercept and participant ID within each morph series as random effect for the slope. We then plot the posterior predictions from the model for each ordered stimulus pair separately.

3.2.5.3 Similarity judgments

In the similarity judgment data, we also wanted to investigate potential differences in perceived similarity across stimulus pairs, whether an overall category boundary effect was present, and whether directional asymmetries emerged. To investigate the presence of differences in perceived similarity across stimulus pairs, we fitted a hierarchical Bayesian linear regression model to the by-participant-standardized similarity judgments with stepsize as fixed effect, for each morph series separately, and with trial stimuli and participant ID within each morph series as random effects for intercept and participant ID within each morph series as random effect for the slope. We plot the posterior predictions from the model for stepsizes 1 to 5 and separately for within- and between-category pairs. To investigate the presence of category boundary effects, we plot the difference in the expectation of the posterior predictive distributions for between-category compared to within-category pairs, per stepsize and morph series. To investigate whether the effect of stepsize on perceived similarity was stronger for the recognizable than for the non-recognizable morph series, we plot the posterior distributions for the effect of stepsize from the model and compare the slope strengths across morph series by plotting contrast distributions for the slopes. To further investigate other trends in the data, we also plot the posterior predictions from the model for each stimulus pair separately.

To further investigate the presence of an overall category boundary effect (i.e., difference in perceived similarity for equidistant between-category versus within-category pairs), we fitted a hierarchical Bayesian linear regression model to the by-participant-standardized similarity judgments for stepsizes 1 to 5 (i.e., the stepsizes which had both within- and between-category pairs) with stepsize as fixed effect, for each morph series and trial type separately, and with trial stimuli and participant ID within each morph series as random effects for intercept and participant ID within each morph series as random effect for the slope. To investigate whether the category boundary effect on perceived similarity was stronger for the recognizable than for the non-recognizable morph series, we plot the posterior distributions for the main effect of trial type as well as for the interaction between stepsize and trial type from the model and compare the slope strengths across morph series by plotting contrast distributions for the slopes.

To investigate the presence of directional asymmetries, we fitted a hierarchical Bayesian linear model to the by-participant-standardized similarity judgments with stepsize as fixed effect, for each morph series separately, and with ordered trial stimuli and participant ID within each morph series as random effects for intercept and participant ID within each morph series as random effect for the slope. We then plot the posterior predictions from the model for each ordered stimulus pair separately.

3.3 Results

3.3.1 Categorization task

3.3.1.1 Categorization responses

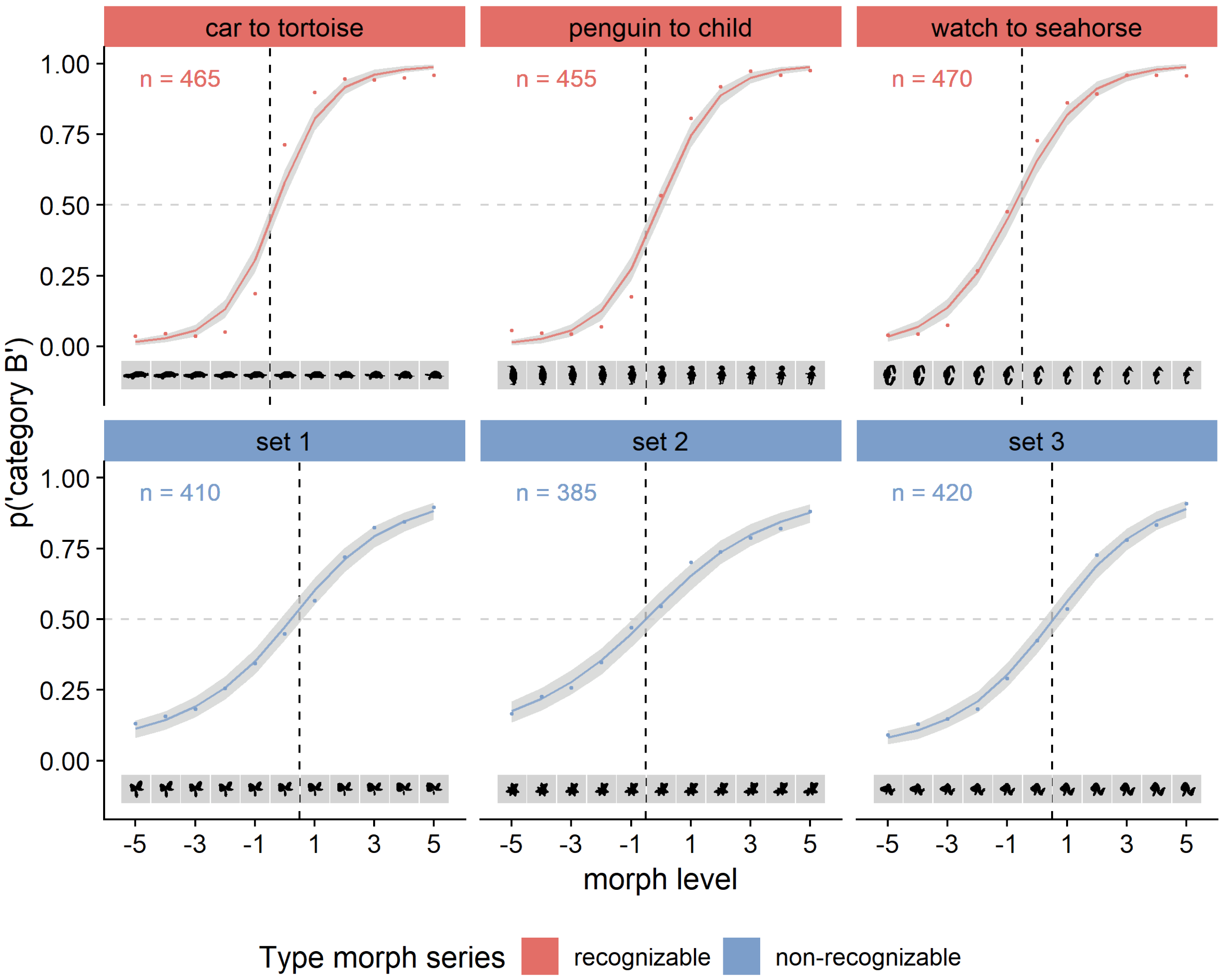


Figure 3.5: Proportion of 'category B' responses for each morph level and morph series separately, averaged across participants. Colored dots indicate the empirical proportions for responding 'category B'. Colored lines indicate the mean posterior predictions from the model and shaded regions indicate 95% highest density continuous intervals of the posterior predictive distributions. The black dashed vertical line indicates the category boundary as defined based on the empirical proportions. In this figure, the proportion 'category B' responses increases more steeply across morph levels for the recognizable than for the non-recognizable morph series. *Note.* In the interactive version of this figure, you can hover over the data points to see the related stimuli as well as the exact percentage category B responses and the number of trials related to each data point.

As can be seen in Figures 3.5 and 3.6, the recognizable morph series show a considerably larger effect of the morph level presented, and thus show stronger categorization. Estimated pairwise differences between the posterior distributions for the effect of morph level on categorization as category B show strong evidence for a larger effect of morph level in the recognizable morph series compared to the non-recognizable morph series (see Figure 3.6). The posterior probability for the recognizable morph series to have a more positive slope across morph level than the non-recognizable morph series was 100%, regardless of the specific morph series compared. In other words, categorization (and the tendency towards reference points) was stronger for morph series with clear, pre-existing reference points (i.e., the recognizable ones) compared to morph series without clear reference points (i.e., the non-recognizable ones). Based on the estimated posterior distributions for categorization responses, categorization strength was highest for the recognizable morph series car-tortoise and penguin-child and lowest for the non-recognizable morph series set 2 (see Figure 3.6).

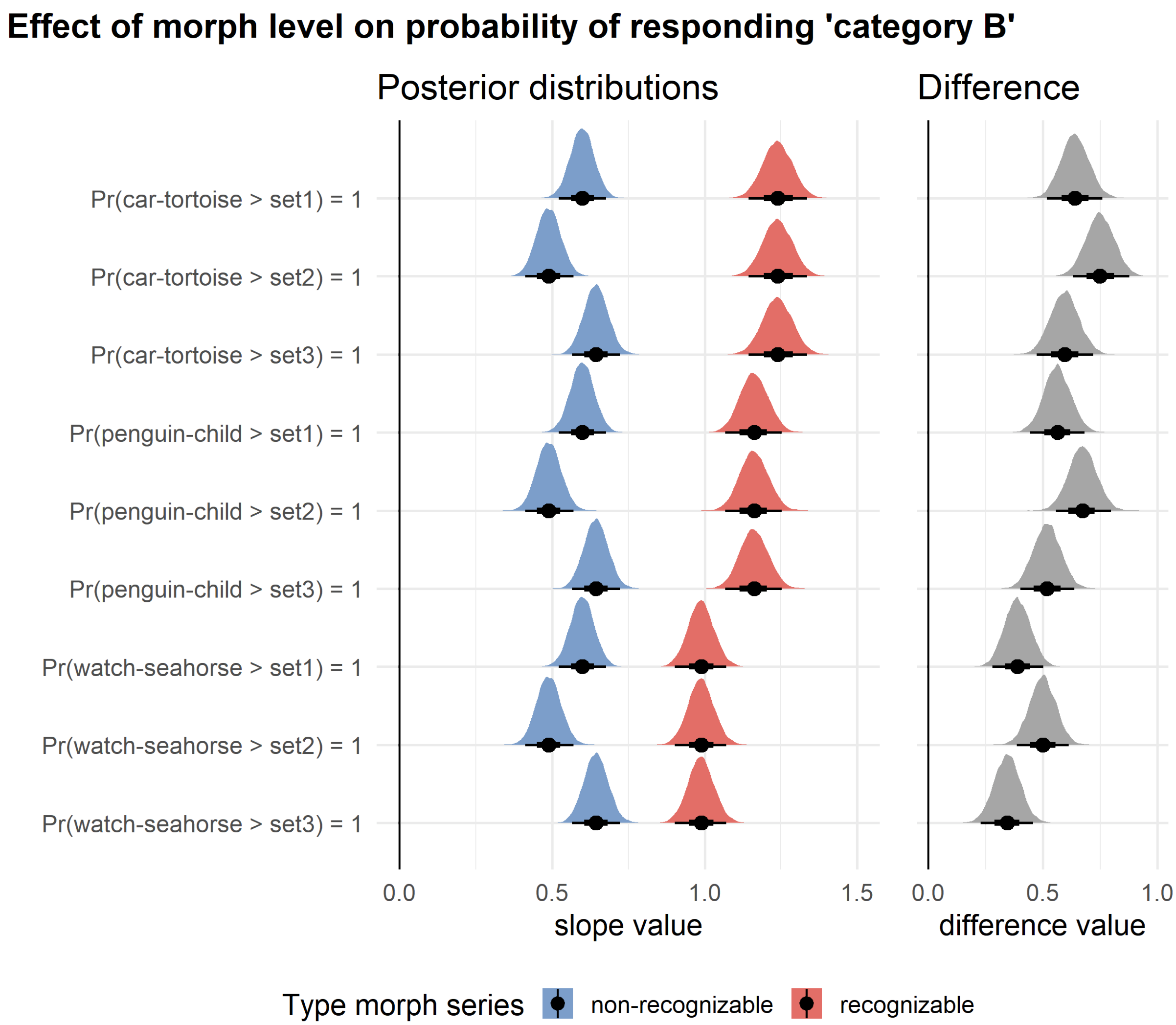


Figure 3.6: Estimated pairwise differences between the posterior distributions for the effect of morph level on the categorization responses for each of the different recognizable and non-recognizable morph series combinations, in logodds units. Black dots and intervals represent the mean, 66%, and 95% highest density continuous interval (HDCI) for each slope or difference value. The black vertical line indicates a slope or difference in slope of zero. This figure confirms that all morph series show a positive slope across morph levels, and that the effect of morph level is larger for each of the recognizable morph series than for the non-recognizable morph series. *Note.* In the interactive version of this figure, you can hover over the intervals to see the related mean and 95% HDCI for each distribution.

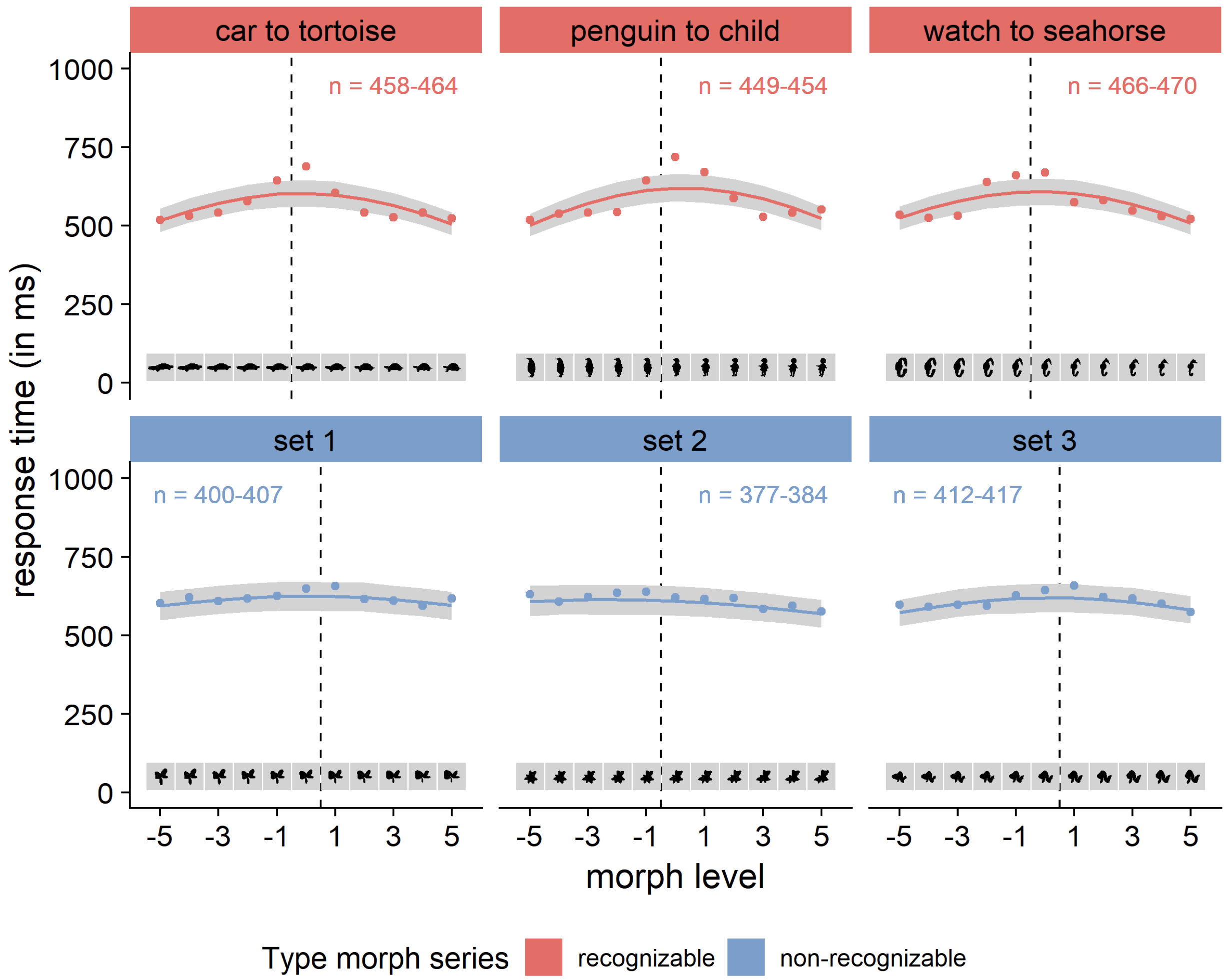


Figure 3.7: Categorization response times for each morph level and morph series separately, averaged across participants. Colored dots indicate the empirical mean response times in the categorization task. Colored lines indicate the mean posterior predictions from the model and shaded regions indicate 95% highest density continuous intervals of the posterior predictive distributions. The black dashed vertical line indicates the category boundary as defined based on the empirical categorization proportions. In this figure, the recognizable series show stronger differences in mean response time across morph levels than the non-recognizable morph series. *Note.* In the interactive version of this figure, you can hover over the data points to see the related stimuli as well as the exact mean response time and the number of trials related to each data point.

3.3.1.2 Categorization response times

As visible from Figures 3.7 and 3.8, the recognizable morph series show a considerably larger effect of the morph level presented on the categorization response times than the non-recognizable morph series: In the recognizable series, response times are faster for exemplars closer to the reference points than for more ambiguous exemplars. This fits with earlier findings relating 'goodness' to ease of encoding and categorization. Estimated pairwise differences between the posterior distributions for the quadratic effect of morph level on categorization response times show strong evidence for a larger quadratic effect of morph level (i.e., more negative, indicating smaller response times for the morph levels closer to the reference points than for the middle morph levels) in the recognizable morph series compared to the non-recognizable morph series (see Figure 3.8). The posterior probability for the non-recognizable morph series to have a quadratic effect across morph level closer to zero than the recognizable morph series was 100%, regardless of the specific morph series compared. In other words, also in the response times, categorization (and the tendency towards reference points) was stronger for morph series with clear, pre-existing reference points than for morph series without clear reference points. Based on the estimated posterior distributions for categorization response time, categorization strength was highest for the recognizable morph series penguin-child and car-tortoise and lowest for the non-recognizable morph series set 2 and set 1 (see Figure 3.8).

3.3.1.3 Definition of category boundary

The category boundary was defined as the middle between the two morph levels where the 50%-boundary (i.e., point of subjective equivalence) was crossed. Although this is an arbitrary decision, we only depend on it when plotting the typical category boundary effect in the successive discrimination and similarity judgment tasks (i.e., adding all physically equidistant within-category pairs and between-category pairs together), not when plotting the results per stimulus pair. The category boundaries used are indicated in Figures 3.5 and 3.7.

3.3.2 Successive discrimination task

3.3.2.1 Discrimination responses

Figure 3.9 shows the empirical proportions and posterior predictive distributions for responding 'different' in the successive discrimination task, per stepsize, trial type (i.e., between-category vs. within-category), and morph series, averaged across participants. The posterior predictive results shown are based on a model fit using the discrimination data of all stepsizes (0-11), with stepsize as a predictor, a separate intercept and slope per morph series, and random effects of stimulus pair and participant within each morph series (see Methods section and Appendix A). Stepsize indicates the absolute difference in morph level between the two morph stimuli presented in a trial, with a minimum of zero (for same trials) and a maximum of eleven (when both extremes or reference points of the morph series are presented). Only stepsizes 1 to 5 are shown, as higher stepsizes only included between-category pairs and no within-category pairs, and stepsize 0 only included same trials (for a visualization including all stepsizes, see Supplementary Figure A.1).

Despite approximately matching the physical similarity between neighboring stimuli in the non-recognizable morph series with the physical similarity between neighboring stimuli in the recognizable morph series, the effect of stepsize on the probability of 'different' responses was lower for the non-recognizable compared to the recognizable

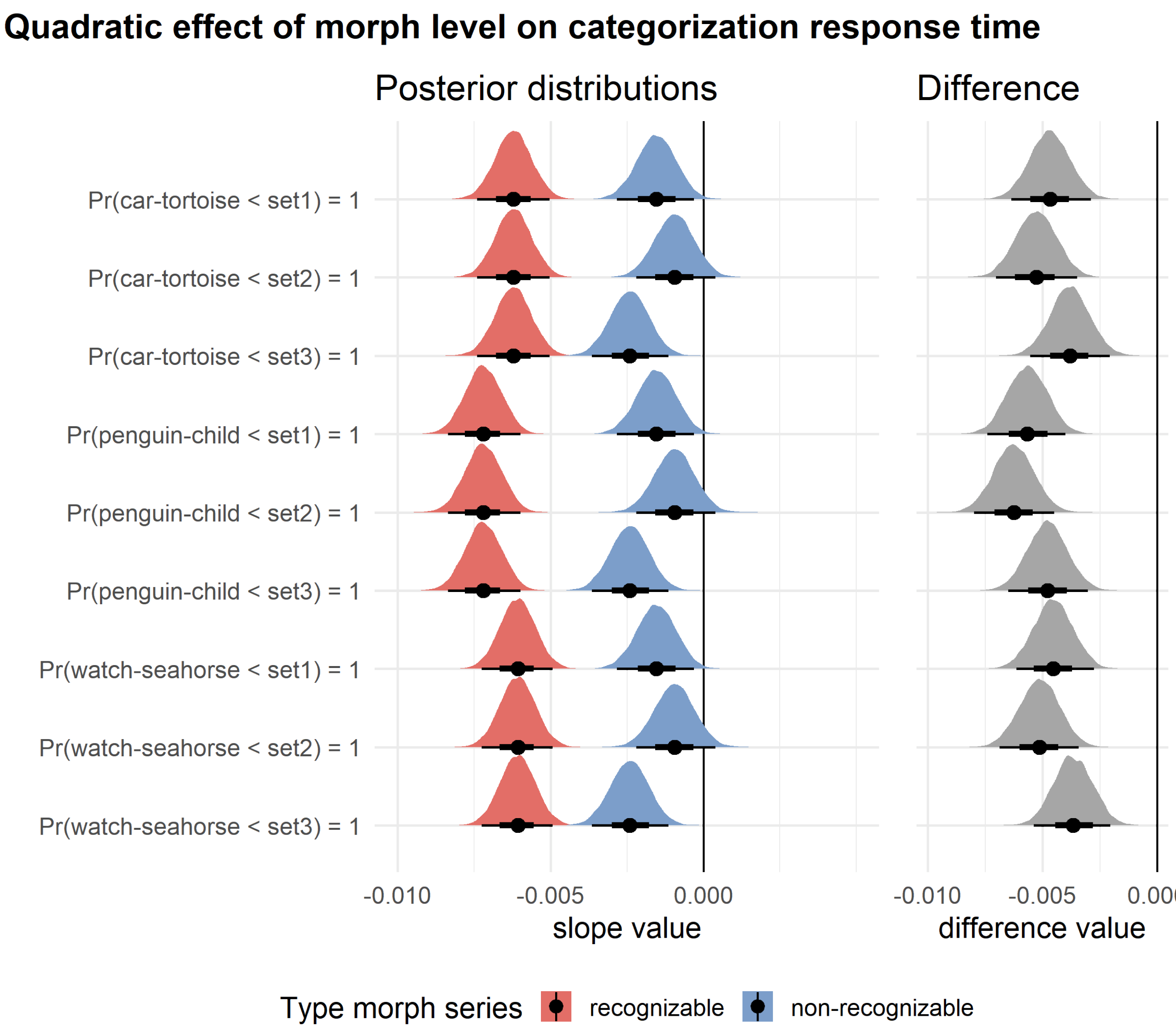


Figure 3.8: Estimated pairwise differences between the posterior distributions for the quadratic effect of morph level on categorization response times for each of the different recognizable and non-recognizable morph series combinations. Black dots and intervals represent the mean, 66%, and 95% highest density continuous interval (HDCI) for each slope or difference value. The black vertical line indicates a slope or difference in slope of zero. This figure confirms that all recognizable morph series show a quadratic effect of morph level different from zero, and that the quadratic effect of morph level is larger (i.e., more negative, indicating slower response times for the middle morph levels than for the extremes) for each of the recognizable morph series than for the non-recognizable morph series. *Note.* In the interactive version of this figure, you can hover over the intervals to see the related mean and 95% HDCI for each distribution.

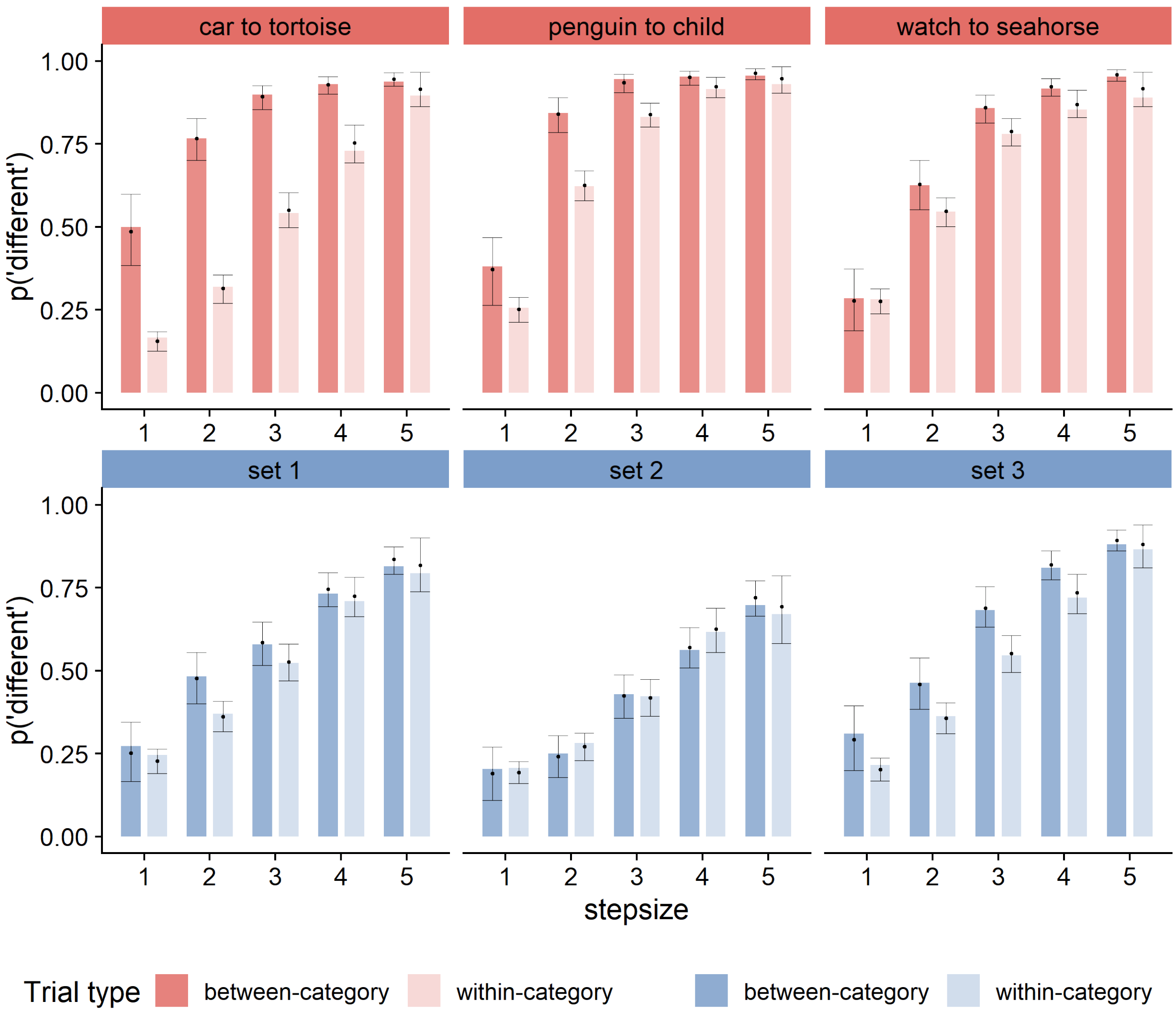


Figure 3.9: Proportion of 'different' responses in the successive discrimination task for each stepsize, trial type (i.e., between-category vs. within-category), and morph series separately, averaged across participants. Bars indicate the empirical proportions for responding 'different'. The black dots indicate the mean posterior predictions from the model and the error bars indicate the 95% highest density continuous intervals (HDCI) of the posterior predictive distributions. In this figure, the difference between the darker and the lighter bars (i.e., the category boundary effect: more 'different' responses for between-category compared to within-category pairs, keeping stepsize equal) is on average higher for the recognizable than for the non-recognizable morph series. *Note.* In the interactive version of this figure, you can hover over the colored bars to see the exact percentage different responses, the mean and 95% HDCI of the posterior predictive distributions, and the number of trials related to each bar.

morph series, regardless of the specific recognizable and non-recognizable morph series that were compared (see Figures A.2B and A.3). In other words, discrimination sensitivity for a fixed stepsize was higher for the recognizable compared to the non-recognizable morph series.

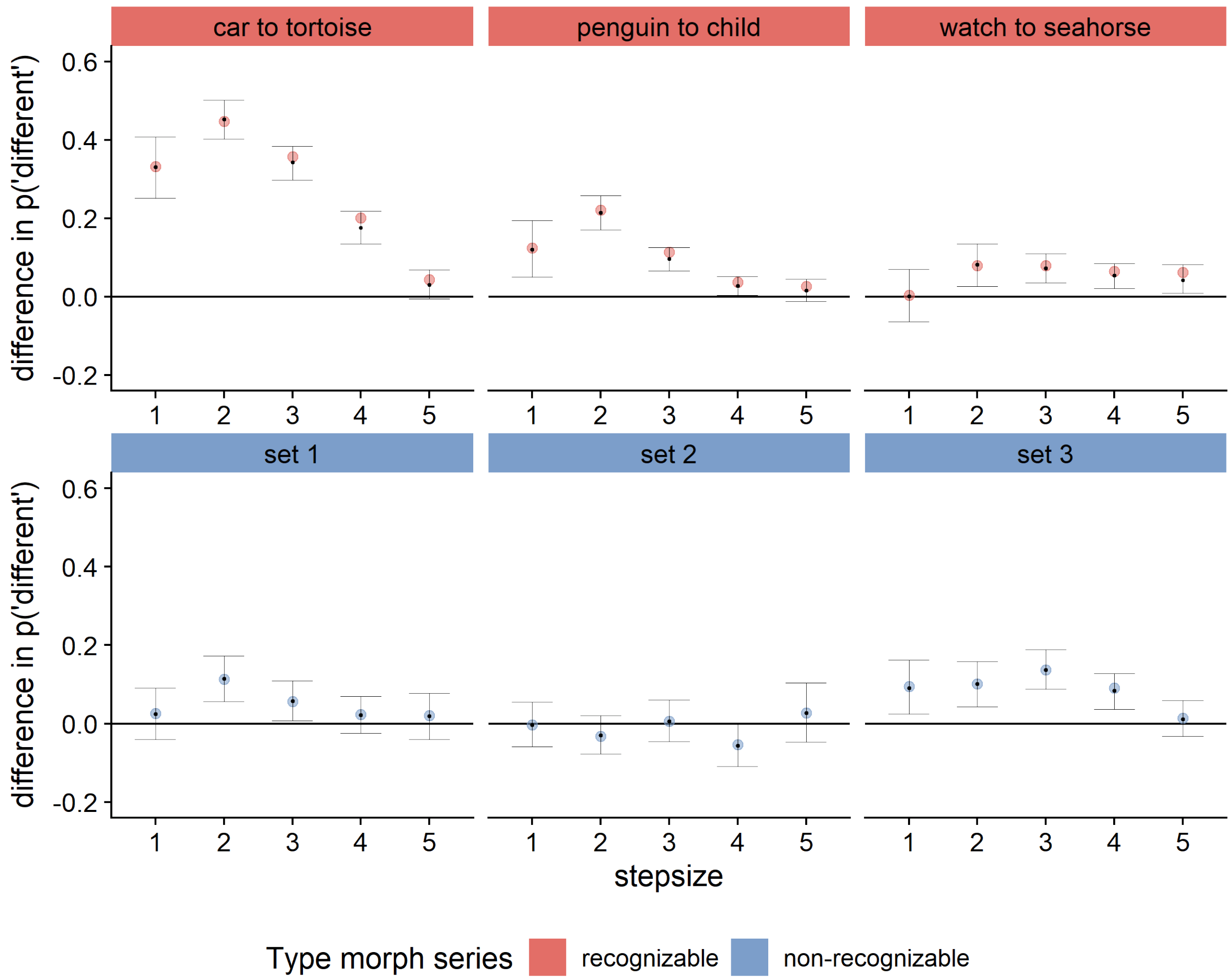


Figure 3.10: Difference in proportion of 'different' responses in the successive discrimination task when comparing between-category and within-category pairs, per stepsize and morph series, averaged across participants. Colored dots are the empirical differences in the proportion of 'different' responses. The black dots and intervals indicate the mean and the 95% highest density continuous intervals (HDCI) of the expected values for the posterior predictive distributions. In this figure, the category boundary effect (i.e., more 'different' responses for between-category compared to within-category pairs, keeping stepsize equal) is on average higher for the recognizable than for the non-recognizable morph series. *Note.* In the interactive version of this figure, you can hover over the data points to see the exact difference in percentage as well as the mean and 95% HDCI of the expected values for the posterior predictive distributions related to each data point.

In Figures 3.9 and 3.10, the typically reported category boundary effect (i.e., the difference between the darker and lighter bars per stepsize) seems larger in the recognizable morph series compared to the non-recognizable morph series. To estimate the size of the category boundary effect for each morph series, we fitted a separate model only including trials with stepsizes 1 to 5 and with trial type (i.e., within-category vs. between-category) and the interaction between trial type and stepsize as additional predictors (for model details, see the Methods section and Appendix A). In this separate model, all morph series except for set 2 on average showed a category boundary effect (see Figure A.4C and recognizable morph series car-tortoise and penguin-child showed a stronger main category boundary effect (i.e., effect of trial type) than all non-recognizable morph series (see Figure A.4C and Figure 3.11C). Recognizable series penguin-child showed a more negative interaction effect between stepsize and trial type than the non-recognizable series: for the series penguin-child the category boundary effect decreased more strongly with increasing stepsize than for the other morph series (see Figure 3.11D and Figure A.4D).

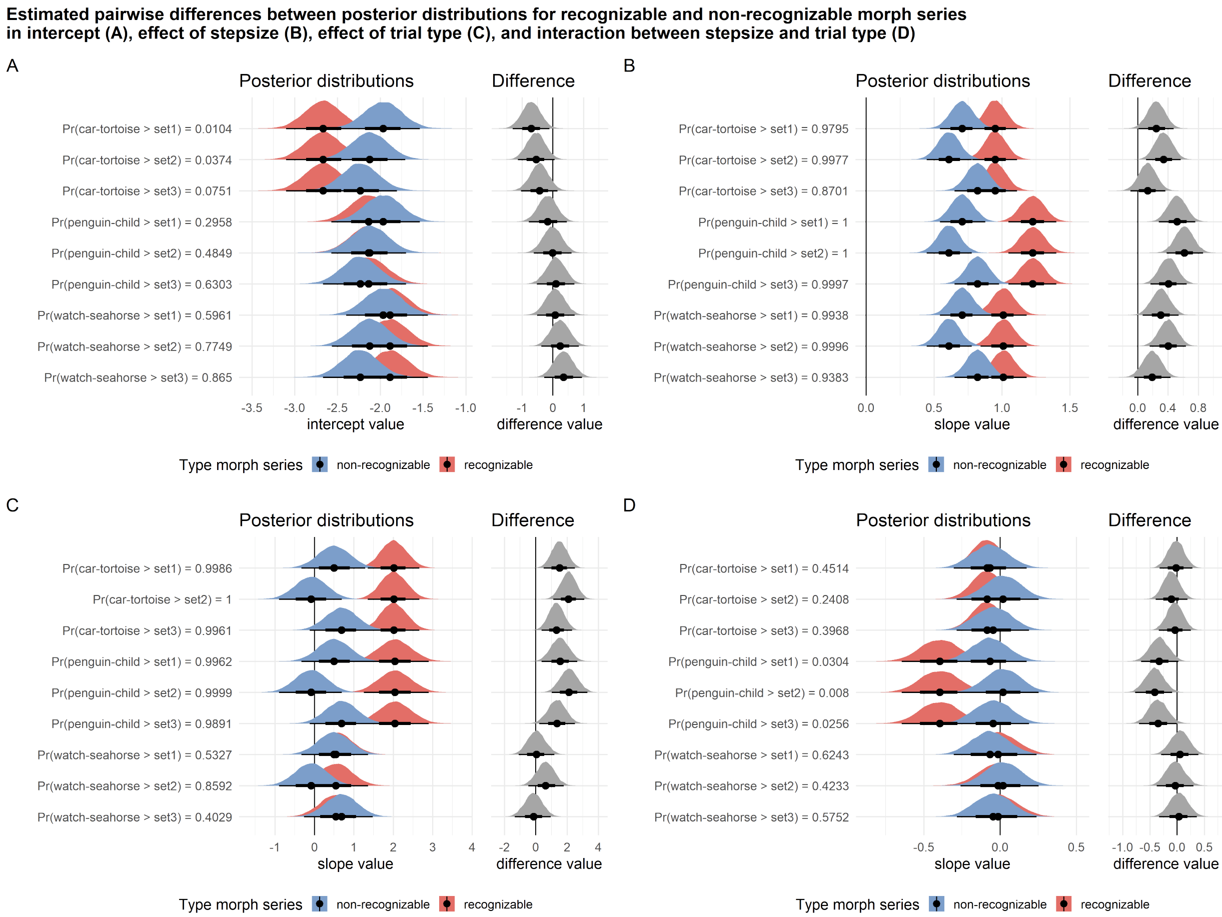


Figure 3.11: Estimated pairwise differences between the posterior distributions for the intercept (A), effect of stepsize (B), effect of trial type (within-category = 0, between-category = 1; C), and interaction between stepsize and trial type (D) on the probability of responding 'different' in the successive discrimination task, for each of the different recognizable and non-recognizable morph series combinations, in logodds units. Black dots and intervals indicate the mean, 66%, and 95% highest density continuous interval (HDCI) for each slope or difference value. The black vertical line indicates a difference in slope of zero. In this figure, the estimated effect of stepsize is larger for the recognizable than for the non-recognizable morph series (B). The main effect of trial type is larger for the recognizable series car-tortoise and penguin-child than for all non-recognizable morph series (C). The interaction effect between stepsize and trial type is more negative for the recognizable series penguin-child than for all non-recognizable series (D). *Note.* In the interactive version of this figure, you can hover over the intervals to see the related mean and 95% HDCI for each distribution.

The results for the discrimination task fit with the categorization results: the stronger the categorization for a morph series, the stronger the difference in discrimination performance between within- and between-category pairs (i.e., the main effect of trial type), or in other words, the stronger the category boundary effect.[7] This is the case even though for each morph series, the categorization and discrimination data were collected from different

[7] Categorization was strongest for recognizable morph series car-tortoise and penguin-child and weakest for non-recognizable morph series 2.

subgroups of participants, and none of the participants in the discrimination task had any prior experience with or information about the morph series in question.

What is remarkable, however, is what happens when plotting the data of the original model including all discrimination data for each stimulus pair within each stepsize separately (see Figure 3.12). A clear peaked pattern emerges: for physically equidistant stimulus pairs, the closer the stimuli in the pair are to one of the reference points, the worse discrimination performance becomes. Rather than a general effect of being a within- or between-category comparison, the effect is much more gradual and related to characteristics of the individual stimuli involved in the comparison. In other words, it is important to not treat the category boundary effect as a binary difference of within-category versus between-category pairs, but rather as a gradual effect of distance from the reference points for the individual stimuli in the pair. This is necessary as discrimination performance is heavily influenced by the clarity of the individual stimuli in the pair, as can be seen in Figure 3.12. For non-recognizable morph series, this peaked pattern is much less pronounced.

To further demonstrate the importance of the individual 'goodness' of the stimuli in the pair for discrimination performance, we plotted predictions of discrimination performance based solely on the predicted categorization probabilities of individual stimuli in a pair to be categorized as A or B (see Figure 3.13). For the recognizable morph series, these predictions show the same gradual, peaked pattern as can be viewed from the actual discrimination data, at least in a qualitative sense, indicating that the observed pattern is likely to be a consequence of properties (i.e., 'goodness') of the individual stimuli involved in the comparison rather than a consequence of stimulus similarity.

When modelling and plotting the probability of responding 'different' in the successive discrimination task for all stimulus pairs in both presentation orders separately, no clear directional asymmetries were present (see Figure A.5).

3.3.3 Similarity judgment task

3.3.3.1 Similarity judgment responses

Figure 3.14 shows the standardized empirical similarity scores and the posterior predictive distributions for the responses to the successive similarity judgment task per stepsize, trial type (i.e., between-category vs. within-category), and morph series, averaged across participants. The posterior predictive results shown are based on a model fit using the similarity judgment data of all stepsizes (0-11), with stepsize as a predictor, a separate intercept and slope per morph series, and random effects of stimulus pair and participant within each morph series (see Methods section and Appendix A). Only stepsizes 1 to 5 are shown, as higher stepsizes only included between-category pairs and no within-category pairs, and stepsize 0 only included same trials (for a visualization including all stepsizes, see Supplementary Figure A.6).

Similar to the results for the successive discrimination task, the effect of stepsize on the perceived similarity between the stimuli in a pair was smaller for the non-recognizable than for the recognizable morph series, despite approximately matching the physical similarity between neighboring stimuli across morph series (see Figures A.8

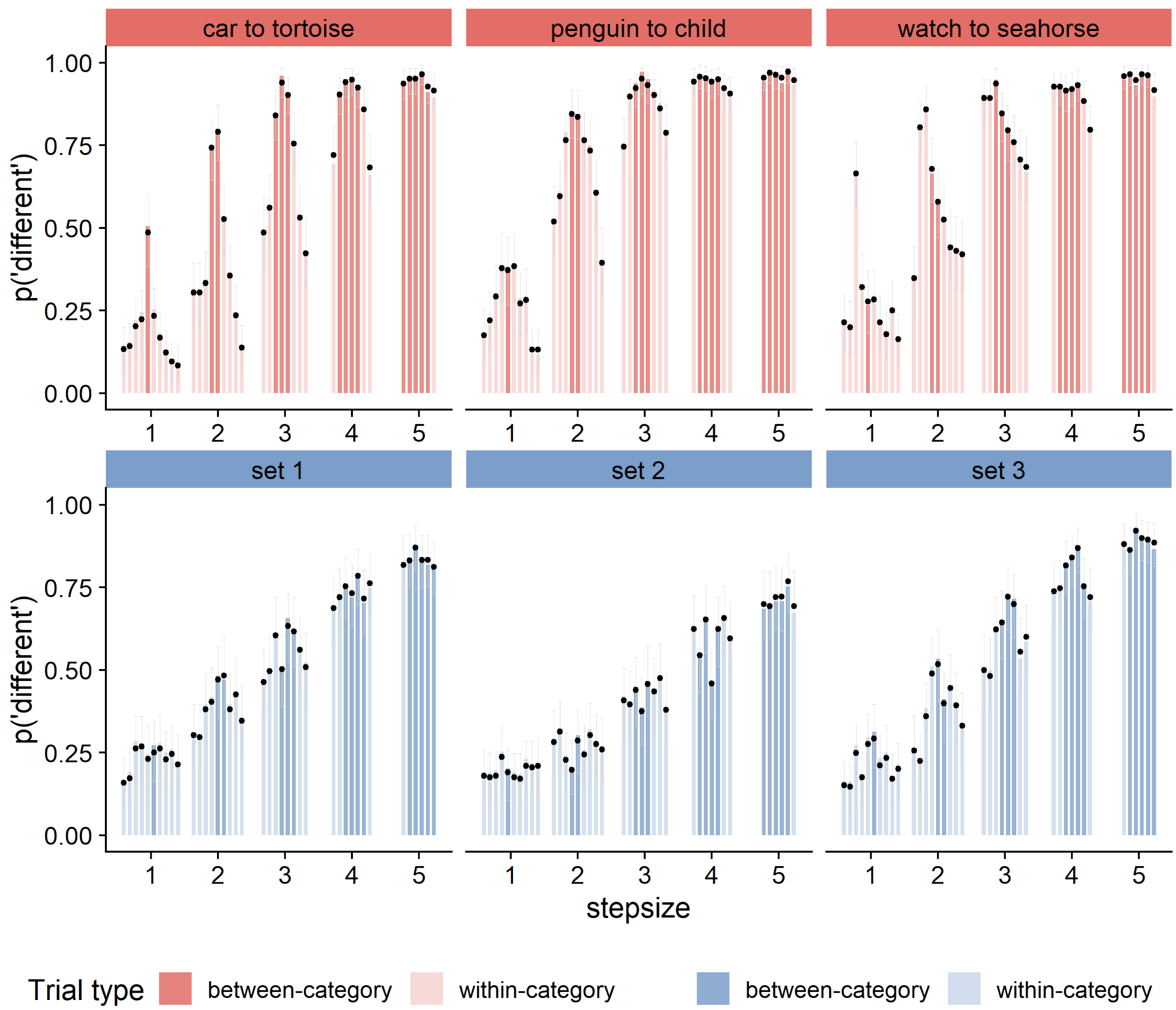


Figure 3.12: Proportion of 'different' responses in the successive discrimination task for each stepsize, stimulus pair, and morph series separately, averaged across participants and stimulus order within the pair. Stimulus pairs are ordered per stepsize and from left to right in the morph series as presented in Figure 3.3. Bars indicate the empirical proportions for responding 'different'. The black dots indicate the mean posterior predictions from the model and the grey error bars indicate the 95% highest density continuous intervals (HDCI) of the posterior predictive distributions. In this figure, a clear, gradual, 'peaked' pattern can be observed for the recognizable morph series, where pairs that include stimuli close to the reference points lead to a lower probability of responding 'different' than pairs that include stimuli further away from the reference points, while keeping the physical distance between the stimuli in the pair (i.e., stepsize) equal. This peaked pattern is much less pronounced for the non-recognizable morph series. *Note.* In the interactive version of this figure, you can hover over the colored bars to see the stimuli involved in the pair, the exact percentage different responses, the mean and 95% HDCI of the posterior predictive distributions, and the number of trials related to each bar.

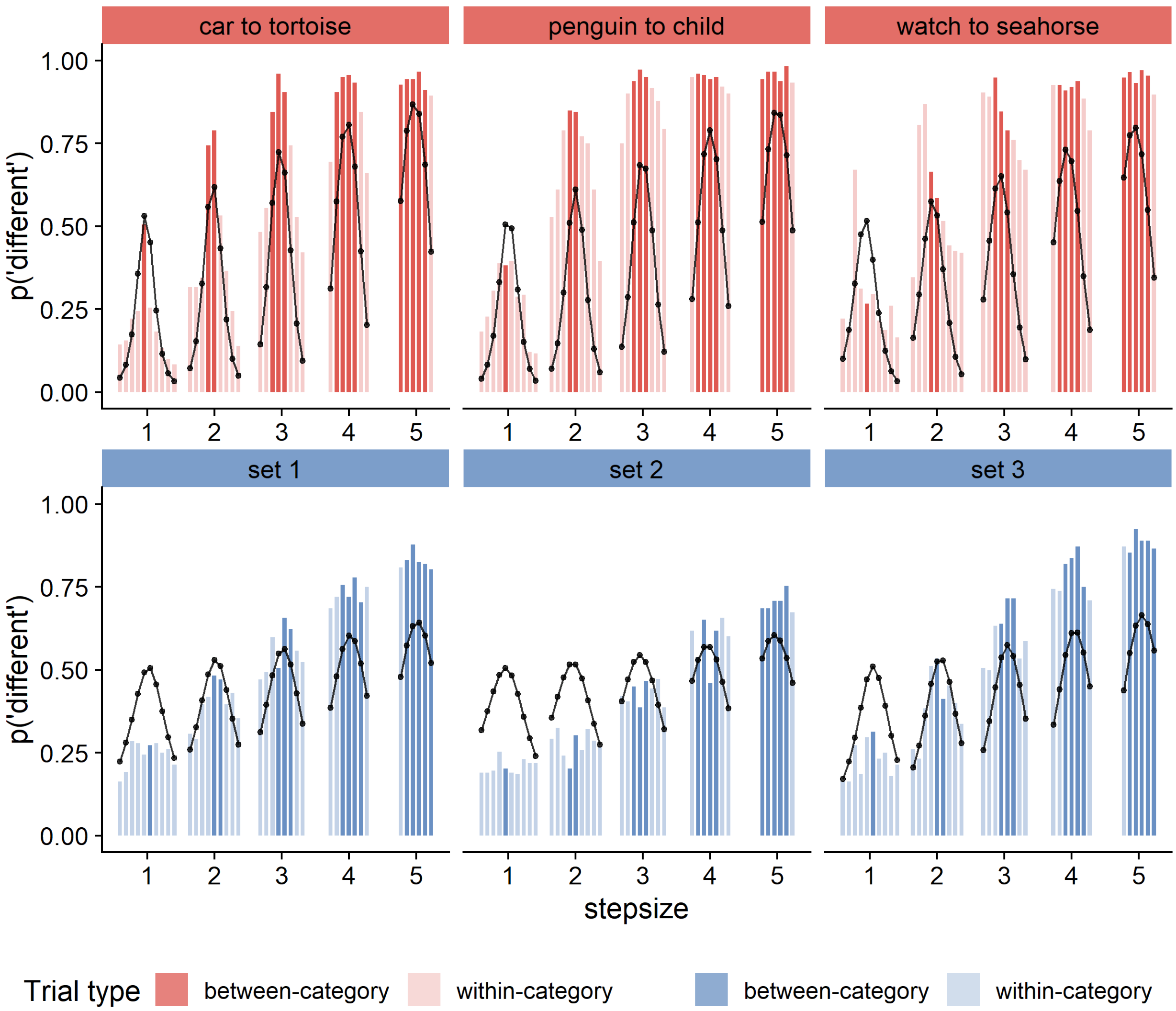


Figure 3.13: Prediction of probability of responding 'different' in the successive discrimination task based on categorization probabilities of individual stimuli in a pair in the categorization task. Bars are the empirical proportions for responding 'different' per stepsize, stimulus pair, and morph series, averaged across participants and stimulus order within the pair. The black dots and lines indicate the predictions based on the estimated mean categorization probabilities for the stimuli in a pair. *Note.* In the interactive version of this figure, you can hover over the colored bars to see the stimuli involved in the pair, the exact percentage different responses, the mean and 95% HDCI of the posterior predictive distributions, and the number of trials related to each bar.

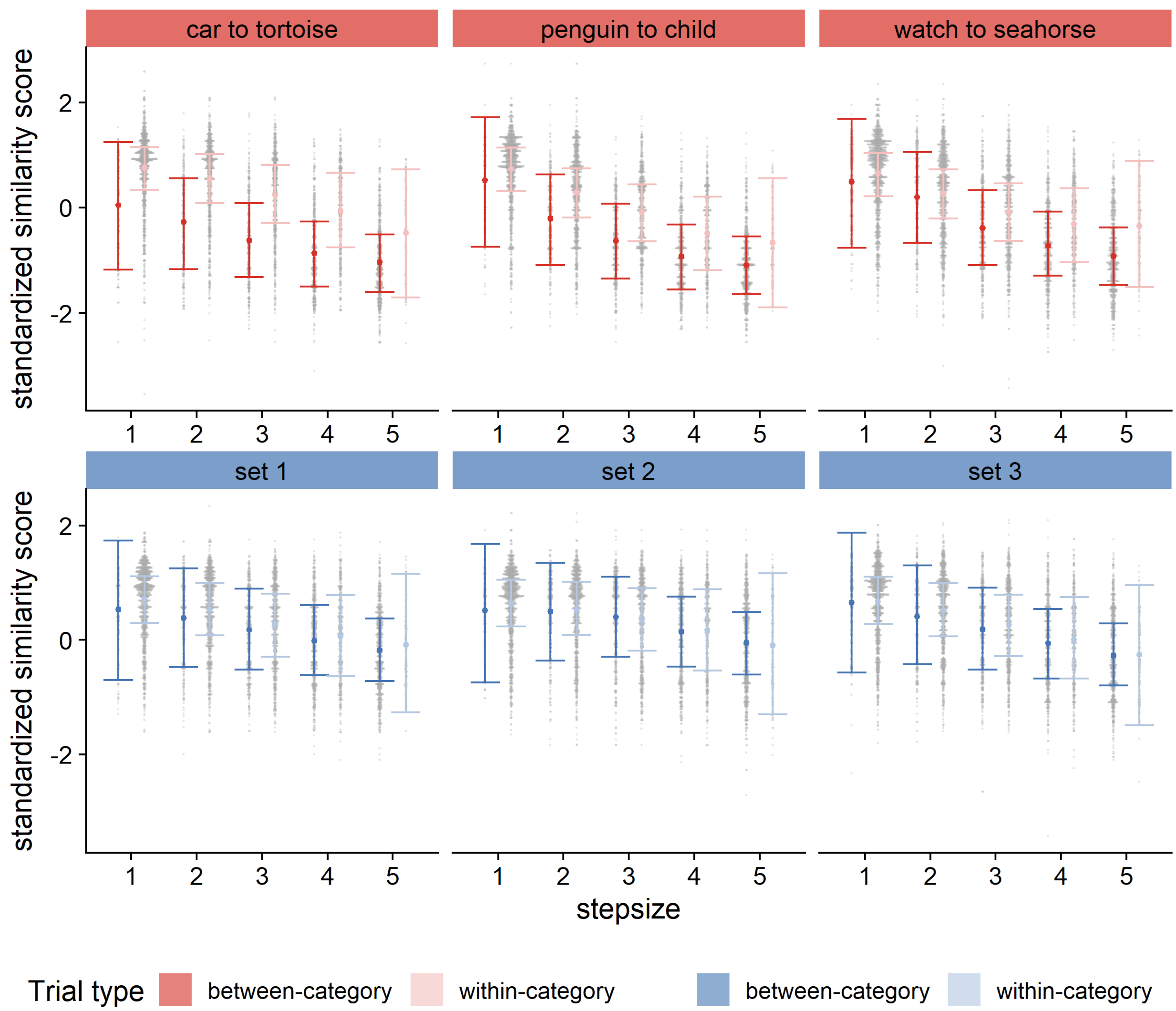


Figure 3.14: Standardized similarity scores for each stepsize, trial type (i.e., between-category vs. within-category), and morph series separately, averaged across participants. Grey dots indicate the raw standardized similarity scores. For the conditions that contain less trials, these grey dots are not always clearly visible. The colored dots and error bars indicate the mean posterior predictions from the model and the 95% highest density continuous interval (HDCI) of the posterior predictive distributions. In this figure, the difference between the darker and the lighter intervals (i.e., the category boundary effect: between-category pairs rated as less similar than within-category pairs, keeping stepsize equal) is on average larger for the recognizable than for the non-recognizable morph series. *Note.* In the interactive version of this figure, you can hover over the intervals to see the exact similarity score, the mean and 95% HDCI of the posterior predictive distributions, and the number of trials related to each interval.

and A.7B). In other words, perceived similarity for a fixed stepsize was lower for the recognizable compared to the non-recognizable morph series.

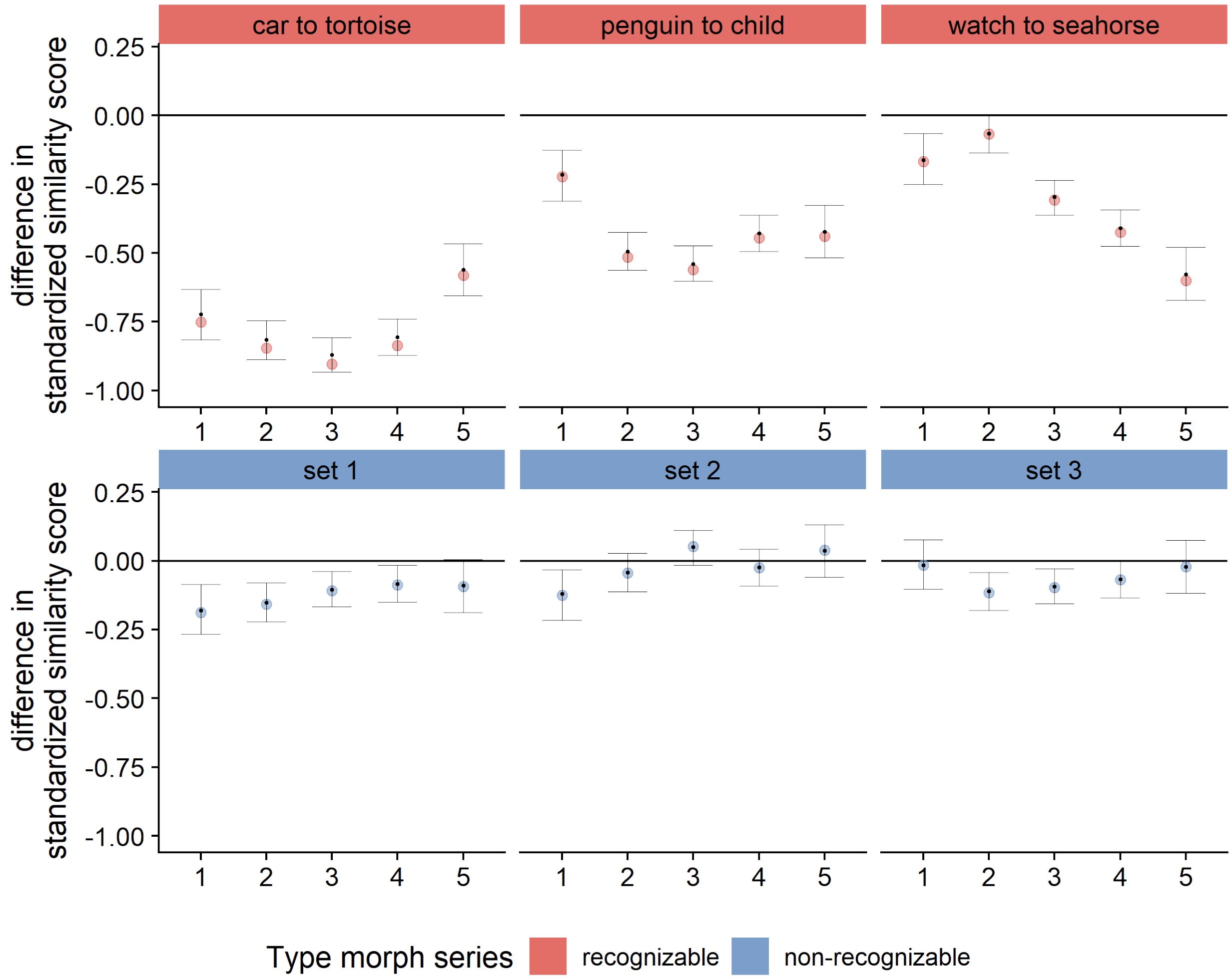


Figure 3.15: Difference in standardized similarity score when comparing between-category and within-category pairs, per stepsize and morph series, averaged across participants. Colored dots are the empirical differences in the standardized similarity scores. The black dots and intervals indicate the mean and the 95% highest density continuous intervals (HDCI) of the expected values for the posterior predictive distributions. In this figure, the category boundary effect (i.e., lower similarity scores for between-category compared to within-category pairs, keeping stepsize equal) is on average higher for the recognizable than for the non-recognizable morph series. *Note.* In the interactive version of this figure, you can hover over the data points to see the exact difference in similarity as well as the mean and 95% HDCI of the expected values for the posterior predictive distributions related to each data point.

In Figures 3.14 and 3.15, the category boundary effect in the similarity judgment task (i.e., the difference between the darker and lighter intervals per stepsize) seems larger for the recognizable morph series compared to the non-recognizable morph series. To estimate the size of the category boundary effect for each morph series, we fitted a separate model only including trials with stepsizes 1 to 5 and with trial type (i.e., within-category vs. between-category) and the interaction between trial type and stepsize as additional predictors (for model details, see the Methods section and Appendix A). In this separate model, all morph series on average showed a category boundary effect (albeit slightly; see Figure A.9C) and recognizable morph series car-tortoise and penguin-child showed a stronger main category boundary effect (i.e., effect of trial type) than all non-recognizable morph series (see Figures A.9C and 3.16C). Recognizable series watch-seahorse showed a more negative interaction effect between stepsize and trial type than the non-recognizable series: for watch-seahorse, the category boundary effect increased more strongly with increasing stepsize than for the non-recognizable series (see Figures 3.16D and A.9D).

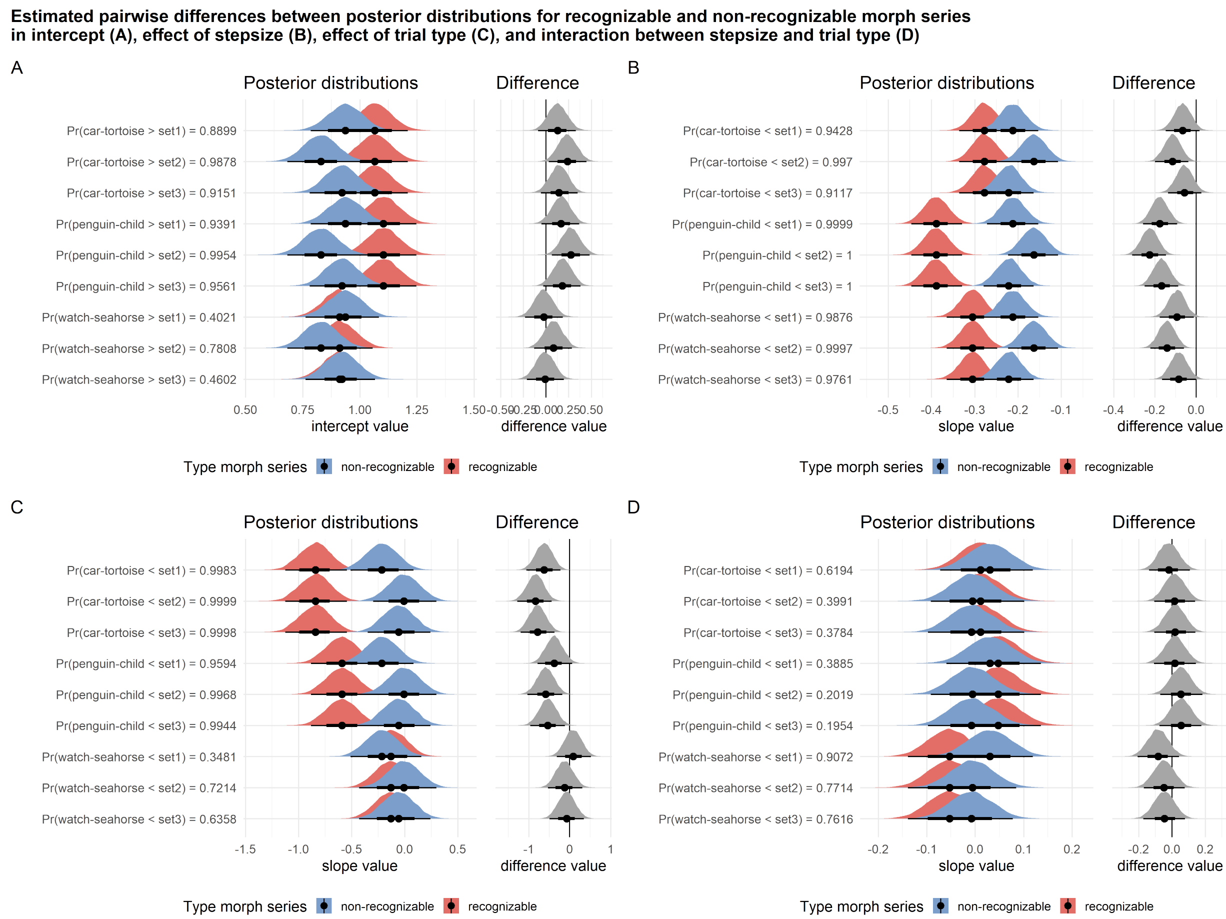


Figure 3.16: Estimated pairwise differences between the posterior distributions for the intercept (A), effect of stepsize (B), effect of trial type (within-category = 0, between-category = 1; C), and interaction between stepsize and trial type (D) on the standardized similarity scores, for each of the different recognizable and non-recognizable morph series combinations. Black dots and intervals indicate the mean, 66%, and 95% highest density continuous interval (HDCI) for each slope or difference value. The black vertical line indicates a difference in slope of zero. In this figure, the estimated effect of stepsize is larger for the recognizable than for the non-recognizable morph series (B). The main effect of trial type is larger for the recognizable series car-tortoise and penguin-child than for all non-recognizable morph series (C). The interaction effect between stepsize and trial type is larger for the recognizable series watch-seahorse than for all non-recognizable series (D). *Note.* In the interactive version of this figure, you can hover over the intervals to see the related mean and 95% HDCI for each distribution.

As for the discrimination task, we can relate these results for the similarity judgment task to the results for the categorization task: the stronger the categorization for a morph series, the stronger the difference in perceived (dis)similarity between within- and between-category pairs (i.e., the main effect of trial type), or its interaction with stepsize. In other words, the stronger the categorization for a morph series, the stronger the category boundary effect in the similarity judgment task.[8] As mentioned before, this is the case even though for each morph series, the categorization and similarity judgment data were collected from different subgroups of participants, and none

[8] Categorization was strongest for recognizable morph series car-tortoise and penguin-child and weakest for non-recognizable morph series 2.

of the participants in the similarity judgment task had any prior experience with or information about the morph series in question.

More importantly, when plotting the data for each stimulus pair within each stepsize separately (see Figure 3.17), the same clear gradual, inversely peaked pattern emerges as in the discrimination data for the recognizable morph series: for physically equidistant stimulus pairs in the recognizable morph series, the closer the stimuli in the pair are to one of the reference points, the more similar participants perceive the stimuli to be. Rather than a general effect of being a within- or between-category comparison, the effect is gradual in nature and relates to the characteristics of the individual stimuli involved in the comparison. In other words, also in a similarity judgment task, it is important not to treat the category boundary effect as a binary difference of within-category versus between-category pairs, but rather as a gradual effect of distance from the reference points for the individual stimuli in the pair. This is necessary as perceived similarity is heavily influenced by the clarity of the individual stimuli in the pair, as can be seen in Figure 3.17. For non-recognizable morph series, this peaked pattern is much less pronounced.

To further demonstrate the importance of the individual 'goodness' of the stimuli in the pair for perceived similarity judgments, we plotted predictions of perceived similarity based solely on the predicted categorization probabilities of individual stimuli in a pair to be categorized as A or B (see Figure 3.18). Although the predicted values do not match the absolute empirical values, these predictions show the same inversely peaked pattern as can be viewed from the actual similarity data. This indicates, just like in the results for the successive discrimination task, that the observed pattern is likely to be a consequence of properties (i.e., 'goodness') of the individual stimuli involved in the comparison rather than a consequence of stimulus similarity.

When modelling and plotting the standardized similarity scores for all stimulus pairs in both presentation orders separately, no clear directional asymmetries were present (see Figure A.10).

3.4 Discussion

In this study, we replicated the typically reported 'category boundary' effect in both a successive discrimination task and a similarity judgment task. Our analyses differed from those in many previous studies of the category boundary effect, as we did not average all equidistant pairs together in groups of within-category vs. between-category pairs. This alternative analysis enabled us to show a gradual rather than a binary pattern in the discrimination and similarity judgment data. No sudden jumps in discrimination performance occurred when either crossing the category boundary or not, but a gradual change in discrimination performance became visible. This gradual change depended on the distance of the individual stimuli compared to the reference points on the dimension. In other words, the 'category boundary' effect is not an effect of the category boundary[9]. We show that the category boundary effect is not a result of increased discrimination sensitivity across the category boundary, but rather related to the clarity of the individual stimuli involved in the comparison. Keeping physical distance between the stimuli in the pair equal, the closer the individual stimuli in the pair are to the reference points, the more perceptually similar the stimuli become and the more difficult discrimination becomes. This explanation is evidenced further by showing that the peaked qualitative pattern of observed discrimination responses can be predicted by solely

[9] This conclusion touches on some of the conclusions made by McMurray (2022) in the field of speech perception.

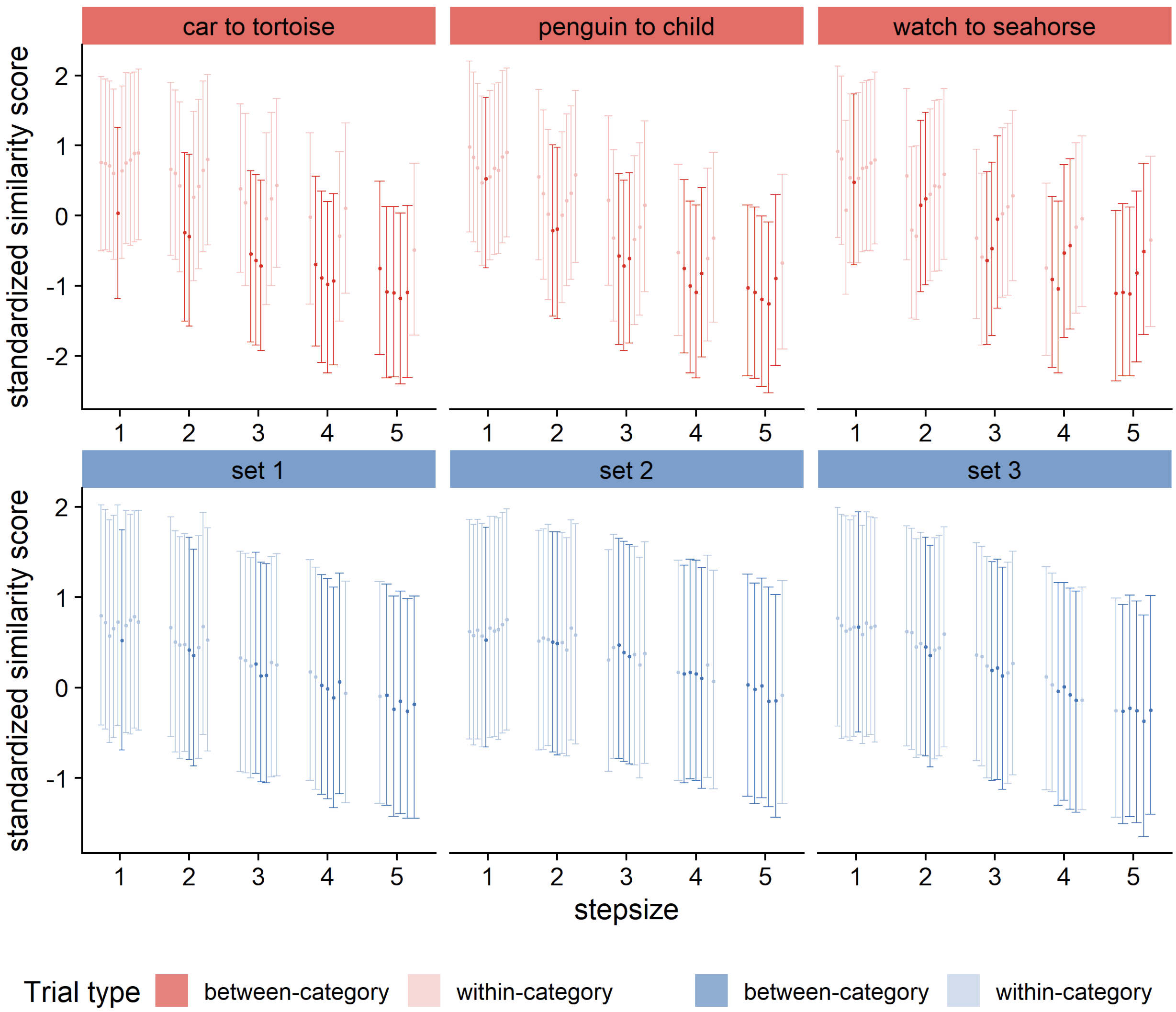


Figure 3.17: Standardized similarity scores for each stepsize, stimulus pair, and morph series separately, averaged across participants and stimulus order within the pair. The colored dots and error bars indicate the mean posterior predictions from the model and the 95% highest density continuous interval (HDCI) of the posterior predictive distributions. In this figure, a clear, gradual, inversely peaked pattern can be observed for the recognizable morph series, where pairs that include stimuli close to the reference points lead to higher similarity scores than pairs that include stimuli further away from the reference points, while keeping the physical distance between the stimuli in the pair (i.e., stepsize) equal. This inversely peaked pattern is much less pronounced for the non-recognizable morph series *Note.* In the interactive version of this figure, you can hover over the intervals to see the stimuli involved in the pair, the exact similarity score, the mean and 95% HDCI of the posterior predictive distributions, and the number of trials related to each interval.

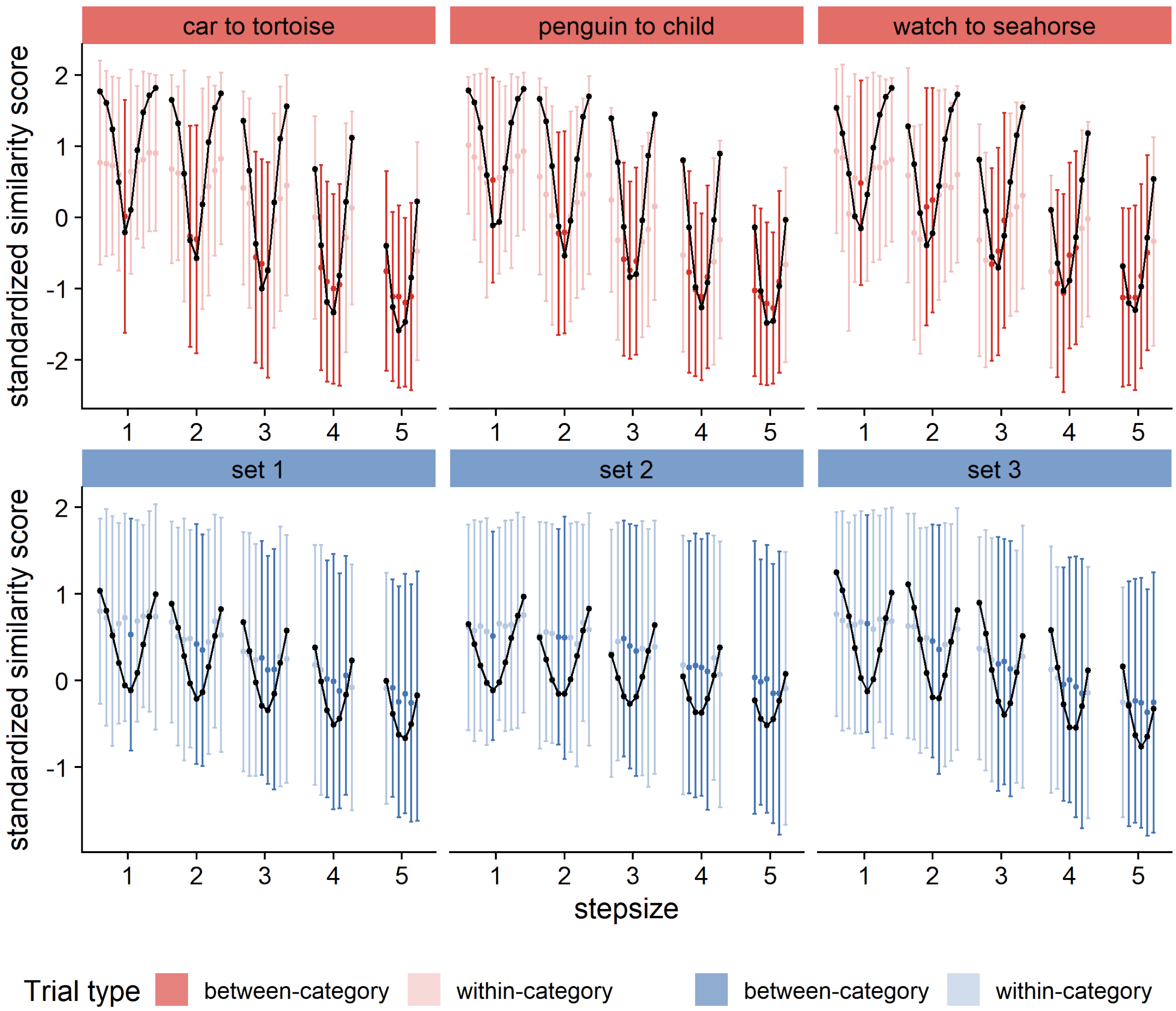


Figure 3.18: Prediction of standardized similarity scores based on categorization probabilities of indvidual stimuli in the pair in the categorization task. The colored dots and intervals indicate the mean and 95% highest density continuous interval (HDCI) for the posterior predictive distributions per stepsize, stimulus pair, and morph series, averaged across participants and stimulus order within the pair. The black dots and lines indicate the predictions based on the estimated mean categorization probabilities for the stimuli in a pair. *Note.* In the interactive version of this figure, you can hover over the intervals to see the stimuli involved in the pair, the exact similarity score, the mean and 95% HDCI of the posterior predictive distributions, and the number of trials related to each interval.

using the categorization probabilities for the individual stimuli in the pair, even when collected from different participants.[10] Just as Nosofsky (1991) indicated earlier for asymmetric similarities, also the category boundary effect needs to be explained by looking at properties of the individual stimuli involved in the comparison rather than by looking at the relation between the stimuli in the pair. The fact that we find a similar peaked pattern for both tasks, strengthens our interpretation of the findings.

We see the same influence of categorization strength not only within each stimulus dimension, but also across morph series. As predicted, both categorization and the category boundary effect were more clearly visible for some of the recognizable morph series than for some of the non-recognizable morph series. In addition, also categorization response times were more influenced by distance from the reference points in the recognizable than in the non-recognizable morph series.

Although we expected to find directional asymmetries, we did not find them. We assume their absence may be due to the short presentation durations of both stimuli in the pair and the short interstimulus interval used (i.e., 500 ms): given that both stimuli were only briefly presented, and given that none of the stimuli was present when making the comparison, the difference in perceptual presence between the stimuli in the pair may have been too small to be consequential (i.e., too little room for memory effects to be larger for the first than for the second stimulus). Future research could investigate whether the reported results hold for different tasks, stimuli, presentation durations, or interstimulus intervals.

Looking at the results for the morph series watch to seahorse, one might find a mismatch between the lowest discrimination performance and the distinction between within- and between-category pairs. This difference in placement of the category boundary between tasks may be the result of differences in the point of subjective equivalence across participants, as categorization and discrimination data for a particular morph series were collected from different participants. Although the placement of the category boundary was arbitrarily defined based on the collected categorization data, our main results do not depend on the placement of the category boundary. Only the typically reported category boundary effect depends on this placement.

One might notice that, although the predictions for discrimination performance and perceived similarity based on the categorization data follow the observed peaked pattern, the absolute predicted values do not closely match the empirical ones. One potential factor in explaining this divergence is the fact that the data were collected in another task from different participants. More specifically, we would expect categorization to be stronger during the categorization task than when performing the discrimination or similarity judgment task. Before starting the categorization task, participants were presented with four ‘clear’ exemplars per category. Although minimally, these example stimuli inform participants about the categorization of the stimuli and stimulate the use of these categories in their answers. In addition, the categorization task by nature focuses on categorizing the stimuli, potentially improving categorization strength. In the discrimination and similarity judgment task, participants were not informed about this categorization, and were not presented with ‘clear’ exemplars before starting the task.

[10] As mentioned earlier, for each morph series the categorization, discrimination, and similarity judgment data were collected from a different subgroup of the participants. The predictions of discrimination performance or perceived similarity based on categorization performance are thus based on data from a different group of participants.

These differences may contribute to a difference in categorization strength for the same morph series across tasks. In addition, more advanced models of discrimination data (e.g., Feldman et al., 2009) include an additional term related to the estimated perceived distance between the stimuli in the pair conditional on the categorization of the stimuli, which could improve absolute prediction performance. Regardless of these differences in absolute values, the typical peaked pattern could be predicted from the categorization data alone, even though these data came from a completely independent task, was collected from different participants per morph series, and the model did not include any information about physical or perceived distance between the stimuli in the pair.

The fact that the tasks were done on different morph series within each participant, and that there was no training before participating in the tasks (except for the four clear exemplars per category shown before starting the categorization tasks), probably contributed to the results for the non-recognizable morph series (they were not categorized as strongly in the categorization task, and the effects in the discrimination and similarity tasks were smaller).

3.4.1 Constraints on generality

This paper puts forward two main results that we expect to generalize to different shape morph series as well as morph series in other stimulus domains (e.g., color, orientation, faces, speech sounds). Firstly, the stronger the categorization for a morph series, the stronger the overall category boundary effect in discrimination performance and similarity judgment. Secondly, for physically equidistant stimulus pairs belonging to morph series that show evidence for categorization, the stimuli in the pair will be perceived as more similar and will be more difficult to discriminate the closer the stimuli in the pair are to one of the reference points. Generalization of the current findings to these other morph series and stimulus domains is however subject to further investigation.

As mentioned earlier, the results in the current study are only directly applicable to settings in which participants have no prior experience with the particular morph series under study. In other words, participants may be able to recognize and categorize shapes correctly without any prior experience, but they should not have had any prior task or prior learning involving the specific stimuli used in the discrimination or similarity judgment task. This constraint on generality is especially important concerning the non-recognizable morph series, as prior experience with the stimuli may lead to implicit categorization and increase categorization strength for these series. Other aspects of the study design important to take into account when assessing its generalizability are the presentation times, intertrial interval durations, and successive presentation format used.

As the current participant sample was a Western European student sample, also generalization to other, especially non-Western participant populations is subject to further investigation. We have no reason to believe that the results depend on other than the above mentioned characteristics of the participants, materials, or context.

3.5 Conclusion

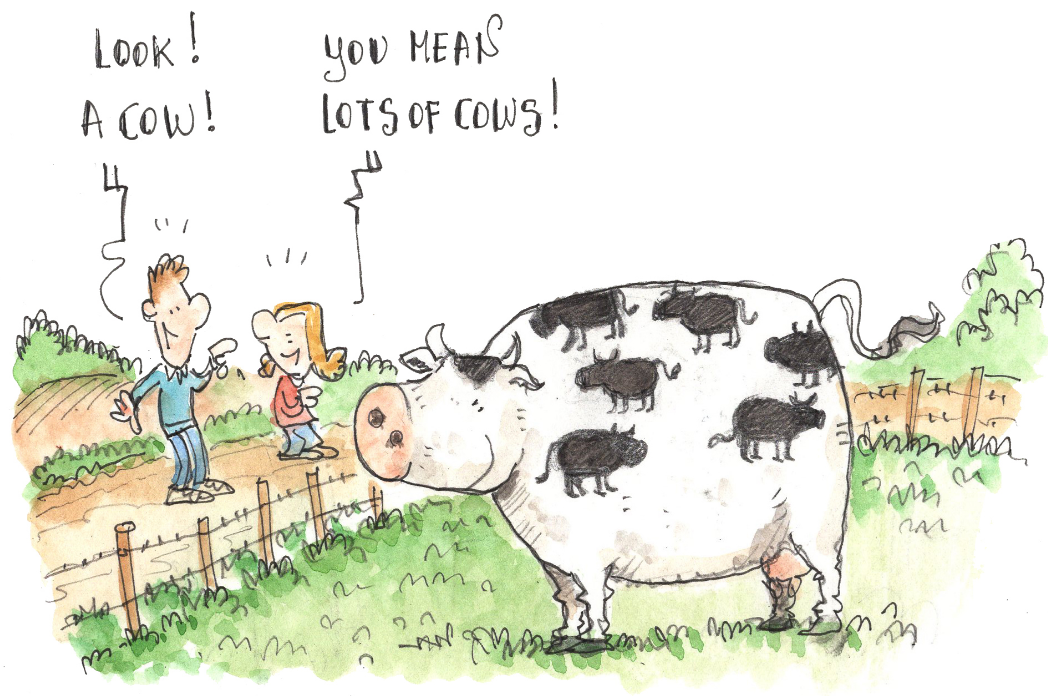
To summarize, we replicated the existence of the category boundary effect in both a successive discrimination and a similarity judgment task, and showed that the strength of the category boundary effect is related to the strength of the reference points involved. The general category boundary effect is not a category boundary effect, however: Although a general category boundary effect results from averaging across physically equidistant stimulus pairs, one needs to take into account distance from the reference points for each of the individual stimuli in the pair to actually predict discrimination performance and similarity judgments. We showed how the observed qualitative pattern of discrimination performance and similarity judgments across stimulus pairs could be predicted by the 'goodness' of the individual stimuli in the pair. In other words, in the current study, the qualitative pattern of relative differences in discrimination performance and similarity judgments for physically equidistant stimulus pairs from the recognizable morph series (i.e., the morph series with the clearest reference points) could largely be explained by categorization strength for the individual stimuli alone, and was largely unrelated to the specific stimulus pairs presented. These results provide evidence that the existence and strength of reference points on a stimulus dimension has tangible consequences for how we perceive, categorize, and react to stimuli on that dimension. Moreover, the results remind us of the danger of averaging without looking at the underlying data patterns, and of the gains that can be made by seriously exploring consistent variation in extensive data sets.

3.6 Open and reproducible practices statement

This manuscript was written in R Markdown using the papaja package (Aust & Barth, 2022) with code for data analysis integrated into the text. The data, materials, and analysis and manuscript code for the experiment are available at https://doi.org/10.17605/osf.io/ugcd8.

Part IV

Hysteresis and adaptation



Also the *immediate context* in which we experience the incoming visual stimulation can influence how we organize this stimulation.

Chapter 4

Same stimulus, same temporal context, different percept? Individual differences in hysteresis and adaptation when perceiving multistable dot lattices

4

How we perceptually organize a visual stimulus depends not only on the stimulus itself, but also on the temporal and spatial context in which the stimulus is presented and on the individual processing the stimulus and context. Earlier research found both attractive and repulsive context effects in perception: tendencies to organize visual input similarly to preceding context stimuli (i.e., hysteresis, attraction) co-exist with tendencies that repel the current percept from the organization that is most dominant in these contextual stimuli (i.e., adaptation, repulsion). These processes have been studied mostly on a group level (e.g., Schwiedrzik et al., 2014). Using a Bayesian hierarchical model comparison approach, the present study (N = 75) investigated whether consistent individual differences exist in these attractive and repulsive temporal context effects, with multistable dot lattices as stimuli. In addition, the temporal stability of these individual differences in context effects was investigated, and it was studied how the strength of these effects related to the strength of individual biases for absolute orientations. The results demonstrate that large individual differences in the size of attractive and repulsive context effects exist. Furthermore, these individual differences are highly consistent across timepoints (one to two weeks apart). Although *almost everyone* showed both effects in the expected direction, *not every* single individual did. In sum, the study reveals differences in how individuals combine previous input and experience with current input in their perception, and more generally, this teaches us that different individuals can perceive identical stimuli differently, even within a similar context.

Van Geert, E., Moors, P., Haaf, J. M., & Wagemans, J. (2022). Same stimulus, same temporal context, different percept? Individual differences in hysteresis and adaptation when perceiving multistable dot lattices. *i-Perception, 13*(4). https://doi.org/10.1177/20416695221109300

Paper | Postprint | Preregistration | Materials, data, and code

Same stimulus, same temporal context, different percept? Individual differences in hysteresis and adaptation when perceiving multistable dot lattices

4.1 Introduction

When we visually experience the world, our experience consists of organized wholes rather than many separate sensations (Wagemans, 2018). Perceptual organization of the visual input we receive from the world is an active process, including perceptual grouping and figure-ground segregation. Although the Gestalt principles of perceptual organization are often described as ‘laws’, which seems to imply a deterministic character, individual differences exist in sensitivity to several grouping principles such as grouping by proximity and grouping by similarity (Wagemans et al., 2018). Furthermore, when individuals perceive multistable stimuli, individual biases can exist, for instance, in the probability to perceive one orientation more often than another objectively equiprobable orientation (Kubovy & Berg, 2002).

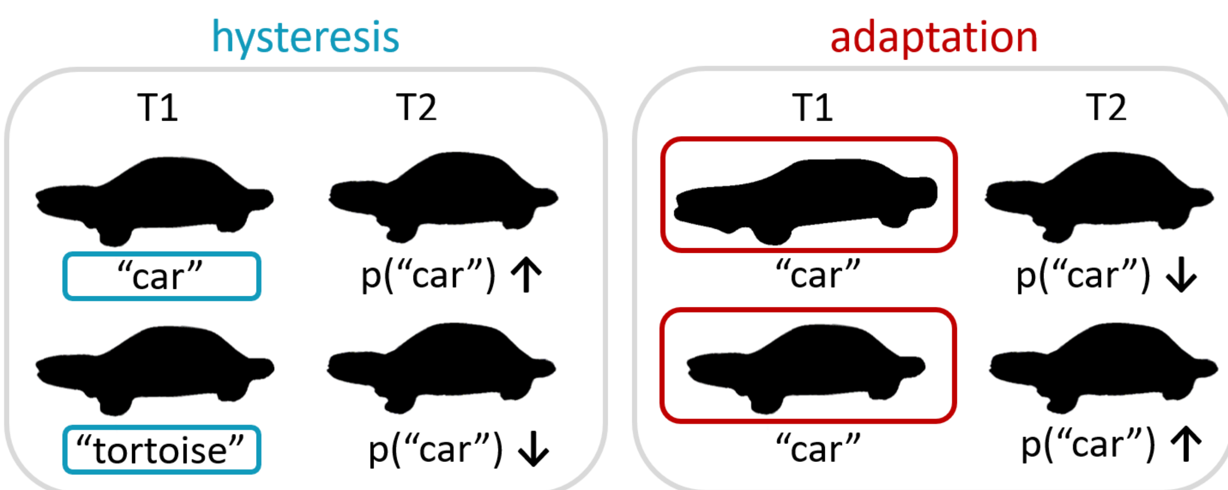


Figure 4.1: Illustration of attractive and repulsive context effects. Left side: attraction effect (hysteresis). When the stimulus is perceived as a car at time 1 (T1), the probability that another stimulus at time 2 (T2) will be perceived as a car is higher than when the stimulus at T1 was interpreted as a tortoise. Right side: repulsion effect (adaptation). When the stimulus at T1 is a very clear example of a car, the probability that another stimulus at T2 will be perceived as a car is lower than when the stimulus at T1 was a more ambiguous example of a car.

Perceptual organization of current visual input can however also be influenced by its temporal context, including previously presented stimuli and their perceived organization. Earlier research has found both attractive and repulsive context effects in perception (Snyder et al., 2015). Attractive context effects (also called hysteresis, stabilization, facilitation, etc.) entail that individuals tend to organize current visual input in a similar way as preceding or simultaneous context stimuli (see left side of Figure 4.1): When people perceive a certain organization in the context stimulus, they are more likely to perceive the same organization in the test stimulus. The repulsive context effect (also known as negative hysteresis, adaptation, contrast, differentiation, etc.) entails that perception tends to repel or move away from the organization that is dominant in the contextual stimuli (see right side of

Figure 4.1): When a lot of evidence for a certain organization is present in the context stimulus, people are less likely to perceive that organization in the test stimulus. Attractive and repulsive tendencies are concurrently present. Schwiedrzik et al. (2014) found evidence for two separate mechanisms underlying hysteresis and adaptation, as they mapped into distinct cortical networks. Whether they are part of the same process or separate processes is still under debate, however (e.g., Gepshtein & Kubovy, 2005).

4.1.1 Hysteresis and adaptation in multistable dot lattices

Gepshtein & Kubovy (2005) presented a paradigm that allows to disentangle attractive and repulsive context effects on perception. They used multistable dot lattices as context and test stimuli, and investigated the influence of (a) the perceived organization of the context stimulus (i.e., which organization was reported) and (b) the stimulus support for a certain organization in the context stimulus (dependent on the stimulus' aspect ratio) on the perception of a second, test stimulus.

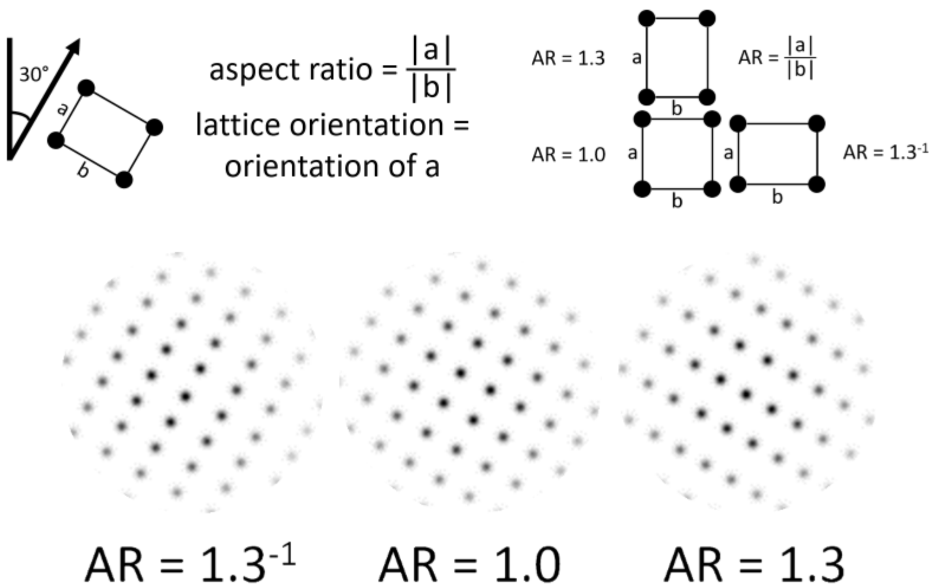


Figure 4.2: Explanation regarding the aspect ratio of a multistable rectangular dot lattice. In rectangular dot lattices, four different orientations can be perceived, of which two are more prevalent (as the dots are closer together along these orientations). The relative dominance of the a orientation relative to the b orientation is expressed in the aspect ratio of the dot lattice (AR = |a| / |b|).

Multistable dot lattices are arrays of aligned dots in which multiple orientations can be perceived (see Figure 4.2). The closer the dots are spaced along a particular orientation, the more likely they are grouped together and that orientation will be perceived (cf. the Gestalt law of proximity, Kubovy, Holcombe, & Wagemans, 1998). This relative grouping strength has been shown to follow a decreasing exponential function of the relative inter-dot distance in that orientation (Kubovy et al., 1998).

In rectangular dot lattices (see left side of Figure 4.3), four different orientations can be perceived, of which two are more prevalent (as the dots are closer together along these orientations). The relative dominance of the a orientation relative to the b orientation is expressed in the aspect ratio of the dot lattice (AR = |a| / |b|) [1]. For a lattice with AR = 1, the distance between the dots in the a and b orientation is equal. For a lattice with AR < 1, the distance between the dots is smaller in the a than in the b orientation. For a lattice with AR > 1, the distance between the dots is smaller in the b than in the a orientation. In hexagonal dot lattices (see right side of Figure 4.3), three prominent orientations are present and equally plausible, which makes it a very ambiguous or unstable lattice type. In both types of lattices we will define the axis orientation of the dot lattice as a whole by the orientation of a, which we will call the 0° orientation. In the rectangular dot lattices, we will call the b orientation the 90° orientation.

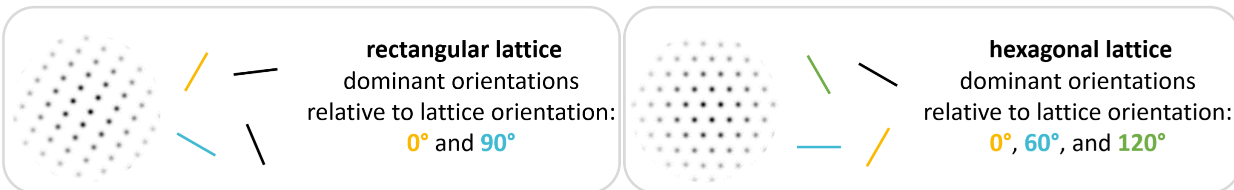


Figure 4.3: Dominantly perceived orientations in multistable rectangular and hexagonal dot lattices.

Gepshtein & Kubovy (2005) used rectangular dot lattices with a randomly varying lattice orientation as context stimuli and more ambiguous hexagonal dot lattices with the same random lattice orientation as test stimuli. The stimulus support for a particular organization in the context stimulus was manipulated by varying the aspect ratio of the rectangular lattice (i.e., the distance between the dots in the a vs. the b orientation). They then investigated the influence of (a) the perceived orientation and (b) the aspect ratio in the context stimulus on the perceived orientation in the test stimulus.

Hysteresis was present when participants perceived the same orientation in both the context and test stimulus (i.e., the a or 0° orientation). Adaptation was present when participants perceived a different orientation in the test stimulus than the one for which there was most support in the context stimulus.

Probabilities for perceiving a particular organization in the test stimulus increased when the same organization was perceived in the context stimulus compared to when an alternative organization was perceived in the context stimulus (i.e., hysteresis effect, see left side of Figure 4.4). At the same time, the stronger the stimulus support was for a certain organization in the context stimulus (i.e., the closer the dots were together in one dominant orientation compared to the other dominant orientation), the lower the probability was that the same organization was perceived in the test stimulus (i.e., adaptation effect, see right side of Figure 4.4). The effects of hysteresis and adaptation were found to combine multiplicatively (in a logistic regression model they related to the current percept independently). Schwiedrzik et al. (2014) used a very similar paradigm as Gepshtein & Kubovy (2005), tested more participants, and added brain imaging to investigate the neural underpinnings of both effects. They found similar behavioral results to those reported by Gepshtein & Kubovy (2005), and the fMRI data provided evidence for two separate mechanisms underlying adaptation and hysteresis effects, as the effects mapped into distinct cortical networks. In addition, Schwiedrzik et al. (2014) reported interindividual variability in the size of the hysteresis effect,

[1]Aspect ratio is sometimes defined as |a| / |b| with |a| and |b| with a fixed orientation but changing length (e.g., Schwiedrzik et al., 2014), and sometimes as |b| / |a|, where |a| varies in orientation but is always the orientation with the shortest inter-dot distance (e.g., Gepshtein & Kubovy, 2005). Here we will use |a| / |b| with |a| and |b| with a fixed orientation but changing length.

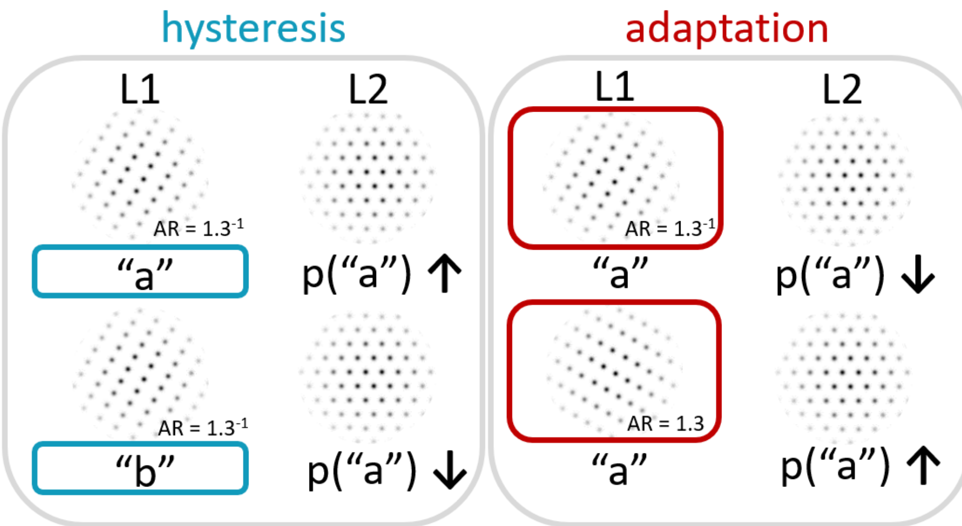


Figure 4.4: Illustration of attractive and repulsive context effects in dot lattices. Left side: attraction effect (hysteresis). When the first lattice (L1) is perceived as orientation a, the probability that the second lattice (L2) will be perceived as orientation a is higher than when L1 was interpreted as orientation b. Right side: repulsion effect (adaptation). When strong support for orientation a is present in L1, the probability that L2 will be perceived as a orientation a is lower than when L1 had less support for orientation a.

and these individual differences were correlated with differences in activation between hysteresis and no hysteresis trials for the right dorsomedial prefrontal cortex.

Does every individual show these attractive and repulsive context effects, and if so, to the same extent? Although the studies of Gepshtein & Kubovy (2005) and Schwiedrzik et al. (2014) demonstrated the existence of these effects when based on *averaged* data, none of these studies focused on *individual differences* in (the strength of) these temporal context effects.

Earlier work has shown that looking at averaged data alone can be misleading (Kanai & Rees, 2011) and that investigating individual differences can contribute to a richer understanding of visual perception (Mollon et al., 2017). More specifically, when testing the presence of an effect by looking at averaged data alone, one ignores the possibility for large consistent variation between individuals (Kanai & Rees, 2011). Interindividual differences are treated as noise, and it is assumed that the effect would be present for all individuals in case no measurement error would occur. Finding evidence for an average effect however does not guarantee the *true* effect for each individual to be of the same size or in the same direction. The average effect could even be purely an artifact from the averaging procedure (Van der Hulst et al., 2022). Haaf & Rouder (2019) proposed a model comparison approach to tackle exactly these questions: (a) whether the data provide evidence for true, consistent individual differences in the size of an effect, and (b) whether the estimated true effects are in the same direction for all tested individuals. To answer the first question, they compare evidence for a model assuming that individuals share a common effect with no individual variability (i.e., common-effect model) with a model that does not place any constraints on the individuals' true effects (i.e., unconstrained model). To answer the second question, the unconstrained model is compared with a model that constrains true individuals' effects to have a particular sign (e.g., to be positive; positive-effects model).

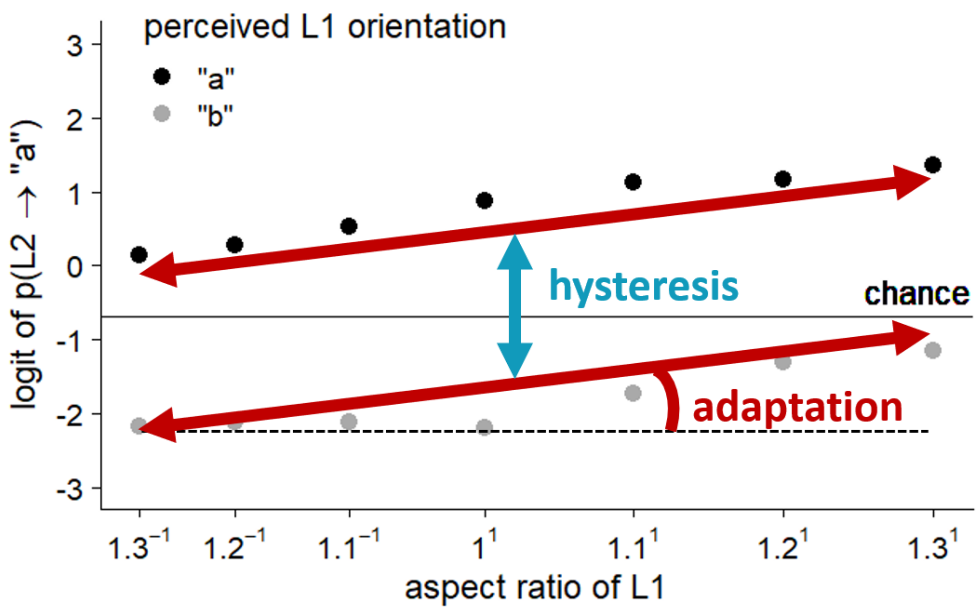


Figure 4.5: Expected average effects of hysteresis and adaptation on the perception of multistable dot lattices, based on the study of Schwiedrzik et al. (2014). Vertical separation of the two lines reflects the size of the perceptual hysteresis effect, the slope of both lines reflects the size of the perceptual adaptation effect.

The current study investigates whether there is evidence for true individual differences in the size of hysteresis and adaptation effects, and whether every tested individual shows true hysteresis and adaptation effects in the expected direction. We do this by implementing the model comparison strategy proposed by Haaf & Rouder (2019). Important in this regard is that these model comparisons bring evidence for whether *true* individual variation exists, rather than whether individual variation is *observed* when conducting a task with a finite number of trials (e.g., Mollon et al., 2017). Observed variation between individuals can be due to multiple factors, including trial-by-trial noise, which would not indicate consistent, true interindividual variation. The hierarchical models used in this study allow for the modeling of trial-by-trial variation as well as variation across individuals, and estimate the true individual effects accurately even with a finite number of trials (in contrast, non-hierarchical sample effects are only estimating the true effects accurately in the large-trial limit, Rouder & Haaf, 2019). Establishing whether true individual differences in the size and/or direction of hysteresis and adaptation effects exist is a necessary first step before investigations in the sources and correlates of these true individual differences can become relevant.

The existence of true individual differences can be of theoretical importance (e.g., Haaf & Rouder, 2019; Miller & Schwarz, 2018), and this is the case for individual differences in hysteresis and adaptation effects as well. For example, in case everyone shows both hysteresis and adaptation, this would suggest both to be fundamental mechanisms in human visual perception. In case individuals differ in the extent to which they show hysteresis and adaptation, and the size of both effects is correlated across individuals, this would suggest at least some common factor affecting the processes underlying both effects. In case of evidence for the absence of a correlation between individual hysteresis and adaptation effects, this would imply clearly independent processes underlying the hysteresis and adaptation effects present in this task.

The results may also be important for our understanding of individual differences in perception in general. Interindividual differences in hysteresis and adaptation strength, if they exist, may cause differences in what individuals will actually perceive, even when the current visual input as well as the context stimuli are equal. In case evidence for a lack of interindividual differences is found, this is evidence against differential use of previous percept and previous stimulus context in the formation of the current percept, and differences in hysteresis and adaptation effects can then not explain perceptual differences between individuals given the same stimulus and context. In other words, the study will provide insight in whether individuals can differ in their perception alone (based on differences in previously encountered stimuli and percepts), or whether they can also differ in the processes underlying their perception: whether context info is differentially used across individuals, or whether everyone combines context and current stimulus in a similar way. Put differently, individual differences could either arise through different context information or previous experiences (i.e., previously encountered stimuli and percepts), or alternatively also through how the same context information is incorporated differently by different individuals. The first would imply that consistent individual differences in perception can be due to differences in external factors alone, and that in case everyone would have the same stimulus and perceptual history, everyone would have the same effects of the previous stimulus and previous percept on their current percept. The second would imply that even in case individuals have exactly the same stimulus and perceptual history, there would still be differences in what they perceive due to differential use of the stimulus and perceptual history when coming to the current percept.

Although earlier research has found evidence for individual differences in several tasks assessing hysteresis, adaptation, or their ratio (e.g., Abrahamyan et al., 2016; Mattar et al., 2018; McGovern et al., 2017; Song, Schwarzkopf, & Rees, 2013; Song, Schwarzkopf, Lutti, et al., 2013), only a few studies have attempted to quantify both effects concurrently at the level of the individual participant, by distinguishing the effects of previous stimulus support and previous percept or response (e.g., Bosch et al., 2020; Urai et al., 2017; Zhang & Alais, 2020). Moreover, in none of these studies individual differences in effects of previous stimulus and perceptual choice were the focus of study.

Urai et al. (2017) asked participants to report whether a test stimulus contained stronger or weaker motion than a reference stimulus, and found robust and idiosyncratic patterns of history biases based on previous stimulus and previous choice, with the weight of the preceding choice generally being stronger than the effect of the preceding stimulus. They also found large interindividual variability in the effect of the previous choice, with a majority of the participants showing hysteresis and some showing alternation.

Zhang & Alais (2020) asked participants to report which orientation (+45° or -45°) they perceived in a grating embedded in noise. In a version of the task where motor response and perceptual choice could not be distinguished, they found large individual differences in the effect of the previous choice or response, but rather consistently no effect of the previous stimulus shown. Based on the results from a task in which motor response and perceptual choice could be distinguished, they suggested that individual differences in the sign of the serial dependence reflect different relative weightings of the hysteresis effect for perceptual choice and the adaptation effect for motor response.

Bosch et al. (2020) examined the effects of choice history and evidence history on subsequent perceptual choices by asking participants to identify a coherent motion test stimulus as more or less coherent than a reference stimulus. They found evidence for a bias toward the previous choice, but, at the same time, they found evidence for a bias away from the direction of evidence on the previous trial, especially when it concerned strong evidence. Although almost all participants showed an attractive choice history bias and all participants showed a repulsive evidence history bias, the size of the choice history bias varied considerably across participants (cf. Supplementary Figure 2 in Bosch et al., 2020).

4.1.2 Hysteresis and adaptation deriving from the same or separate mechanisms?

Whether hysteresis and adaptation effects are the result of the same process or of two separate processes is still under debate. Whereas some argue that both effects can be explained through a single mechanism of sensory integration operating over varying timescales (Mattar et al., 2016), of persistent bias (Gepshtein & Kubovy, 2005), or of neuronal adaptation (Maus et al., 2013), others state that both are separate processes, either in the same neuronal location (e.g., Brascamp et al., 2008) or in distinct cortical networks (Fritsche et al., 2020; Pascucci et al., 2019; Schwiedrzik et al., 2014). Additional arguments for assuming separate mechanisms are differences in the extent to which hysteresis and adaptation are dependent on attention, are modulated by subjective confidence, are modulated by working memory delay, or exhibit clear spatial specificity (for an overview, see Fritsche et al., 2020). Many have also distinguished the effects based on their source being stimulus-related, percept-related, choice-related or motor-related (e.g., Bosch et al., 2020; Carter et al., 2014; Cicchini et al., 2017; Pascucci et al., 2019; Sadil et al., 2021; Zhang & Alais, 2020).

In case individual differences are present in both hysteresis and adaptation, we can also determine the correlation in the size of both effects. A strong correlation between hysteresis and adaptation may suggest at least some common factor affecting the mechanisms underlying both effects, whereas evidence for a correlation close to zero may imply independent processes underlying both effects. Based on a reanalysis of the data from Schwiedrzik et al. (2014)[2], we expect a positive correlation between individual hysteresis and adaptation effects.

4.1.3 Hysteresis as a perceptual or decisional effect

Whereas adaptation is typically seen as a stimulus-related effect (e.g., Fritsche et al., 2017; Pascucci et al., 2019; Sadil et al., 2021), there is more debate on the nature of the hysteresis effect. Whereas some 'serial dependence' research has suggested the attractive history effect to be the result of a perceptual process (e.g., Carter et al., 2014; Cicchini et al., 2017; Manassi et al., 2018; Schwiedrzik et al., 2018), other research has suggested a post-perceptual, decision-related source of the effect (e.g., Bosch et al., 2020; Fritsche et al., 2017; Pascucci et al., 2019).

In the current study we define hysteresis as a percept-related effect, but it cannot be excluded that the nature of the effect could be related to postperceptual decision processes rather than perceptual processes. To control for the possibility of the hysteresis effect being a purely decisional rather than a perceptual effect, we will include the control task presented by Schwiedrzik et al. (2018) as an additional task in our study. In this control task, the rectangular dot lattices used as context stimuli will be replaced by random dot lattices that can not induce the

[2] The results from this reanalysis are available on the Open Science Framework: https://osf.io/xa5ut/

perception of a particular orientation. Participants will then be asked to choose between four simultaneously presented orientations. As in the main task, the test stimuli will be hexagonal dot lattices, and also here participants will choose between four simultaneously presented orientations. In case the hysteresis effect would be a decisional effect rather than a percept-related effect, an influence of the response to the first random dot lattice would still have an effect on the perceived orientation in the test stimulus (i.e., a hysteresis effect would be present). In case the hysteresis effect would be percept-related, no hysteresis effect would be found in this control task.

4.1.4 Making the distinction between stimulus-related, percept-related, and response-related effects

Whereas the debate has mostly focused on attractive history effects being perceptual or post-perceptual (e.g., Cicchini et al., 2017; Fritsche et al., 2017; Manassi et al., 2018; Pascucci et al., 2019), we argue that it is important to make a distinction between stimulus-related effects on the one hand and percept-, response-, or decision-related effects on the other hand. The mixed results in the serial dependence literature are in our view partially due to the use of paradigms that cannot make this distinction between influences of previous stimuli and previous percepts. In addition, in many studies the distinction between percept, response, or decision is difficult to make. The literature suggesting the hysteresis effect to be post-perceptual has typically argued as follows: When the effect was not stimulus-related, they concluded it to be post-perceptual, and when stimulus-related effects were found those were typically reported as 'perceptual'. Making the conceptual distinction between stimulus-related and percept-related effects could help clarify this literature. The earlier findings could potentially be interpreted as evidence for perceptual hysteresis as those studies did not distinguish between percept-related and stimulus-related effects (e.g., Bosch et al., 2020; Fritsche et al., 2017; Pascucci et al., 2019).

4.1.5 Orientation bias

Earlier research reported effects of absolute orientation of stimuli on performance in several perceptual tasks (i.e., the 'oblique effect,' Appelle, 1972) with performance being higher for horizontally or vertically oriented stimuli than for obliquely oriented stimuli. Absolute orientation can not only influence perceptual performance, it may also influence perceptual experience. Kubovy & Berg (2002) reported three main bias categories for absolute orientation in the perception of hexagonal dot lattices: preference for vertical, preference for horizontal, and preference for vertical and horizontal over oblique. Some individuals stayed in one bias pattern consistently, others gradually shifted from one pattern to another. In a study by Claessens & Wagemans (2008) observers generally preferred vertical over horizontal orientations, but the exact orientation bias distribution was subject to individual differences. In the present study, the relation of the strength of individual's absolute orientation bias with the magnitude of their hysteresis and adaptation effects on perception will be investigated. We expect the effects of hysteresis and adaptation to be smaller when a stronger absolute orientation bias is present, as a stronger longer-term absolute orientation bias may overshadow influences of short-term temporal context like hysteresis and adaptation. In other words, we expect that individuals who have a stronger longer-term prior (likely based at least partially on longer-term stimulus history and perceptual history) will be less influenced by shorter-term expectations (i.e., hysteresis) as well as by shorter-term stimulus history (i.e., adaptation).

4.1.6 Temporal stability of individual differences in hysteresis, adaptation, and orientation bias

Although previous research has investigated the temporal stability of some perceptual biases for motion direction and of grouping behavior in multistable dot lattices (e.g., Van der Hulst et al., 2022; Wexler et al., 2015), we do not know of any research on the temporal stability of individual differences in the *magnitude* of short-term history effects or (assumedly) longer-term perceptual absolute orientation biases. Wexler, Duyck, and Mamassian (2015, Experiment 1) found that, although significant changes in structure-from-motion (SFM) and transparency-from-motion (TFM) bias directions occur, most biases are stable even over periods as long as one year. In addition, they found moderate but robust correlations between daily steps in the SFM and TFM biases, both within and between participants (Experiment 3). Van der Hulst and colleagues (2022) investigated the consistency of perceptual grouping behavior across two testing sessions that were one day apart. For most participants, behavior in both sessions was moderately to very strongly correlated, indicating that perceptual grouping behavior remained stable across testing sessions.

In this study, we investigate the temporal stability of individual differences in the magnitude of hysteresis and adaptation effects as well as of differences in the magnitude of the absolute orientation bias, by collecting data from the same participants in two sessions at least a week apart (minimally 7 days, maximally 14 days). As there are reasons to believe that the data for the second session may be less informative (e.g., participants may be less motivated for the second session because they already took part in the tasks before, non-random dropout may occur, etc.), all planned analyses (except for the ones on temporal stability) will be conducted based on the data for the first session. When estimating and testing the temporal stability of the hysteresis and adaptation effects, we will use the hierarchical model approach suggested by Rouder & Haaf (2019), as this approach provides a more accurate estimate of the correlation of individuals' effects between sessions, as it is less affected by design choices (e.g., the number of trials per individual per session) than correlating effects estimated separately for each session (Rouder & Haaf, 2019).

4.1.7 Research questions and hypotheses

This study thus investigates (a) whether the average attractive and repulsive context effects found in the perception of multistable dot lattices replicate (Gepshtein & Kubovy, 2005; Schwiedrzik et al., 2014), (b) whether consistent individual differences exist in the size of these effects, and (c) whether each individual shows both effects in the expected direction. Furthermore, it investigates (d) whether individual differences in hysteresis or adaptation effects in the dot lattice paradigm discussed are correlated, (e) whether the hysteresis effect is a perceptual or a purely decisional phenomenon, and (f) whether individual differences in hysteresis or adaptation effects in the dot lattice paradigm relate to differences in the strength of individuals' absolute orientation biases. Finally, we also investigate (g) whether individual differences in the size of hysteresis and adaptation effects as well as in the magnitude of absolute orientation biases are stable across time.

All research questions and hypotheses can be found in detail in Table 4.1. Firstly, the study serves as a replication of the distinct *average* effects of hysteresis and adaptation on the perception of multistable dot lattices found in Gepshtein & Kubovy (2005) as well as Schwiedrzik et al. (2014). Regarding the hysteresis effect, we predict that (a) the probability of perceiving orientation 0° (the a orientation) in the second lattice will be higher when the first lattice is perceived as orientation 0° than when the first lattice is perceived as orientation 90° (H1). Regarding the adaptation effect, we predict that the probability of perceiving orientation 0° in the second lattice will be lower for smaller aspect ratios in the first stimulus (|a|/|b|): The more the aspect ratio of the first stimulus is in favor of orientation 0°, the less the second stimulus will be perceived as orientation 0° (H2). Similar to those previous studies we also hypothesize that the hysteresis and adaptation effects will combine multiplicatively (H3).

Secondly, *individual* hysteresis and adaptation effects are investigated. Based on the methods developed by Haaf & Rouder (2019), we investigate whether consistent individual differences exist in the size of these hysteresis and adaptation effects (H4), by comparing evidence for a model with a common hysteresis effect (a common adaptation effect) across individuals with a model including a variable hysteresis effect (adaptation effect) for every individual (Haaf & Rouder, 2019). In addition, we investigate whether the evidence is in favor of true hysteresis and adaptation effects in the expected direction for everyone (H5), by comparing evidence for a positive effects model and an unconstrained model (Haaf & Rouder, 2019).

Thirdly, we investigate whether individual differences in hysteresis or adaptation effects in the dot lattice paradigm discussed correlate with each other (H6). Furthermore, we examine whether the hysteresis effect is a perceptual or a purely decisional effect by adding a control task in which the first lattice can not induce the perception of a particular orientation (H7). In this control task, we predict the absence of an attractive effect of the response to the first lattice on the percept of the second lattice (i.e., no hysteresis effect). As longer-term biases may diminish the influence of short-term temporal context effects, we also explore whether individual differences in hysteresis or adaptation effects correlate negatively with the strength of the individual's absolute orientation bias (H8).

Finally, we study the temporal stability of individual differences in the size of individual's hysteresis and adaptation effects (H9), as well as in the magnitude of individuals' absolute orientation biases (H10). We predict individual differences in the magnitude of these effects to be correlated positively across sessions.

4.2 Methods

The data collection for this study is part of the data collection for a larger research project. Here we specify all collected measures that are used in the scope of this specific study.

4.2.1 Participants

Anyone between 18 and 100 years old, with (corrected to) normal vision, and able to understand Dutch instructions was able to participate. Participants were recruited via the faculty's participant pool, personal contacts of the researchers, social media, and offline advertisements in public places and university buildings. Depending on the wish of the participant, either a monetary compensation of 8 euros per hour or one research credit per hour was

Table 4.1: Research questions and hypotheses.

H	Hypothesis	Statistical Test
H1	Perceiving a certain organization in the context stimulus will increase the probability of perceiving that same organization in the test stimulus (i.e., hysteresis effect).	Calculate the Bayes factor in favor of the model including the percept of the first lattice as predictor compared to a model without the percept of the first lattice as predictor, using bridge sampling (Gronau et al., 2017).
H2	The stronger the stimulus support for a certain organization in the context stimulus (i.e., based on aspect ratio), the lower the probability to perceive that organization in the test stimulus (i.e., adaptation effect).	Calculate the Bayes factor in favor of the model including the aspect ratio of the first lattice as predictor compared to a model without the aspect ratio of the first lattice as predictor, using bridge sampling (Gronau et al., 2017).
H3	The hysteresis and adaptation effects described in H1 and H2 will combine multiplicatively and will thus be independent in logit space (i.e., there will be no significant interaction).	Calculate the Bayes factor in favor of the model without interaction between the percept and the aspect ratio of the first lattice as predictor compared to a model with the interaction between the percept and the aspect ratio of the first lattice as predictor, using bridge sampling (Gronau et al., 2017).
H4	Consistent individual differences will exist in the size of the estimated true individual hysteresis and adaptation effects (i.e., a model predicting individual differences in each of these effects will do better than a model predicting the same effect sizes for every participant).	Calculate the Bayes factor in favor of a model including random intercepts and slopes for every participant compared to a model including no random slopes (cf. unconstrained model vs. common effects model in Haaf & Rouder, 2019), using bridge sampling (Gronau et al., 2017). We conduct this model comparison for each effect separately.
H5	Every participant will show the hysteresis and adaptation effects described in H1 and H2 to some extent: Every participant in the study will show an estimated true positive hysteresis effect and an estimated true positive adaptation effect. A model predicting a positive effect size for every participant in the case of both hysteresis and adaptation will do better than a model without constraints on the direction of the effects for every participant.	Calculate the Bayes factor in favor of a model predicting a positive effect size for every participant compared to a model that does not place any order or equality constraints on individuals' effects, using the encompassing approach (cf. positive effects model vs. unconstrained model in Haaf & Rouder, 2019). In the positive effects model, the main hysteresis and the main adaptation effect are both restricted to be positive. We conduct this model comparison for each effect separately, however.
H6	The size of individuals' estimated true hysteresis effect will correlate positively with the size of their estimated true adaptation effect.	Calculate the Bayes factor in favor of a model that assumes the true linear correlation to be positive compared to a model assuming a non-positive true linear correlation, using the Savage-Dickey density ratio method (Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010).
H7	In the control task with a random dot lattice as context stimulus, responding to have perceived a certain organization in the context stimulus will not increase the probability of perceiving that same organization in the test stimulus (i.e., no hysteresis effect).	For the data of the control task, calculate the Bayes factor in favor of the model including the response to the first lattice as predictor compared to a model without the response to the first lattice as predictor, using bridge sampling (Gronau et al., 2017).
H8	The size of individuals' orientation bias will correlate negatively with the size of their estimated true hysteresis and adaptation effects.	Calculate the Bayes factor in favor of a model that assumes the true linear correlation to be negative compared to a model assuming a non-negative true linear correlation, using the Savage-Dickey density ratio method (Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010).
H9	The size of individuals' estimated true hysteresis and adaptation effects at a first timepoint will correlate positively with the size of their estimated true hysteresis and adaptation effects at a second timepoint at least one week later.	Calculate the Bayes factor in favor of a general model that allows for a correlation between individuals' hysteresis (adaptation) effects across sessions compared to a model that assumes uncorrelated individual hysteresis (adaptation) effects per session, using bridge sampling (Gronau et al., 2017). In addition, compare this general model that allows for a correlation between individuals' hysteresis (adaptation) effects across sessions with a model that assumes fully correlated individual hysteresis (adaptation) effects across sessions (cf. Rouder & Haaf, 2019).
H10	The size of individuals' absolute orientation bias as measured at a first timepoint will correlate positively with the size of their absolute orientation bias as measured at a second timepoint at least one week later.	Calculate the Bayes factor in favor of a model that assumes the true linear correlation to be positive compared to a model assuming a non-positive true linear correlation, using the Savage-Dickey density ratio method (Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010).

offered for participation. The only criteria for exclusion from the analyses concerning the first session were (a) incomplete participation to the first session and (b) choosing the diagonal options in the first lattice in the main task in more than 40 percent of the trials (this is interpreted as an indication of random responding, based on the law of proximity, Kubovy et al., 1998). For the analyses including the data for both the first and the second session, the exclusion criteria were (a) incomplete participation to the first or the second session and (b) choosing the diagonal options in the first lattice in the main task in more than 40 percent of the trials in either the first or the second session.

We opted for a sequential Bayes factor design with minimal and maximal n (Schönbrodt & Wagenmakers, 2018). The minimum sample size for the first session was 30 participants. After each 5 additional participants meeting the inclusion criteria, the Bayes factors related to RQ4 and RQ5 was calculated. Data collection would be terminated when either a Bayes factor of 1/6 or 6 was reached for both main research questions (i.e., H4 and H5 in Table 4.1)[3], or when a sample size of 75 participants for the first session was reached (i.e., 2.5 times the original sample size). As we only conduct Bayesian analyses, a sequential stopping rule was allowed and appropriate (Rouder, 2019, 2014; Schönbrodt et al., 2017; Schönbrodt & Wagenmakers, 2018).

Although a Bayes factor of 1/6 or 6 was reached for both main research questions after 55 participants (and we should have stopped according to the preregistered criteria), we continued data collection until data was collected from 75 participants fulfilling the inclusion criteria. The decision to continue was made partly because of logistic reasons (i.e., participation was already scheduled) and partly because of our preference to continue collecting after a sudden direction change in one of the Bayes factors related to H5 going from 50 to 55 participants.[4] The final sample size for the first session therefore consisted of 75 participants between the ages of 18 and 65 years (59 women, 15 men, 1 other, M_{age} = 22.56 years, SD_{age} = 7.92 years). The data of 6 participants were excluded from analyses based on the stated exclusion criteria: 1 participant did not complete the first session and 5 participants chose the diagonal options in the first lattice in the main task in more than 40 percent of the trials. The final sample size for the analyses based on the first and the second session consisted of 72 participants between the ages of 18 and 65 years (57 women, 14 men, 1 other, M_{age} = 22.69 years, SD_{age} = 8.04 years). The data of 9 participants were excluded from analyses based on the stated exclusion criteria: besides the 6 participants who were excluded because of reasons related to the first session, 1 participant did not complete the second session and 2 participants chose the diagonal options in the first lattice in the main task of the second session in more than 40 percent of the trials. As the exclusion criteria for analyses related to the first and second session combined focused on the main

[3]As it was not feasible to use sequential sampling based on all hypotheses, we focused our sampling plan on hypotheses 4 and 5. We view these hypotheses as our main targets for this study. As we believe the conclusiveness of our results depends mostly on the informativeness of the Bayes factors for hypotheses 4 and 5, a Bayes factor less informative than 6 or 1/6 for any of the additional hypotheses would not be as problematic. At least for hypotheses 1-2 and for hypothesis 6, our reanalysis of the data from Schwiedrzik et al. (2014; with our exclusion criteria, N = 27) indicated that this sample size was enough to get a Bayes factor above 6. For hypothesis 7-10, we could not test this as Schwiedrzik et al. (2014) did not include a control task, an absolute orientation bias task, or two separate sessions.

[4]More specifically, the Bayes factor comparing the likelihood of the observed data under the positive effects model and under the unconstrained model for the percept of L1 (i.e., hysteresis effect) went from 3.04 (in favor of positive effects model) to 0.02 (conclusive and in favor of the unconstrained model) when going from 50 to 55 participants. Also the Bayes factor comparing the likelihood of the observed data under the positive effects model and under the unconstrained model for the aspect ratio of L1 (i.e., adaptation effect) suddenly became conclusive going from 50 to 55 participants, albeit without direction change: from 0.55 to 0.05.

task, we did include the data from all 75 participants for the visualizations and analyses relating to the absolute orientation bias task only.

4.2.2 Material

4.2.2.1 Dot lattice stimuli and main task

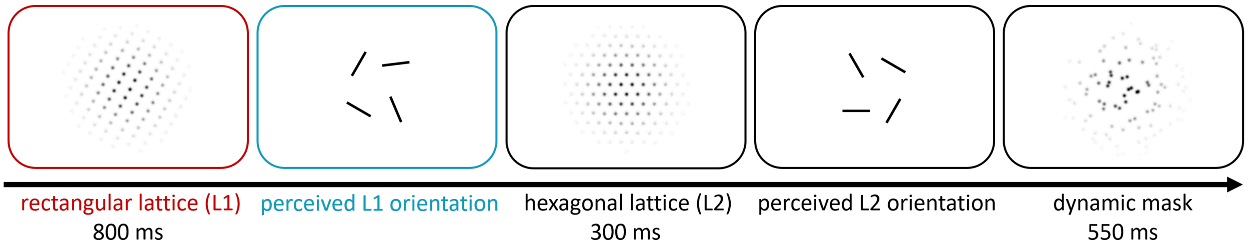


Figure 4.6: Illustration of trial structure. *Note.* For reasons of visibility, the shown trial components in this figure have black dots on a white background. The actual experiment had white dots on a grey background, as indicated in the task description.

A first version of the dot lattice paradigm that as used here was introduced by Gepshtein & Kubovy (2005) and a modified version was used in Schwiedrzik et al. (2014). Each trial (see Figure 4.6) consisted of:

(a) the presentation of a red fixation dot only (1000 ms)

(b) the presentation of a rectangular dot lattice L1 at a randomly chosen 0°-orientation (800 ms), on a gray background. The orientation of the lattice was randomized to minimize the accumulation of perceptual bias across trials. The dot lattice had a diameter of 11.5 degrees of visual angle (dva) and the exact position of the dots in the lattice was jittered between 0 and 1.15 dva to prevent that dots of subsequent displays occupy systematically related portions of space. The 'dots' were white Gaussian blobs with a diameter of 0.25 dva. The inter-dot distance, here defined as center-to-center distance, was kept to +- 1 dva and was varied with aspect ratio so that the product of the distance in the 0°-orientation and in the 90°-orientation ($|a| \times |b|$) was invariant.

(c) a response screen for reporting the percept of L1 (4-AFC; 4 icons with lines parallel to possible organizations: 0°, 90°, and 2 diagonal orientations; duration under observer's control). The position of the response options were randomized across trials. Once the participant had selected one of the four responses by pressing the corresponding key (e/f/i/j), a green circle appeared around the chosen orientation (for 200 ms) and the experiment automatically progressed. This was followed by an additional 100 ms interval, which made the interval between response to the first lattice and presentation of the second lattice 300 ms.

(d) the presentation of a hexagonal dot lattice L2 at the same randomly chosen 0°-orientation as dot lattice L1 (300 ms), on a gray background. The same diameters and inter-dot distances were applied as in (b).

(e) a response screen for reporting the percept of L2 (4-AFC; 4 icons with line parallel to possible organizations: 0°, 60°, 120°, 90° ; duration under observer's control). The position of the response options was randomized across trials. Once the participant had selected one of the four responses by pressing the corresponding key (e/f/i/j), a green circle appeared around the chosen orientation (for 200 ms) and the experiment automatically progressed. This was followed by an additional 100 ms interval, which made the interval between response to the second lattice and presentation of the mask 300 ms.

(f) mask presented on a gray background (550 ms; dynamic random dot mask updated at 25 Hz).

The red fixation dot was continuously present in the center of the screen. Participants were instructed to fixate on the central fixation dot, and to report the first perceived organization in case the percept switched during the presentation period of the target stimulus (either L1 or L2). There were 21 practice trials to get participants acquainted with the task.

The independent variable is the inter-dot distance ratio in the first dot lattice stimulus (i.e., |a|/|b| = aspect ratio of L1). This ratio varied between 1.3^{-1} and 1.3, with values of 1.3^{-1}, 1.2^{-1}, 1.1^{-1}, 1, 1.1, 1.2, and 1.3.

The dependent variables are the individual reports of the percept of the first (L1) and of the second dot lattice (L2) in each trial. Dominant percepts at aspect ratio equal to 1 are parallel to the orientations 0° and 90° in the first lattice, and parallel to orientations 0°, 60°, and 120° in the second lattice.

The 0°-orientation in each trial was randomly chosen, covering 90° in steps of 1°.

As in Schwiedrzik et al. (2014), each participant was asked to complete 9 blocks of 70 trials, with 10 trials for each of the 7 aspect ratios per block. The order of trials was pseudorandomized: each aspect ratio occurred equally often in each block, but otherwise the order within each block was randomized. Furthermore, the location of the four response options within and between trials was also randomized.

4.2.2.2 Control task

To control for the possibility of the hysteresis effect being a purely decisional rather than a perceptual effect, we included the control task presented by Schwiedrzik et al. (2018) as an additional task in our study. This control task was equal to the main task, with the exception of the presentation of the first lattice. In this control task, the first lattice in each trial was a random dot lattice instead of a rectangular dot lattice, as this random dot lattice cannot induce a particular orientation. The response screen for the first lattice in each trial included the relative 0°, 90°, 45° and 135° orientations (i.e., the 2 diagonal orientations for a lattice with an aspect ratio of 1). Each participant was asked to complete 1 block of 90 trials. The order of the trials was randomized, as well as the location of the four response option within and between trials. There were 3 practice trials to get participants acquainted with the task.

4.2.2.3 Absolute orientation bias

As we expected the effects of hysteresis and adaptation to be smaller when a strong absolute orientation bias was present, we included a task with ambiguous hexagonal dot lattices only, varying in absolute orientation with the a orientation from 1° to 60°. In every hexagonal lattice, six different orientations can be perceived, of which three are most and equally dominant in general. Four blocks of 60 trials were presented, with every absolute orientation shown once per block and the presentation order randomized within each block. There were 5 practice trials to get participants acquainted with the task.

Each trial consisted of:

(a) the presentation of a red fixation dot only (750 ms)

(b) the presentation of a hexagonal dot lattice at a randomly chosen 0°-orientation varying between 1° and 60° (500 ms), on a gray background. The same diameters and inter-dot distances were applied as in the main task described above.

(c) a response screen for reporting the percept of the hexagonal lattice (4-AFC; 4 icons with lines parallel to possible organizations: 0°, 60°, 120°, 90° ; duration under observer's control). The position of the response options was randomized across trials. Once the participant had selected one of the four responses by pressing the corresponding key (e/f/i/j), a green circle appeared around the chosen orientation (for 200 ms) and the experiment automatically progressed. This was followed by an additional 200 ms interval, which made the interval between response to the lattice and presentation of the next 1150 ms (200 ms feedback, 200 ms interval, 750 ms fixation dot).

4.2.3 Procedure

The experimental sessions took place in a darkened room using a cathode ray tube monitor ViewSonic G90fB, 1024 by 768 pixels, at 60 cm distance, refresh rate 60 Hz. Participants' stable head position was guaranteed by using a chinrest with forehead support. The dot lattice stimuli were generated in Matlab 2018b using the code of Schwiedrzik et al. (2014). [5] Stimulus presentation and response collection was controlled using Python 3 (Van Rossum & Drake Jr, 1995) and the PsychoPy library (Peirce, 2007). In the first session, participants first completed the orientation bias task, then the main task measuring hysteresis and adaptation, and finally the control task. In the second session, participants completed the orientation bias task and the main task measuring hysteresis and adaptation for the second time. The second session took place at least one week after the first session, with a minimum of 7 days and a maximum of 14 days apart.[6]

4.2.4 Data analysis

We used R [Version 4.0.4; R Core Team (2021)] for all our analyses.[7] All models were fitted using the R package brms (Bürkner, 2017, 2018). The analysis procedure described below (except for the analyses related to H7, H8, and H10) had been worked out and was tested on the data previously collected by Schwiedrzik et al. (2014). [8]

4.2.4.1 Preprocessing

Planned analyses were restricted to the response alternatives with equal likelihood at aspect ratio equal to 1. This means that only trials in which participants responded 0° or 90° for the first lattice and 0°, 60°, or 120° for the second lattice were used. For this reason, we excluded 9026 out of 47250 trials (19.1%) from analyses of the main task in the first session, as well as 3807 out of 6750 trials (56.4%) for the control task in the first session[9], and 6434 out of 45360 trials (14.18%) for the main task in the second session. In the absolute orientation bias task, 1848 out of 18000 trials

[5] We thank the authors for sharing their code with us.

[6] For two participants the sessions were more than 14 days apart (i.e., 17 and 21 days), as the second session had to be rescheduled because of medical reasons. Nevertheless, we included their data for the second session in the final analysis.

[7] For an overview of all R packages used, see Appendix B.

[8] We thank the authors for sharing their data with us for this purpose. The results from this reanalysis are available on the Open Science Framework: https://osf.io/xa5ut/

[9] For the control task, approximately 50% of trials are excluded, which is to be expected given that no orientation was visible in the first lattice (and all responses should be chosen approximately equally).

(10.27%) with 90° responses were excluded in the first session, and 1045 out of 18000 trials (5.81%) in the second session.

For visualization purposes, we computed, per participant and on average, the logit of the probability to perceive the 0° orientation in the first stimulus (i.e., logit[p(l1 → 0°)]) and the logit of the probability to perceive the 0° orientation in the second stimulus given that the first stimulus was perceived as orientation 0° or orientation 90° (i.e., logit[p(l2 → 0°)] for l1 → 0° and for l1 → 90°) to overcome floor effects at high aspect ratios[10]:

$$logit[p(l1 \to 0°)] = ln\left[\frac{p(l1 \to 0°)}{1 - p(l1 \to 0°)}\right]$$

and

$$logit[p(l2 \to 0°)] = ln\left[\frac{p(l2 \to 0°)}{1 - p(l2 \to 0°)}\right].$$

To determine the preferred orientation direction and the size of the individual's absolute orientation bias, we calculated the direction and magnitude of the orientation vector per participant (cf. Curray, 1956). The orientation vector is the vector of all chosen orientations, excluding trials in which participants chose the unlikely 90° orientation in the hexagonal lattices (1848 trials out of 18000 were excluded for this reason in the first session and 1045 trials out of 18000 in the second session). The vector direction can be interpreted as the preferred orientation direction, whereas the vector magnitude, which varies from 0 to 100%, can be interpreted as the strength of the absolute orientation bias. Vector magnitude (L) and direction ($\overline{\theta}$) were calculated as follows (Curray, 1956):

$$L = \frac{\sqrt{(\sum n\, sin\, 2\,\theta)^2 + (\sum n\, cos\, 2\,\theta)^2}}{\sum n} * 100$$

$$\overline{\theta} = \frac{1}{2}\, arctan \frac{\sum n\, sin\, 2\,\theta}{\sum n\, cos\, 2\,\theta}.$$

4.2.4.2 Data visualizations

We plot the average and individual results on probability scale and logit scale for perceiving the first lattice as orientation 0° (Y-axis: logit[p(l1 → 0°)]; X-axis: aspect ratio L1) and for perceiving the second lattice as orientation 0° (Y-axis: logit[p(l2 → 0°)]; X-axis: aspect ratio L1; grouping var = l1 → 0° or l1 → 90°). As relative grouping strength of the dots in a lattice among a certain orientation has been shown to follow a decreasing exponential trend in function of the relative inter-dot distance in that orientation (Kubovy et al., 1998), the logit of the probability is approximately linear. Vertical separation of the two lines reflects the size of the perceptual hysteresis effect; the slope of both lines reflects the size of the perceptual adaptation effect. We also plot the results regarding absolute orientation bias, on average, per individual, and per block.

Regarding the individual estimates of the hysteresis and adaptation effect, we plot mean estimates and 95% highest density continuous intervals for the hysteresis and adaptation effect separately, the correlation between

[10] Gepshtein & Kubovy (2005) and Schwiedrzik et al. (2014) used an alternative calculation of the logodds (cf. p. 489 in Gepshtein & Kubovy, 2005). We provide these alternative visualizations in the supplementary material for this paper to facilitate the comparison to these papers.

individual hysteresis and adaptation effects, as well as the correlation between the individual orientation bias and the size of the estimated individual hysteresis and adaptation effects.

4.2.4.3 Model estimation

The full model used to estimate individual hysteresis and adaptation effects is a Bayesian multilevel binary logistic regression model predicting the percept of the second lattice (Y_{ijkl}), with aspect ratio of the first lattice (AR) and the percept of the first lattice ($R10$) as fixed and random effects. The model thus includes fixed and individual random effects for percept in the first lattice (i.e., hysteresis effect) as well as aspect ratio in the first lattice (i.e., adaptation effect), and individual random intercepts.

4

Y_{ijkl} stands for the response variable, more specifically the percept of the second lattice, for the lth replicate for the ith participant, i = 1, . . . , I in the jth condition for aspect ratio of the first lattice (AR), j = 1, . . . , 7 and the kth condition for the percept of the first lattice ($R10$), k = 1, 2 with l = 1, . . . , L_{ijk}. I is the number of participants in the data. Y_{ijkl} is modelled to follow a Bernouilli distribution with a probability p_{ijkl} of the second lattice being perceived as the 0° orientation. The percept of the first and the second lattice can be 0 (when different from the 0° orientation in the lattice) or 1 (when equal to the 0° orientation in the lattice). Centered aspect ratio was used, which means that a value of zero corresponds to an aspect ratio of 1, a value of 1.1^{-1} -1 (i.e., ≈ -0.09) corresponds to 1.1^{-1}, and a value of 1.1^{1} -1 (i.e., 0.10) to an aspect ratio of 1.1.

$$Y_{ijkl} \sim Bernouilli(p_{ijkl})$$

$$log\left(\frac{p_{ijkl}}{1-p_{ijkl}}\right) = \beta_0 + \beta_j AR + \beta_k R10 + \beta_{i0} + \beta_{ij} AR + \beta_{ik} R10$$

β_0 represents the fixed intercept, whereas β_j and β_k represent the fixed adaptation and hysteresis effect, respectively. β_{i0}, β_{ij}, and β_{ik} represent the individual random intercepts, the individual random slopes of aspect ratio of the first lattice (i.e., adaptation effects), and the individual random slopes of percept of the first lattice (i.e., hysteresis effect), respectively. Another way to formulate the model is:

$$R20 \sim Intercept + AR + R10 + (Intercept + AR + R10 \mid participant).$$

Figure 4.7 visualizes the priors we specified for the fixed effects, for the standard deviation of the random effects, and for the correlation matrix.

We fitted this model of perceived L2 orientation using brms (Bürkner, 2017, 2018). We used 4 chains with 20000 iterations each with the default number of warmup iterations per chain. In case of computational issues we could have decided to deviate from the specified number of iterations, but this was not necessary. We used a delta equal to .8 and a maximum treedepth of 10. For any other sampling specifications we used the default settings when possible.

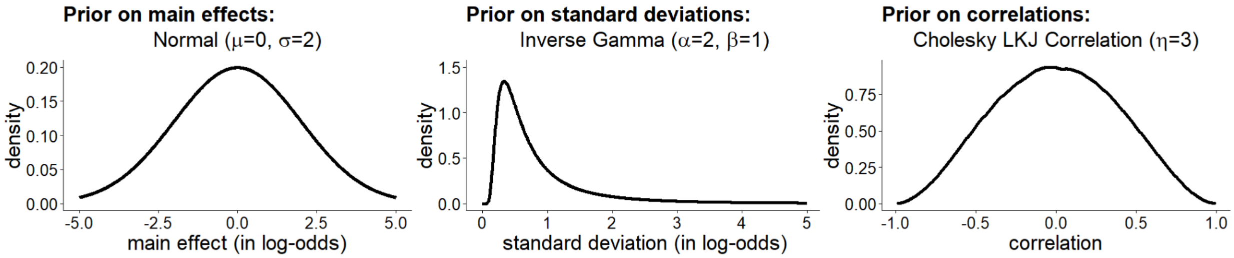


Figure 4.7: Illustration of priors used in the model predicting the percept of L2.

4.2.4.4 Average hysteresis effect (H1)

To test the presence of an average hysteresis effect across individuals, we compared a model including the percept of the first lattice as predictor versus a model without the percept of the first lattice as predictor and calculated the Bayes factor in favor of the model including the hysteresis effect, using bridge sampling (Gronau et al., 2017). In case the Bayes factor was in favor of the model including the hysteresis effect, we report the mean and 95% highest density continuous interval (HDCI) for the coefficient related to the percept of L1 in the full model described above, to have an estimate of the size of the average hysteresis effect.

4.2.4.5 Average adaptation effect (H2)

To test the presence of an average adaptation effect across individuals, we compared a model including the aspect ratio of the first lattice as predictor versus a model without the aspect ratio of the first lattice as predictor and calculated the Bayes factor in favor of the model including the adaptation effect, using bridge sampling (Gronau et al., 2017). In case the Bayes factor was in favor of the model including the adaptation effect, we report the mean and 95% highest density continuous interval (HDCI) for the coefficient related to the aspect ratio of L1 in the full model, to have an estimate of the size of the average adaptation effect.

4.2.4.6 Independence of average hysteresis and adaptation effects (H3)

To test the independence of the average hysteresis and adaptation effects, we compared a model including the interaction between the percept and the aspect ratio of the first stimulus as predictor versus a model without the interaction and calculated the Bayes factor in favor of the model without the interaction, using bridge sampling (Gronau et al., 2017). In case the Bayes factor was in favor of the model including the interaction effect, we report the mean and 95% highest density continuous interval (HDCI) for the interaction coefficient in a full model including the interaction and all random effects, to have an estimate of the size of the average interaction effect.

4.2.4.7 Individual hysteresis and adaptation effects: Do individual effects differ? (H4)

To test whether individual hysteresis and adaptation effects differ in size, we calculated the Bayes factor in favor of a model including random intercepts and slopes for every participant compared to a model including no random slopes (cf. unconstrained model vs. common effects model in Haaf & Rouder, 2019), using bridge sampling (Gronau et al., 2017). We conducted this model comparison for each effect separately.

4.2.4.8 Individual hysteresis and adaptation effects: Does everyone show the effects? (H5)

To test whether every individual participant shows a positive hysteresis or adaptation effect, we calculated the Bayes factor in favor of a model predicting a positive effect size for every participant compared to a model that does not place any order or equality constraints on individuals' effects, using the encompassing approach (cf. positive effects model vs. unconstrained model in Haaf & Rouder, 2019). In the positive-effects model, the main hysteresis and the main adaptation effect are both restricted to be positive. The model comparison was done for each effect separately, however.

4.2.4.9 Does the size of hysteresis and adaptation effects correlate positively across individuals? (H6)

To determine the size of the hysteresis effect, we used the individual estimates for the effect of the percept of the first lattice on the percept of the second lattice. To determine the size of the adaptation effect, we used the individual estimates for the effect of aspect ratio of the first lattice on the percept of the second lattice. These estimates are based on the Bayesian model of the percept of the second lattice described above, with the aspect ratio of the first lattice and the percept of the first lattice as fixed effects, with random intercepts and random slopes for both hysteresis and adaptation effects.

To test whether the size of individuals' hysteresis effect correlates positively with the size of their adaptation effect, we calculated the Bayes factor in favor of a model that assumes the true linear correlation to be positive compared to a model assuming a non-positive true linear correlation using the Savage-Dickey density ratio method (Wagenmakers et al., 2010). As this is a one-sided hypothesis, the Bayes factor is equal to the posterior probability under the hypothesis ($r > 0$) against its alternative ($r <= 0$). To have an estimate of the strength of the correlation, we report the mean and 95% HDCI for the correlation between estimated individual hysteresis and adaptation effects, based on the full model described above.

4.2.4.10 Is the hysteresis effect absent in the control task? (H7)

To test the presence of an average hysteresis effect across individuals in the control task, we compared a model including the response to the first lattice as predictor versus a model without the response to the first lattice as predictor and calculated the Bayes factor in favor of the model without the hysteresis effect, using bridge sampling (Gronau et al., 2017). In case the Bayes factor was in favor of the model including the hysteresis effect, we report the mean and 95% highest density continuous interval (HDCI) for the coefficient related to the response to the first lattice in a model including the response to the first lattice as main and random effect, to have an estimate of the size of the effect.

4.2.4.11 Do individual differences in absolute orientation bias correlate negatively with hysteresis and adaptation effects? (H8)

To test whether the size of individuals' orientation bias correlates negatively with the size of their hysteresis and adaptation effects, we calculated the Bayes factor in favor of a model that assumes the true linear correlation to be negative compared to a model assuming a non-negative true linear correlation, using the Savage-Dickey density ratio method (Wagenmakers et al., 2010). As this is a one-sided hypothesis, the Bayes factor is equal to the posterior probability under the hypothesis ($r < 0$) against its alternative ($r >= 0$). We conducted this model comparison for each effect separately. To have an estimate of the strength of the correlation, we report the mean and 95% HDCI for the correlation between individual orientation bias estimates and individual hysteresis (adaptation) effects.

4.2.4.12 Does the size of individuals' hysteresis and adaptation effects correlate positively across timepoints? (H9)

To test whether the size of individuals' hysteresis effect correlates positively across timepoints, we calculated the Bayes factor in favor of a general model that allows for a correlation between individuals' hysteresis (adaptation) effects across sessions compared to a model that assumes uncorrelated individual hysteresis (adaptation) effects per session (Rouder & Haaf, 2019), using bridge sampling (Gronau et al., 2017). In addition, we compared this general model that allows for a correlation between individuals' hysteresis (adaptation) effects across sessions with a model that assumes fully correlated individual hysteresis (adaptation) effects across sessions (Rouder & Haaf, 2019). We conducted these model comparisons for each effect separately. To have an estimate of the strength of the temporal stability, we report the mean and 95% HDCI for the correlation between individual hysteresis (adaptation) estimates across sessions, based on the winning model (in case the winning model is not the model assuming the absence of a correlation).

4.2.4.13 Does the size of individuals' absolute orientation biases correlate positively across timepoints? (H10)

To test whether the size of individuals' absolute orientation bias correlates positively across timepoints, we calculated the Bayes factor in favor of a model that assumes the true linear correlation to be positive compared to a model assuming a non-positive true linear correlation using the Savage-Dickey density ratio method (Wagenmakers et al., 2010). As this is a one-sided hypothesis, the Bayes factor is equal to the posterior probability under the hypothesis ($r > 0$) against its alternative ($r <= 0$). To have an estimate of the strength of the temporal stability, we report the mean and 95% HDCI for the correlation between individual orientation bias estimates across sessions.

4.3 Results

In Figures 4.8 and 4.9 one can find the results on logit scale on average and per participant respectively. The same figures representing the results on probability scale can be found in Appendix B (see Figures B.1 and B.2). In addition, graphs using the alternative logit calculation as used by Gepshtein & Kubovy (2005) and Schwiedrzik et al. (2014) are provided in Appendix B too (see Figures B.3 and B.4).

4.3.1 Confirmatory analyses

4.3.1.1 Average hysteresis and adaptation effects? (H1-2)

The Bayes factor in favor of the model including the influence of the L1 percept is very large, with the exact value outside of computer precision. This means that the data are more likely under the model with the hysteresis effect. The Bayes factor in favor of the model including the influence of aspect ratio on the second lattice is 8×10^{25}. This means that the data are more likely under the model with the adaptation effect. For a visual representation of the average predicted hysteresis and adaptation effects in the full model, see Figures 4.8 and B.1.

Figure 4.10 shows the posterior distributions of the fixed effects, standard deviation of random effects, and the correlation between the random effects in the model predicting the perceived orientation in the second lattice. Figure 4.10 a shows the posteriors for the effect of the perceived orientation in the first lattice (i.e., hysteresis effect) and the effect of aspect ratio (i.e., adaptation effect) on the perceived orientation in the second lattice. The

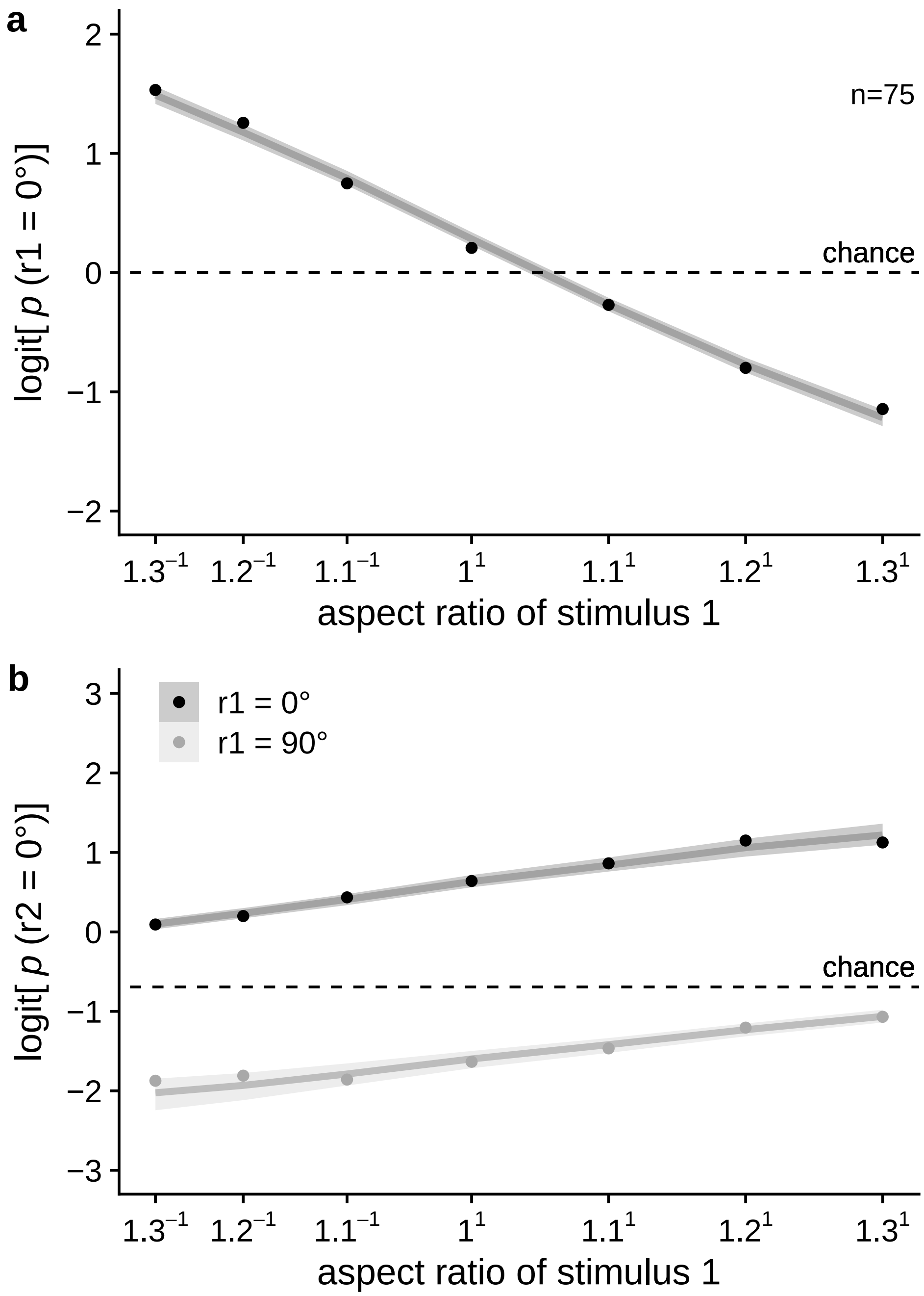


Figure 4.8: (a) Mean response to the first stimulus dependent on aspect ratio (logit). The probability of responding 0° to the first stimulus decreases with aspect ratio (|a|/|b|). The value of aspect ratio increases with increasing distance in the 0°-orientation, leading to more 90° responses. (b) Mean response to the second stimulus dependent on aspect ratio (logit). The probability of responding 0° to the second stimulus increases with aspect ratio (|a|/|b|; i.e., adaptation effect), and increases when the first stimulus was perceived as 0° rather than 90° (i.e., hysteresis effect). Dots indicate observed values. In addition, mean posterior predictions and their 95% highest density continuous intervals are shown.

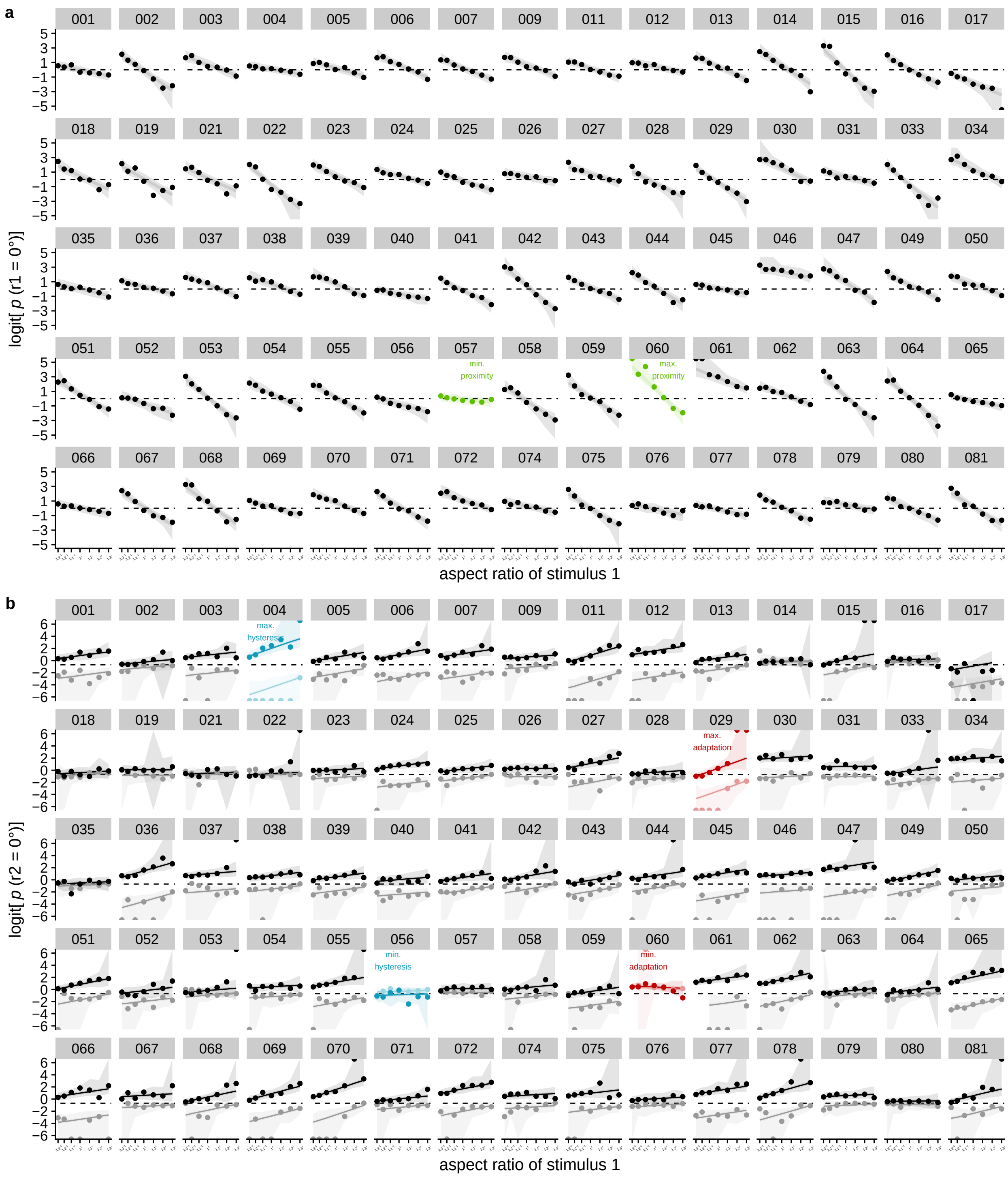


Figure 4.9: (a) Mean individual responses to the first stimulus dependent on aspect ratio (logit). The probability of responding 0° to the first stimulus decreases with aspect ratio (|a|/|b|). The value of aspect ratio increases with increasing distance in the 0°-orientation, leading to more 90° responses. Dots indicate observed values. In addition, mean posterior predictions and their 95% highest density continuous intervals are shown. Plots for participants with the smallest and largest estimated proximity effect are indicated in green. (b) Mean individual responses to the second stimulus dependent on aspect ratio (logit). The probability of responding 0° to the second stimulus increases with aspect ratio (|a|/|b|; i.e., adaptation effect), and increases when the first stimulus was perceived as 0° rather than 90° (i.e., hysteresis effect). Dots indicate observed values. In addition, mean posterior predictions and their 95% highest density continuous intervals are shown. Plots for participants with the smallest and largest estimated hysteresis effect are indicated in blue, participants with the smallest and largest estimated adaptation effect are indicated in red.

95% highest density continuous interval for the main hysteresis effect ranges from 2.01 to 2.64. The 95% highest density continuous interval for the main adaptation effect ranges from 1.64 to 2.38. Figure 4.11 shows the estimated individual effects of perceived L1 orientation and aspect ratio of L1 in the model predicting perceived L2 orientation.

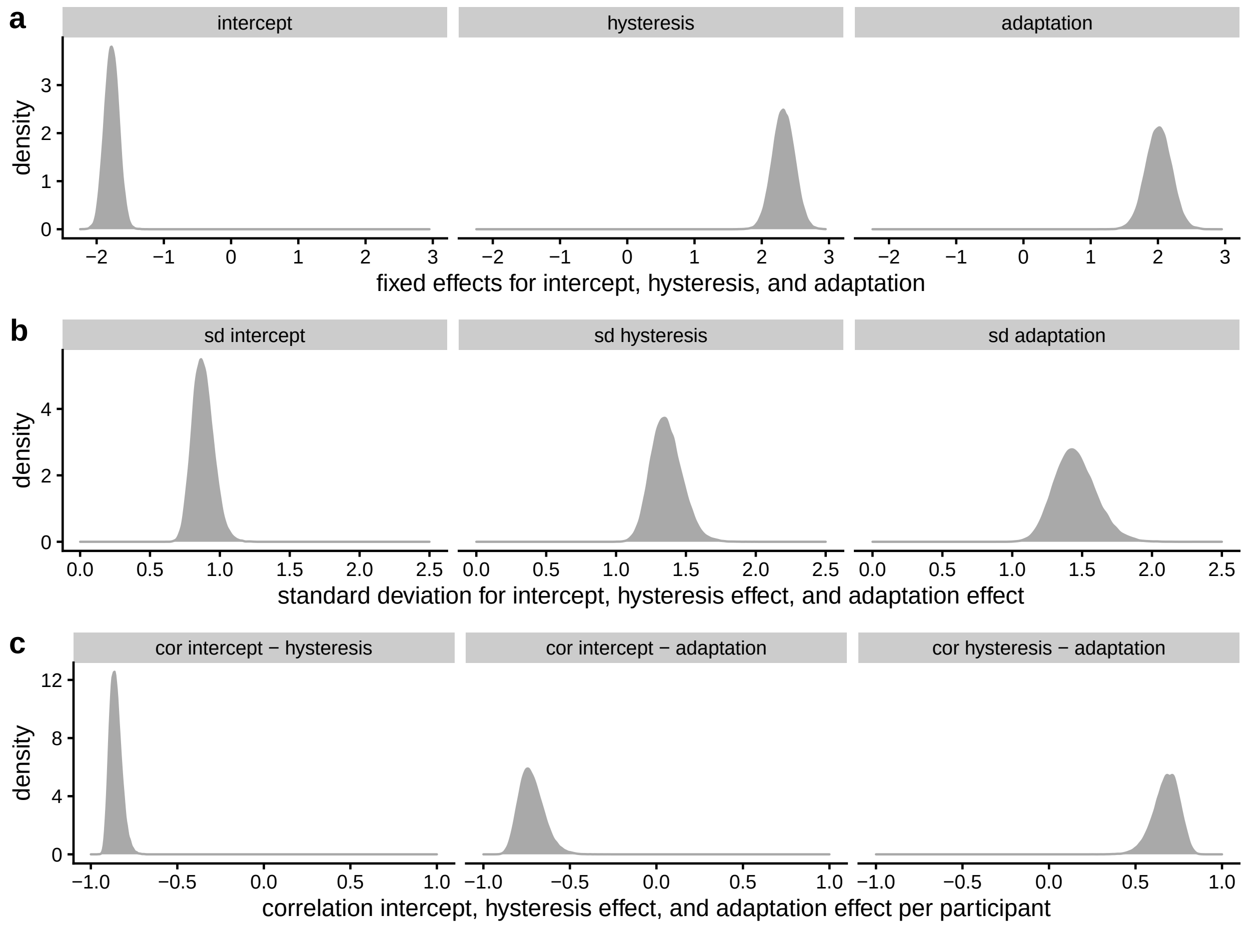


Figure 4.10: Posterior distributions of fixed effects, standard deviation of random effects, and the correlation between the random effects for the model of perceived L2 orientation.

4.3.1.2 Absence of interaction effect between hysteresis and adaptation? (H3)

The Bayes factor in favor of the model including no interaction compared to the model including an interaction is 7.3039. This means that the data are more likely under the model without the interaction between the hysteresis and adaptation effect.

4.3.1.3 Are there individual differences in the size of hysteresis and adaptation effects? (H4)

The Bayes factor in favor of the model with a random effect for percept L1 (i.e., a random hysteresis effect) compared to the common effects model is very large, with the exact value outside of computer precision. This means that the observed data are more likely under the unconstrained model than under the common effects model. The Bayes factor in favor of the model with a random effect for aspect ratio (i.e., a random adaptation effect) compared to the common effects model is 2×10^{45}. This means that the observed data are more likely under the unconstrained model than under the common effects model. These Bayes factors indicate that it is much more likely to assume

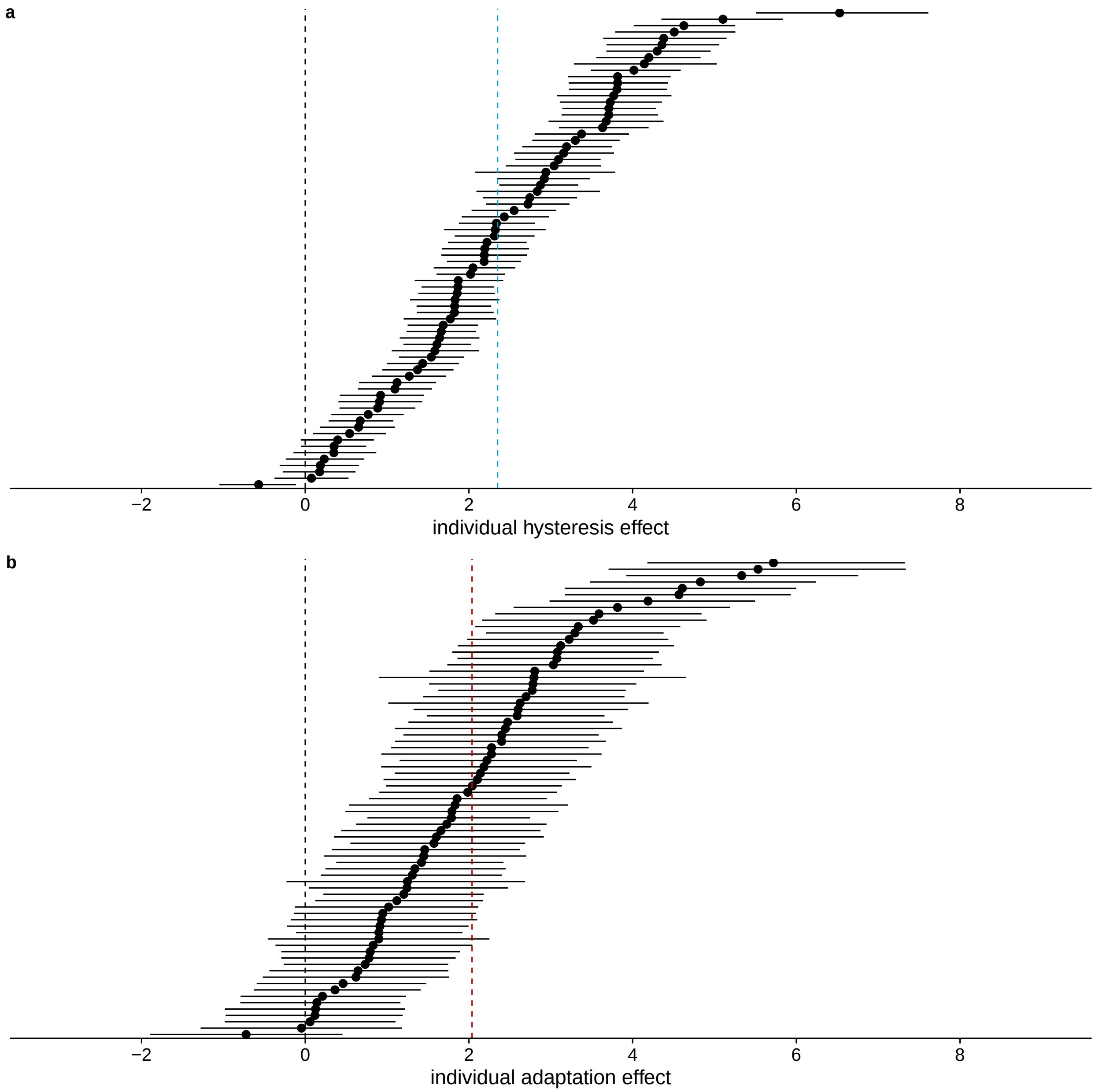


Figure 4.11: Slopes for the effect of perceived L1 orientation and aspect ratio on perceiving the 0° orientation in L2 per participant. Mean and 95% highest density continuous intervals are shown. The colored line indicates the average mean effect across participants. The black line indicates a slope of zero.

individual differences in both the hysteresis and adaptation effects than to assume everyone to show the same effect sizes.

4.3.1.4 Does everyone show hysteresis and adaptation? (H5)

The Bayes factor comparing the likelihood of the observed data under the positive effects model and under the unconstrained model for the percept of L1 (i.e., hysteresis effect) is 0.0228 (inverse BF: 43.8232). This means that the observed data are less likely under the positive effects model than under the unconstrained model. The Bayes factor comparing the likelihood of the observed data under the positive effects model and under the unconstrained model for aspect ratio of L1 (i.e., adaptation effect) is 0.0145 (inverse BF: 69.1914). This means that the observed data are less likely under the positive effects model than under the unconstrained model. These Bayes factors indicate that it is more likely to assume that not everyone shows a hysteresis or adaptation effect than to assume that everyone shows these effects.

4.3.1.5 Correlation between individual hysteresis and adaptation effects? (H6)

Figure 4.12 shows the correlation between the individual slopes for aspect ratio and perceived L1 orientation in the model predicting perceived L2 orientation. The Bayes factor in favor of a model that assumes the true linear correlation to be larger than zero compared to a model assuming a true linear correlation smaller than or equal to zero is larger than 1×10^4. This means that the observed data are more likely under the model assuming a positive linear correlation between individual hysteresis and adaptation effects than under the model assuming a non-positive linear correlation. The 95% highest density continuous interval for the correlation between individual effects of perceived L1 orientation and aspect ratio on the perceived L2 orientation ranges from 0.53 to 0.81.

4.3.1.6 Absence of hysteresis effect in the control task? (H7)

The Bayes factor in favor of the model including the influence of the L1 percept for the data of the control task is 2×10^{29}. This means that the data are more likely under the model with the hysteresis effect. The 95% highest density continuous interval for the perceived L1 orientation coefficient in the model including a fixed and random hysteresis effect per participant in the control task ranges from 0.67 to 1.23. Although this means that the hysteresis effect is present in the control task, the effect is remarkably smaller than in the experimental hysteresis and adaptation task (see Figure 4.13). In addition, several participants do not show an irrefutably positive hysteresis effect in the control task. For an overview of the individual estimated hysteresis effects in the experimental and control task, see Figure 4.14.

4.3.1.7 Correlation with strength of absolute orientation bias? (H8)

The direction and magnitude of the orientation bias per participant can be found in Appendix B (see Figures B.11 to B.21). Figure 4.15 shows the correlation between the magnitude of the absolute orientation bias per individual and the individual slopes for aspect ratio and perceived L1 orientation in the model predicting perceived L2 orientation for the first session. The Bayes factor in favor of a model that assumes the true linear correlation between the individual hysteresis effects and the magnitude of the absolute orientation bias for the first session to be smaller than zero compared to a model assuming a true linear correlation larger than or equal to zero is 0.006 (inverse BF: 165.6667). This means that the observed data are less likely under the model assuming a negative linear correlation

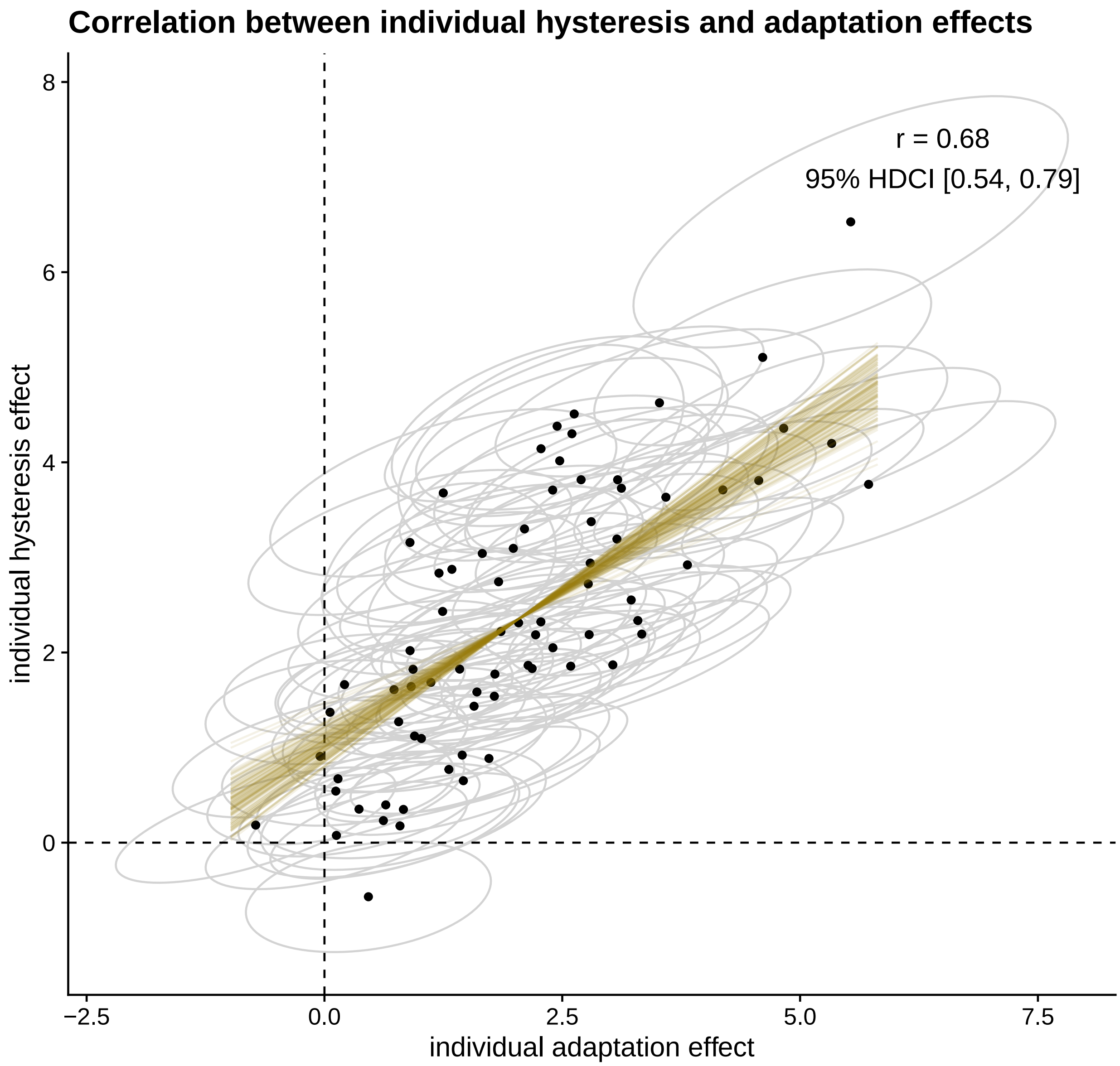


Figure 4.12: Correlation between individual slopes for the effect of aspect ratio and perceived L1 orientation on perceiving the 0° orientation in L2. Mean and 80% highest density continuous intervals per individual are shown. The black lines indicate a slope of zero. The colored lines give examples of plausible correlation estimates. *Note.* As the estimated correlation value shown comes from a hierarchical model including both estimates of the hysteresis and the adaptation effect, potential attenuation of the correlation as a result of noise is already taken into account.

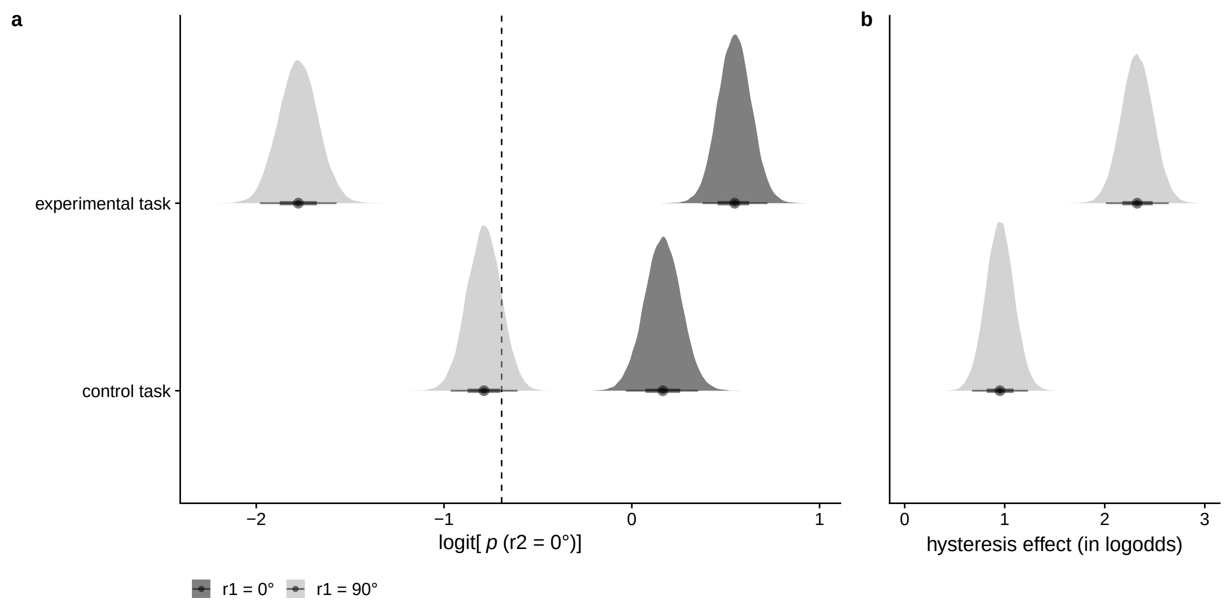


Figure 4.13: (a) Posterior distribution for the percept of the second lattice separately for each percept of the first lattice in the control task and in the experimental task. (b) Estimated hysteresis effect in the control task and the experimental task. Mean, 66%, and 95% highest density continuous intervals are shown. The vertical black line indicates chance level.

between individual hysteresis and absolute orientation bias effects than under the model assuming a non-negative linear correlation. The 95% highest density continuous interval for the correlation between the individual hysteresis effects and the magnitude of the absolute orientation bias ranges from 0.06 to 0.51, with a mean of 0.29. In addition to the planned analysis above, we calculated the Bayes factor in favor of a model assuming the true linear correlation between the individual hysteresis effects and the magnitude of the absolute orientation bias for the first session to be larger than zero compared to a model assuming a true linear correlation smaller than or equal to zero. This Bayes factor is equal to 165.6667, meaning that the observed data are more likely under the model assuming a positive linear correlation between individual hysteresis and absolute orientation bias effects than under the model assuming a non-positive linear correlation.

The Bayes factor in favor of a model that assumes the true linear correlation between the individual adaptation effects and the magnitude of the absolute orientation bias for the first session to be smaller than zero compared to a model assuming a true linear correlation larger than or equal to zero is 0.0085 (inverse BF: 118.0476). This means that the observed data are less likely under the model assuming a negative linear correlation between individual hysteresis and absolute orientation bias effects than under the model assuming a non-negative linear correlation. The 95% highest density continuous interval for the correlation between the individual adaptation effects and the magnitude of the absolute orientation bias ranges from 0.05 to 0.49, with a mean of 0.27.

In addition to the planned analysis above, we calculated the Bayes factor in favor of a model assuming the true linear correlation between the individual adaptation effects and the magnitude of the absolute orientation bias for the first session to be smaller than zero compared to a model assuming a true linear correlation larger than or equal to zero. This Bayes factor is equal to 118.0476, meaning that the observed data are more likely under the model

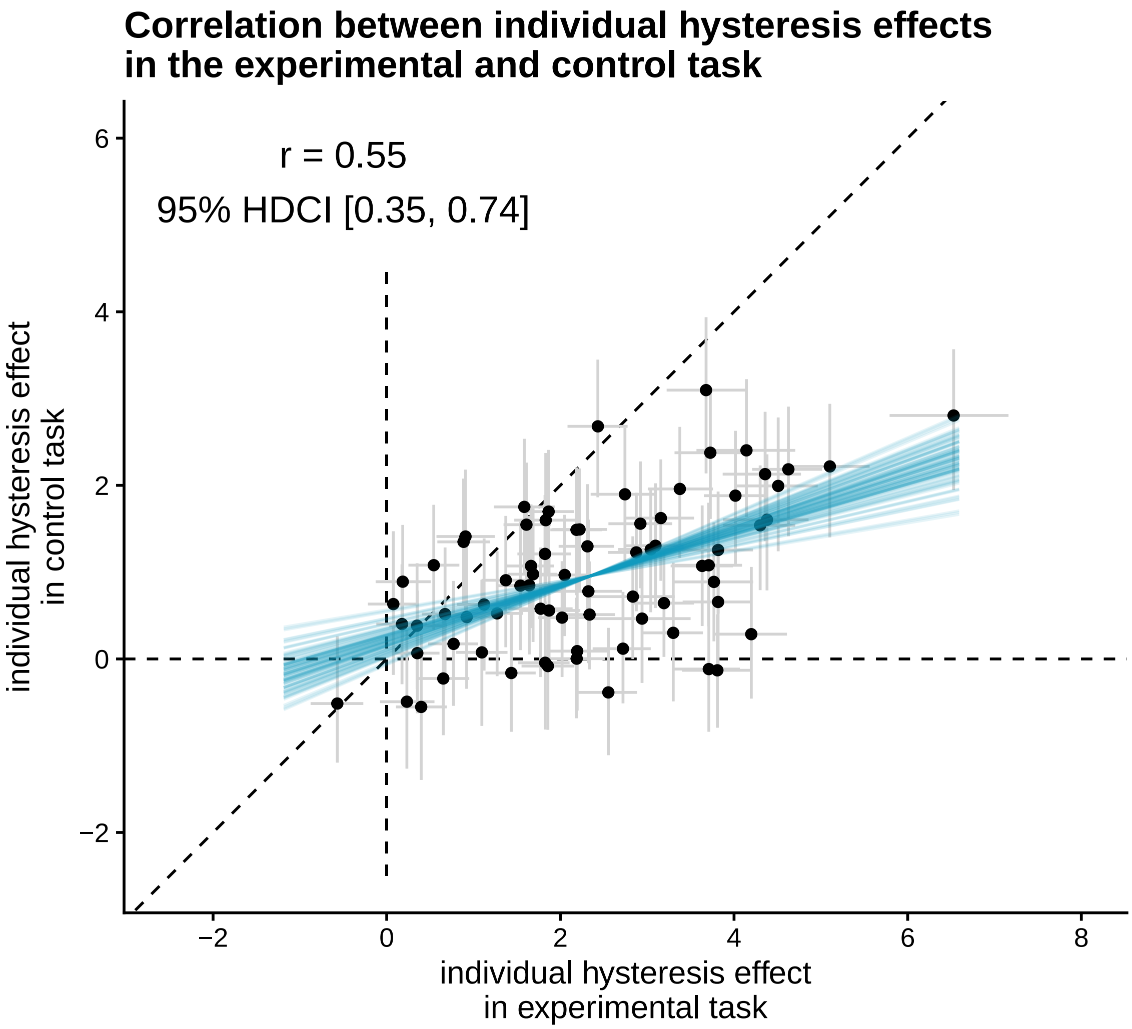


Figure 4.14: Correlation of estimated individual hysteresis effects in the experimental and control task. Mean and 80% highest density continuous intervals are shown. The diagonal black line indicates equal effects in control and experimental task. The horizontal and vertical black lines indicate a hysteresis effect of zero. The blue lines give examples of plausible correlation estimates.

assuming a positive linear correlation between individual adaptation and absolute orientation bias effects than under the model assuming a non-positive linear correlation.

Furthermore, we explored whether a quadratic model could better fit the data than a positive linear relation. For the hysteresis effect the Bayes factor of the model assuming a quadratic relation compared to a model assuming a linear relation is equal to 0.2584 (inverse BF: 3.8693), meaning that the observed data are less likely under the model assuming a quadratic relation between individual hysteresis and absolute orientation bias effects than under the model assuming a linear relation. For the adaptation effect the Bayes factor of the model assuming a quadratic relation compared to a model assuming a linear relation is equal to 0.4654 (inverse BF: 2.1487), meaning that the observed data are less likely under the model assuming a quadratic relation between individual adaptation and absolute orientation bias effects than under the model assuming a linear relation.

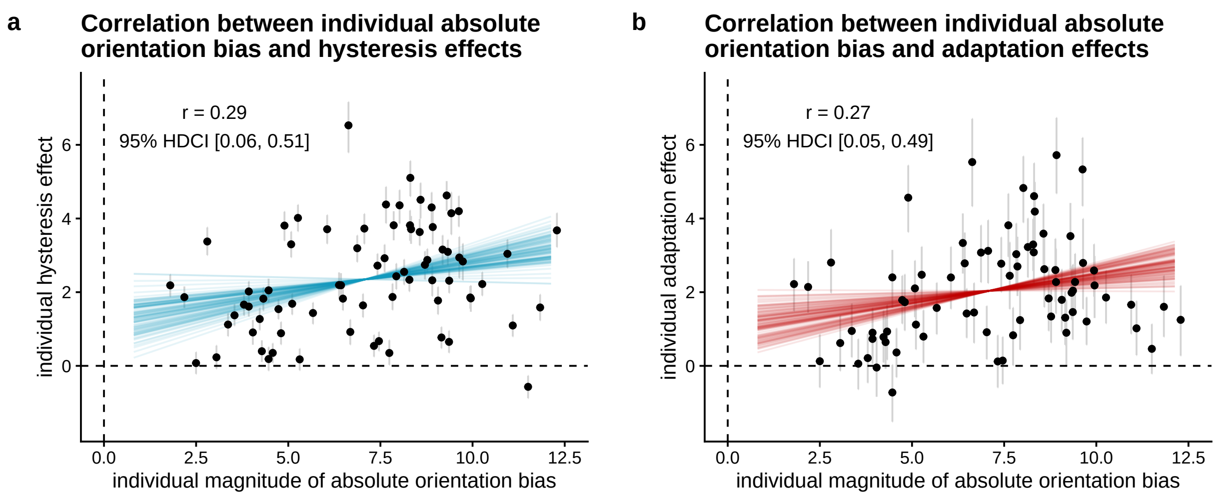


Figure 4.15: (a) Correlation between individual slopes for the effect of the aspect ratio of L1 on perceiving the 0° orientation in L2 and magnitude of the absolute orientation bias in the first session. (b) Correlation between individual slopes for the effect of perceived L1 orientation on perceiving the 0° orientation in L2 and magnitude of the absolute orientation bias in the first session. Mean and 80% highest density continuous intervals for the hysteresis and adaptation effects are shown. The black lines indicate an effect of zero. The colored lines give examples of plausible correlation estimates.

4.3.1.8 Temporal stability of individual differences in strength of hysteresis and adaptation effects? (H9)

In Supplementary Figures B.5, B.6, and B.7 one can find the results on logit scale on average and per participant for both sessions separately. The same figures representing the results on probability scale can be found in the Supplementary Figures B.8, B.9, and B.10.

The Bayes factor in favor of the model that allows for a correlation between individuals' hysteresis effects across sessions compared to a model assuming uncorrelated individual hysteresis effects is 2×10^{19}. This means that the observed data are more likely under the model allowing for a correlation between individual hysteresis effects across sessions than under the model assuming uncorrelated effects. The Bayes factor in favor of the model that allows for a correlation between individuals' hysteresis effects across sessions compared to a model assuming fully correlated individual hysteresis effects is 5×10^{83}. This means that the observed data are more likely under the model allowing for a correlation between individual hysteresis effects across sessions than under the model assuming fully correlated effects.

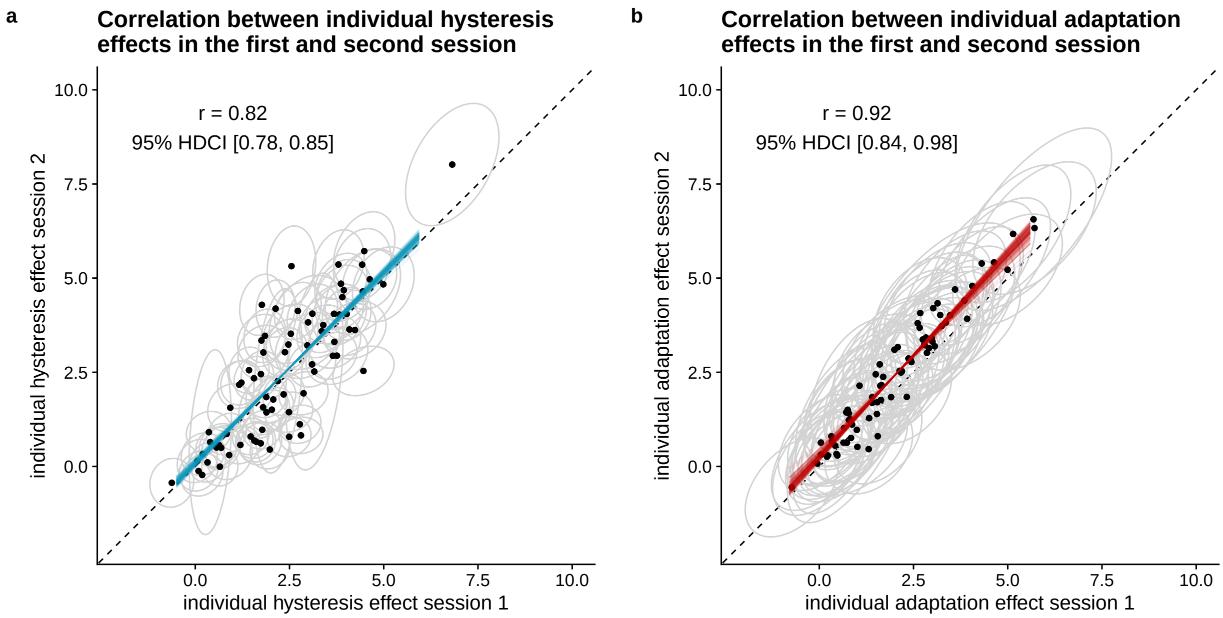


Figure 4.16: (a) Correlation between individual slopes for the effect of aspect ratio of L1 on perceiving the 0° orientation in L2 in the first and the second session. (b) Correlation between individual slopes for the effect of perceived L1 orientation on perceiving the 0° orientation in L2 in the first and the second session. Mean and 80% highest density continuous intervals per individual are shown. The black diagonal line indicates equally sized effects in both sessions. The colored lines give examples of plausible correlation estimates. *Note.* As the estimated correlation values shown come from a hierarchical model including both the estimates from the first and the second session, potential attenuation of the correlation as a result of noise is already taken into account.

The Bayes factor in favor of a model that allows for a correlation between individuals' adaptation effects across sessions compared to a model assuming uncorrelated individual adaptation effects is 4×10^{15}. This means that the observed data are more likely under the model allowing for a correlation between individual adaptation effects across sessions than under the model assuming uncorrelated effects. The Bayes factor in favor of the model that allows for a correlation between individuals' adaptation effects across sessions compared to a model assuming fully correlated individual adaptation effects is 4×10^{-7} (inverse BF: 2×10^{6}). This means that the observed data are less likely under the model allowing for a correlation between individual adaptation effects across sessions than under the model assuming fully correlated effects.

Figure 4.16 shows the correlation between the first and second session individual slopes for aspect ratio and perceived L1 orientation in the model predicting perceived L2 orientation that allows for a correlation in the effects across sessions.[11]

4.3.1.9 Temporal stability of individual differences in strength of absolute orientation bias effects? (H10)

The Bayes factor in favor of a model that assumes the true linear correlation between the magnitude of the absolute orientation biases for the first and second session to be positive compared to a model assuming a true linear correlation smaller than or equal to zero is 1999. This means that the observed data are more likely under the model assuming a positive linear correlation between the magnitudes of the absolute orientation bias effects across

[11] Different from what we had preregistered, we do not show the mean and 95% HDCI for the correlation in the winning model for the adaptation effect, as the winning model was the model assuming fully correlated effects across sessions and thus assumes a correlation equal to 1.

sessions than under the model assuming a non-positive linear correlation. Figure 4.17a shows the correlation between the magnitude of the absolute orientation bias per individual in the first and second session. Figure 4.17b shows the correlation between the magnitude of the absolute orientation bias per individual in the first and second session.

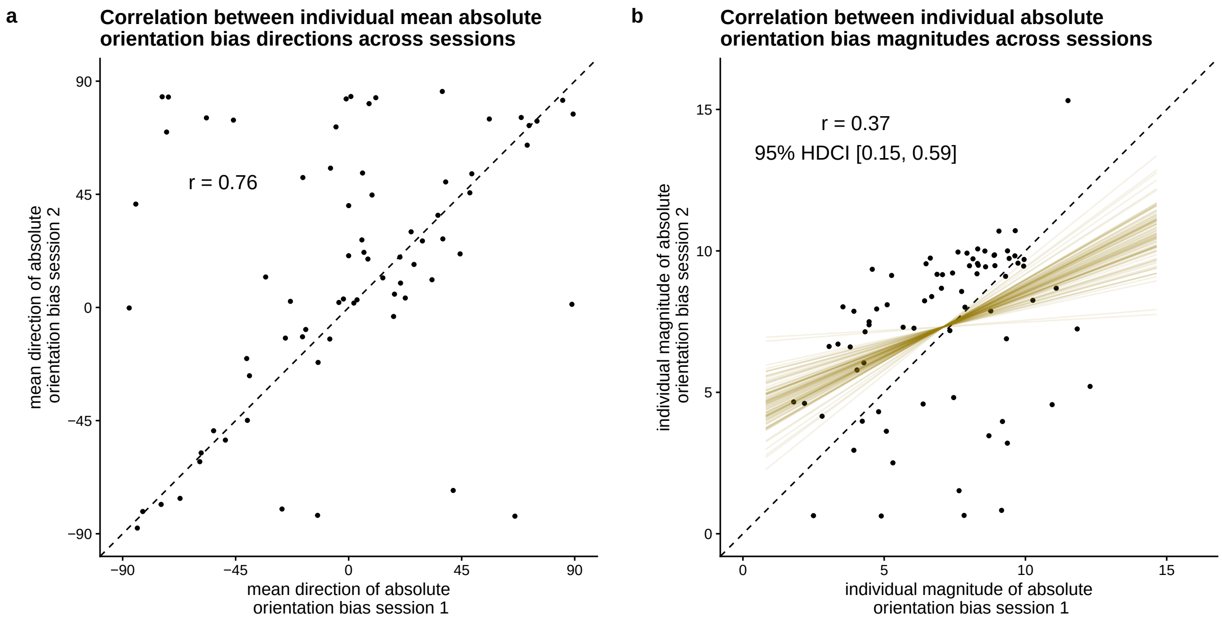


Figure 4.17: (a) Correlation between the mean direction of the absolute orientation bias in the first and second session per individual. The circular-circular correlation coefficient as defined in Mardia & Jupp (2000) is given. The black diagonal line indicates equal mean directions for the first and second session. (b) Correlation between the mean magnitude of absolute orientation bias in the first and second session per individual. The black diagonal line indicates equally sized magnitudes for the first and second session. The colored lines give examples of plausible correlation estimates.

4.3.2 Additional exploratory analyses

4.3.2.1 Individual differences in the proximity effect?

We explored whether the current dataset provided formal evidence for consistent individual differences in the proximity effect, i.e., the direct effect of the aspect ratio in the first lattice on which orientation was perceived in the first lattice. The Bayes factor in favor of the model with a random effect for proximity compared to the common effects model is 3×10^{253}. This means that the observed data are more likely under the unconstrained model than under the common effects model. This Bayes factor indicates that it is much more likely to assume individual differences in the proximity effect than to assume everyone to show the same effect size. Figure 4.18 shows the estimated individual effects of aspect ratio of L1 (i.e., proximity effect) in the model predicting perceived L1 orientation.

In addition, we explored whether the current data provided evidence for the hypothesis that everyone shows the proximity effect in the expected direction. The Bayes factor comparing the likelihood of the observed data under the negative effects model and under the unconstrained model for the proximity effect is 4.5621. This means that the observed data are more likely under the negative effects model than under the unconstrained model. This Bayes

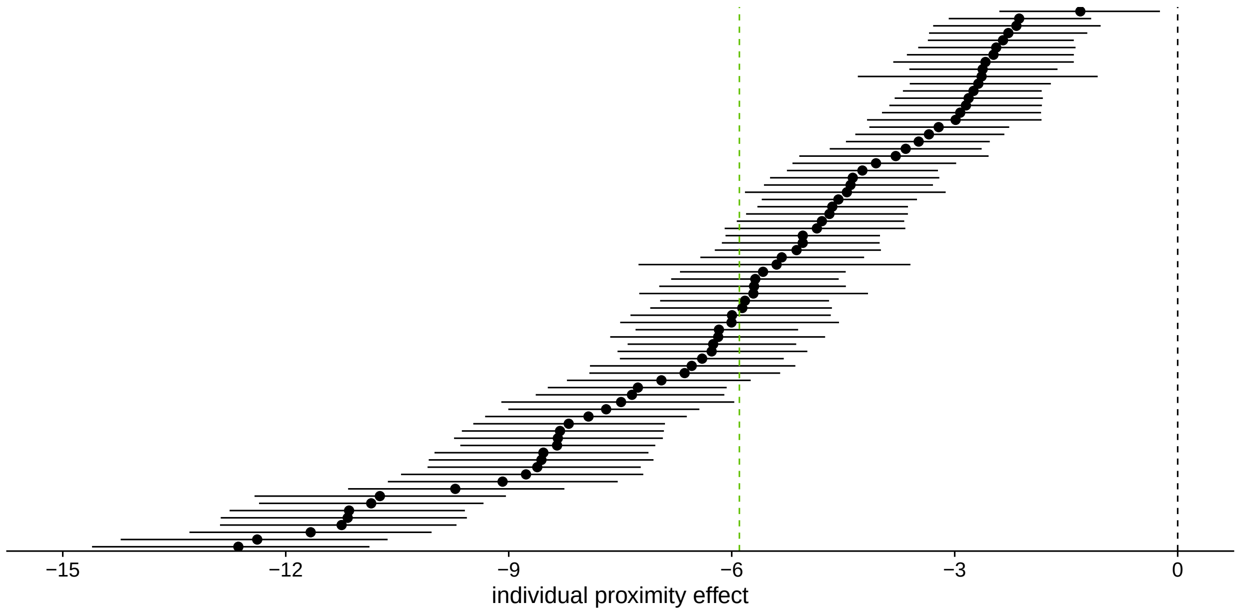


Figure 4.18: Slopes for the effect of aspect ratio on perceiving the 0° orientation in L1 per participant (i.e., proximity effect). Mean and 95% highest density continuous intervals are shown. The colored line indicates the average mean effect across participants. The black line indicates a slope of zero.

factor indicates that it is more likely to assume that everyone shows a proximity effect in the expected direction, than to assume that not everyone shows this effect in the expected direction.

4.3.2.2 Temporal stability of individual proximity effects?

Figure 4.19 shows the correlation between the first and second session individual slopes for aspect ratio in the model predicting perceived L1 orientation. It is clear from the figure that the correlation between individual proximity effects for both sessions is very high: individuals with a strong proximity effect in the first session tend to also have a strong proximity effect in the second session. In addition, except for one participant, all proximity effects are in the expected direction. The absolute size of the proximity effect per individual tended to be slightly larger in the second session.

The Bayes factor in favor of a model that allows for a correlation between individuals' proximity effects across sessions compared to a model assuming uncorrelated individual proximity effects is 6×10^{13}. This means that the observed data are more likely under the model allowing for a correlation between individual proximity effects across sessions than under the model assuming uncorrelated effects. The Bayes factor in favor of the model that allows for a correlation between individuals' proximity effects across sessions compared to a model assuming fully correlated individual proximity effects is 4×10^{63}. This means that the observed data are more likely under the model allowing for a correlation between individual proximity effects across sessions than under the model assuming fully correlated effects.

4.3.2.3 Relation between individual proximity effects and context effects?

Given that individual proximity effects and temporal attractive and repulsive context effects (i.e., hysteresis and adaptation) show very stable across sessions, we were interested in the relation between the direct effect of aspect

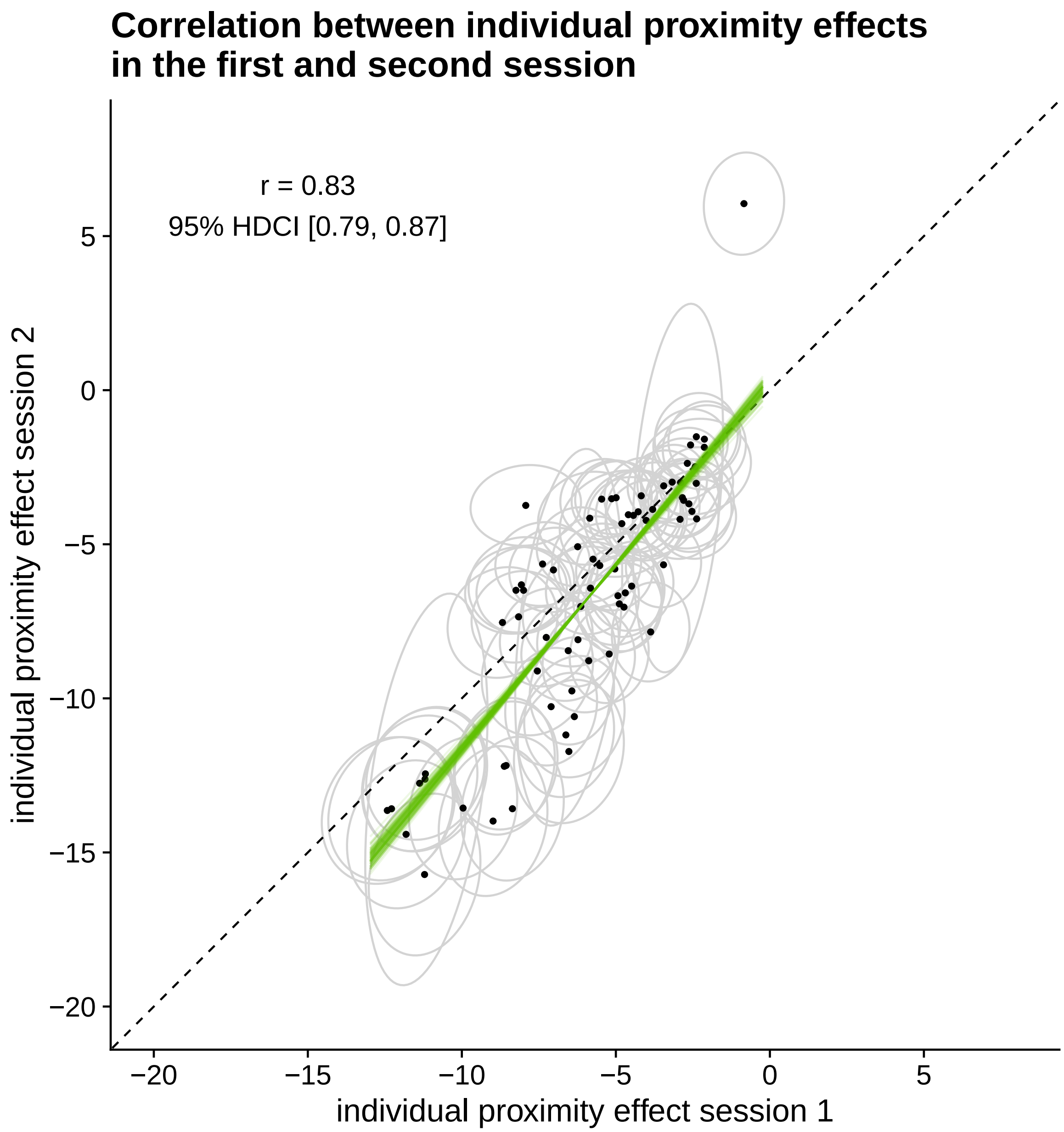


Figure 4.19: Correlation between individual slopes for the effect of aspect ratio on perceiving the 0° orientation in L1 in the first and the second session. Mean and 80% highest density continuous intervals are shown. The black lines indicate a slope of zero. The green lines give examples of plausible correlation estimates. *Note.* As the estimated correlation value shown comes from a hierarchical model including both the estimates for the first and the second session, potential attenuation of the correlation as a result of noise is already taken into account.

ratio on (more often) perceiving the 0° orientation in the first lattice (i.e., proximity effect) and the indirect effect of aspect ratio on (less often) perceiving the 0° orientation in the second lattice (i.e., adaptation effect). In addition, we computed the correlation between individual proximity effect and hysteresis effects.

Figure 4.20 shows the correlation between the proximity effect and the temporal context effects per individual. The correlation of individual proximity effects and individual adaptation effects was negligible (see Figure 4.20b): knowing the size of an individual's proximity effect does not tell us much about the size of an individual's adaptation effect. The size of individual proximity effects and individual hysteresis effects was negatively correlated (see Figure 4.20a)[12], but also the differences in variance across the range of hysteresis effects needs to be taken into account: whereas individuals with a strong influence of their previous percept on their current percept have a larger probability of having a small proximity effect, individuals with a small hysteresis effect do not necessarily have a strong direct effect of aspect ratio on their percept (i.e., a strong proximity effect).

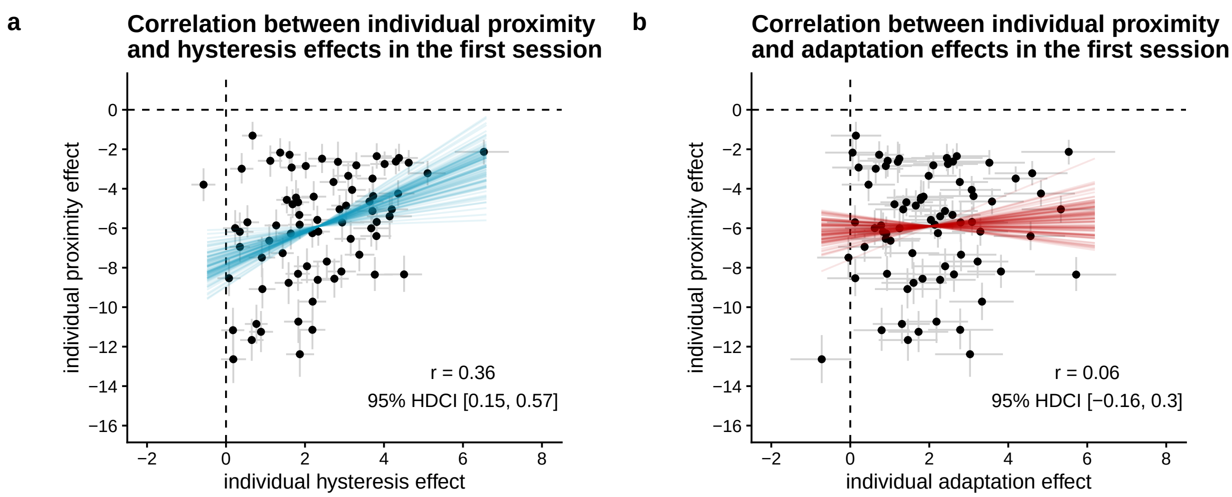


Figure 4.20: (a) Correlation of estimated individual hysteresis effects concerning the second lattice with estimated individual proximity effects concerning the first lattice. (b) Correlation of estimated individual adaptation effects concerning the second lattice with estimated individual proximity effects concerning the first lattice. Mean and 80% highest density continuous intervals are shown. The black lines indicate a slope of zero. The colored lines give examples of plausible correlation estimates.

4.3.2.4 Relation of proportion of non-dominant responses, left-right response bias, and context effects?

We explored the relation of an individual's probability to give non-dominant responses to the first and second lattice as well as their asymmetry of choosing a response option for the second lattice requiring a response with the left or the right hand with the size of the individual's attractive and repulsive context effects. First, the probability of giving a diagonal response in the first lattice was positively correlated across sessions, as was the probability of giving (impossible) 90° responses to the second lattice (see Figure 4.21a). Although for the magnitude of the left-right response asymmetry most participants showed only slight deviations from chance, participants with strong deviations from chance level did at least sometimes show this deviation in both sessions (see Figure 4.21a). Furthermore, the probability of giving a diagonal response in the first lattice correlated considerably with the probability of giving an (impossible) 90° response to the second lattice as well as the difference in proportion of

[12] Note that a more negative proximity effect indicates a stronger effect of proximity on perception. Consequently, the positive slope in Figure 4.20a implies a negative correlation.

left and right responses to the second lattice compared to chance level (see Figure 4.21b). When correlating an individuals' probability of giving non-dominant responses to their estimated hysteresis and adaptation effects, a consistent pattern arises: Whereas participants with a small number of non-dominant responses vary widely in the size of their hysteresis and adaptation effects, having more non-dominant responses seems to relate to smaller hysteresis and adaptation effects. A similar pattern is visible for the relation between the size of the left-right response asymmetry to the second lattice and the size of the hysteresis and adaptation effects: individuals with a large left-right response asymmetry typically have small hysteresis and adaptation effects, whereas the range of possible hysteresis and adaptation effect sizes is much wider for individuals with only a small left-right response asymmetry.

4.4 Discussion and conclusion

With this Registered Report, we investigated (a) whether we could replicate the average attractive and repulsive context effects found in the perception of multistable dot lattices, (b) whether consistent differences in the size of these effects could be found between individuals, and (c) whether every individual showed both effects in the expected direction. In addition, we investigated (d) whether individual differences in both context effects were positively correlated, (e) whether the hysteresis effect could be ascribed to perceptual or decisional causes, (f) whether individual differences in both context effects were correlated with the strength of individual's absolute orientation biases, and (g) whether individual differences in attractive and repulsive context effects as well as in the magnitude of absolute orientation biases were stable across time. In addition, we exploratorily investigated (h) whether consistent differences in the size of the proximity effect exist between individuals and whether every individual showed a proximity effect in the expected direction, (i) whether individual differences in the proximity effect were stable across time, (j) whether individual differences in the proximity effect correlated with individual differences in the hysteresis and adaptation effects, and (k) how individual differences in proportions of non-dominant responses and left-right motor response biases related to individual differences in the hysteresis and adaptation effects.

4.4.1 Summary of the main findings

4.4.1.1 Average results on attractive and repulsive context effects replicate (H1-3)

When looking at the results averaged across participants (see Figures 4.8 and B.1), we successfully replicated the attractive effect of the previous percept (i.e., perceived L1 orientation; cf. H1) and the repulsive effect of the previous stimulus (i.e., aspect ratio of L1; cf. H2) on the current percept (i.e., perceived L2 orientation), as well as the absence of evidence for an interaction between both effects (cf. H3). The Bayes factors, indicating how to update our belief in one model relative to the alternative model, were strongly in favor of including both the hysteresis and the adaptation effect. The Bayes factor comparing a model with and without interaction between the hysteresis and the adaptation effect was in favor of the model without the interaction. This study thus fully replicates the average results from Schwiedrzik et al. (2014) and Gepshtein & Kubovy (2005).

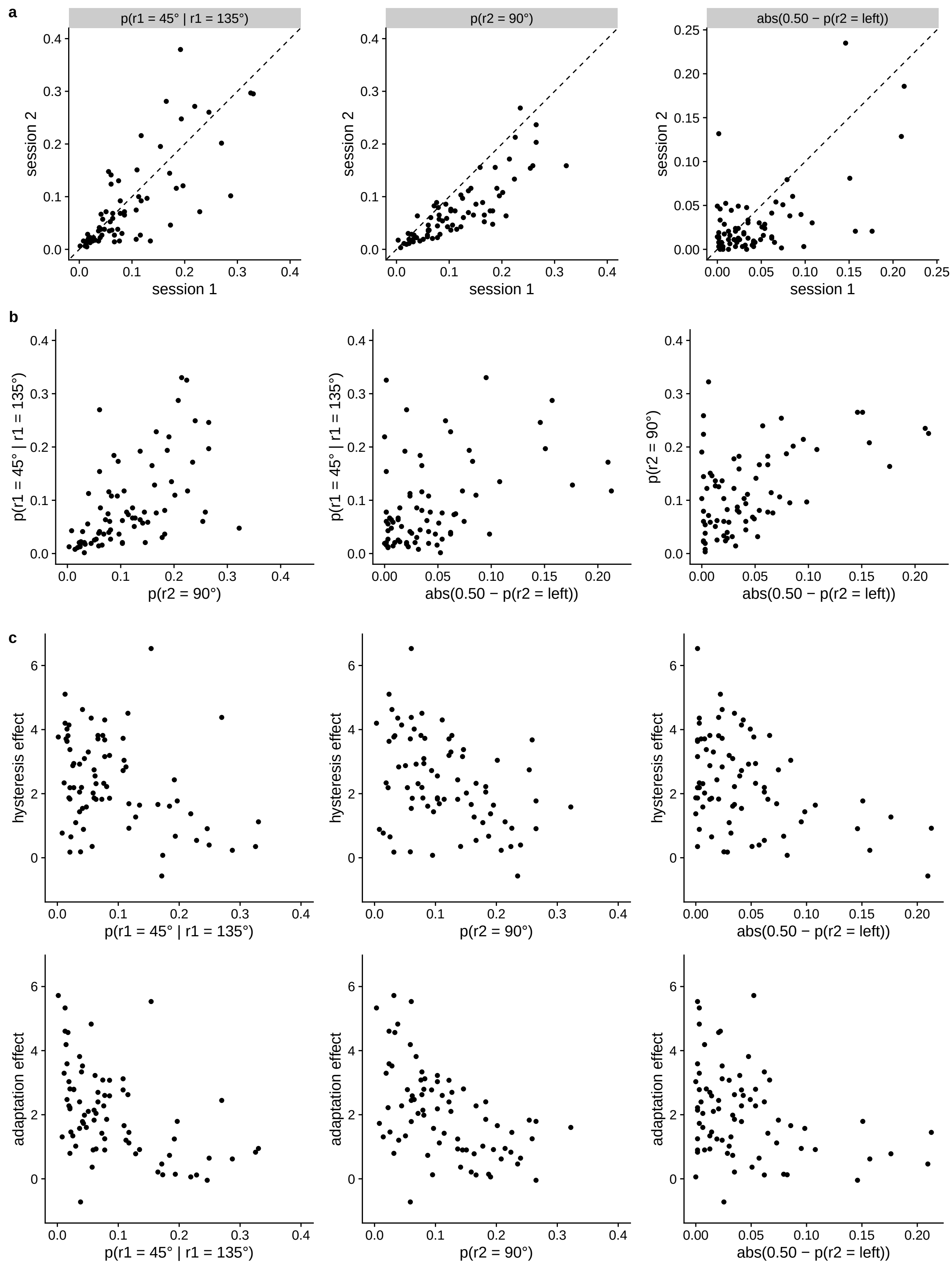


Figure 4.21: (a) Correlation across sessions for the probability of diagonal L1 responses, the probability of 90° responses for L2, and the absolute difference from chance level in selecting a left or right response option for L2 per individual. The black diagonal lines indicate equal probabilities. (b) Correlations between the probability of diagonal L1 responses, the probability of 90° responses for L2, and the absolute difference from chance level in selecting a left or right response option for L2 per individual. (c) Correlation of estimated individual hysteresis and adaptation effects concerning the second lattice with individual probabilities of diagonal L1 responses, probabilities of 90° responses for L2, and individual magnitudes of the left-right response asymmetry to the second lattice.

4.4.1.2 Consistent individual differences exist in the magnitude of attractive and repulsive context effects (H4)

The results averaged across participants do not tell the complete story, however: finding evidence for an average effect does not guarantee individuals' true effects to be of the same size or in the same direction. When inspecting individual results for the experimental task, it is clear that individuals differ in how strongly the aspect ratio of the first lattice and their percept of the first lattice influence their percept of the second lattice (see Figures 4.9b and B.2b). Bayes factors strongly preferred the unconstrained models above the common effects models, providing evidence for true individual differences in both the size of the hysteresis effect and the size of the adaptation effect.

This evidence for true individual differences in the size of attractive and repulsive temporal context effects is of theoretical importance: It tells us that individuals cannot only differ in their perception because of differences in previously encountered stimuli and percepts, but can also differ in the way context is incorporated into perception: different individuals use context information concerning the previous stimulus and the previous percept to a different extent. In other words, even when individuals would have exactly the same stimulus history and perceptual history, they could still differ in what they perceive due to differential use of the stimulus history and perceptual history when forming a new percept.

4.4.1.3 Not *everyone* clearly shows attractive and repulsive context effects (H5)

As the Bayes factors concerning H5 indicated a preference for the unconstrained models above the positive effects models, these results indicate that neither do all individuals show a clear attractive effect of the previous percept, nor do all individuals show a clear repulsive effect of the previous stimulus. Nevertheless, *almost* everyone showed clear attractive and repulsive context effects in the expected direction. The number of participants with an estimated true non-positive hysteresis effect and/or an estimated true non-positive adaptation effect was very low (see Figure 4.11). Importantly, also individuals with an estimated non-positive hysteresis or adaptation effect showed consistency across sessions, indicating that the non-positive estimate was not a strange oddity (see Figure 4.16). In addition, we explored possible differences between the participants with somewhat extreme results and the other participants but found no consistent differences regarding the demographics (e.g., age) or the number of days in-between the two test sessions.

Even though the current findings indicate that not *everyone* shows hysteresis and adaptation effects in the expected direction, the results do correspond well with the results from the reanalysis of the data collected by Schwiedrzik et al. (2014). As having a true non-positive hysteresis and/or adaptation effect seems to be very rare, it is reasonable that no non-positive effects were found in that reanalysis, which only included 27 participants. The current sample thus gives a more complete and nuanced picture on the range of plausible hysteresis and adaptation effects, but also confirms that *almost* everyone shows attractive and repulsive context effects in the expected direction.

Also the finding that almost everyone shows an attractive effect of the previous percept and a repulsive effect of the previous stimulus evidence has theoretical implications. Future research can aim to shed light on why these context effects show this direction for almost all individuals. Nevertheless, it is equally important for future research to take the full range of individual differences present into account when attempting to explain these effects, including the presence of at least some individuals with a true effect in the opposite direction. In addition, it is important to investigate whether existing models of attractive and repulsive temporal context effects can incorporate the variability found, as a good model should not only be able to predict the mean, but also plausible variation in the effect's size and direction.

4.4.1.4 At least a common factor affecting both hysteresis and adaptation (H6)

The results indicate a strong positive correlation between estimated individual hysteresis and adaptation effects. This positive correlation thus suggests that there may at least be a common factor affecting the processes underlying both effects. It is unclear however what exactly may explain this high positive correlation between the magnitude of both effects. One way to understand the high positive correlation is that hysteresis and adaptation are both context effects, and that individuals can be contrasted based on how strongly they are influenced by context in general versus how strongly they are influenced by the direct perceptual evidence present. The current results can thus not exclude the hypothesis that both context effects stem from the same underlying mechanism, and thus also seem to support the hypothesis that both effects have at least some common underlying factor. This conclusion is similar to the conclusion of Snyder et al. (2019), who found a positive correlation between individual differences in inhibition across contrast and assimilation tasks, indicating at least some common factor influencing the size of both context effects.

4.4.1.5 The attractive context effect is partially percept-related, partially decision-related (H7)

As the experimental task could not distinguish between a perceptual or a postperceptual nature of the hysteresis effect, a control task was included in which perceptual factors were ruled out. Even in the control task there was an attractive effect of the previous response, although this effect was considerably smaller than in the experimental task (and absent for at least some participants). This suggests that the attractive context effect is neither solely percept-related (i.e., dependent on actually perceiving the orientation in question), nor solely decision-related (i.e., dependent on the choice for a specific orientation without involving perception). In this way, the results nuance earlier perspectives stating 'serial dependence' to be either a fully percept-related or a fully decision-related effect (e.g., Bosch et al., 2020; Carter et al., 2014; Cicchini et al., 2017; Fritsche et al., 2017; Manassi et al., 2018; Pascucci et al., 2019; Schwiedrzik et al., 2018). Also, individuals seem to differ in the extent to which their hysteresis effect is percept- or decision-related: several participants do not show an indisputably positive hysteresis effect in the control task, indicating a more perceptual basis for their hysteresis effect. The difference in the size of the hysteresis effect between the experimental and the control task could potentially also be interpreted as related to decision confidence[13]: biases based on past decisions could be expected to be larger in cases in which decision confidence was higher because the past decision was based on perceptual evidence, compared to cases in which no perceptual evidence was present. Even when following this interpretation in terms of decision confidence however, the actual reason for the difference stays perceptual.

[13] We thank an anonymous reviewer of our First Stage Registered Report for this suggestion.

4.4.1.6 The magnitude of individual's absolute orientation bias and their attractive and repulsive context effects correlate positively (H8)

In contrast to our expectation, the magnitude of individual's absolute orientation bias did correlate positively rather than negatively with the strength of individual's attractive and repulsive context effects. It has to be noted, however, that Bayes factors provide an evidence ratio between two specific models, in this case being the model assuming a negative linear correlation versus the model assuming a non-negative linear correlation. Consequently, a high Bayes factor does not guarantee the winning model to provide a good fit to the observed data. From the scatterplots (see Figure 4.15), it is unclear whether a linear model provides a good fit for the data. Although we tested whether a quadratic model could better predict the data pattern, relative evidence for a model assuming an inverted U-shaped curve compared to a linear model was slightly in favor of the linear model.

The results provide slight evidence for a positive relation between the size of individuals' absolute orientation bias and their hysteresis and adaptation effects. This could be interpreted as a slight positive relation between different types of biases, but the actual reason for this positive correlation is unclear. Furthermore, although the linear model was preferred over the quadratic model, it is unclear whether the linear model does provide a good fit for the data.

4.4.1.7 Individual differences in attractive and repulsive context effects are stable over time (H9)

Individual differences in both attractive and repulsive context effects in the used multistable dot lattice paradigm show to be very stable, at least across a period of 7 to 14 days. For hysteresis (i.e., attractive context effect of the previous percept), the winning model was the model assuming a correlation but no full correlation between the hysteresis effects in both sessions. For adaptation (i.e., repulsive context effect of the previous stimulus shown), the winning model was the model assuming a full correlation between the adaptation effects in both sessions. These results indicate that individual differences in the size of hysteresis and adaptation effects are reliable indices of individual differences across time, at least in the current multistable dot lattices paradigm, and it can be useful to investigate their relations with other individual difference factors as well as with estimates of individual hysteresis and adaptation effects assessed using different stimuli and tasks. Our results indicating a strong but not full correlation of individual differences in attractive context effects are in line with the results of Kondo et al. (2022), who found a high degree of consistency within individual observers when assessing attractive serial dependence in orientation perception.

4.4.1.8 Differences in the magnitude of individual's absolute orientation bias are stable over time (H10)

Although a large number of the participants in the sample showed a very consistent mean absolute orientation bias strength, this was not the case for all participants. Post-hoc analyses showed that the mean direction of the absolute orientation bias stayed relatively stable across time, at least for most participants (see Figure 4.17).

4.4.1.9 Consistent individual differences exist in the magnitude of proximity effects, and everyone shows the proximity effect

When exploring individual results for the proximity effect, it became clear that individuals differ in how strongly the aspect ratio of the first lattice influences their percept of the first lattice (see Figures 4.9a and B.2a). The Bayes factor indicated a strong preference for the unconstrained model above the common effects model, indicating that there is evidence for true individual differences in the size of the proximity effect. When exploring whether everyone shows the proximity effect in the expected direction, the Bayes factor indicated a preference for the negative effects model above the unconstrained model. This supports the idea that all individuals show a proximity effect in the expected direction.

4.4.1.10 Individual differences in the proximity effect are stable over time

Post-hoc analyses indicated high stability for individual differences in how strongly participants are affected by proximity in their percept of the first lattice (see Figure 4.19). The absolute size of the proximity effect per individual tended to be slightly larger in the second session. Furthermore, the size of individuals' proximity effects was negatively related to the size of individuals' hysteresis effects: the larger an individual's hysteresis effect, the smaller the range of plausible values for their proximity effect, and the smaller their proximity effect. To the contrary, the size of an individual's proximity effect was uncorrelated to the size of their adaptation effect in the current sample (see Figure 4.20). Although at first sight proximity seems to be differentially related to hysteresis and adaptation, this result should be replicated and further investigated before making firm conclusions. In case the differential relationship of hysteresis and adaptation with proximity holds, this would suggest a dissociation between hysteresis and adaptation.

4.4.1.11 Proportion of non-dominant responses and left-right response bias relate negatively to attractive and repulsive context effects

Post-hoc visualizations (see Figure 4.21) indicated stable individual differences in the probability of choosing a non-dominant response option for the percept of both the first and the second lattice. Although most participants showed only slight deviations from chance, participants with strong deviations from chance level when choosing a response option for the second lattice requiring a response with the left or the right hand did at least sometimes show this deviation in both sessions. High probabilities of choosing non-dominant responses related negatively to hysteresis and adaptation effects: Whereas participants with a small number of non-dominant responses vary widely in the size of their hysteresis and adaptation effects, having more non-dominant responses seems to be related to smaller hysteresis and adaptation effects. In addition, individuals with a large left-right asymmetry in their L2 responses showed smaller hysteresis and adaptation effects. These exploratory results may indicate that more attentive participants show a larger range of possible hysteresis and adaptation effect sizes, whereas less attentive participants have smaller effects of previous percept and previous stimulus on the current percept.

4.4.2 Suggestions for future research

4.4.2.1 Factors correlating with individual differences in hysteresis and adaptation effects

By providing strong empirical evidence for the existence of consistent differences in individuals' true attractive and repulsive context effects, this work can form a starting point for future work exploring potential factors to explain these individual differences. Some earlier research already suggested relations between the reduced use or differential weighting of stimulus history and some clinical conditions, using different tasks (not distinguishing between stimulus history and perceptual history). Stein et al. (2020) found that a reduced influence of previous stimuli on working memory contents in patients with schizophrenia and anti-NMDAR encephalitis. Lieder et al. (2019) showed a differential use of previous sensory information in individuals with autism and dyslexia: whereas individuals with autism relied more on longer-term statistics, individuals with dyslexia relied more on information about the immediate past. Future research can explore relations with the multistable dot lattices paradigm in different clinical conditions, but can also explore other potential correlates of individual differences in the use of stimulus history and perceptual history across tasks (e.g. personality differences). In addition, special attention needs to be paid to the individuals showing negative hysteresis and adaptation effects. Future research can investigate what underlies the unexpected direction of the effects in these individuals, and needs to take the existence of those negative effects into account whenever attempting to explain individual differences.

4.4.2.2 Explaining the strong positive correlation between hysteresis and adaptation effects

As the current study found evidence for a strong positive correlation between the size of attractive and repulsive context effects across individuals, future research can further investigate the source of this strong positive correlation. Especially theoretical and modeling work, in combination with empirical validation, can be useful to get a more concrete insight in the process underlying this positive correlation.

4.4.2.3 Processes underlying hysteresis and adaptation effects

Why does almost everyone show attractive effects of the previous percept and decision, and repulsive effects of the previous stimulus? Also regarding this question, future theoretical work and modeling efforts, in combination with empirical validation, can contribute to a better understanding of the underlying processes. In addition, the existence of true individual differences in the size and direction of an effect has consequences for the models and theories aiming to explain or predict these effects: It is important to verify whether existing models and theories can reproduce or explain the range of variability found in the effects' size and direction across individuals, as a good model should not predict the mean alone, but also plausible variation in the effect's size and direction.

4.4.2.4 Individual differences in the presence of an interaction effect between hysteresis and adaptation

Although the results from our study suggest the model without an interaction effect between hysteresis and adaptation to be preferred above the model including an interaction, individual differences seem to exist in the presence of this interaction, with most participants not showing an interaction, but some clearly showing an interaction between the two (e.g., participants 011, 029, and 081 in the current dataset). Future research could investigate whether it is worth including an interaction effect for a subsample of the participants to more accurately estimate their context effects.

4.4.2.5 Generalizability of individual differences in hysteresis and adaptation effects to different stimuli and tasks

The current results indicate highly stable individual differences in attractive and repulsive context effects across time, at least when assessed using this specific multistable dot lattices paradigm. Future research needs to examine whether the stable individual differences in attractive and repulsive context effects found in the current task correlate with similar individual differences assessed using different tasks or stimuli.

4.4.2.6 Further disentangling hysteresis as a perceptual and a decisional effect

The current study finds support for a partially perceptual and partially decisional nature of the attractive context effect. In addition, the current results suggest that individuals may differ in the extent to which their attractive context effect is related to perception or decision: Whereas some individual's effects are almost equal in size in both control and experimental task (i.e., indicating a mainly decisional nature of the effect), most individuals show a considerably smaller effect in the control task (i.e., indicating a combination of perceptual and decisional nature), and some individuals do show no evidence for a hysteresis effect in the control task (i.e., hinting at a fully perceptual nature). Future studies can focus more specifically on individual differences in the nature of these effects, and in that way disentangle individuals' perception- and decision-related attractive context effects.

4.4.2.7 Further disentangling stimulus-, percept-, decision-, and response-related effects

One of the advantages of using the current multistable dot lattices paradigm is the explicit distinction that can be made between effects of the previous stimulus and those of the previous percept / decision / response. We believe this distinction is crucial to enhance clarity in the research literature. Although previous work has often distinguished between stimulus and decision or stimulus and response, any non-stimulus related effect has typically been reported as 'postperceptual'. We want to clarify that the fact that an effect is non-stimulus related does not directly imply that the effect is postperceptual, but that the effect could be the consequence of the way the stimulus was experienced (i.e., the percept), rather than being put away as purely decisional or response-related. That said, explicitly studying the distinct contributions of all different factors, including stimulus, percept, decision, and motor response, is a relevant topic for future research.

4.4.2.8 Replicating and explaining the relation between absolute orientation biases and hysteresis and adaptation effects

The results concerning the relationship between absolute orientation bias and hysteresis and adaptation effects found in the current study were ambiguous and require further study and replication by other researchers. Given that the current results replicate, it is worth explaining how the complex relation between these individual difference factors emerges.

4.4.3 Take home message

In this study, we replicated the average attractive effect of a previous percept on the current percept and the repulsive effect of previously presented stimulus evidence on the current percept. Large individual differences in the size of these attractive and repulsive context effects exist, however, and these individual differences are consistent across timepoints (one to two weeks apart). Although *almost everyone* shows both effects in the expected direction, *not every* single individual does. Furthermore, individual differences in the size of attractive and repulsive context effects are strongly positively correlated, suggesting at least a common factor influencing the processes underlying both effects. In addition, the attractive context effect is shown to be partially percept-related and partially decision-related, nuancing earlier debates on the origin of this effect.

In sum, the study provides insight in how individuals differ in how they combine previous input and experience with current input in their perception, and more generally, this tells us that different individuals can perceive identical stimuli differently, even within a similar context.

4.5 Open and reproducible practices statement

This manuscript was written in R Markdown using the papaja package (Aust & Barth, 2022) with code for data analysis integrated into the text. The data, materials, and analysis and manuscript code for the experiment are available at https://doi.org/10.17605/osf.io/wae6k. The preregistration for this experiment is available at https://doi.org/10.17605/osf.io/qmgca.

Chapter 5

An efficient Bayesian observer model for temporal context effects when perceiving multistable dot lattices

5

In multistable dot lattices, the orientation we perceive is attracted towards the orientation we perceived in the immediately preceding stimulus and repelled from the orientation for which most evidence was present previously (Van Geert, Moors, et al., 2022). Theoretically-inspired models have been proposed to explain the co-occurrence of attractive and repulsive context effects in multistable dot lattice tasks, but these models artificially induced an influence of the previous trial on the current one without detailing the process underlying such an influence (Gepshtein & Kubovy, 2005; Schwiedrzik et al., 2014). We conducted a simulation study to test whether the observed attractive and repulsive context effects could be explained with an efficient Bayesian observer model (Wei & Stocker, 2015). This model assumes variable encoding precision of orientations in line with their frequency of occurrence (i.e., efficient encoding) and takes the dissimilarity between stimulus space and sensory space into account. An efficient Bayesian observer model including both a stimulus and a perceptual level was needed to explain the co-occurrence of both attractive and repulsive temporal context effects. Furthermore, this model could reproduce the empirically observed strong positive correlation between individuals' attractive and repulsive effects (Van Geert, Moors, et al., 2022), by assuming a positive correlation between temporal integration constants at the stimulus and the perceptual level. To conclude, the study brings evidence that efficient encoding and likelihood repulsion on the stimulus level can explain the repulsive context effect, whereas perceptual prior attraction can explain the attractive temporal context effect when perceiving multistable dot lattices.

Van Geert*, E., Ivančir*, T., & Wagemans, J. (under review). An efficient Bayesian observer model for temporal context effects when perceiving multistable dot lattices. Preprint available from https://doi.org/10.31234/osf.io/g64fp

* Eline Van Geert and Tina Ivančir contributed equally to this work.

Preprint | Data and code

An efficient Bayesian observer model for attractive and repulsive temporal context effects when perceiving multistable dot lattices

5.1 Introduction

What we perceive is not only influenced by the current stimulus we have in front of our eyes, but also by the recent stimulus and perceptual history. Many recent studies have confirmed the existence of both attractive and repulsive effects of immediate temporal context on perception (Bosch et al., 2020; Fritsche et al., 2020; Pascucci et al., 2019; Sadil et al., 2021; Snyder et al., 2015; Van Geert, Moors, et al., 2022). Also in the perception of multistable dot lattices, both attractive and repulsive context effects are at play (Gepshtein & Kubovy, 2005; Schwiedrzik et al., 2014; Van Geert, Moors, et al., 2022). On the one hand, the perceived orientation in these lattices is attracted towards the orientation perceived in the immediately preceding lattice (i.e., hysteresis, attractive effect of previous percept). On the other hand, the perceived orientation is repelled from the orientation for which most evidence was present in the previous lattice (i.e., adaptation, repulsive effect of previous stimulus evidence, Gepshtein & Kubovy, 2005; Schwiedrzik et al., 2014; Van Geert, Moors, et al., 2022).

Several theoretically-inspired models have been proposed to explain the co-occurrence of these context effects when perceiving multistable stimuli (Gepshtein & Kubovy, 2005; Schwiedrzik et al., 2014), but these models either artificially induce a direct influence of the previous stimulus evidence on the likelihood distribution for the current stimulus (Schwiedrzik et al., 2014) or induce a randomly determined shift in prior bias from the previous to the current percept (Gepshtein & Kubovy, 2005). From these models, it is not clear why such a direct influence would occur, or which underlying process would determine a random shift in bias. In this simulation study, we investigate whether an efficient Bayesian observer model based on Wei & Stocker (2015) can explain the co-occurrence of both attractive and repulsive temporal context effects in multitstable dot lattice perception. Earlier variants of the efficient Bayesian observer model have successfully been used to model effects in different tasks involving non-ambiguous stimuli (Fritsche et al., 2020; Langlois et al., 2021; Wei & Stocker, 2015). In this study, we assess the viability of explaining temporal context effects on multistable dot lattice perception using an efficient Bayesian observer model. As part of this investigation, we test whether the model can not only successfully account for the average temporal context effects observed in Van Geert, Moors, et al. (2022), but also for the observed range and strong positive correlation of interindividual variation in both effects.

5.1.1 Attractive and repulsive temporal context effects: separate but related mechanisms?

Whereas repulsive temporal context effects are often seen as resulting from the previous stimulus evidence, attractive temporal context effects are seen as resulting from the previous percept, response, and/or decision (Bosch et al., 2020; Sadil et al., 2021; Schwiedrzik et al., 2014; Van Geert, Moors, et al., 2022). This is often related to repulsion being a more 'low-level' phenomenon, showing larger spatial and featural specificity than the 'higher-level' attraction (Fritsche et al., 2020; Gepshtein & Kubovy, 2005; Schwiedrzik et al., 2014). Schwiedrzik et al. (2014) also found both effects to map into distinct cortical networks. Although many researchers thus state that attractive and repulsive effects result from separate processes (Brascamp et al., 2008; Fritsche et al., 2020; Pascucci et al., 2019; Schwiedrzik et al., 2014), others argue that both effects share a common underlying mechanism (Gepshtein & Kubovy, 2005; Mattar et al., 2016; Maus et al., 2013). As Van Geert, Moors, et al. (2022) found a strong positive correlation between the size of individuals' hysteresis (i.e., attractive effect of previous percept) and adaptation effects (i.e., repulsive effect of previous stimulus evidence), there needs to be at least some common factor influencing both effects. We therefore hypothesize that both effects stem from separate but related mechanisms, and in this simulation study an efficient Bayesian observer model is put forward to model the processes underlying both context effects in a theoretically coherent way.

5.1.2 Attractive and repulsive temporal context effects in multistable dot lattice perception

Gepshtein & Kubovy (2005) proposed a paradigm to distinguish between attractive and repulsive context effects on perception. They investigated the influence of (a) the perceived organization of the preceding stimulus (i.e., which organization was reported) and (b) the stimulus support for a certain organization in the preceding stimulus (dependent on the stimulus' aspect ratio) on the perception of a second, current stimulus, using multistable dot lattices as stimuli.

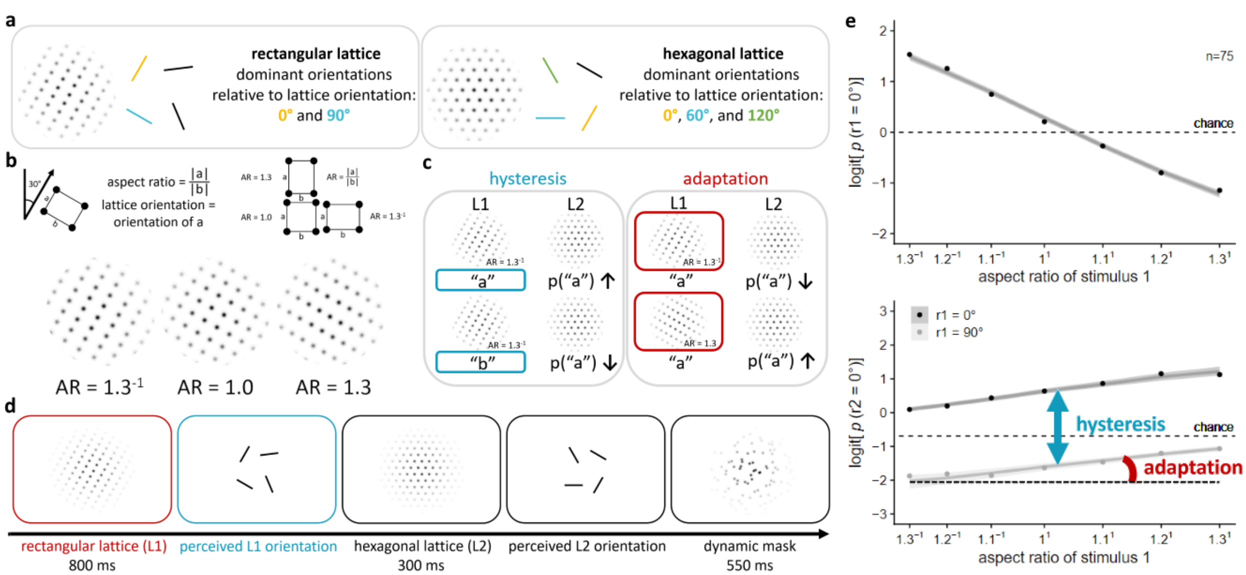


Figure 5.1: (a) Dominantly perceived orientations in multistable rectangular and hexagonal dot lattices. (b) Explanation regarding the aspect ratio of a multistable rectangular dot lattice. In rectangular dot lattices, four different orientations can be perceived, of which two are more prevalent (as the dots are closer together along these orientations). The relative dominance of the a orientation relative to the b orientation is expressed in the aspect ratio of the dot lattice (AR = |a| / |b|). (c) Illustration of attractive and repulsive context effects in dot lattices. Left side: attraction effect (hysteresis). When the first lattice (L1) is *perceived* as orientation a (indicated by "a"), the probability that the second lattice (L2) will be perceived as orientation a is higher than when L1 was interpreted as orientation b (indicated by "b"). Right side: repulsion effect (adaptation). When strong support for orientation a is present in L1, the probability that L2 will be perceived as a orientation a is lower than when L1 had less support for orientation a. (d) Illustration of trial structure. For reasons of visibility, the shown trial components in this figure have black dots on a white background. The actual experiment had white dots on a grey background. (e) Mean empirical logit probability of perceiving the relative 0° orientation in the first and the second lattice dependent on aspect ratio. The probability of responding 0° to the first lattice decreases with aspect ratio (|a|/|b|). The value of aspect ratio increases with increasing distance in the 0°-orientation, leading to more 90° responses. The probability of responding 0° to the second lattice increases with aspect ratio (|a|/|b|; i.e., adaptation effect), and increases when the first stimulus was perceived as 0° rather than 90° (i.e., hysteresis effect). Reprinted and adapted from Van Geert, Moors, et al. (2022).

Multistable dot lattices are aligned dot arrays in which multiple orientations can be perceived. In rectangular dot lattices, four different orientations can be perceived (cf. left part of Figure 5.1a), two of which are more prevalent.

In hexagonal dot lattices (cf. right part of Figure 5.1a), three equally plausible orientations are most prominent. According to the Gestalt law of proximity (Kubovy et al., 1998), the closer the dots are together along a particular orientation, the more likely they will be grouped together, and consequently, the more likely that orientation will be perceived. Relative grouping strength has been shown to decrease exponentially in accordance with the relative inter-dot distance (Kubovy et al., 1998). For two orientations a and b, the aspect ratio of a dot lattice (AR = |a|/|b|) expresses the a orientation's relative dominance over the b orientation (cf. Figure 5.1b). For a lattice with AR = 1, the inter-dot distance in the a and b orientations is equal. For a lattice with AR < 1, the inter-dot distance is smaller in the a than in the b orientation. For a lattice with AR > 1, the inter-dot distance is smaller in the b than in the a orientation. In both rectangular and hexagonal dot lattices, we define the axis orientation of the dot lattice as a whole by the a orientation, which we will refer to as the 0° orientation. In the rectangular dot lattices, we will refer to the b orientation as the 90° orientation.

Gepshtein & Kubovy (2005) introduced a multistable dot lattice paradigm to concurrently assess attractive and repulsive immediate temporal context effects on perception, that was later adapted by Schwiedrzik et al. (2014; cf. Figure 5.1d). They used rectangular dot lattices with randomly varying absolute lattice orientation as context stimuli (presented for 800 ms) and more ambiguous hexagonal dot lattices with the same random absolute lattice orientation as test stimuli (presented for 300 ms). Participants indicated which orientation they perceived in each lattice using a four-alternative forced-choice task (always including the most dominant orientations). To manipulate the stimulus support for the 0° orientation, the aspect ratio of the rectangular dot lattice was varied. They then assessed how (a) perceived orientation and (b) aspect ratio in the first, rectangular lattice affected perceived orientation in the second, hexagonal lattice. An attractive (i.e., hysteresis) effect of the previous percept was present, as well as a repulsive (i.e., adaptation) effect of the previous stimulus evidence: the probability of perceiving a particular orientation in the second lattice increased when the same orientation was perceived in the first lattice, and the stronger the evidence for a specific orientation in the first lattice, the lower the probability to perceive that orientation in the second lattice (cf. Figure 5.1c). These results by Gepshtein & Kubovy (2005) and Schwiedrzik et al. (2014) were replicated and extended by Van Geert et al. (2022; cf. Figure 5.1e). Whereas Gepshtein & Kubovy (2005) and Schwiedrzik et al. (2014) mainly explored the existence of these effects on the group level, Van Geert, Moors, et al. (2022) tested whether individual differences existed in the size of these effects, and whether every individual participant showed both effects in the expected direction. The results confirmed the presence of large, consistent differences in the size of attractive and repulsive context effects across individuals, and these differences stayed stable across one to two weeks time. Furthermore, almost every participant showed both effects in the expected direction, although not every single participant did. As indicated above, the results of Van Geert, Moors, et al. (2022) provided evidence for at least some common factor underlying both effects, as individual differences in attractive and repulsive context effects were highly positively correlated (cf. Figure 5.9a). Also, hysteresis (i.e., the attractive effect of the previous percept) showed to be a partially percept-related and a partially decision-related effect, nuancing earlier debates on the origin of this effect (Bosch et al., 2020; Cicchini et al., 2017; Fritsche et al., 2017; Manassi et al., 2018; Pascucci et al., 2019; Schwiedrzik et al., 2018).

Now that the existence of both temporal context effects on multistable dot lattice perception has been firmly established, including consistent variation in the size of the effects across individuals, one way forward is to further our understanding of the processes underlying these effects by developing models and verifying whether they can reproduce and explain the range of variability in effect size and direction across individuals. Importantly, a good process model should not only be able to predict a mean response, but also plausible variation in the effect size and direction (Van Geert, Moors, et al., 2022).

5.1.3 Models of temporal context effects on multistable dot lattice perception

Two earlier models of the multistable dot lattice paradigm for hysteresis and adaptation exist (Gepshtein & Kubovy, 2005; Schwiedrzik et al., 2014). In the model of Gepshtein & Kubovy (2005), combinations of attraction strengths (related to inter-dot distances in the first, rectangular lattice and sensitivity to the inter-dot distance) and a randomly determined persistent intrinsic bias (i.e., a higher probability to perceive some orientation more than others, which stays similar but not identical from first to second lattice) determine the perceived orientation in the second, hexagonal lattice. When the intrinsic bias exceeds the stimulus support (i.e., attraction strength), the multistable lattice is perceived inconsistently with the stimulus support. In this model, the direction of the intrinsic bias is determined randomly, and it is not clear where the bias comes from or which process determines the direction of the bias.

Schwiedrzik et al. (2014) developed a Bayesian model to account for the co-occurrence of hysteresis and adaptation in multistable dot lattice perception. They model both effects independently, with the perceived orientation in the first lattice directly impacting the prior distribution of possible perceived orientations for the second lattice, and the reduction in stimulus support due to neuronal adaptation to the aspect ratio of the first lattice directly impacting the likelihood distribution for the second lattice. In this model, the process behind the direct change in the likelihood is not included and its size is determined arbitrarily.

Based on the model by Wei & Stocker (2015), we propose an alternative Bayesian model that can explain the co-occurrence of hysteresis and adaptation as separate but related processes influencing multistable dot lattice perception. A similar model has been applied by Fritsche et al. (2020) to explain attractive and repulsive stimulus history effects in orientation perception. They did look at the influence of previous stimuli and went more than one stimulus back, but did not distinguish between effects of previous percepts and previous stimulus evidence, and only used non-ambiguous Gabor stimuli.

5.1.4 Efficient Bayesian observer models of perception

A key assumption of an efficient Bayesian observer model is that available coding resources are limited, and that those feature values that occur more frequently will be more accurately encoded or represented (i.e., the principle of efficient coding, Wei & Stocker, 2015). In case of orientation perception, this variable encoding precision will thus lead to frequent orientations being encoded more accurately than less frequent orientations. A second key feature of the model is that it takes the dissimilarity between stimulus space and sensory space into account (as in psychophysics). This leads to differential predictions of adding external stimulus noise, that is, noise related to uncertainty in a specific stimulus feature (e.g., variance in orientation) or internal sensory noise, that is, noise related to uncertainty in encoding and processing (e.g., due to presentation duration, stimulus size, or luminance contrast). Whereas external stimulus noise will only widen the likelihood distribution and increase the overall influence of the prior (leading to stronger prior attraction, 'Bayesian' percepts), uniform internal sensory noise will make the likelihood distribution asymmetric in stimulus space and hence create the possibility for biases away from the peak of the prior distribution (i.e., likelihood repulsion, 'anti-Bayesian' percepts, Wei & Stocker, 2015). The relative amount of stimulus versus sensory noise will determine which effect will show behaviorally (i.e., attraction or repulsion).

More concretely, frequency of occurrence will jointly influence prior and likelihood in the model. It influences the prior distribution directly: more frequently occurring orientations are also expected to occur more often. However, frequency of occurrence also influences the mapping between stimulus and sensory space: it influences the accuracy with which different orientations will be encoded, and consequently also the width and form with which different orientations will be represented in the likelihood. In other words, the frequency of occurrence will determine the prior distribution as well as how the currently encountered stimulus will be encoded. The stimulus-to-sensory mapping is given by the cumulative density function of the encoding accuracy distribution, and the sensory-to-stimulus mapping is given by the inverse of that cumulative density function.

In the original model, frequency of occurrence was seen on the long term. For example, in most daily environments, cardinal orientations are more prevalent than oblique ones (Coppola et al., 1998; Girshick et al., 2011). In later versions of the model, however, it has been shown that frequency of occurrence can also be defined on the short term, e.g., with the frequency of occurrence changing during the experiment (Fritsche et al., 2020; Ni & Stocker, 2023; Noel et al., 2021). These changes in the frequency of occurrence in the short term may then be used to model short-term temporal context effects on perception.

Fritsche et al. (2020) modeled attractive and repulsive biases of stimulus history using an efficient Bayesian observer model. In this model, they did not disentangle effects of previous stimuli and previous percepts, but treated all effects as related to the previous stimulus evidence. In the empirical study they conducted, they found evidence for short-term attraction and long-term repulsion. When fitting different models to the empirically collected data, a model with distinct transition distributions and different integration time constants for prior and likelihood performed better than a model that used the same parameters for prior and likelihood. In the prior distribution, only the most recently presented stimuli mattered, and updating was fast. In the likelihood distribution, information was integrated over longer timescales, and updates happened more slowly.

In a recent extension of the original model proposed by Wei & Stocker (2015), Mao & Stocker (2022) described perception as a holistic inference process, where the percept of a stimulus is jointly represented at different levels of a representational hierarchy. To adequately model the (variation in) behavioral data of an earlier study (Tomassini et al., 2010), it was necessary to take the higher-level representation into account (i.e., categorization of orientation in this case).

5.1.5 An efficient Bayesian observer model for temporal context effects on multistable dot lattice perception

In this study, we develop an efficient Bayesian observer model for the multistable dot lattices paradigm used by Gepshtein & Kubovy (2005), Schwiedrzik et al. (2014), and Van Geert, Moors, et al. (2022) to assess hysteresis and adaptation effects. In addition, we investigate how several versions of the model compare to the empirical results obtained by Van Geert, Moors, et al. (2022). More specifically, we test whether model implementations can explain co-occurring attractive and repulsive context effects, as well as a range of plausible variation in effect size and direction across 'individuals' (in this case across simulations with different parameter values). Furthermore, we test whether the model can reproduce a positive correlation between the size of both effects, as was empirically observed in Van Geert, Moors, et al. (2022).

Different from the original Wei & Stocker (2015) model, we will not (only) implement a long-term orientation prior, but (also) take the short-term context into account: the prior and stimulus-to-sensory mapping for the second lattice will be updated based on the stimulus evidence present for and the percept of the first lattice. Different from the implementation by Fritsche et al. (2020), the model will distinguish attractive influences of the previous percept and repulsive influences of the previous stimulus evidence. Given that a mask was present in the dot lattice paradigm to avoid longer-term context effects, we only take the previous lattice into account and do not model longer-term context influences (different from what was the case in Fritsche et al., 2020). Furthermore, the dot lattice paradigm concerns multistable stimuli resulting in multi-peaked likelihood distributions, whereas previous implementations of the efficient Bayesian observer model focused on non-ambiguous stimuli (e.g., Fritsche et al., 2020; Wei & Stocker, 2015). In sum, our model builds on earlier models but makes at least three innovative contributions.

5

5.2 Methods

5.2.1 Efficient Bayesian observer model

In this study, we model the perception of two successive dot lattices within one trial of the paradigm. All model simulations were performed in R [Version 4.0.4; R Core Team (2021)].[1] All code related to this paper is openly available on the Open Science Framework: https://doi.org/10.17605/OSF.IO/48ESD.

The first lattice is a rectangular dot lattice with varying aspect ratio across trials, the second a hexagonal dot lattice with three equally dominant orientations. First, we develop the model for the percept of the first lattice. Then, we update the prior and stimulus-to-sensory mapping to predict hysteresis and adaptation effects in the perception of the second lattice. In this model, the adaptation effect will be due to efficient encoding and likelihood repulsion on the stimulus level, the hysteresis effect will be due to prior attraction on the perceptual level. Therefore, we will describe this model as *hierarchical*. The size of the adaptation effect will depend on the relative amount of stimulus noise and sensory noise present, but the size of both context effects will depend mostly on the weights given to the stimulus evidence and percept in the previous trial compared to the long-term context.

[1] For an overview of all R packages used, see Appendix C.

5.2.1.1 Perception of the first lattice

5.2.1.1.1 Prior distribution for the first lattice

In the prior distribution for the first lattice, the long-term stimulus distribution for orientation is represented. We try two variants of the prior. In a first variant, we use the same natural stimulus distribution as in Wei & Stocker (2015): $p(\theta) = c_0(2 - |sin(\theta)|)$, where c_0 is a normalization constant (cf. Figure C.1a). This natural stimulus distribution reflects the fact that horizontal and vertical orientations are more common in the natural environment than oblique orientations. On the other hand, within the dot lattice paradigm the long-term stimulus distribution is uniform: every absolute lattice orientation occurs equally frequently. Therefore, we implemented a second variant of the model, with a uniform prior distribution for the first lattice (cf. Figure 5.2a). The distribution used in this prior for the first lattice will also affect the stimulus-to-sensory mapping that is used in the calculation of the likelihood distribution for the first lattice.

5.2.1.1.2 Likelihood distribution for the first lattice

Given that the first dot lattice is rectangular, it has two dominant orientations, of which the relative dominance is dependent on the aspect ratio (AR) of the lattice. In case AR = 1, we expect the stimulus support to be equal for both orientations, which will be represented by an equal weight for both likelihood peaks in the distribution. In case AR $\neq$ 1, one of the peaks will have a stronger representation in the likelihood than the other. To arrive at a double-peaked likelihood distribution (cf. Figure 5.2b), we combine separate von Mises (i.e., circular normal) distributions for both peaks on the 180° (i.e., halfcircular) stimulus orientation space. Both distributions are generated with the same level of stimulus noise, inversely represented with the κ_{stimL1} parameter (i.e., stimulus precision), but depending on the aspect ratio, the peaks are weighted differently in the mixture distribution. In other words, aspect ratio does not influence the width of the likelihood peaks (i.e., precision κ_{stimL1}), but the relative height or weight of the 0° compared to the 90° likelihood peak. The size of the aspect ratio effect on the relative height of the 0° and 90° peaks is determined by a constant (i.e., c_{stim}). The level of sensory noise (included in the model as sensory precision: κ_{sensL1}) is assumed to be equal for both distributions.

5.2.1.1.3 Posterior distribution and percept for the first lattice

To arrive at the posterior probability distribution for the perceived orientation of the first lattice (cf. Figure 5.2c), prior and likelihood distributions are combined. From this posterior distribution, the probability of perceiving the relative 0° or 90° orientation can directly be deduced. In case one wants to derive perceptual responses, one of the two dominant orientations can be sampled with the relative posterior probability at these orientations.

5.2.1.2 Perception of the second lattice

5.2.1.2.1 Prior distribution for the second lattice

In the current version of the model, we assume two different priors: a stimulus prior affecting the stimulus-to-sensory mapping and a perceptual prior used in combination with the likelihood to form the posterior distribution. A version of the model that used the stimulus prior only could explain the occurrence of adaptation, but not the occurrence of the hysteresis effect.

5.2.1.2.1.1 Stimulus prior for the second lattice

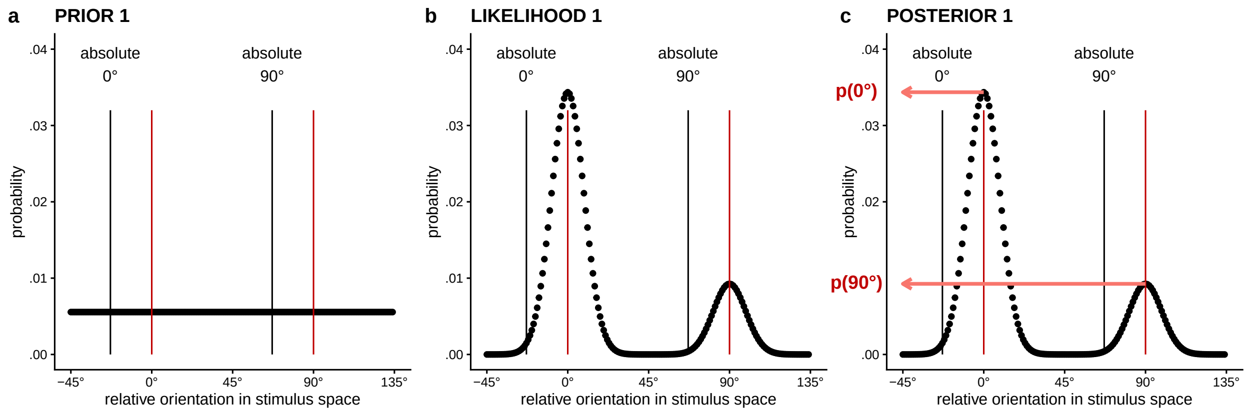


Figure 5.2: (a) Uniformly distributed prior for the first lattice. (b) Likelihood distribution defined in the stimulus space, for a first lattice with an absolute lattice orientation of 23° and AR = 1.3^{-1}, which favors the relative 0° orientation. Consequently, the two likelihood peaks are skewed away from the colored lines. (c) Posterior distribution for the first lattice. Based on the difference in height of the peaks for the relative 0° and 90° orientation, i.e., p(0°) and p(90°), the probability of a 0° or 90° response can be determined. *Note.* The red vertical lines in the graph are placed at the two dominant relative 0° and 90° orientations in the lattice. The black vertical lines label the absolute 0° and 90° orientations.

5

The stimulus prior for the second lattice (cf. Figure 5.3a) is defined as a weighted mixture between the posterior for the first lattice (representing short-term context influences based on the stimulus evidence present) and the prior for the first lattice (representing longer-term context influences of the natural stimulus distribution). If the weight of the posterior compared to that of the prior is increased (i.e., higher w_{stimL1}), the stimulus prior will update more heavily based on the immediate stimulus history.

5.2.1.2.1.2 Perceptual prior for the second lattice

The perceptual prior for the second lattice (cf. Figure 5.3b) is defined as a weighted mixture between a flat, uniform distribution and a single-peaked von Mises distribution around the perceived orientation of the first lattice. If the weight of the single-peaked von Mises distribution compared to that of the uniform distribution is increased (i.e., higher w_{percL1}), the perceptual prior will update more heavily based on the immediate perceptual history. Different from the stimulus prior, the perceptual prior thus includes direct information about the percept/decision/response concerning the first lattice. We assume the precision of the single-peaked von Mises distribution part of the perceptual prior (i.e., κ_{percL1}) to be smaller than the stimulus or sensory precision for the second lattice.

5.2.1.2.2 Likelihood distribution for the second lattice

Given that the second dot lattice is hexagonal, it has three equally dominant orientations. Therefore, we assume the stimulus support to be equal for all three orientations, which will be represented by an equal weighting of all three likelihood peaks in the distribution. To arrive at a triple-peaked likelihood distribution (cf. Figure 5.3c), we combine separate von Mises (i.e., circular normal) distributions for all three peaks on the 180° (i.e., halfcircular) stimulus orientation space, based on the stimulus-to-sensory mapping resulting from the *stimulus* prior distribution for the second lattice. All three distributions are generated with the same level of stimulus precision (κ_{stimL2}). The level of stimulus precision is kept similar to the stimulus precision for the first rectangular lattice. The level of sensory precision (κ_{sensL2}) is also assumed to be equal for all three distributions. Given that the second lattice was presented more briefly than the first lattice (300 ms vs 800 ms), we assume the sensory precision for the second lattice to be lower than the precision for the first lattice.

5.2.1.2.3 Posterior distribution and percept for the second lattice

To arrive at the posterior probability distribution for the perceived orientation of the second lattice (cf. Figure 5.3d), the *perceptual* prior distribution and the likelihood distribution are combined. From this posterior distribution, the probability of perceiving the relative 0°, 60°, or 120° orientations can directly be deduced. In case one wants to derive perceptual responses, one of the three dominant orientations can be sampled with the relative posterior probability at these orientations.

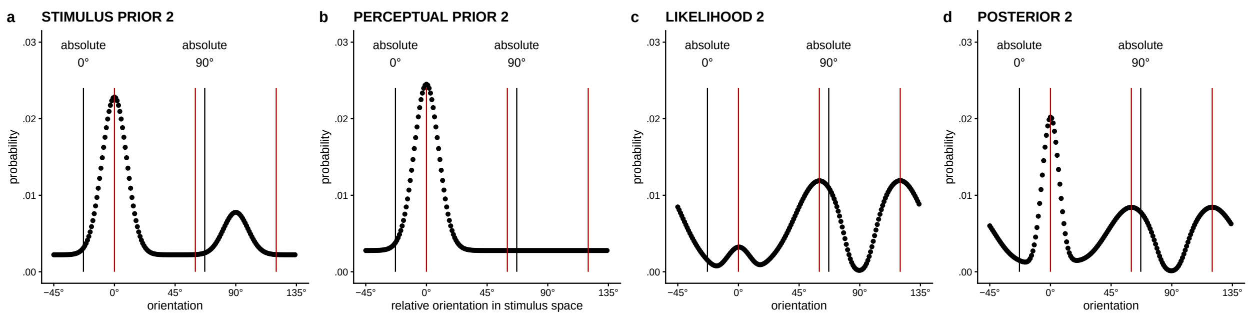


Figure 5.3: (a) Stimulus prior for the second lattice given a first lattice with AR = 1.3^{-1}, which favors the relative 0° orientation. (b) Perceptual prior for the second lattice, given the relative 0° orientation was perceived in the first lattice. (c) Likelihood distribution defined in the stimulus space for the second lattice. This distribution is influenced by the stimulus prior for the second lattice (and hence the aspect ratio of the first lattice) via the stimulus-to-sensory mapping. (d) Posterior distribution for the second lattice, combining perceptual prior and likelihood for the second lattice. Based on the difference in height of the peaks for the relative 0°, 60°, and 120° orientation, the probability of a 0°, 60°, or 120° response can be determined. *Note.* The red vertical lines in the graph are placed at the three dominant relative 0°, 60°, and 120° orientations in the lattice. The black vertical lines label the absolute 0° and 90° orientations.

5.2.2 Free parameters in efficient Bayesian observer model

$c_{\mathbf{stim}}$ influences the strength of the effect of aspect ratio on the relative difference in height between the 0° and 90° peaks in the likelihood distribution for the first lattice. When c_{stim} is increased, aspect ratio more heavily influences the difference in height for the 0° and the 90° peak in the likelihood distribution for the first lattice.

$\kappa_{\mathbf{stimL1}}$ (i.e., stimulus precision for the first rectangular lattice) and $\kappa_{\mathbf{stimL2}}$ (i.e., stimulus precision for the second hexagonal lattice) influence the general precision of the likelihood peaks for the first and the second lattice, respectively. Stimulus precision does not alter the asymmetry of the likelihood distributions in stimulus space. When κ_{stim} is decreased, lower stimulus precision or, in other words, more external stimulus noise is present.

$\kappa_{\mathbf{sensL1}}$ (i.e., sensory precision for the first lattice) and $\kappa_{\mathbf{sensL2}}$ (i.e., sensory precision for the second lattice) influence the asymmetry of the likelihood distributions for the first and the second lattice (in stimulus space), respectively. When κ_{sens} is decreased, lower sensory precision or thus more internal sensory noise is present. Given the difference in presentation time (i.e., 800 ms for the first and 300 ms for the second lattice), we assume κ_{sensL1} to be higher than κ_{sensL2}.

$w_{\mathbf{stimL1}}$ (i.e., the weight of the posterior of the first lattice on the stimulus prior for the second lattice) determines the relative influence of the short-term effect of the first lattice on the stimulus prior for the second lattice compared to the influence of the long-term natural stimulus distribution.

$\boldsymbol{w_{\mathrm{percL1}}}$ (i.e., the weight of the percept of the first lattice on the perceptual prior for the second lattice) determines the relative influence of the percept of the first lattice on the perceptual prior for the second lattice compared to a uniform distribution.

$\boldsymbol{\kappa_{\mathrm{percL1}}}$ (i.e., the precision of the peak for the percept of the first lattice) reflects the precision of the von Mises distribution used in determining the perceptual prior for the second lattice.

To visualize the prior, likelihood, and posterior distributions for one trial under different parameter settings, the reader can try out the Shiny application accompanying this manuscript (cf. Figure C.7).

5.2.3 Model calculations and analyses

To investigate the effect of different model choices and parameters, we calculated the probabilities of perceiving the relative 0° orientation in the first and the second lattice for different versions of the general model described above. For each version of the model that we investigated, we calculated the probabilities for each possible trial, with the trial defined by a combination of the aspect ratio of the first lattice (i.e., 1.3^{-1}, 1.2^{-1}, 1.1^{-1}, 1.0, 1.1, 1.2, 1.3), and the percept of the first lattice (i.e., relative 0° or relative 90° orientation). When using a non-uniform natural stimulus distribution in the prior for the first lattice, we also calculated the probabilities for each absolute lattice orientation (i.e., from 1° to 180° in steps of 1°).[2]

Our first aim was to find a model and parameter combination that matched well with the average behavioral results found in Van Geert, Moors, et al. (2022). Once this model version and parameter values was found, we manipulated each of the model parameters separately to investigate their effect on the expected probabilities of perceiving the relative 0° orientation in the first and the second lattice.

Our second aim was to introduce variation in some of the parameter values, to approximate the interindividual variation in effect size and direction found in the behavioral data for the dot lattices paradigm (Van Geert, Moors, et al., 2022). We varied (a) the constant influencing the relation between aspect ratio and differential height of the 0° and 90° peak in the likelihood for the first lattice (c_{stim}), (b) the weight of the posterior of the first lattice on the stimulus prior for the second lattice (w_{stimL1}), and (c) the weight of the percept of the first lattice on the perceptual prior for the second lattice (w_{percL1}). To investigate whether we could reproduce the strong positive correlation between individuals' hysteresis and adaptation effects found in Van Geert, Moors, et al. (2022), we drew 75 individual parameter combinations for c_{stim}, w_{stimL1}, and w_{percL1} from a truncated multivariate normal distribution with means 5, 6.5, 5, a lower boundary of zero for all three parameters, an upper boundary of 10 for w_{stimL1} and w_{percL1}, and the following variance-covariance matrix: $\begin{bmatrix} 25 & 0 & 0 \\ 0 & 9 & 8.95 \\ 0 & 8.95 & 9 \end{bmatrix}$. The w_{stimL1} and w_{percL1} parameters were then rescaled with a maximum of one instead of ten to match the zero-to-one range. We then calculated the probabilities of perceiving the relative 0° orientation in the first and the second lattice for all 75 parameter combinations and calculated the expected frequencies of each response given those probabilities.

[2] When using a uniform prior distribution for the first lattice, the absolute lattice orientation does not influence the posterior probabilities for the first or the second lattice.

To compare the variation in hysteresis and adaptation effects in the models to the variation in the behavioral results from Van Geert, Moors, et al. (2022), and also to compare the observed correlation between individual hysteresis and adaptation effects, we conducted similar Bayesian analyses as in Van Geert, Moors, et al. (2022) to the simulated data. More specifically, we estimated individual hysteresis and adaptation effects using a Bayesian multilevel binomial regression model predicting the percept of the second lattice, with aspect ratio of the first lattice (AR) and the percept of the first lattice ($R10$) as fixed and random effects. To estimate the direct proximity effect, we used a Bayesian multilevel binomial regression model predicting the percept of the first lattice, with aspect ratio of the first lattice (AR) as fixed and random effect. For more details on these Bayesian analyses, please consult Appendix C as well as Van Geert, Moors, et al. (2022).

5

5.3 Results

5.3.1 Approximation of average attractive and repulsive temporal context effects

After exploration of several parameter combinations, we were able to approximate the average behavioral results of Van Geert, Moors, et al. (2022) with both a uniform (cf. Figure 5.4a) or a natural stimulus distribution prior for the first lattice (cf. Figure C.3) and the following parameter values: $c_{\text{stim}} = 5$, $\kappa_{\text{stimL1}} = 20$, $\kappa_{\text{sensL1}} = 20$, $\kappa_{\text{stimL2}} = 20$, $\kappa_{\text{sensL2}} = 18$, $\kappa_{\text{percL1}} = 10$, $w_{\text{stimL1}} = 0.60$, and $w_{\text{percL1}} = 0.50$. Whether a uniform prior distribution or a natural stimulus distribution was used as prior for the first lattice did not visibly influence the results.

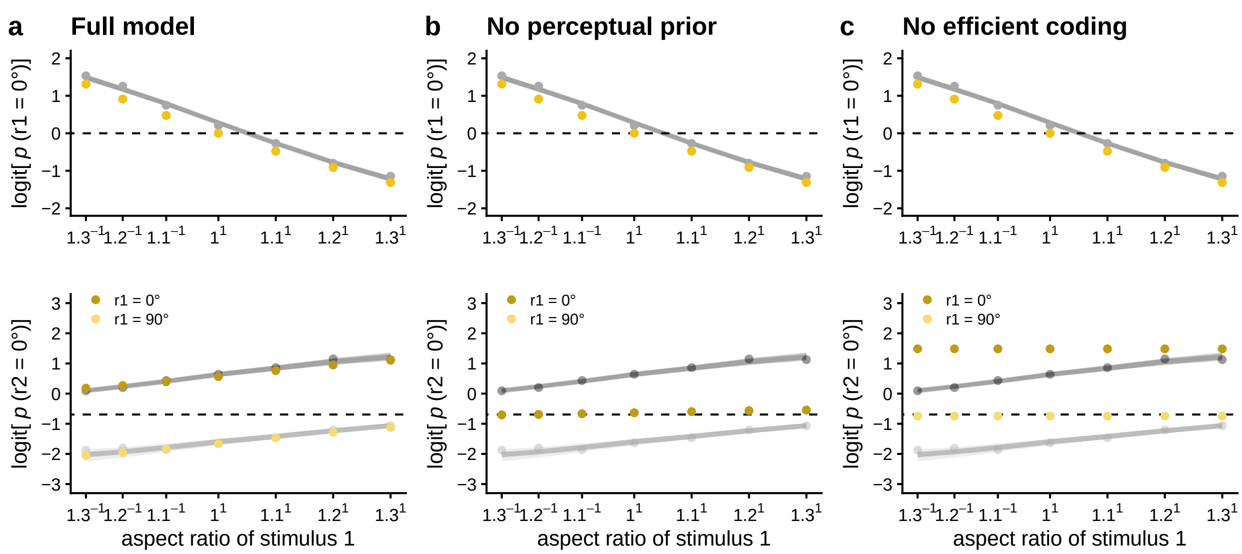


Figure 5.4: Visualization of the logit probability to perceive the relative 0° orientation in the first lattice and the second lattice, based on (a) an efficient Bayesian observer model with a flat prior distribution for the first lattice and the following parameters: $c_{\text{stim}} = 5$, $\kappa_{\text{stimL1}} = 20$, $\kappa_{\text{sensL1}} = 20$, $\kappa_{\text{stimL2}} = 20$, $\kappa_{\text{sensL2}} = 18$, $\kappa_{\text{percL1}} = 10$, $w_{\text{stimL1}} = 0.60$, and $w_{\text{percL1}} = 0.50$, (b) the same model as in (a) but using a stimulus prior rather than a perceptual prior for the second lattice, and (c) the same model as in (a) but without efficient encoding. The yellow dots indicate the expected probabilities based on the model. The behavioral results and the estimated effects based on the behavioral results of Van Geert, Moors, et al. (2022), averaged across participants, are indicated in grey.

A version of the model using the stimulus prior in combination with the likelihood for the second lattice instead of the perceptual prior was able to predict a repulsive context effect of the previous stimulus evidence, but not the attractive effect of the previous percept (cf. Figure 5.4b). Although the predicted repulsive effect is only weak in this particular version, this is a consequence of the parameter settings: as in this version of the model, the same distribution is used in the prior (resulting in prior attraction) and in the likelihood (resulting in likelihood repulsion), and w_{stimL1} and w_{percL1} are almost equal, attractive and repulsive effects largely cancel each other out. If w_{stimL1} is increased and w_{percL1} is decreased, a stronger repulsive effect is visible (cf. Figure C.4).

Is efficient encoding necessary to reproduce the behavioral results? A version of the model without efficient encoding was able to predict an attractive context effect of the previous percept (as the perceptual prior was still combined with the likelihood for the second lattice), but not the repulsive effect of the previous stimulus evidence, as that effect depends on the impact of the first lattice on the stimulus-to-sensory mapping and the likelihood of the second lattice (cf. Figure 5.4c).

5.3.2 Effects of free parameters on attractive and repulsive temporal context effects

Here we start from the final efficient Bayesian observer model with a uniform prior for the first lattice and the parameters specified above and explore the effect of each parameter separately on the expected probabilities of perceiving the relative 0° orientation in the first and the second lattice. Under these settings, c_{stim} is the only parameter influencing the size of the direct proximity effect (i.e., the effect of aspect ratio on the percept of the first lattice; cf. Figure 5.5a). Through its influence on the likelihood for the first lattice, c_{stim} also indirectly influences the size of the repulsive context effect of aspect ratio on the second lattice (cf. Figure 5.6a).

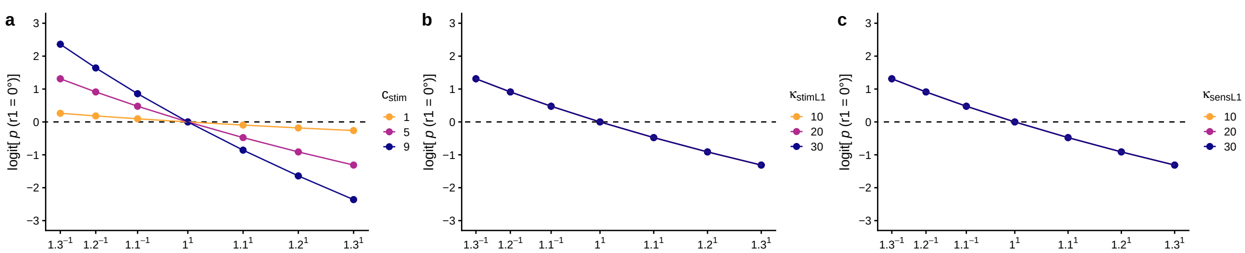


Figure 5.5: Effects of parameter variations on the logit probability of perceiving the relative 0° orientation in the first lattice, for an efficient Bayesian observer model with a uniform prior for the first lattice and baseline parameter values: $c_{\text{stim}} = 5$, $\kappa_{\text{stimL1}} = 20$, and $\kappa_{\text{sensL1}} = 20$. Under these settings, κ_{stimL1} and κ_{sensL1} do not influence the size of the direct proximity effect (i.e., the effect of aspect ratio on the percept of the first lattice).

As κ_{stimL1} only lowers overall precision of the likelihood distribution for the first lattice (which increases the influence of the prior on the posterior) and a uniform prior distribution is used, a change in κ_{stimL1} does not have an influence on the relative posterior probabilities for the 0° and 90° orientation in the first lattice. Therefore, κ_{stimL1} does not influence the size of the proximity effect in case a uniform prior is used for the first lattice (cf. Figure 5.5b). In the expected probabilities for the second lattice, a higher stimulus precision for the first lattice (i.e., κ_{stimL1}) results in slightly lower probabilities of perceiving the relative 0° orientation in the second lattice, especially for lower aspect ratios (i.e.., in favor of the relative 0° orientation). In other words, a higher κ_{stimL1} thus results in a slightly stronger repulsive effect of the previous stimulus evidence (cf. Figure 5.6b).

Given a uniform prior distribution, also κ_{sensL1} does not influence the relative posterior probabilities for the 0° and 90° orientation in the first lattice. In other words, κ_{sensL1} does not influence the size of the proximity effect in case a uniform prior is used for the first lattice (cf. Figure 5.5c). In the expected probabilities for the second lattice, a higher sensory precision for the first lattice (i.e., κ_{stimL1}) results in a slightly stronger repulsive effect of the previous stimulus evidence (cf. Figure 5.6c).

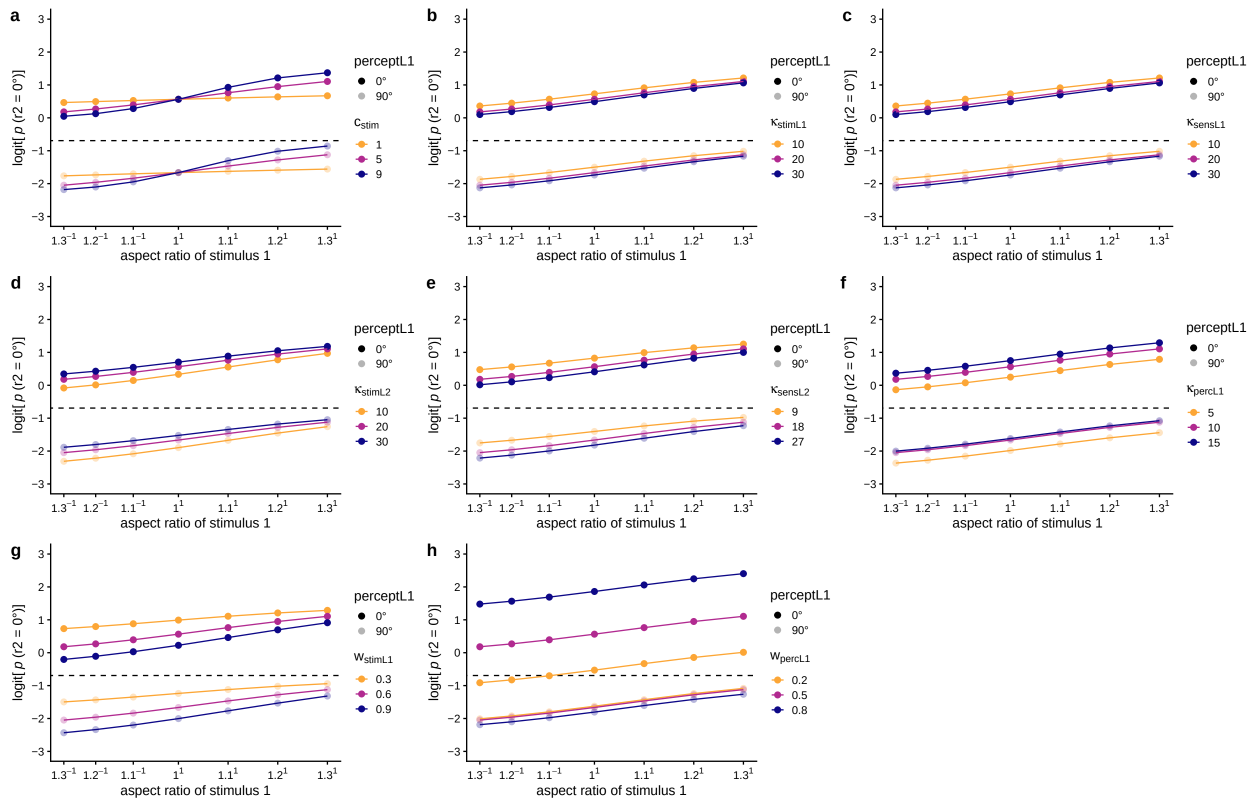


Figure 5.6: Effects of parameter variations on the logit probability of perceiving the relative 0° orientation in the second lattice, for an efficient Bayesian observer model with a uniform prior for the first lattice and baseline parameter values: $c_{\mathrm{stim}} = 5$, $\kappa_{\mathrm{stimL1}} = 20$, $\kappa_{\mathrm{sensL1}} = 20$, $\kappa_{\mathrm{stimL2}} = 20$, $\kappa_{\mathrm{sensL2}} = 18$, $\kappa_{\mathrm{percL1}} = 10$, $w_{\mathrm{stimL1}} = 0.60$, and $w_{\mathrm{percL1}} = 0.50$.

As can be seen in Figure 5.6d, increasing κ_{stimL2} slightly increases the expected probabilities for perceiving the 0° orientation in the second lattice overall, but more so for lower aspect ratios. Hence, a higher κ_{stimL2} results in a slightly shallower adaptation effect (i.e., repulsive effect of the previous stimulus evidence). Increasing κ_{sensL2} leads to the opposite effect (cf. Figure 5.6e): the higher the sensory precision for the second lattice, the stronger the adaptation effect.

The more precise the peak in the perceptual prior for the second lattice, the higher the overall probability of perceiving the relative 0° orientation in the second lattice. Although the effect of κ_{percL1} is present regardless of the percept for the first lattice being the relative 0° or the relative 90° orientation, the effect of κ_{percL1} is larger for conditions in which the relative 0° orientation was perceived in the first lattice (cf. Figure 5.6f).

Increasing the weight of the previous stimulus evidence compared to the long-term uniform stimulus distribution (i.e., w_{stimL1}) increases the size of the adaptation effect (cf. Figure 5.6g). Increasing the weight of the previous percept compared to the long-term uniform perceptual history (i.e., w_{percL1}) increases the size of the hysteresis effect (cf. Figure 5.6h).

5.3.3 Interindividual variation in proximity, hysteresis, and adaptation

When introducing interindividual variation in the parameter values for c_{stim}, w_{stimL1}, and w_{percL1}, interindividual variation in proximity, hysteresis, and adaptation effects results. With the currently used parameter combinations, the size of the hysteresis and adaptation effects varied less in the simulation data than in the empirical data, but the simulated variation is plausible given the empirical data (cf. Figure 5.7 for average results and Figure 5.8 for individual simulation results). Furthermore, the same relation between hysteresis and adaptation effects is visible as in the empirical data: By generating w_{stimL1} and w_{percL1} in a positively correlated manner, we were able to reproduce the empirically found positive correlation between individuals' attractive and repulsive temporal context effects (cf. Figure 5.9b). Different from the empirical results in Van Geert, Moors, et al. (2022), the adaptation effect showed a strong negative correlation with the direct proximity effect in the simulation results and the hysteresis effect showed no correlation with the direct proximity effect (cf. Figure C.6).

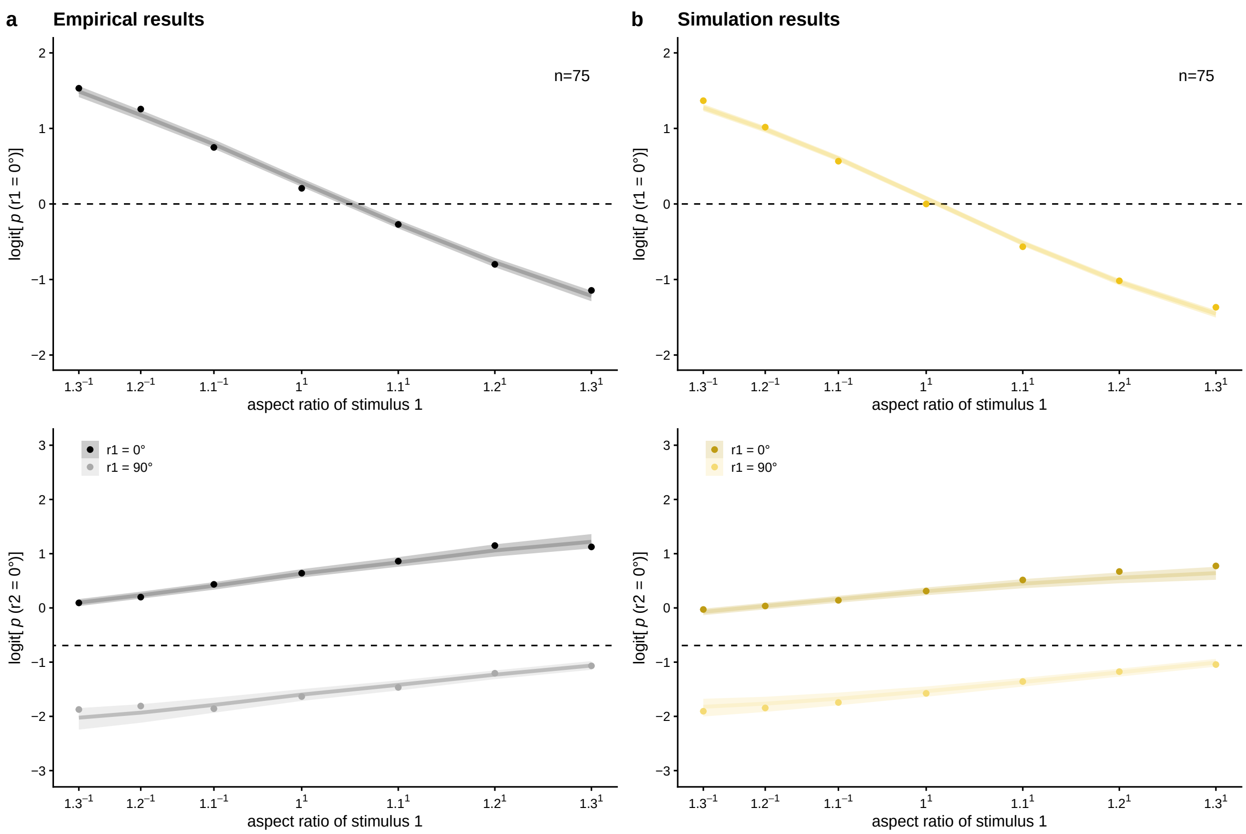


Figure 5.7: (a) Mean empirical logit probability of perceiving the relative 0° orientation in the first and the second lattice dependent on aspect ratio. The probability of responding 0° to the first lattice decreases with aspect ratio (|a|/|b|). The value of aspect ratio increases with increasing distance in the 0°-orientation, leading to more 90° responses. The probability of responding 0° to the second lattice increases with aspect ratio (|a|/|b|; i.e., adaptation effect), and increases when the first stimulus was perceived as 0° rather than 90° (i.e., hysteresis effect). (b) Mean simulated logit probability of perceiving the relative 0° orientation in the first and the second lattice dependent on aspect ratio. *Note.* Dots indicate mean values. In addition, mean posterior predictions and their 95% highest density continuous intervals are shown.

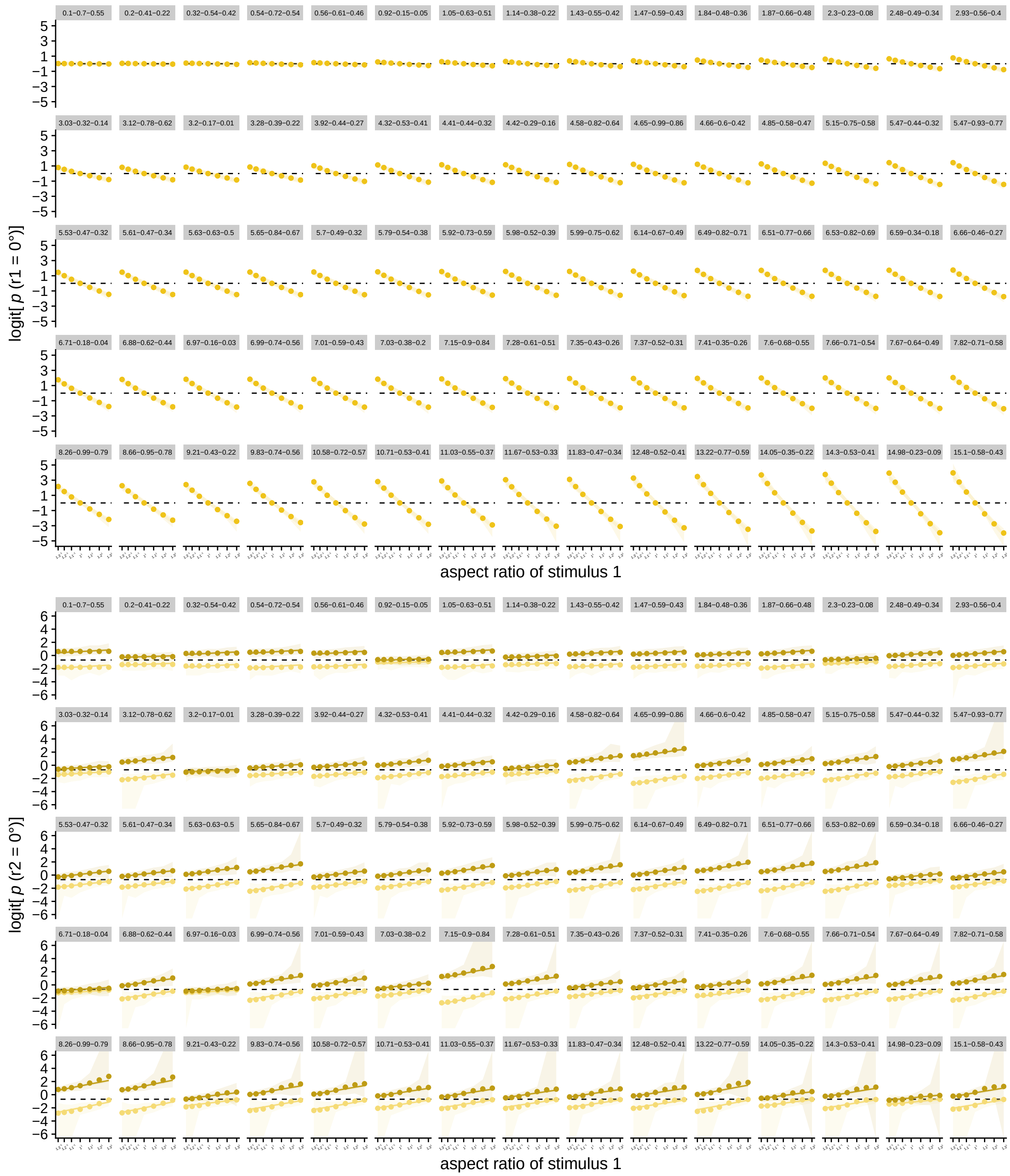


Figure 5.8: (a) Mean simulated individual responses to the first stimulus dependent on aspect ratio (logit). The probability of responding 0° to the first stimulus decreases with aspect ratio (|a|/|b|). The value of aspect ratio increases with increasing distance in the 0°-orientation, leading to more 90° responses. Dots indicate observed values. In addition, mean posterior predictions and their 95% highest density continuous intervals are shown. (b) Mean simulated individual responses to the second stimulus dependent on aspect ratio (logit). The probability of responding 0° to the second stimulus increases with aspect ratio (|a|/|b|; i.e., adaptation effect), and increases when the first stimulus was perceived as 0° rather than 90° (i.e., hysteresis effect). Dots indicate observed values. In addition, mean posterior predictions and their 95% highest density continuous intervals are shown. *Note.* Labels indicate parameter values for c_{stim}, w_{stimL1}, and w_{percL1} per simulated participant, respectively.

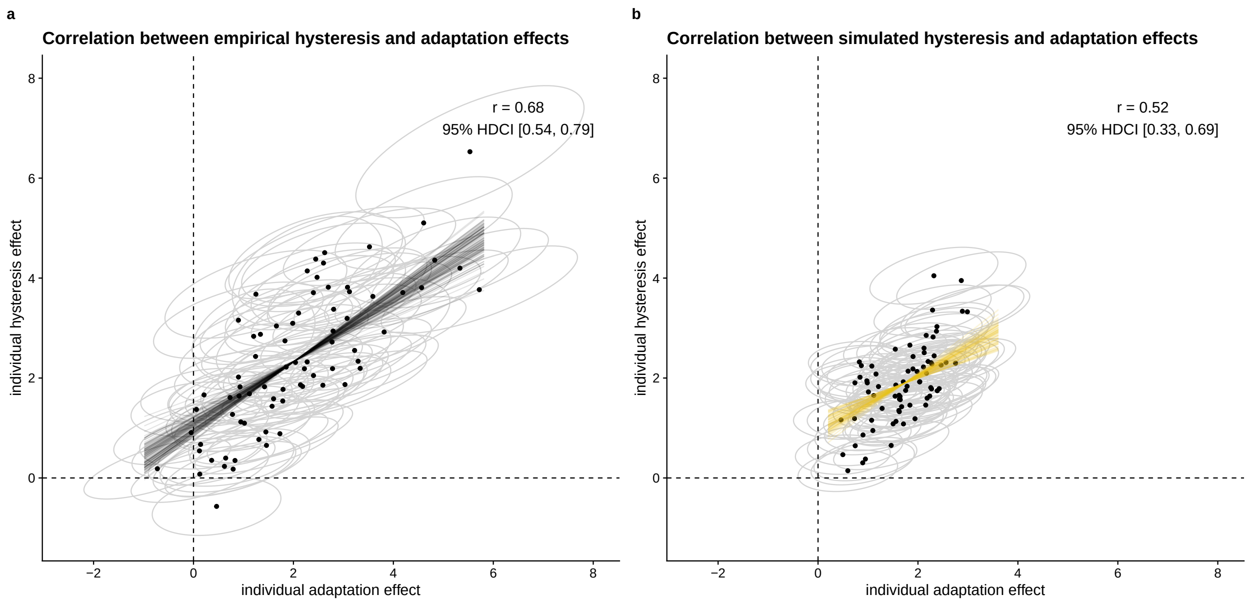


Figure 5.9: Correlation between individual slopes for the effect of aspect ratio and perceived L1 orientation on perceiving the 0° orientation in L2, for (a) the empirical data collected in Van Geert, Moors, et al. (2022) and (b) the simulated data. Mean and 80% highest density continuous intervals are shown. The dashed lines indicate a slope of zero. The black and yellow lines give examples of plausible correlation estimates.

5.4 Discussion and conclusions

We tested whether the observed attractive and repulsive temporal context effects could be explained by an efficient Bayesian observer model (Wei & Stocker, 2015), which has previously been successfully applied to many different study designs involving non-ambiguous stimulus perception (e.g., Fritsche et al., 2020; Langlois et al., 2021; Wei & Stocker, 2015). The efficient Bayesian observer model assumes variable encoding precision of orientations in line with their frequency of occurrence (i.e., efficient encoding) and takes the dissimilarity between stimulus space and sensory space into account, which leads to asymmetric likelihood distributions as a result of uncertainty induced by internal sensory noise, and consequently the possibility for 'anti-Bayesian' percepts biased away from the observer's prior beliefs.

A hierarchical efficient Bayesian observer model including both a stimulus and a perceptual level was needed to explain the co-occurrence of both attractive and repulsive temporal context effects. The aspect ratio of the first lattice (i.e., the previous stimulus evidence) affected the percept of the second lattice via the stimulus-to-sensory mapping (i.e., efficient encoding) and the likelihood (i.e., likelihood repulsion) of the second lattice. The previous percept affected the perceptual prior for the second lattice and as a consequence the posterior probability of perceiving the relative 0° orientation in the second lattice (i.e., prior attraction). In other words, efficient encoding and likelihood repulsion on the stimulus level could explain the repulsive context effect, whereas perceptual prior attraction could explain the attractive temporal context effect when perceiving multistable dot lattices. This reasoning was confirmed based on simulations from model variants without efficient coding (showing only hysteresis) and without a perceptual prior (showing only adaptation). The conclusion that a hierarchical model including both a stimulus and a perceptual level is needed, is in line with Mao & Stocker (2022), who suggested the need to take higher-level representations into account to adequately model human subjects' orientation percepts.

Not only the mean attractive and repulsive temporal context effects present in Van Geert, Moors, et al. (2022) could be reproduced using a hierarchical efficient Bayesian observer model, also plausible variation in effect size and direction could be derived by varying (a) the constant influencing the relation between aspect ratio and differential height of the 0° and 90° peak in the likelihood for the first lattice (c_{stim}), (b) the weight of the posterior of the first lattice on the stimulus prior for the second lattice (w_{stimL1}), and (c) the weight of the percept of the first lattice on the perceptual prior for the second lattice (w_{percL1}). Furthermore, the hierarchical efficient Bayesian observer model could reproduce the empirically observed strong positive correlation between individuals' attractive and repulsive effects (Van Geert, Moors, et al., 2022), by assuming a positive correlation between temporal integration constants at the stimulus and the perceptual level. That is, individuals who weight the previous stimulus evidence more highly in relation to the long-term stimulus context will also weight the previous percept more highly in relation to the long-term perceptual context than individuals who weight the previous stimulus evidence less highly. Different from the successful reproduction of the high positive correlation between attractive and repulsive temporal context effects, the correlations between the temporal context effects and the direct proximity effect did not match those observed in the empirical data. Follow-up research may aim to find parameter combinations that provide a closer match to those aspects of the empirical data.

Whereas earlier models induced a direct effect of the previous stimulus evidence on the likelihood distribution for the second lattice (Schwiedrzik et al., 2014), or posited a persistent bias for an absolute orientation but did not model the origin of the bias (Gepshtein & Kubovy, 2005), the current efficient Bayesian observer model provides a more complete process model of how previous percept and stimulus evidence can influence multistable dot lattice perception. We consider changes in the frequency of occurrence (in this case on the short term) and consequently in the prior distribution and the stimulus-to-sensory mapping a conceptually plausible explanation for the co-occurrence of separate but related short-term attractive and repulsive temporal context effects. The currently proposed model thus integrates explanations for both mechanisms in one coherent (hierarchical) theory, which was not the case in the earlier models of the dot lattice paradigm (Gepshtein & Kubovy, 2005; Schwiedrzik et al., 2014). Assuming separate but related processes underlying both context effects present at multiple hierarchical levels (i.e., likelihood repulsion on the stimulus level and prior attraction on the perceptual level, related through their dependence on the posterior for the first lattice), provides an intermediate position, in between researchers positing one single mechanism underlying both effects and researchers confirming differences in the characteristics of both effects. Furthermore, the idea of separate but related processes inherent in the efficient Bayesian observer model is highly compatible with the empirically observed high correlation between individuals' attractive and repulsive temporal context effects as observed in Van Geert, Moors, et al. (2022).

It is highly likely that our visual system takes more previous stimulus evidence or percepts into account than only one stimulus back. Another possible follow-up is therefore to take changes throughout the entire experiment into account when modeling the behavioral data resulting from the multistable dot lattice paradigm. However, the presence of a mask in between trials makes modeling this process conceptually more complex. Furthermore, when modeling the results for the current paradigm, going only one trial back, was enough to replicate the behavioral effects. It was thus not necessary to go more than one trial back to successfully account for the co-occurrence of both effects.

The current computational model can serve to inspire new experimentation. The model can generate quantitative predictions that can be tested in new experiments: for example, sensory noise can be manipulated using exposure time or stimulus contrast, the alignment of the dots in the lattice can be decreased to lower stimulus precision, or a longer inter-stimulus interval could be introduced to weaken the influence of the first lattice. Also, the same modeling approach can be adapted to other tasks measuring temporal context effects with different multistable stimuli.

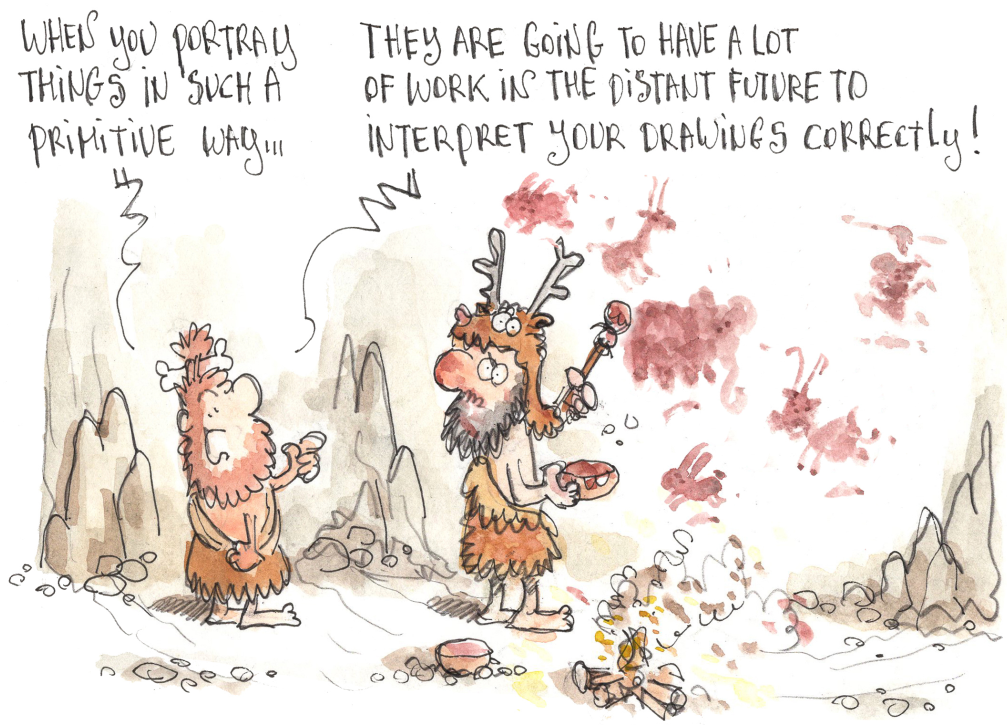
To conclude, a hierarchical efficient Bayesian observer model including both a stimulus and a perceptual level can explain repulsive temporal context effects in multistable dot lattice perception via efficient encoding and likelihood repulsion, and attractive effects via perceptual prior attraction. This conclusion is in line the conclusion of Mao & Stocker (2022), who suggested the need for considering the complex hierarchical structure of the brain, by also taking the higher-level representation into account to adequately model human subjects' response behavior.

5.5 Open and reproducible practices statement

This manuscript was written in R Markdown using the papaja package (Aust & Barth, 2022) with code for data analysis integrated into the text. The data, materials, and analysis and manuscript code for the experiment are available at https://doi.org/10.17605/osf.io/48esd.

Part V

Simplification and complication



To come to a clearer percept, we cannot only remove unnecessary details but also emphasize characteristic features.

Chapter 6

Towards the most prägnant Gestalt: Leveling and sharpening as contextually dependent adaptive strategies

Gestalt psychologists posited that we always organize our visual input in the best way possible under the given conditions. Both weakening or removing unnecessary details (i.e., leveling) and exaggerating distinctive features (i.e., sharpening) can contribute to achieve a better organization. When will a feature be leveled or sharpened, however? We investigated whether the importance of a feature for discrimination among alternatives influences which organizational tendency occurs. Participants were simultaneously presented with four figures composed of simple geometrical shapes, and asked to reconstruct one of these figures in such a way that another participant would be able to recognize it among the alternatives. The four figures differed either qualitatively or only quantitatively (i.e., far or close context). Regarding quantitative differences, two feature dimensions were varied, with one manifesting a wider range of variability across the alternatives than the other. In case of a smaller variability range, the target figure was either at the extreme of the range or had an in-between value. As expected, the results indicated that sharpening occurred more often for the feature with an extreme value, for the feature exhibiting more variability, and for the features of figures presented in the close context, than for the feature with a non-extreme value, exhibiting less variability, or in the far context. In line with Metzger's (1941) definition of prägnant Gestalts, the essence of a Gestalt is context-dependent, and this will influence whether leveling or sharpening of a feature will lead to the best organization in the specific context.

6

Van Geert*, E., Frérart*, L., & Wagemans, J. (under review). Towards the most prägnant Gestalt: Leveling and sharpening as contextually dependent adaptive strategies. Preprint available from https://doi.org/10.31234/osf.io/t3bzw

* Eline Van Geert and Liesse Frérart contributed equally to this work.

Preprint | Materials, data, and code

Towards the most prägnant Gestalt: Leveling and sharpening as contextually dependent adaptive strategies

6.1 Introduction

In the beginning of the 20th century, the Gestalt psychologist Max Wertheimer coined the law of *Prägnanz der Gestalt* or the law of a tendency towards establishing clear, simple Gestalts or experienced organizations (Schumann, 1914). The best-known English definition of the Prägnanz tendency was given by Koffka (1935, p. 110): "psychological organization will always be as 'good' as the prevailing conditions allow". Prägnanz or goodness of organization is a multifaceted concept (Metzger, 1941; Rausch, 1966; Van Geert & Wagemans, 2023). Rausch (1966) indicated that for an organization to be experienced as 'good' or 'prägnant', the organization should be experienced as lawful (i.e., non-random). Furthermore, a psychological organization will be 'better' when the organization is experienced as autonomous rather than derived, complete rather than disturbed, simple of structure rather than complicated of structure, element rich rather than meager, expressive rather than lacking expressiveness, and meaningful rather than meaningless (Rausch, 1966; cf. also Van Geert & Wagemans, 2023). These specifications of Prägnanz indicate that the goodness of an organization can increase with increasing lawfulness and regularity as well as with increasing intricacy and complexity (Rausch, 1966; Van Geert & Wagemans, 2023). Two main aspects are important to fully understand the meaning of Prägnanz, according to Metzger (1941). On the one hand, prägnant organizations show an outstanding and consequently persistent figural order (i.e., unity, regularity). On the other hand, psychological organizations are experienced as prägnant when they are completely specified structures in which the *essence* of the organization is most pure and compelling. Thus, both form-related and content-related aspects may play a role in the goodness of a psychological organization (Metzger, 1941; Rausch, 1966; Van Geert & Wagemans, 2023).

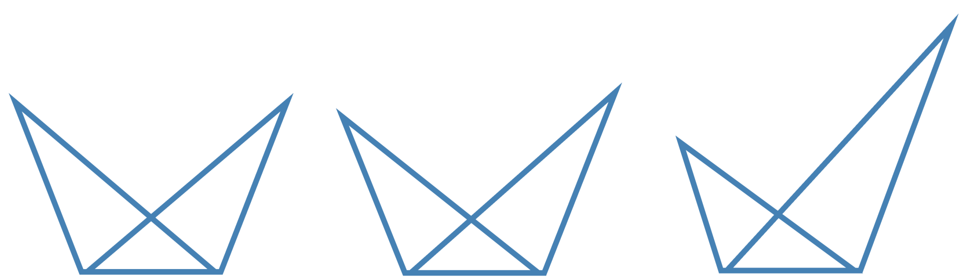


Figure 6.1: Illustration of leveling and sharpening of a figure inspired by Arnheim (1974). To imagine the original figure in a context with more symmetric figures, hide the rightmost figure with your hand. To imagine the original figure in a context with more asymmetric figures, hide the leftmost figure with your hand.

Taking the tendency towards better psychological organizations as a starting point, *how* can we clarify the incoming stimulation to come to a better psychological organization? Koffka (1935) specified two variants of the tendency towards prägnant Gestalts: either *as much* or *as little* will happen as the prevailing conditions allow. The former is referred to as maximum simplicity or the simplicity of perfect articulation, the latter as minimum simplicity or the simplicity of uniformity. For example, given a figure with slightly unequal tips as the middle one in Figure 6.1, one can either downplay this inequality and make the figure more symmetric (i.e., minimum simplicity, leveling), or one can intensify this inequality and make the figure more asymmetric (i.e., maximum simplicity, sharpening).

Importantly, which organization will be experienced as 'better' will depend on the prevailing conditions: the incoming stimulation, the perceiving individual, and the context in which the stimulation is encountered (Koffka, 1935; Van Geert & Wagemans, 2023). In this example, the experienced organization is compared to an *internal* reference (e.g., a figure with equal, symmetrical tips). The differences in comparison to this internal reference can be diminished (i.e., leveling), or they can be exaggerated (i.e., sharpening; Van Geert & Wagemans, 2023). However, if the same figure would be presented under different conditions, for example, in spatial or temporal proximity to similar figures with either more or less equal tips, a *local* reference may be used. That is, the original figure may be evaluated in the context of these other figures in the display (cf. Figure 6.1).

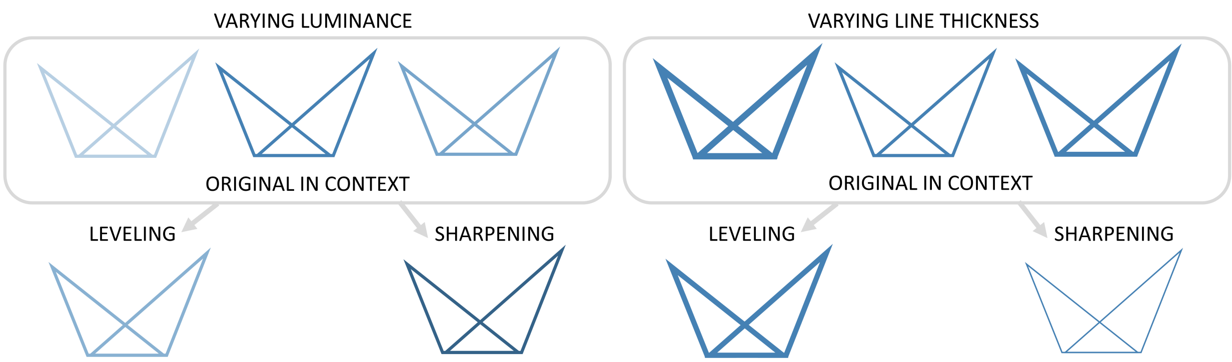


Figure 6.2: Illustration of how leveling and sharpening can be applied to different feature dimensions of a figure.

Although Koffka (1935) viewed minimum and maximum simplicity (i.e., leveling and sharpening) as alternative tendencies, Arnheim (1986) considered them as antagonistic but complementary tendencies concurrently present in every perceptual event. That is possible because perception is inherently multidimensional. In our example, we can *not only* evaluate the figure on the equality of its tips, but we can *also* evaluate the size of the figures, the thickness or luminance of the lines, etc. (cf. Figure 6.2). To clarify the incoming stimulation, tension-reducing tendencies (i.e., leveling, simplification, minimum simplicity) will increase the experienced regularity in the figure and remove distracting, unessential details, while tension-enhancing tendencies (i.e., sharpening, complication, maximum simplicity) will intensify the figure's characteristic features (Arnheim, 1986; Van Geert & Wagemans, 2023). In this way, leveling and sharpening can concurrently contribute to the Prägnanz (clarity) of a Gestalt (Arnheim, 1986; Hubbell, 1940; Köhler, 1951/1993; Van Geert & Wagemans, 2023).

Recent work on consistent, specific false memories for some images from popular iconography confirmed the idea that both leveling and sharpening[1] can lead to a better psychological organization (Prasad & Bainbridge, 2022). For example, Pikachu (from the Pokémon franchise) is consistently remembered to have a yellow tail with a black tip (i.e., adding a characteristic feature: Pikachu's ears have black tails) instead of an almost completely yellow tail with a patch of brown at the base (i.e., removing an unessential detail: the patch of brown is experienced as unessential).

Importantly, leveling and sharpening tendencies towards the best psychological organization given the prevailing conditions can occur at two levels (Hüppe, 1984; Van Geert & Wagemans, 2023). *Primary Prägnanz tendencies* occur from the stimulus to the percept: when we perceive a stimulus, our percept already deviates from the stimulus, and this deviation is unconscious to the observer (Hüppe, 1984; Van Geert & Wagemans, 2023). *Secondary Prägnanz tendencies* on the other hand operate at the level of experience: although one may be able to *perceptually discriminate* the experienced organization from the reference, one may *cognitively evaluate* the organization in relation to the reference (Hüppe, 1984; Van Geert & Wagemans, 2023; Wertheimer, 1923). In our example, one may be able to visually perceive the difference between the figure with equal and slightly unequal tips, but one may cognitively evaluate the slightly unequal figure as "almost equal" (i.e., secondary leveling). This evaluation can sometimes occur consciously (Hüppe, 1984).

Furthermore, these secondary leveling and sharpening tendencies can also be voluntarily applied as communicative strategies. This is the case, for example, in the work of artists, who will go beyond the physical stimulation they receive in the direction of Prägnanz, to more purely and compellingly represent an essence (Arnheim, 1975; Metzger, 1941). Scott McCloud (1993) pointed to the use of leveling and sharpening strategies in cartoons in his work *'Understanding comics'*. He described cartooning as a form of 'amplification through simplification': cartooning is said to be not so much about eliminating details but about focusing on specific details, and this to strip an image down to its essential meaning (McCloud, 1993, p. 30). This idea corresponds closely to the idea of leveling and sharpening as co-occurring tendencies to come to a better psychological organization (Arnheim, 1986). Secondary Prägnanz tendencies are not completely independent of primary Prägnanz tendencies, however: to be able to consciously evaluate and/or manipulate the closeness of an experienced organization to a reference, the organization first needs to be perceived.

Also in more recent scientific literature on the origin of errors in drawing, this distinction between primary and secondary Prägnanz tendencies comes back. Cohen & Bennett (1997) distinguished four psychological sources for drawing errors: (1) misperception of the object, (2) misperception of the drawing, (3) motor skills, and (4) representational decisions (cf. also Chamberlain & Wagemans, 2016). From a series of drawing experiments isolating different components of the drawing process, Cohen & Bennett (1997) concluded that *misperception* of the object was most crucial for drawing errors. Also later studies found evidence for the idea that there is close link between internal representations of objects and how they are drawn (Chamberlain & Wagemans, 2016; Cohen, 2005; Fan et al., 2018; Matthews & Adams, 2008). For example, Ostrofsky et al. (2015) and Mitchell et al. (2005) found reliable relations between illusion strength and drawing accuracy. On the other hand, also misperception of

[1] Leveling and sharpening are here interpreted in the broader sense: not only including weakening or exaggerating features, but also adding or removing features.

the drawing and *representational decisions* play a role in drawing errors (Chamberlain & Wagemans, 2016; Cohen & Bennett, 1997). Concerning representational decisions, Ostrofsky et al. (2012) found that artists were better at producing recognizable minimal line drawings than non-artists, probably because the artists included more features necessary for object recognition. Also Chamberlain & Wagemans (2016) concluded that artists probably know better which parts of a visual image are crucial, and this is important for later identification of the drawn image. Misperceptions of the object and/or the drawing may here be attributed to primary Prägnanz tendencies leading to deviations from stimulus to percept, while the representational decisions can be related to secondary Prägnanz tendencies.

Given a specific figure, which primary and secondary Prägnanz tendencies will occur? That will depend on the prevailing internal and external conditions (Koffka, 1935). *External* conditions are created in the receptor organs by the physical stimulation they receive, *internal* conditions are inherent to the human nervous system, which can be more permanent or more temporary (cf. also Van Geert & Wagemans, 2023). When the external conditions are weak (i.e., limited visibility due to, e.g., low contrast, brief presentation, or small size), primary Prägnanz tendencies will get more room and lead to tangible deviations of the percept compared to the external stimulation (Koffka, 1935). When the external conditions are stronger, primary Prägnanz tendencies will be more constrained, but secondary Prägnanz tendencies can still play an important role in how the organization is experienced.

6.1.1 Empirical evidence for Prägnanz tendencies

Wulf (1922) asked participants to reproduce abstract figures from memory after different time intervals, going from 30 seconds to two months. He found empirical evidence for both leveling (i.e., weakening one or more features of a figure) and sharpening (i.e., exaggerating one or more features of a figure). Allport (1930) conducted a similar study in large groups of children. His findings indicated a tendency towards symmetry, and several changes in the reproductions could be categorized as leveling and sharpening. Fehrer (1935) also noticed leveling and sharpening tendencies in reproductions, but under different conditions: participants were repeatedly exposed to these figures for short periods of time. In addition, she found evidence for related tendencies such as simplification and complication (i.e., the use of fewer or more parts in the reproduction compared to the original), and for tendencies towards symmetry.

Building on Wulf's (1922) study, Gibson (1929) studied memory reproductions in figures that could be changed in more diverse ways, which made it more difficult to categorize changes as either leveling or sharpening. Similar to Wulf (1922), he found a 'normalizing' tendency: when a figure had previously been associated with a familiar object, reproductions of the figure deviated in the direction of this familiar object. We could thus say that the participant's internal representation of that familiar object served as a reference. Also Granit (1922) found evidence for normalizing tendencies when he asked both children and adults to reproduce figures to which they were exposed very briefly. This normalizing tendency was interpreted by Wulf (1922) as one way in which reproductions can be leveled. Furthermore, Gibson (1929) also described leveling tendencies in comparison to a local reference: when two figures had previously been associated, one figure's reproduction frequently changed in the direction of the other figure.

Hubbell (1940) asked participants to freely change given geometrical figures until they considered them 'good' figures. Different from the above reported studies, in this study by Hubbell (1940) the viewing conditions were not limited. Therefore, this study tapped more into secondary Prägnanz tendencies, while the studies reported above focused on studying primary Prägnanz tendencies. Although a considerable proportion of the figures was simplified (i.e., removing parts), participants most frequently changed the figures in the direction of greater differentiation (i.e., complication, adding parts), but importantly, these changes enhanced the unity and coherence of the original figure. Similar tendencies were found by Wohlfahrt (1932), when he presented participants with abstract line figures in a very small size. Participants' reproductions showed tendencies towards regularity: their reproductions were more meaningful and complex than the figures from which they originated (i.e., sense making through complication). In sum, the complication of figures (i.e., addition of lines or dots) did not indicate a more complex end result in these studies; by adding lines, simplicity was achieved (Hubbell, 1940).

Also more recent work confirmed biases towards a better psychological organization. Feldman (2000) asked participants to draw a triangle and a quadrilateral, and the produced shapes were biased towards equilateral triangles and squares, respectively. Miller and Gazzaniga (1998) found evidence for false recognition of schema-consistent items in visual scenes (e.g., a beach ball in a beach scene; cf. also Brewer & Treyens (1981) and Roediger III & McDermott (1995) for the same tendency for real scenes and for words, respectively). This can be interpreted as a sharpening tendency in memory (i.e., the beach becomes more 'beachy' by adding the ball). Also, some studies reported more false recognition for gist-consistent versus gist-inconsistent pictures (Koutstaal & Schacter, 1997) and for high-frequency compared to low-frequency category exemplars (Seamon et al., 2000). Furthermore, this difference in false recognition rates increased over time (Seamon et al., 2000).

Rosielle & Hite (2009) showed that participants sharpened the small but noticeable difference in size between two or three simultaneously presented simple shapes in their drawings from memory. Also when directly copying the figure, participants showed a sharpening tendency. The authors strongly argued against explaining this sharpening as a conscious strategy (i.e., secondary Prägnanz tendency) in their study, given that almost all participants showed the sharpening tendency in spite of being told repeatedly to draw the stimuli as veridically as possible, and given that expert drawers showed reduced sharpening compared to novices. Rosielle & Hite (2009) named this effect the 'caricature' effect in drawing, after the investigations showing better and faster recognition for caricature drawings of faces compared to veridical face drawings, both for familiar and unfamiliar faces (e.g., Lee et al., 2000; Mauro & Kubovy, 1992; Rhodes et al., 1987; Robert, 1999; Stevenage, 1995). Furthermore, Rodríguez et al. (2009) showed that being familiarized with caricature faces during training improved recognition for veridical drawings in a test phase. In addition, caricature effects have also been found for other stimulus categories than faces. Rhodes & McLean (1990) observed that for both experts and non-experts, transformations of bird drawings that increased distinctiveness (i.e., sharpening, caricature) led to faster identification and higher recognition than anticaricatures (i.e., leveling). Furthermore, for experts there was also a caricature advantage, with caricatures of birds in a highly homogeneous and familiar class being identified more quickly than uncaricatured veridical drawings (Rhodes & McLean, 1990).

6.1.2 Conditional dependencies of Prägnanz tendencies

The above described studies indicate that both primary and secondary leveling and sharpening tendencies can occur to clarify an experienced organization, and that this clearer organization can lead to better identification and recognition performance. But when will leveling or sharpening of a particular feature occur? Which tendency will lead to a better psychological organization, will depend on the prevailing conditions (Koffka, 1935; Van Geert & Wagemans, 2023). These can be roughly categorized in stimulus-, person-, and context-level conditions.

Concerning *stimulus*-level conditions, Allport (1930) indicated that broken lines, acute angles and marked asymmetry are features likely to be sharpened. Goldmeier (1972) coined the idea of 'singular' properties (i.e., properties sensitive to change) as being the properties more likely to be sharpened.

Besides stimulus-level differences, also *person*-level differences may play a role in whether a feature will be sharpened or leveled. For example, Koffka (1935) suggested that, in a state of high energy, individuals will tend towards maximum simplicity (i.e., sharpening), while they will tend towards minimum simplicity (i.e., leveling) when in a low energy state. Furthermore, Wulf (1922) indicated that the same feature could be leveled or sharpened by different participants, depending on how they view or understand a figure. Different views corresponding to the same objective figure are in different respects 'imperfect' or 'bad' and will therefore be changed into different directions (Wulf, 1922). Also the study of Carmichael et al. (1932) confirmed that how a figure is interpreted by the viewer is crucial for how it will be drawn} (cf. also Bartlett, 1932). More specifically, they found that the presentation of linguistic labels together with a figure influenced how the figures were drawn.

In a more recent drawing study, Long et al. (2021) found evidence that with age, children improved in their ability to produce drawings including diagnostic visual information about the intended category, and this coincided with an improved ability to use this diagnostic information for recognition of other children's drawings. Furthermore, these skills were correlated at the category level: in case dogs were drawn better, they were on average also recognized better than other categories (Long et al., 2021). In addition, changes in children's drawings across development showed to be surprisingly similar for drawings from observation and from memory (Long et al., 2022). Hence, also a person's developmental stage can influence how objects are perceived and reproduced.

Moreover, participants may also differ in a more stable way. Based on a personality test, Holzman & Gardner (1960) divided their participants in levelers and sharpeners, and asked both groups to recall a story. Sharpeners' recall of the story was superior over that of levelers: they retained the overall theme more, and their stories were better organized and less vague than the ones of levelers.

A third group of factors influencing whether a feature will be leveled or sharpened are *context*-level conditions. Ostrofsky et al. (2015) asked participants to reproduce an indicated angle that was presented as part of a simple geometric figure, both using an adjustment-based task and a drawing task. In both tasks, the average size of the reproduced angle was significantly influenced by the figure in which the angle was embedded (Hammad et al., 2008; cf. also Kennedy et al., 2008). Mitchell et al. (2005) showed similar evidence for the influence of context on perceptual distortions and drawing errors: the experienced perceptual distortion of Shepard stimuli as a pair of tables resulted in larger illusion and larger drawing errors than the same stimuli with the table legs removed. Blake et al. (2015) asked participants to recognize and draw the famous Apple logo, and both recognition in a forced-choice task and produced drawings were surprisingly poor in details. This may be due to the fact that under a naturalistic setting (i.e., in a naturalistic context), the misremembered details are superfluous and unimportant to recognize the logo (Blake et al., 2015). Only in the context of the experiment, this simplified internal representation seemed insufficient.

Fan et al. (2020) paired participants in an online game, where one participant was assigned the role of viewer and the other participant the role of sketcher. Both participants were shown four real-world objects, and the sketcher had to draw the indicated object such that the viewer could pick the same object out of the four objects presented, based on the drawing. In one condition of the experiment, the four objects from which the viewer had to choose belonged to the same category (e.g., four birds; i.e., close context condition). In the other condition, the four objects belonged to different object categories (i.e., far context condition). Fan et al. (2020) found that sketchers used more time and ink in the close context condition compared to the far context condition. This result indicates that simplification occurred in a context where greater detail was superfluous to capture the object. In the close context, certain features might have been complicated or sharpened to better capture the essence of the object in that context. Also Yang & Fan (2021) pointed to the importance of task context for explaining how people use drawings to communicate visual concepts in different ways.

6.1.3 Current study

The current study builds on the methodology of Fan et al. (2020) to answer the question whether the importance of a feature for discrimination within a specific (task) context influences which Prägnanz tendency (i.e., leveling or sharpening) will occur. In other words, is the essence of a Gestalt dependent on the context, and does it therefore influence which Prägnanz tendency will occur (and thus whether a feature will be leveled or sharpened)?

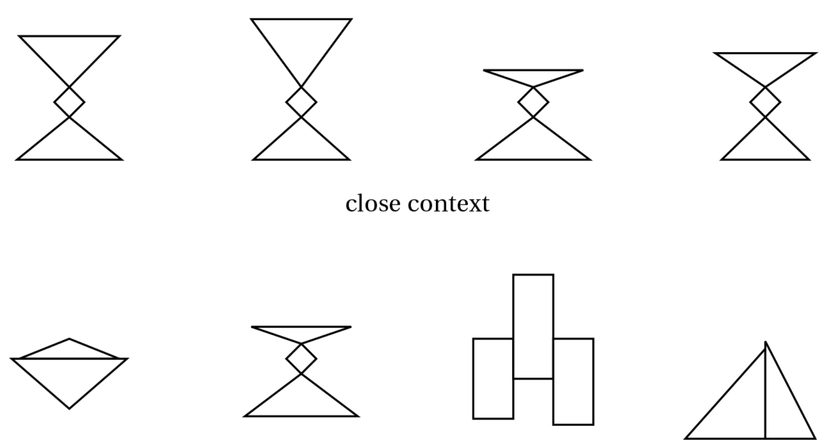


Figure 6.3: Example of stimulus design 1 in a close and far context condition. In this example given for the close context, the height of the upper triangle has major variability, and the width of the lower triangle has minor variability.

It was hypothesized that features important for discrimination of a figure from other figures would more often be sharpened than features that are less important for discrimination. In this online study, participants were presented with four figures and asked to reconstruct one of these, using basic geometric shapes. The target figure either appeared in the context of three other figures differing only quantitatively from the target figure on two feature dimensions (i.e., close context), or in the context of three qualitatively different figures (i.e., far context; see Figure 6.3). It was assumed that the quantitative feature differences would become rather unimportant for discrimination in the far context, leading to more sharpening of the features in the close than in the far context, and more leveling of the features in the far than in the close context. To make one of the two feature dimensions more important for discrimination, the variability of the feature's values across the four figures was four times larger than for the other feature dimension. It was hypothesized that high variability would make the feature more important for discrimination, and that the feature would therefore more often be sharpened in the case of high compared to low variability. On the feature dimension with low variability, the target figure could either have an extreme value in the context of the four figures present, or an in-between value, the latter making the feature even less useful for differentiation of the target figure from the distractors. One would therefore expect more sharpening for features with an extreme rather than an in-between value on the feature dimension.

The results were analyzed both qualitatively, i.e., looking at the proportion of times leveling or sharpening occurred in a certain context, and quantitatively, i.e., looking how strongly a tendency occurred. Additionally, we explored the effect of individual and stimulus differences on the context dependency of the Prägnanz tendencies.

6.2 Methods

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

6.2.1 Participants

553 Dutch-speaking first-year psychology students from KU Leuven participated in the study, of which 546 participants completed at least one drawing and 541 completed all 24 drawings. The number of participants was dependent on the number of students subscribing and actually taking part in the online study. Data collection was combined with data collection for another online study (following after completing participation in the current study). Data was collected in three separate waves between March and November 2021. Participants were granted 0.5 research credits for a complete participation. Of the 546 participants used in the analyses, 473 (86.63%) were female. Age of the participants varied between 17 and 60 years (M_{age} = 18.69 years, SD_{age} = 2.95 years).

6.2.2 Material

6.2.2.1 Stimuli

A new set of 24 stimulus designs (see Figure 6.4) was created to meet the goals of the study. These designs were black-and-white abstract line figures constructed out of basic geometric shapes (i.e., rectangles, rectangular triangles, and equilateral triangles). The number of basic shapes from which a design was constructed varied between two and five basic shapes (4 designs with two basic shapes, 8 with three basic shapes, 8 with four basic shapes, and 4 with five basic shapes). Within each design, two feature dimensions were varied. These dimensions were selected to be variable in a quantitative way, because they needed to vary in intensity rather than content (Gati & Tversky, 1982), and to be perceptually separable, because selective attention to separate dimensions was necessary to make

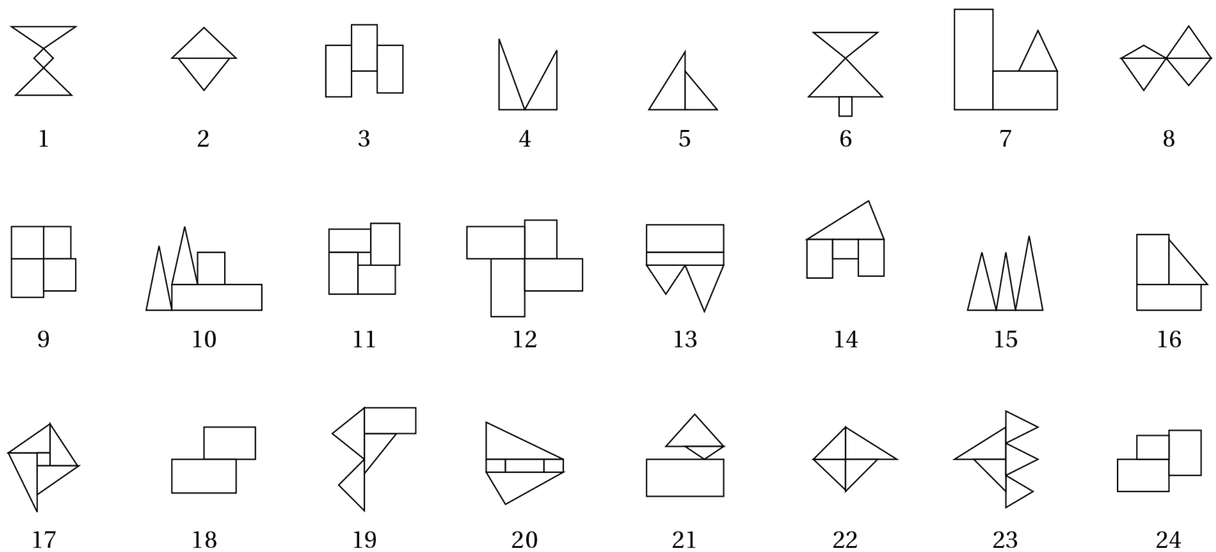


Figure 6.4: Illustration of all 24 stimulus designs.

independent decisions about a single feature dimension (Dunn, 1983). The possible feature dimensions included: height or width of one of the basic shapes, or the position of a point of a triangle relative to its opposite side. By means of varying these feature dimensions quantitatively, for each design two series of four stimuli were constructed, an A and a B series (see Figure 6.5). In the A series, the variability on the first feature dimension was higher than the variability on the second feature dimension, whereas in the B series the variability on the second feature dimension was higher that the variability on the first feature dimension. Specific rules were followed to determine the range of the feature values and to select the possible target stimuli (see Appendix D).

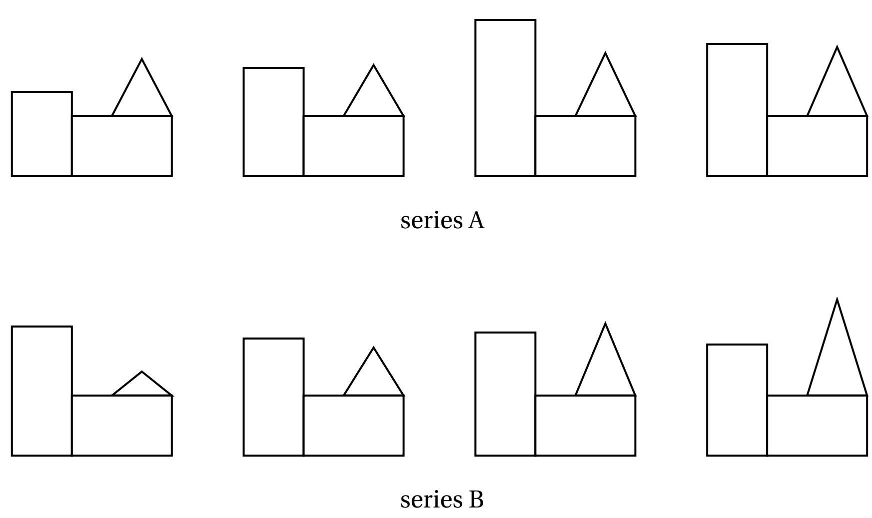


Figure 6.5: Series A and series B for design 7. In series A, the height of the left rectangle has major variability, and the height of the triangle on the right has minor variability. This is opposite for series B.

6.2.3 Procedure

The study was conducted online, and was programmed in HTML, CSS, and JavaScript, using the jsPsych 6.1 and Fabric.js 4.2 JavaScript libraries (de Leeuw, 2015; http://fabricjs.com/). At the start of the experiment, participants were briefly informed about the general procedure, and were asked for informed consent to participate in the experiment. To fix the size of the stimuli in the experiment independent of the screen resolution of the participant, a resizing procedure was introduced, in which the participant resized a rectangular container on the screen to match the size of a particular physical object (i.e., student card, identity card, or bank card). The scaling factor was then applied to the rest of the experiment. Participants were instructed to sit at an arm's length distance directly in front of the screen, and to keep that distance as constant as possible throughout the experiment. After providing their gender and age, participants were redirected to a practice trial. In the practice trial, participants were given the task instructions and the online interface was explained. Before continuing to the experimental task, participants were able to try out the interface (see Figure 6.6) themselves.

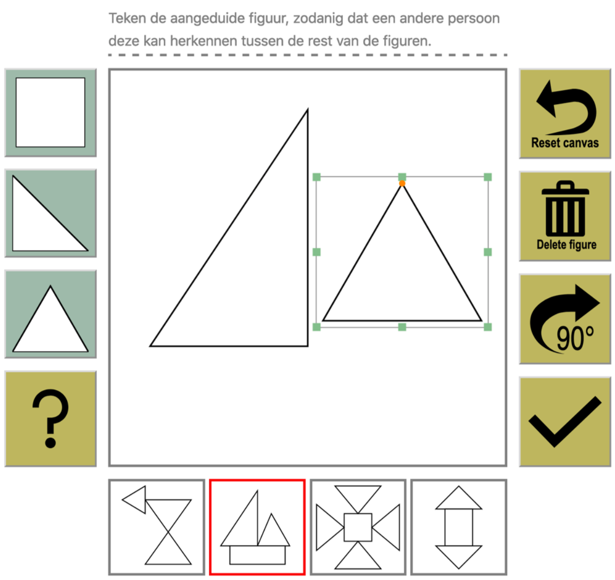


Figure 6.6: Interface of the online drawing experiment. *Note.* In English, the Dutch instruction can be translated as: "Draw the indicated figure, in such a way that another person can recognize it among the rest of the figures".

In each trial, participants were presented with a set of four stimuli (100 x 100 user units each), simultaneously shown in boxes at the bottom of the screen. The box around the target figure was outlined in red. The goal of the participant in each trial was to construct the target figure by means of basic geometric shapes, in such a way that someone else would be able to recognize it among the other figures shown next to it.

20 out of 24 designs consisted of one or more additional shapes that did not contain a relevant feature dimension. When the target was of one of those designs, the additional shapes (from now on referred to as *background shapes*) were already depicted on the canvas, to make the task less demanding and more controlled (line width = 2 user units, 500 x 500 user units). To allow for additional space to sharpen the relevant feature dimension, the ratio of the size of these additional shapes to the size of the stimuli presented was 4:1 (whereas the canvas size was 5:1).

The basic shapes (i.e., rectangles, rectangular triangles, and equilateral triangles) were presented as buttons on the left side of the screen. When they were clicked, the requested shapes appeared on the left side of the canvas. Participants were able to reposition and rescale a shape, as well as make it higher, lower, wider and narrower. The position of the top point of the equilateral triangle could be moved horizontally. By means of separate buttons, participants could also rotate the shape in steps of 90 degrees, remove it from the canvas, or clear the canvas entirely. Participants received unlimited time to construct the figures. When participants were unsure about the instructions during the experiment, they could click a help button that made the instructions reappear. When they completed a drawing, participants had to click a button to go to the next trial. It was only possible to proceed when at least two shapes were added to the canvas, otherwise they received the following warning message: "Attention! The drawing is not yet complete".

Across trials, the similarity of the target to the distractors was manipulated, yielding two types of contexts, the close and the far context (see Figure 6.3). In the close context, target and distractors differed only quantitatively on the two relevant feature dimensions, whilst they were qualitatively the same, making the values on the dimensions potentially important for discrimination. In the far context, target and distractors were qualitatively different, making the values on the dimensions rather unimportant for discrimination.

Independently for each participant, the 24 designs were randomly assigned a context, i.e., close or far, such that 12 designs were presented in a far and 12 designs were presented in a close context. Within each context condition, for six designs series A and for six designs series B was presented. Within each series and context condition, one of the two possible targets was randomly chosen, such that three times the target with the minimum value on the dimension with the largest variability was used, and three times the target with the maximum value on the dimension with the largest variability was used. The trial order as well as the position of target and distractor stimuli in a series was randomized. Participation took approximately 30 minutes. After completion, participants could indicate whether they wanted to receive the debriefing about the main goals and/or the results of the study.

6.2.4 Data analysis

We used R [Version 4.0.4; R Core Team (2021)] for all our analyses.[2]

6.2.4.1 Preprocessing

Several preprocessing steps were required before we could calculate the relevant measures of interest (see Appendix D). Difference scores were then calculated between the drawn feature values and the target values, with negative values indicating leveling and positive values indicating sharpening.

6.2.4.2 Qualitative analyses

In the qualitative analyses of the data, we used drawn feature values relative to the available drawing space on the canvas (for sharpening and leveling combined) to calculate the proportion of times a feature dimension was leveled or sharpened depending on the context and variability conditions.[3] We defined leveling and sharpening in multiple ways. Firstly, we defined leveling as a drawn value on the same side of the target value as the mean, and sharpening

[2] For an overview of all R packages used, see Appendix D.

[3] We also conducted the analyses based on the absolute feature values. These additional results can be found in Figure D.4 and Figure D.11 in Appendix D.

as a drawn value equal to the target value or more extreme (i.e., on the other side of the target value compared to the mean). Secondly, we included a third, 'neutral' category around the target value which could be seen as equivalent to the target value (only possible when using the relative feature values, i.e., relative to the available space for drawing a specific feature dimension). Thirdly, we explored an alternative definition of leveling as a drawn feature value closer to the mean than to the target, and sharpening as a drawn feature value being closer to the target than to the mean. The results for this last interpretation are available in Appendix D.

To visualize the data, we plotted the proportion of times a feature value was leveled or sharpened, separate for each context and variability condition, with feature dimensions as data points. Using the qualitative data, we were interested in investigating whether the proportion of sharpening was higher (and the proportion of leveling was lower) for (a) drawings in the close context compared to the far context; (b) feature dimensions exhibiting major variability rather than minor variability (especially in the close context); and (c) feature values on an extreme of the dimension rather than exhibiting an in-between value (especially in the close context). To investigate our research questions, we fitted hierarchical regression models using the brms (Bürkner, 2017, 2018) package in R (see below as well as Appendix D for more information).

For each of the definitions of sharpening, we fitted a hierarchical Bayesian binomial regression model to the proportion of times a dimension was sharpened, with context, variability, and their interaction as fixed effects and feature dimension and participant ID as random effects for both intercept and slopes. Given the second definition of sharpening, i.e., including a neutral range around the target value, we also fitted hierarchical Bayesian binomial regression models to the proportion of times a dimensions was leveled or in the 5% range around the target value.

To investigate whether the proportion of sharpening, leveling, or being in target range was higher for drawings in the close compared to the far context, for major compared to minor variability dimensions, and for extreme compared to non-extreme feature values, we plotted the posterior estimates from the model and compared the slope strengths across conditions by plotting contrast distributions for the slopes.

6.2.4.3 Quantitative analyses

In the quantitative analyses, we used the drawn feature values relative to the available drawing space on the canvas. We explored whether the *extent* to which a feature was leveled or sharpened differed between context and variability conditions. It is important to keep in mind that given the definition of the feature values relative to the total available drawing space (for leveling and sharpening combined), the available range for sharpening was bigger for dimensions with less variability, leading to a conservative estimate for the variability and extremeness effects.

To visualize the data, we plotted the difference between the drawn feature value and the target feature value relative to the available drawing space for that feature dimension, separate for each context and variability condition, with feature dimensions as data points. Using the quantitative data, we were interested in investigating whether the extent to which a feature was sharpened was on average larger for (a) drawings in the close context compared to the far context; (b) feature dimensions exhibiting major variability rather than minor variability (especially in the close context); and (c) feature values on an extreme of the dimension rather than exhibiting an in-between value (especially in the close context). Also for the quantitative data, we fitted hierarchical regression models using the brms (Bürkner, 2017, 2018) package in R to answer our research questions (see below as well as Appendix D for more information).

We fitted a hierarchical Bayesian regression model to the difference in feature value between drawing and target, with context, variability, and their interaction as fixed effects and feature dimension and participant ID as random effects for both intercept and slopes. To investigate whether the extent of sharpening was larger for drawings in the close compared to the far context, for major compared to minor variability dimensions, and for extreme compared to non-extreme feature values, we plotted the posterior estimates from the model and compared the slope strengths across conditions by plotting contrast distributions for the slopes.

6.3 Results

6.3.1 Qualitative analysis

6.3.1.1 Traditional interpretation of leveling and sharpening (binarized; relative values)

Figure 6.7 shows the empirical percentage of drawings in which each of the 48 feature dimensions was sharpened, separately for each context (i.e., close or far) and variability (i.e., major extreme, minor extreme, minor non-extreme) condition, as well as the distribution across feature dimensions. In this figure, sharpening is defined as a drawn value equal to or more extreme than the target value (i.e., on the other side of the target value compared to the mean). As expected, Figure 6.7 indicates that sharpening was more likely for the major variability dimension than for the minor variability dimensions, and that sharpening was more common in the close than in the far context. Furthermore, it is clear that there was much more variability in the percentage of sharpening across feature dimensions in the minor variability conditions than in the major variability condition.

To investigate the effects of context, variability, and extremeness in more detail, we plotted the posterior estimates from the model and compared the slope strengths across conditions by plotting contrast distributions for the slopes.

Figure 6.8 visualizes the estimated context effect per variability condition. Especially in the major variability condition, there was a clear effect of context on the probability that the feature was sharpened: in all posterior samples, the probability for the major feature dimension to be sharpened was higher in the close than in the far context. Given that the model is a good approximation of the data, the data provide evidence for a clear context effect in the major variability condition. For the minor variability conditions, the context effect was less clear.

Figure 6.9 pictures the estimated variability effect per context condition. Within the close context, there was a clear effect of the range of variability present on a dimension on the probability that the feature was sharpened: in all posterior samples, the probability for an extreme value on the major feature dimension to be sharpened was higher than the probability for an extreme value on a minor feature dimension. Given that the model is a good approximation of the data, the data provide evidence for a clear effect of variability in the close context condition. In the far context condition, the posterior probability of the proportion of sharpening for an extreme value on a major feature dimension to be larger than for an extreme value on a minor feature dimension was 89%.

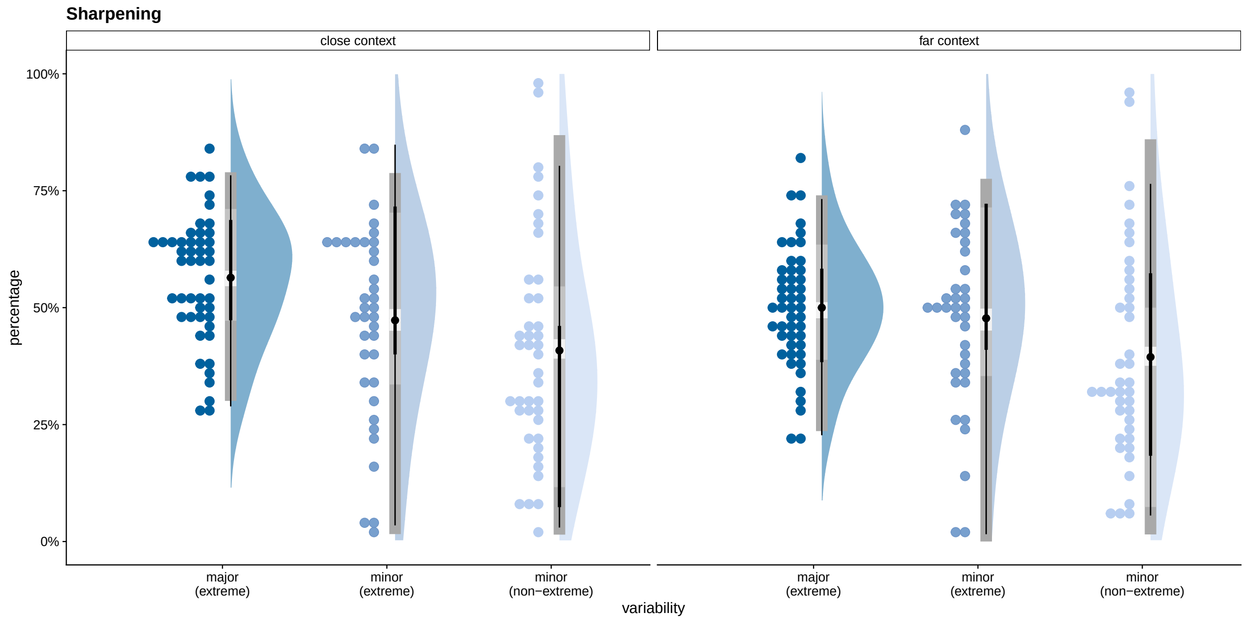


Figure 6.7: Distribution of percentage of times a feature dimension was sharpened, per context and variability condition, with feature dimensions as individual data points. The black point and intervals indicate mean and 66% and 95% highest density continuous intervals for the data distribution. The grey intervals indicate the 66% and 95% highest density continuous intervals for the posterior predictive distribution, based on the qualitative model of sharpening as binary variable using the relative feature values. In addition, the white interval indicates the 95% highest density continuous interval for the mean posterior prediction.

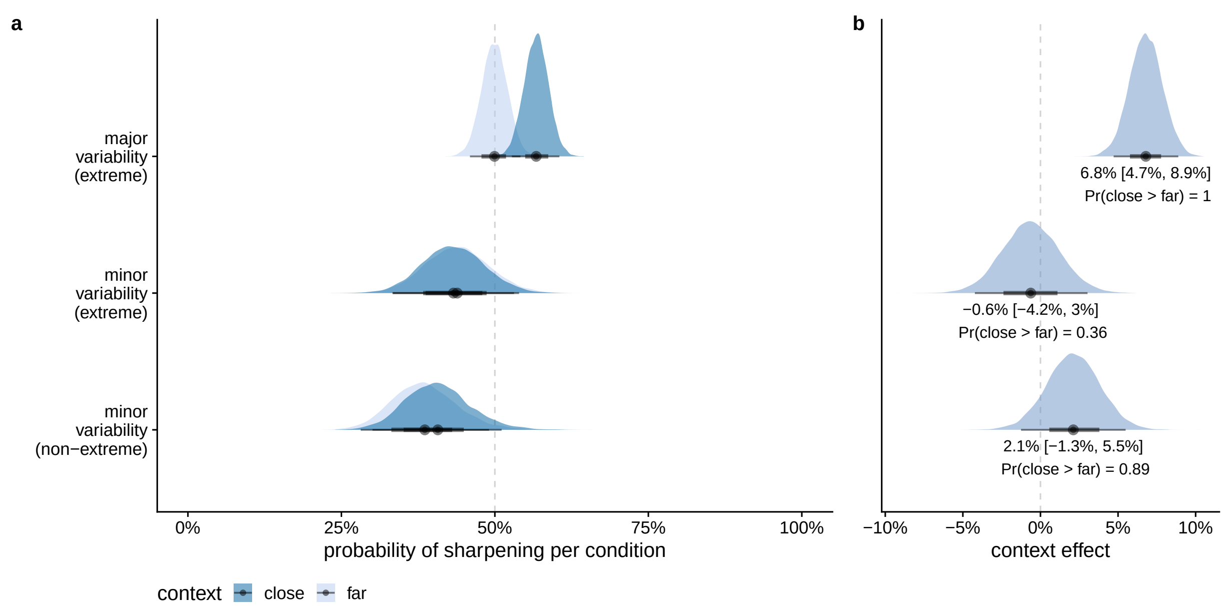
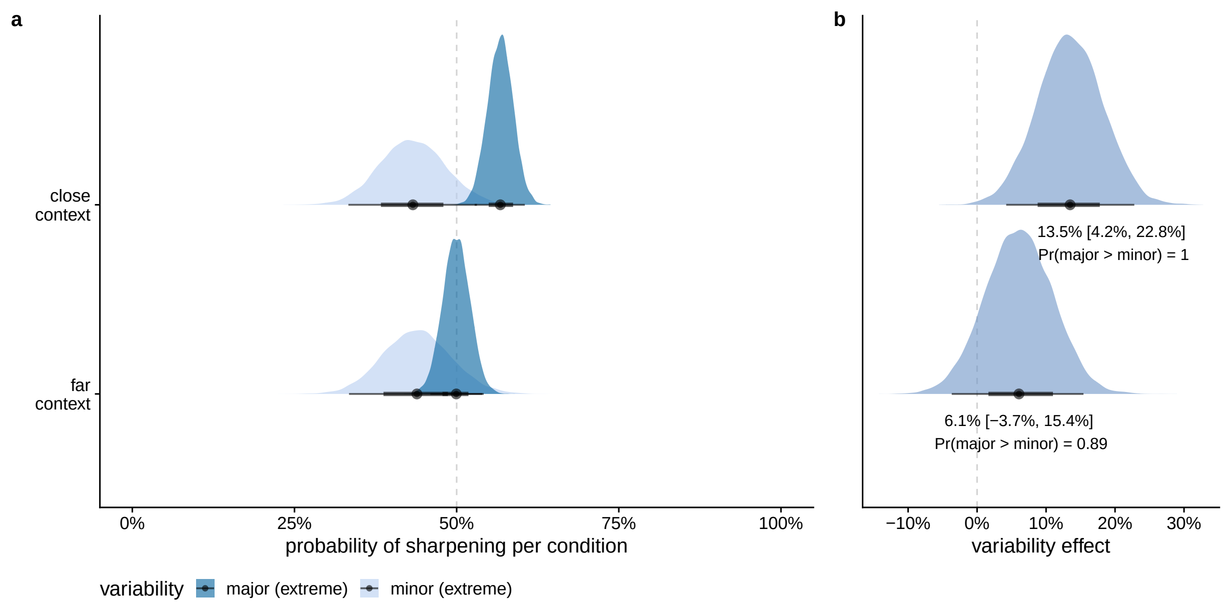


Figure 6.8: (a) Posterior distribution for the probability of drawing a feature equal to the target value or more extreme, separately for each context and variability condition. (b) Estimated context effect in each variability condition. Mean, 66%, and 95% highest density continuous intervals are shown. The text also indicates the mean estimated difference with its 95% highest density continuous interval, as well as the posterior probability of the percentage of sharpening being higher in the close than in the far context, given a particular variability condition.



6

Figure 6.9: (a) Posterior distribution for the probability of drawing a feature equal to the target value or more extreme, separately for the major (extreme) and minor (extreme) variability and context conditions. (b) Estimated variability effect in each context condition. Mean, 66%, and 95% highest density continuous intervals are shown. The text also indicates the mean estimated difference with its 95% highest density continuous interval, as well as the posterior probability of the percentage of sharpening being higher in case the feature exhibits major compared to minor variability, given a particular context condition.

Figure 6.10 shows the estimated extremeness effect per context condition. Within both the close and the far context, there was a tendency for extrema on a feature dimension to be sharpened more often than non-extreme values on feature dimensions with an equal variability range, but in both conditions the 95% HDCI for the difference between an extreme and a non-extreme feature value on the minor dimension included zero as well as differences in the opposite direction.

Additional results concerning the random effects per feature dimension and individual are provided in Appendix D.

Independent of the fitted model, we plotted the empirical percentage of drawings in which each of the 48 feature dimensions was sharpened, separately for each context (i.e., close or far), variability (i.e., major extreme, minor extreme, minor non-extreme), and target (maximum/above mean, minimum/below mean) condition, as well as the distribution across feature dimensions. From visual exploration of the figure, there seems to be no overall effect of target condition on the proportion of times a feature was sharpened (see Figure 6.11).

6.3.1.2 Traditional interpretation of leveling and sharpening including neutral category (relative values)

Figure 6.12 shows the empirical percentage of drawings in which a feature dimension was leveled, in a 5% range around the target, or sharpened, averaged across all feature dimensions and participants. From this figure, it is clear that the number of drawn values in the target range increased for the minor non-extreme variability condition compared to the minor extreme variability condition (i.e., extremeness effect), and for the minor extreme variability condition compared to the major extreme variability condition (i.e., variability effect). The opposite trend was observed for the number of sharpened feature values: sharpening was more prominent in the major extreme

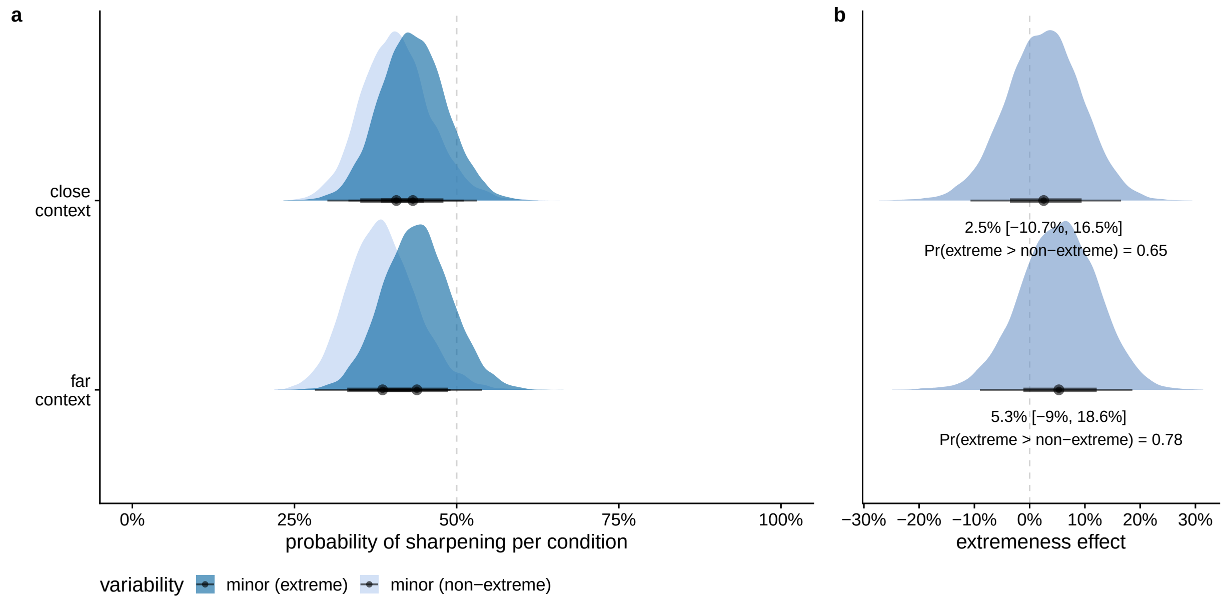


Figure 6.10: (a) Posterior distribution for the probability of drawing a feature equal to the target value or more extreme, separately for the minor (extreme) and minor (non-extreme) variability and context conditions. (b) Estimated variability effect in each context condition. Mean, 66%, and 95% highest density continuous intervals are shown. The text also indicates the mean estimated difference with its 95% highest density continuous interval, as well as the posterior probability of the percentage of sharpening being higher in case the feature value is one of the extrema on the dimension, given a particular context condition.

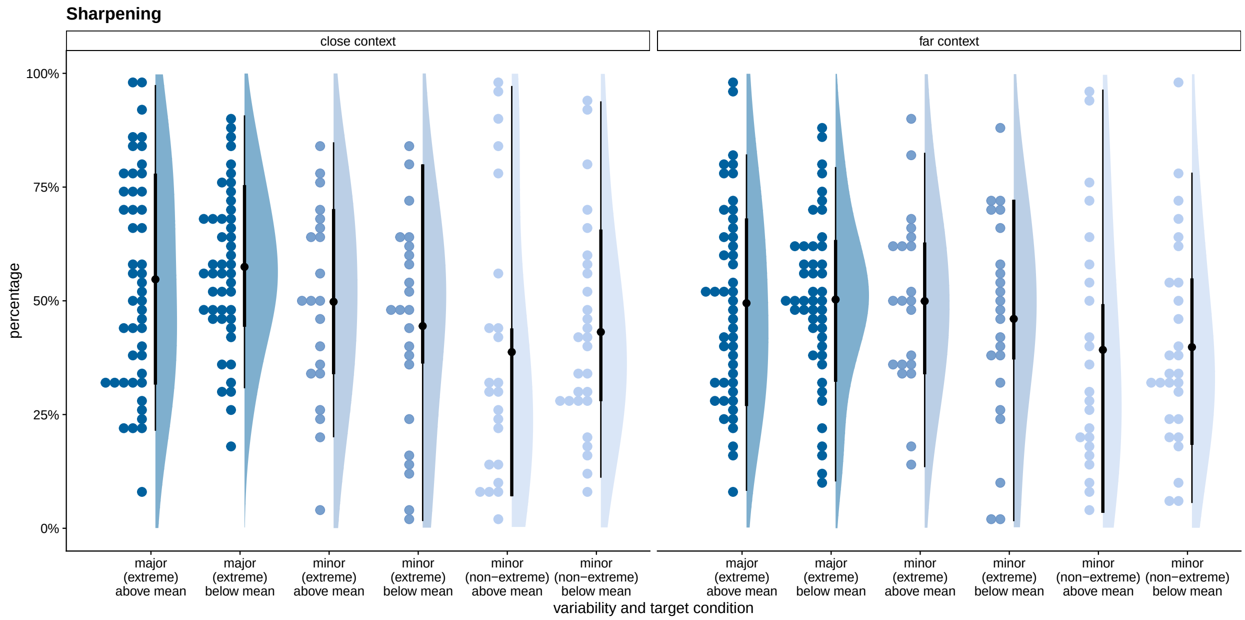


Figure 6.11: Distribution of percentage of times a feature dimension was sharpened, per context, variability, and target condition, with feature dimensions as individual data points. The black point and intervals indicate mean and 66% and 95% highest density continuous intervals for the data distribution.

6

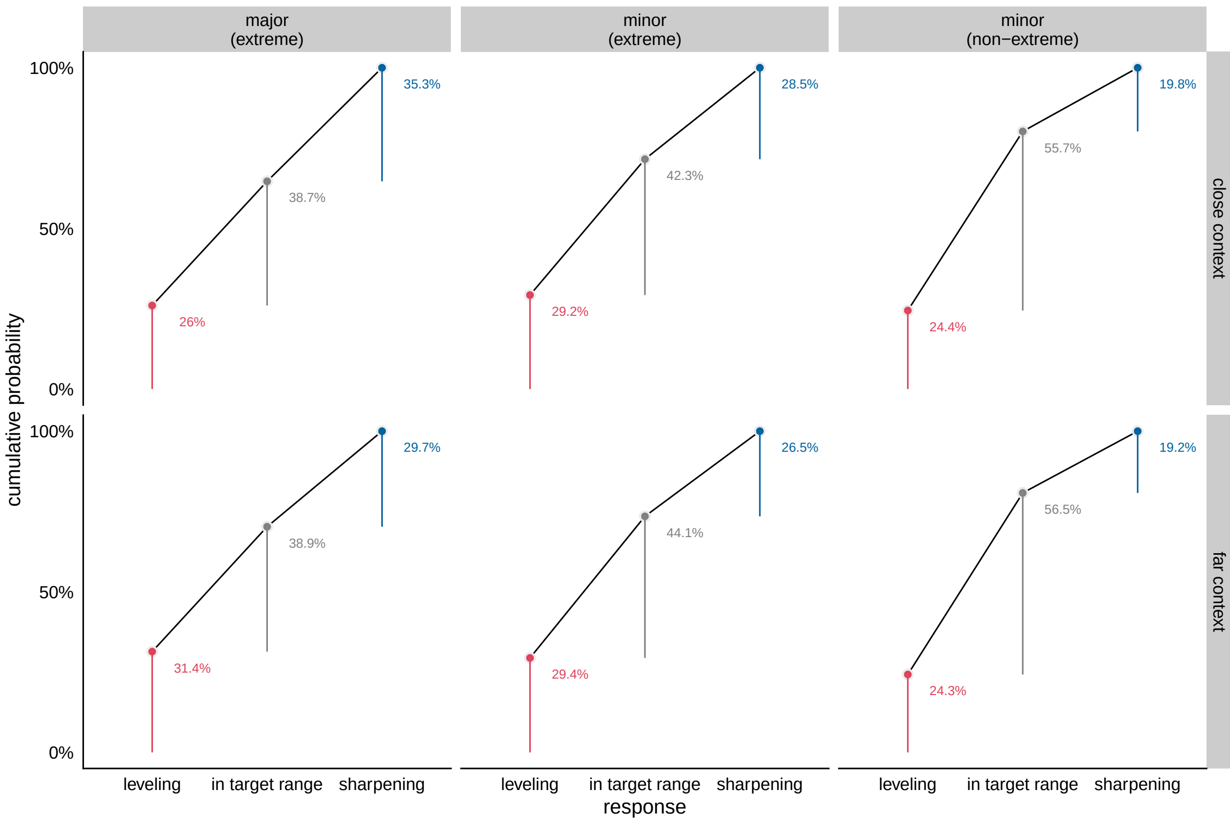


Figure 6.12: Empirical percentages of times a feature value was leveled, in target range, or sharpened (defined as a drawn value at least 2.5% larger or smaller than the target and more extreme), based on the relative feature values, and averaged across feature dimensions and participants.

condition than in the minor extreme condition (i.e., variability effect) and in the minor extreme condition than in the minor non-extreme condition (i.e., extremeness effect). In addition, the number of sharpened drawings seems to be larger in the close than in the far context, especially for the major variability dimension (i.e., context effect). Furthermore, the number of leveled feature values seems to be higher for the extreme variability conditions than for the non-extreme variability condition (i.e., extremeness effect).

The top left graph in Figure 6.13 shows the empirical percentage of drawings in which each of the 48 feature dimensions was sharpened, separately for each context (i.e., close or far) and variability (i.e., major extreme, minor extreme, minor non-extreme) condition. In this figure, sharpening is defined as a drawn value 2.5% away from the target value and more extreme than the target value (i.e., on the other side of the target value compared to the mean). As for the binary definition of sharpening, the figure indicates that sharpening was more likely for the major variability dimension than for the minor variability dimensions, and that sharpening was more common in the close than in the far context. Moreover, there was slightly more variability in the percentage of sharpening across feature dimensions in the minor variability conditions than in the major variability condition.

To investigate the effects of context, variability, and extremeness on the probability of sharpening a feature in more detail, we plotted the posterior estimates from the model and compared the slope strengths across conditions by plotting contrast distributions for the slopes.

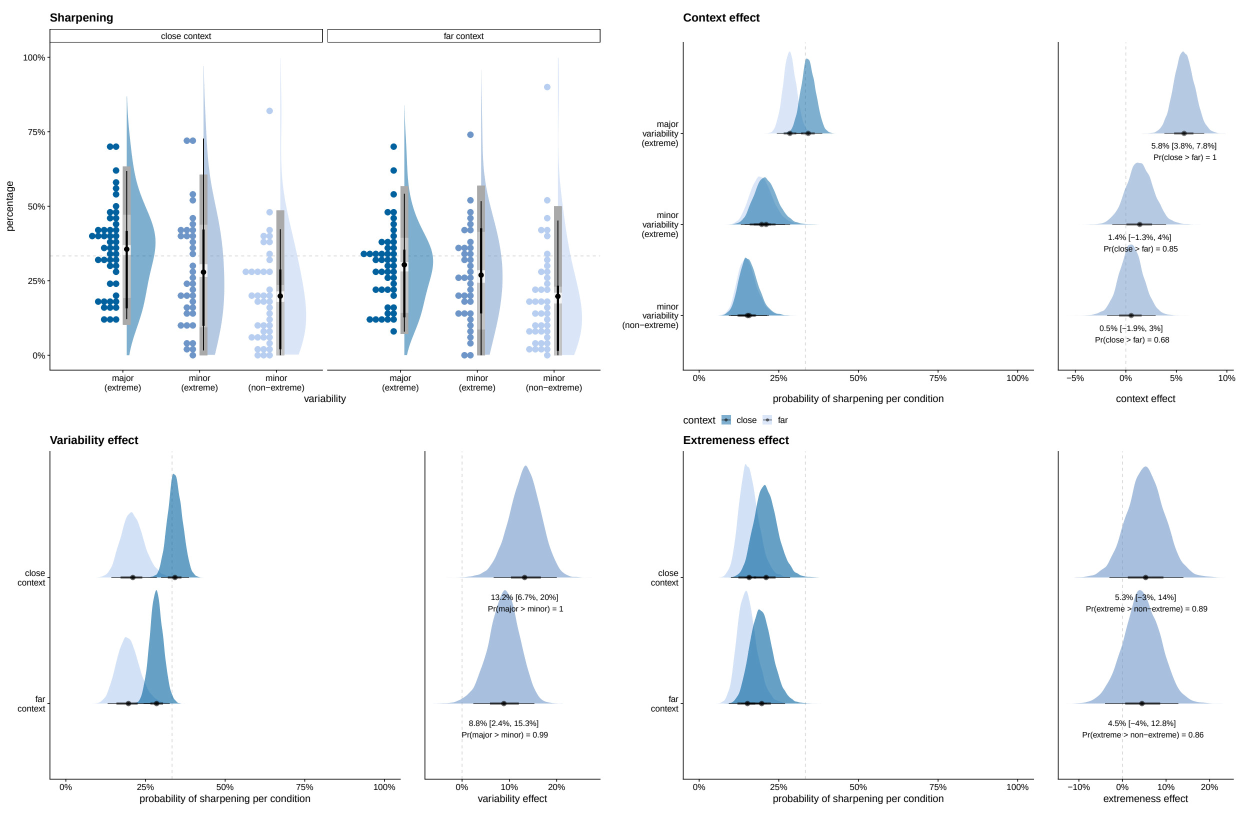


Figure 6.13: Results of the qualitative model of sharpening, defined as a drawn value at least 2.5% larger or smaller than the target and more extreme, based on the relative feature values.

The top right graph in Figure 6.13 visualizes the estimated context effect per variability condition. Concerning the major feature dimension, there was a clear effect of context on the probability that the feature was sharpened: in all posterior samples, the probability of sharpening was higher in the close context than in the far context condition. Within both the close and the far context, there was a clear effect of the range of variability present on a dimension on the probability that the feature was sharpened (see bottom left graph in Figure 6.13): in 100% and 99% of the posterior samples respectively, the probability for an extreme value on the major feature dimension to be sharpened was higher than the probability for an extreme value on a minor feature dimension to be sharpened. Within both the close and the far context, there was a slight tendency for extrema on a feature dimension to be sharpened more often than non-extreme values, on feature dimensions with an equal variability range. In the close and far context conditions, the posterior probability of the proportion of sharpening for an extreme value on a minor feature dimension to be larger than for a non-extreme value on a minor feature dimension were 89% and 86%, respectively.

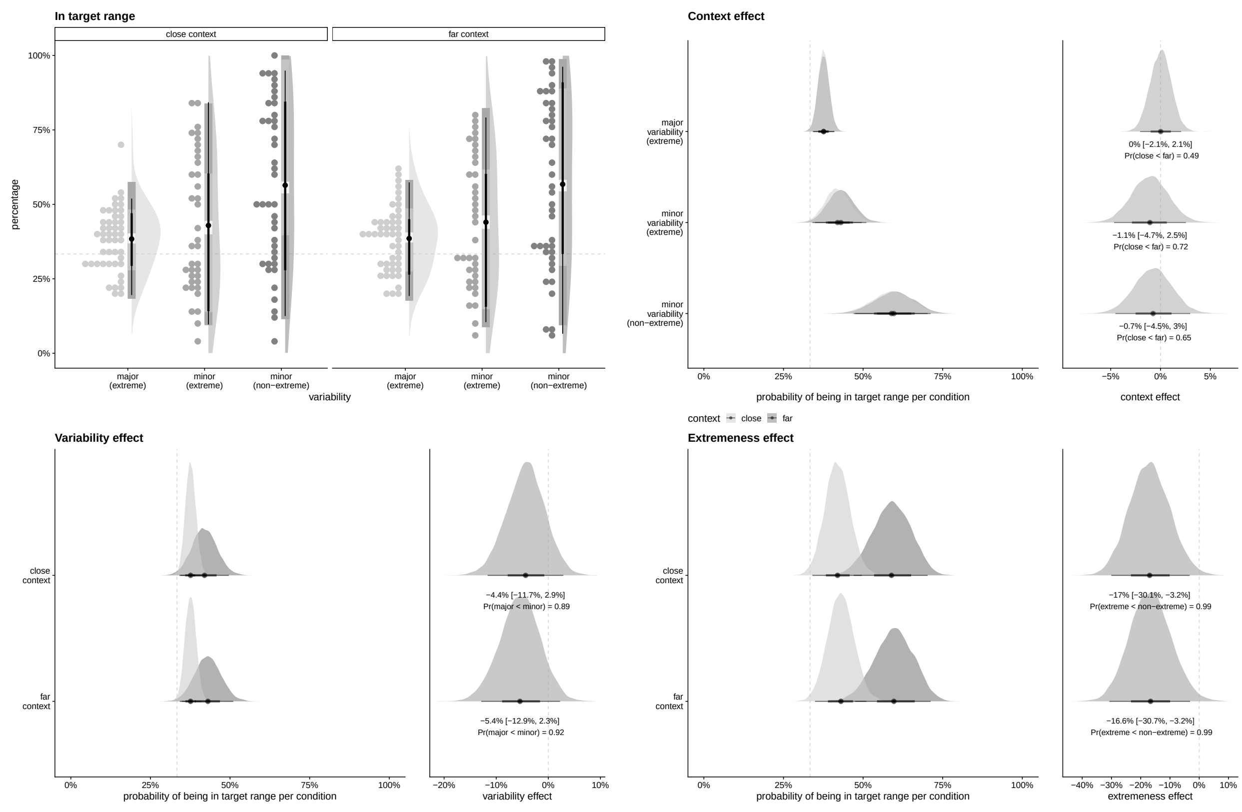


Figure 6.14: Results of the qualitative model of being in the target range, defined as a drawn value within 5% range around target, based on the relative feature values.

Figure 6.15 visualizes the estimated effects of context, variability, and extremeness for the probability of leveling a feature value. Concerning the major feature dimension, there was a clear effect of context on the probability that the feature was leveled: in all posterior samples, the probability of leveling was higher in the far context than in the close context condition. No clear effect of the range of variability present was found on the probability that a feature was leveled (see bottom left graph in Figure 6.15). Within both the close and the far context, there was a tendency

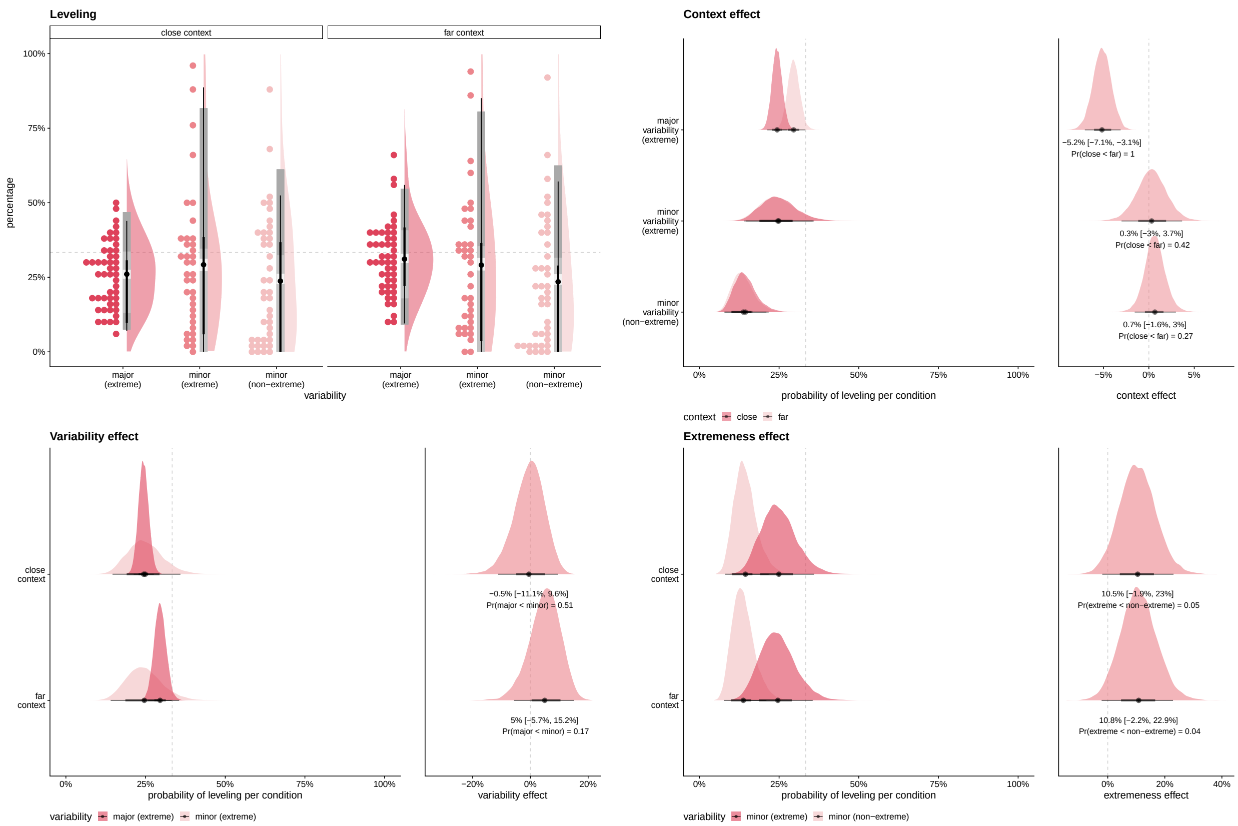


Figure 6.15: Results of the qualitative model of leveling, defined as a drawn value at least 2.5% larger or smaller than the target and less extreme, based on the relative feature values.

6

for extrema on a feature dimension to be leveled more often than non-extreme values on feature dimensions with an equal variability range. In the close and far context conditions, the posterior probability of the proportion of leveling for an extreme value on a minor feature dimension to be larger than for a non-extreme value on a minor feature dimension were 95% and 96%, respectively.

Figure 6.14 visualizes the estimated effects of context, variability, and extremeness for the probability of drawing a feature within the target range. There was no clear effect of context on the probability that the feature was in the target range. Within both the close and the far context, there was a tendency for the minor variability dimension to be in target range more often than the major variability dimension (see bottom left graph in Figure 6.14). Within both the close and the far context, there was a clear effect of extremeness: non-extreme feature values on a feature dimension were more often in target range than extreme values, on feature dimensions with an equal variability range. In both the close and the far context condition, the posterior probability of the proportion of being in target range for a non-extreme value on a minor feature dimension to be larger than for an extreme value on a minor feature dimension was 99%.

6.3.2 Quantitative analysis

Figure 6.16 shows the distribution of difference between drawn and target value (relative to the available drawing space for that feature dimension), per context and variability condition, with values for individual trials as data points. Figure 6.17 shows the mean empirical difference between drawn and target feature values relative to the available drawing space for each of the 48 feature dimensions, separately for each context (i.e., close or far) and variability condition (i.e., major extreme, minor extreme, minor non-extreme), as well as the distribution across feature dimensions. The figure indicates that the extent of sharpening was on average slightly larger for the major variability dimension than for the minor variability dimensions, especially in the close context.

To investigate the effects of context, variability, and extremeness in more detail, we plotted the posterior estimates from the model and compared the slope strengths across conditions by plotting contrast distributions for the slopes.

Figure 6.18 visualizes the estimated context effect per variability condition. In the major variability condition, there was a clear effect of context on the extent to which the feature was sharpened: in 99% of the posterior samples, the expected sharpening for the major feature dimension was larger in the close than in the far context. Given that the model is a good approximation of the data, the data provide evidence for a clear effect of context in the major variability condition. For the minor variability conditions, there was no clear effect of context.

Figure 6.19 pictures the estimated variability effect per context condition. Within the close context, there was a clear effect of the range of variability present on a dimension on the extent to which the feature was sharpened: in 98% of the posterior samples, the expected sharpening for an extreme value on the major feature dimension was higher than for an extreme value on a minor feature dimension. For the far context condition, the variability effect was in the expected direction in 84% of the posterior samples.

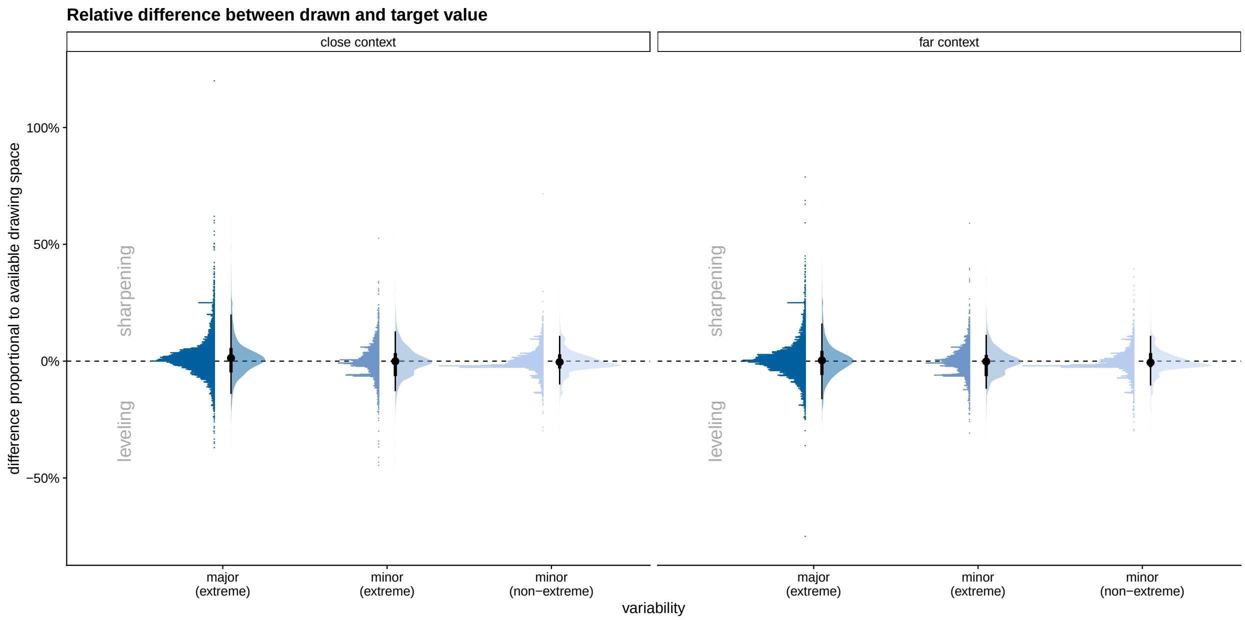


Figure 6.16: Distribution of difference between drawn and target value (relative to the available drawing space for that feature dimension), per context and variability condition, with values for individual trials as data points. The black point and intervals indicate mean and 66% and 95% highest density continuous intervals for the data distribution.

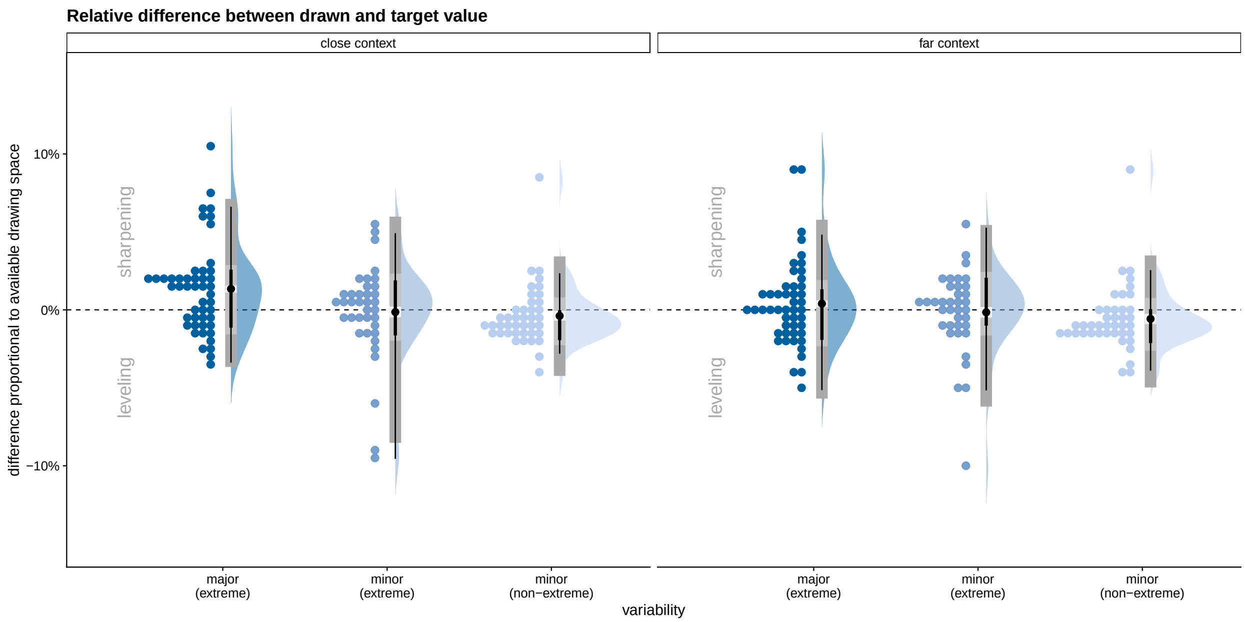


Figure 6.17: Distribution of difference between drawn and target value (relative to the available drawing space for that feature dimension), per context and variability condition, with feature dimensions as individual data points. The black point and intervals indicate mean and 66% and 95% highest density continuous intervals for the data distribution. The grey intervals indicate the 66% and 95% highest density continuous intervals for the posterior predictive distribution, based on the quantitative model using the relative feature values. In addition, the white interval indicates the 95% highest density continuous interval for the mean posterior prediction.

6

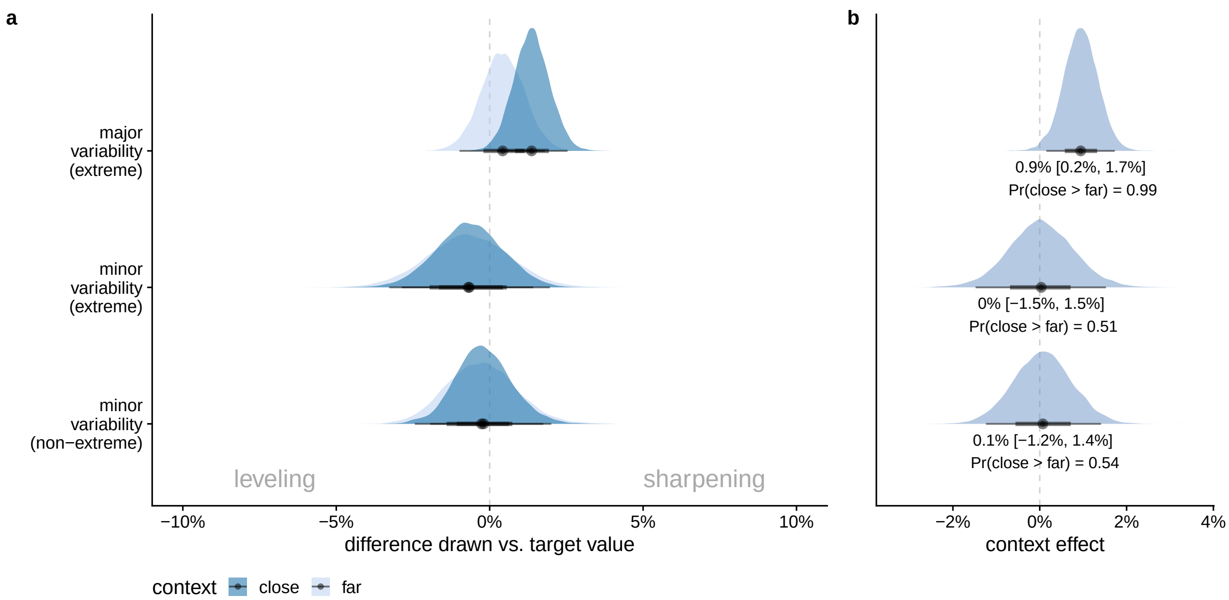


Figure 6.18: (a) Posterior distribution for the signed difference between drawn and target value, separately for each context and variability condition. (b) Estimated context effect in each variability condition. Mean, 66%, and 95% highest density continuous intervals are shown. The text also indicates the mean estimated difference with its 95% highest density continuous interval, as well as the posterior probability of the extent of sharpening being higher in the close than in the far context, given a particular variability condition.

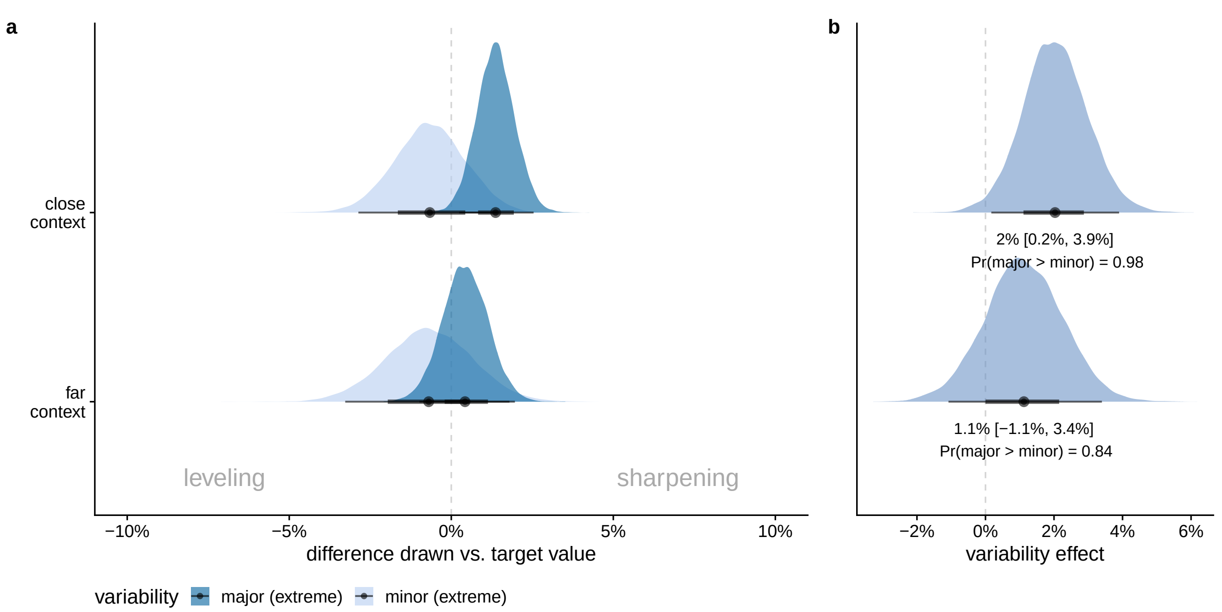


Figure 6.19: (a) Posterior distribution for the signed difference between drawn and target value, separately for the major (extreme) and minor (extreme) variability and context conditions. (b) Estimated variability effect in each context condition. Mean, 66%, and 95% highest density continuous intervals are shown. The text also indicates the mean estimated difference with its 95% highest density continuous interval, as well as the posterior probability of the percentage of sharpening being higher in case the feature exhibits major compared to minor variability, given a particular context condition.

Figure 6.20 shows the estimated extremeness effect per context condition. Within both the close and the far context, there was no evidence for extrema on a feature dimension to be sharpened more often than non-extreme values, on feature dimensions with an equal variability range.

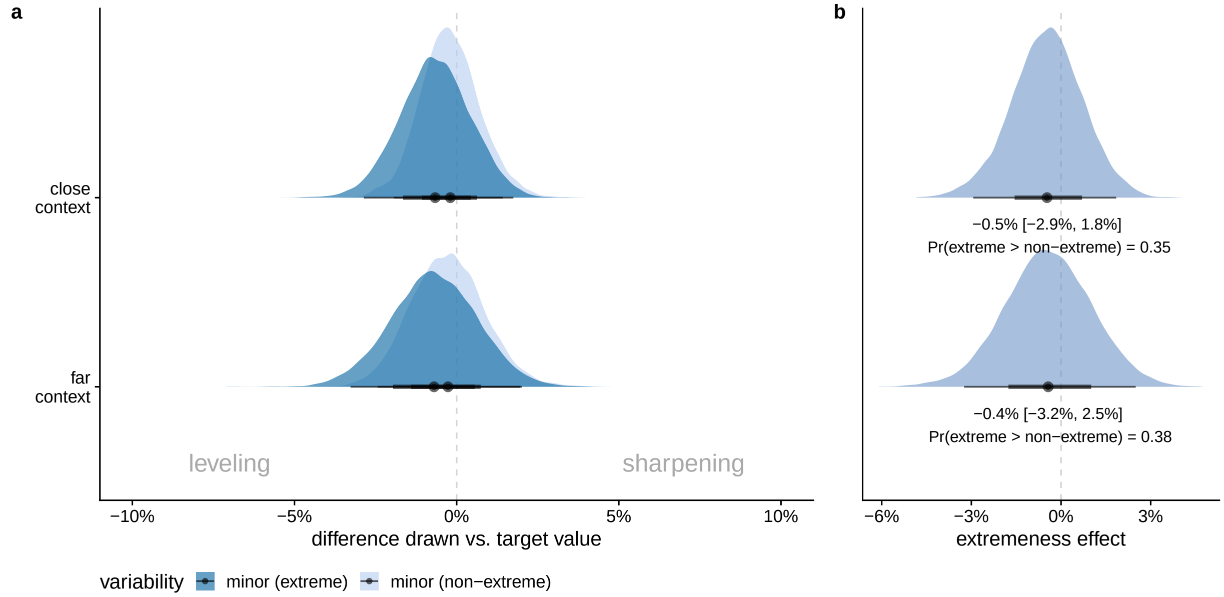


Figure 6.20: (a) Posterior distribution for the signed difference between drawn and target value, separately for the minor (extreme) and minor (non-extreme) variability and context conditions. (b) Estimated variability effect in each context condition. Mean, 66%, and 95% highest density continuous intervals are shown. The text also indicates the mean estimated difference with its 95% highest density continuous interval, as well as the posterior probability of the percentage of sharpening being higher in case the feature value is one of the extrema on the dimension, given a particular context condition.

Additional results concerning the random effects per feature dimension and individual are provided in Appendix D.

Independent of the fitted model, we plotted the mean empirical difference difference between drawn and target feature values relative to the available drawing space for each of the 48 feature dimensions, separately for each context (i.e., close or far), variability (i.e., major extreme, minor extreme, minor non-extreme), and target (maximum/above mean, minimum/below mean) condition, as well as the distribution across feature dimensions. From visual exploration of the figure, there seems to be an effect of target condition on the extent to which a feature was sharpened in the major extreme and minor extreme variability conditions (see Figure 6.21). More specifically, the extent of sharpening seems larger in the maximum target conditions (i.e., major extreme above mean and minor extreme above mean) than in the minimum target conditions (i.e., major extreme below mean and minor extreme below mean).

6.4 Discussion and conclusion

6.4.0.1 Summary of the experiment and main results

In this study, we investigated whether the importance of a feature for discrimination within a specific (task) context influences which Prägnanz tendency (i.e., leveling or sharpening) will occur. More specifically, we hypothesized that

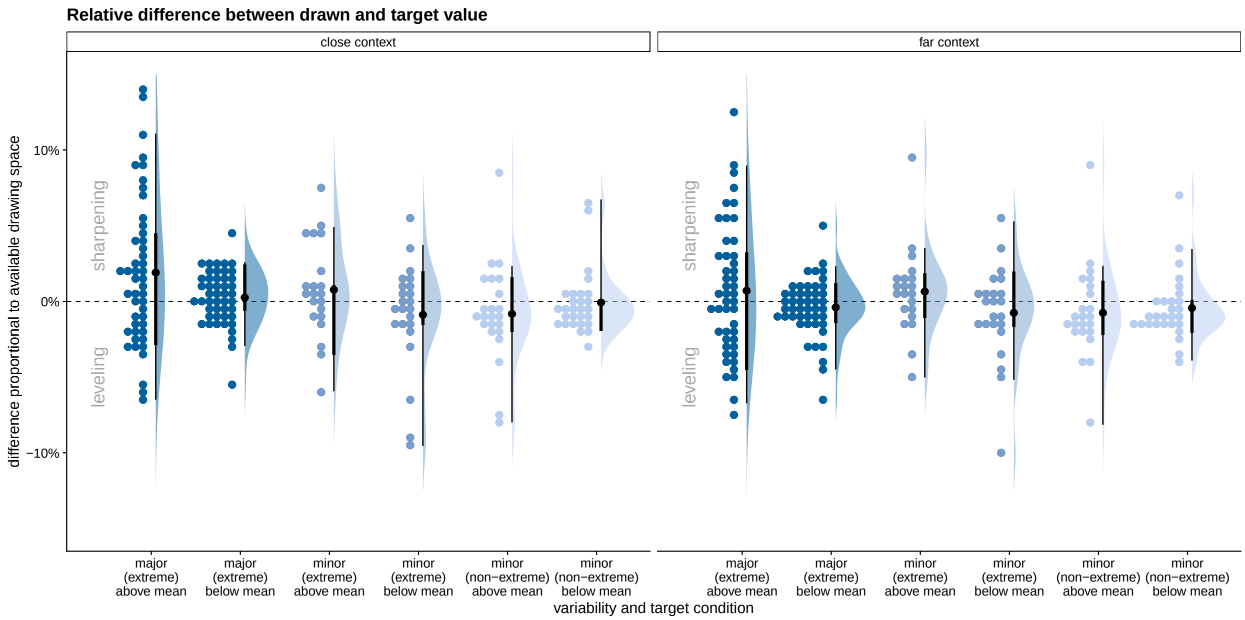


Figure 6.21: Distribution of difference between drawn and target value (relative to the available drawing space for that feature dimension), per context, variability, and target condition, with feature dimensions as individual data points. The black point and intervals indicate mean and 66% and 95% highest density continuous intervals for the data distribution.

when a feature is important for discrimination of a target figure from alternative figures, the feature will more often be sharpened (and less often be leveled) than features less important for discrimination within the given context.

Building on the methodology of Fan et al. (2020), we presented participants with four figures consisting of basic geometric shapes and asked them to reconstruct one of these. The figures either differed qualitatively from each other (i.e., far context condition), or only quantitatively on two feature dimensions (i.e., close context condition). Assuming that the quantitative feature differences would play a bigger role in discrimination from the other figures in the close context condition, we expected more sharpening in the close compared to the far context condition, and more leveling of features in the far compared to the close context condition.

Of the two feature dimensions showing variability across the four simultaneously presented figures, one feature dimension showed larger differences than the other feature dimension (i.e., major vs. minor variability). More specifically, the range of feature values was four times wider for the major variability compared to the minor variability condition. Assuming that the major feature dimensions would be more important for discrimination of the target from the alternatives than the minor feature dimension, we expected more sharpening for the major compared to the minor feature dimension, and more leveling for the minor compared to the major feature dimension.

Although the target figure was always either the figure with a minimum or a maximum value on the major feature dimension, target figures varied in whether the value on the minor feature dimension was an extreme value as well or rather an in-between value. Assuming that having an extreme value on a feature dimension is more useful for discrimination from the alternatives than having an in-between value, it could be expected that extreme values would be sharpened more than non-extreme values, and that more leveling would occur for non-extreme than for extreme feature values.

After extensive preprocessing of the data, we investigated both the percentage of times a feature dimension was leveled or sharpened (i.e., qualitative analysis) and the extent to which a feature dimension was leveled or sharpened (i.e., quantitative analysis) using Bayesian hierarchical regression models with context, variability, and their interaction as fixed effects and as random effects per feature dimension and participant.

For the qualitative analysis, we compared different definitions of sharpening and leveling and how they impacted the results. In case sharpening was defined as a drawn value equal or more extreme than the target value (i.e., on the other side of the target value compared to the mean), results indicated a strong effect of context on the percentage of times a feature was sharpened, at least in the major variability condition: sharpening was more likely for major feature values in the close than in the far context. Similarly, in the close context, the major feature dimension was sharpened more often than the minor variability dimension.

When including a neutral category and thus limiting sharpening to drawn values at least 2.5% more extreme than the target value, again a strong effect of context for the major variability condition was found: sharpening was more likely for major feature values in the close than in the far context. Using this definition including a neutral range, the probability of sharpening was influenced by the range of variability present on the dimension in both the close and the far context: sharpening was more likely for dimensions showing major variability than for dimensions showing only minor variability, both in the close and in the far context condition. Furthermore, we found more leveling in the far than in the close context, at least in the major variability condition, but also more leveling for extreme rather than non-extreme feature values. The number of drawings within the target range was higher for non-extreme feature values than for extreme ones.

When defining sharpening as a drawn value closer to the target value than to the mean value on the dimension (see Appendix D), the effect of context diminished, and the effect of variability and extremeness on the percentage of times a feature is sharpened increased. In other words, the drawn value will much more often be closer to the target than to the mean in case it concerns a major variability dimension (compared to a minor variability dimension). In addition, an extreme value on a minor variability dimension will more often be closer to the target than to the mean compared to a non-extreme value on a minor variability dimension.

For the quantitative analysis we compared the difference between drawn value and target value proportional to the available drawing space (for leveling and sharpening combined). The Bayesian hierarchical linear regression model led to similar conclusions as the qualitative analysis: the extent to which a feature was sharpened in the major variability condition was higher in the close compared to the far context condition. In addition, in the close context, the major variability dimension was sharpened more strongly than the minor variability dimension. It is important to take into account that the estimated size of the variability effect is a conservative estimate, given that the measure does not take into account that there is more drawing space available for a minor variability feature to be sharpened than for a major variability feature. Visual inspection of the quantitative data grouped by context, variability, and target condition (i.e., target above or below the mean) hinted at an additional effect of the target type (minimum/below mean vs. maximum/above mean) on the extent to which a feature was sharpened, at least in the major extreme and minor extreme conditions. More specifically, the extent to which a feature was sharpened seemed to be larger in the maximum target conditions than in the minimum target conditions. In the qualitative analysis, no such effect of target type seemed to be present.

In general, we found more evidence for variability between feature dimensions than between participants (see Appendix D), but this result can very likely be attributed to the limited number of data points available per participant (2-48 data points per participant) compared to the number of data points available per feature dimension (400-541 data points per feature dimension). When using the traditional definition of sharpening, we found more uncertainty for the estimates for the effect of context per feature dimension than for the estimated differences between the variability conditions per feature dimension, which is also potentially due to a difference in the number of data points involved in the comparisons.

As repeating the analyses with different data exclusion criteria, and where possible, with absolute scores rather than scores relative to the available drawing space, resulted in similar conclusions and effect sizes as presented in the main text (see Appendix D), the given conclusions are consistent and seem to be robust against such arbitrary choices. Using different definitions of leveling and sharpening led to effects in the same direction, but not always of the same size. Where differences in effect size occurred, presenting the different results gives a more nuanced view on how the results can be interpreted depending on the definitions used.

6.4.0.2 Main conclusions

6.4.0.2.1 Features important for discrimination will more often be sharpened

When a target figure is compared to other figures that only differ from the target figure in a quantitative way, the main distinctive feature will more often be sharpened (i.e., drawn as more extreme than the target value) than when the target is compared to qualitatively different figures (i.e., context effect), especially when it concerns a feature dimension with high variability. Moreover, the main distinctive feature will more often be sharpened than the feature showing only minor variability across figures (i.e., variability effect), especially when figures differ only quantitatively from each other. Similarly, the *extent* of sharpening is larger in the close compared to the far condition (given a feature showing major variability) and for the major compared to the minor variability dimension (given a close context condition). From these results, we can conclude that a feature will be sharpened more often when this feature is important for discrimination of the target figure from alternatives, compared to when it concerns a feature less important for discrimination. In other words, whether a feature of a figure is important for discrimination, and thus what comprises the essence of the figure, will depend on the context (i.e., the other figures present) in which this figure is presented. As a consequence, the context will also influence which Prägnanz tendency will occur for a specific feature (i.e., leveling or sharpening). The results of this study are in line with recent work by Feldman (2021), who concluded that the improvement in perceptual discrimination of a feature is proportional to the informativeness of a feature in the specified context. Related to Metzger's (1941) definition of Prägnanz, we can conclude that the essence of a Gestalt is context-dependent, and this will influence which tendency (i.e., leveling or sharpening) will lead to the best organization in the specified (task) context.

6.4.0.2.2 Close context increases sharpening, variability and extremeness increase identifiability

Drawing a feature value as more extreme than a target value was influenced more by the other figures present in a certain context (i.e., difference between close and far context) and particularly when it concerns a dimension with major variability across figures. In contrast, drawing a feature value as closer to the target value than to the mean value was more heavily influenced by differences in the range of variability present for that feature (i.e., differences between feature dimensions with major and minor variability) and differences in the position of the feature value on the range of variability (i.e., differences between extreme and non-extreme feature values) than by differences in context. Therefore, one could conclude that the presence of close alternatives increases sharpening (as it is traditionally defined), whereas variability and extremeness increase the percentage of times a feature will be drawn in an identifiable way: not necessarily more extreme than the target value, but at least closer to the target value than to the mean of the dimension.

6.4.0.3 Limitations and suggestions for future research

6.4.0.3.1 Degrees of freedom in study design and analysis

A lot of arbitrary design choices had to be made when setting up this study, including the choice of the figure designs, feature dimensions to vary, feature values, and assigned targets. These choices were standardized as much as possible, and randomized when standardization was not an option. Moreover, also in the analysis pipeline, including both preprocessing and regression analyses, many choices had to be made. Often multiple alternative choices were available. We therefore chose to report several defensible preprocessing and analysis options, often on the conservative side, and in that way conducted a mini-multiverse analysis (Steegen et al., 2016). As such, we tested whether our results are consistent across these options and independent of the errors or choices made. The main results did seem to hold across the different analysis options, and where they didn't, the differences taught us important nuances concerning the effects.

6.4.0.3.2 Limited information per participant

Although we collected information from a large group of participants, the information received from each participant was rather limited (maximally 24 trials per participant, one per design). This resulted in very large credible intervals for the random effects per participant and limited information about individual differences between participants. Future research can increase the number of trials per participant to gain more knowledge about the presence and/or size of individual differences in leveling and sharpening tendencies, including repeated measures per design to compute reliability.

6.4.0.3.3 Benefits of leveling and sharpening for recognition

The current study assumes sharpening of a feature important for discrimination and leveling of features unimportant for discrimination in a specific (task) context to be beneficial for recognizing the target figure among the alternatives. Future research can put this assumption to the test by presenting a second set of participants with the leveled or sharpened drawings of the first set of participants and compare recognition performance in the specific context in which the drawings were generated with alternative contexts (for inspiration, see also Fan et al., 2020). We predict recognition performance (i.e., speed and accuracy) to be higher when the conditions in which the drawing was created and is presented for recognition match (i.e., presented with the same alternatives in both the drawing and the recognition task) compared to when the conditions in drawing and recognition task are different.

6.4.0.3.4 Distinction between leveling and sharpening and simplification and complication

In the current study, we focused on the exaggerating or weakening of a feature (i.e., sharpening and leveling). Early work has distinguished these concepts from the addition or removal of parts in a figure (i.e., simplification and complication; Wulf, 1922). Future research can investigate whether this is an important distinction to make, or whether both concept pairs undergo the same tendencies dependent on (task) context.

6.4.0.4 Distinction between primary and secondary Prägnanz tendencies

In the current study, viewing conditions for participants were unlimited. Participants could freely inspect the target figure and distractors during drawing, although they were presented at relatively small size (cf. the Methods section). Based on our approach, it is not possible to disentangle primary and secondary Prägnanz tendencies: it is impossible to say whether participants *purposefully adapted* their drawings to be different from their percept of the target figure (i.e., secondary Prägnanz tendency), or whether they drew the figure differently because they *perceived* the target figure differently depending on the context manipulations (i.e., primary Prägnanz tendency). Future research can aim to investigate whether the context effects on leveling and sharpening tendencies we report here are mainly caused by perceptual distortions or by conscious communicative strategies.

6.4.1 Take home message

Gestalt psychologists posited that we will always organize our visual input in the best way possible under the given conditions. Both the weakening or removal of unnecessary details (i.e., leveling) and the exaggeration of distinctive features (i.e., sharpening) can contribute to achieve a better organization. We hypothesized that the importance of a feature for discrimination among alternatives influences which organizational tendency will occur. As expected, the results indicated that sharpening in the sense of drawing a feature as more extreme than the target value, occurred more often for features exhibiting more rather than less variability in the close context (with alternatives differing only quantitatively from the target figure), and in the close rather than in the far context (with alternatives differing qualitatively from the target figure), especially for the features of figures exhibiting major variability across the alternatives presented. Furthermore, major variability present on a feature dimension (compared to minor variability) as well as whether a target value was one of the extrema on the feature dimension (rather than an in-between value) increased the percentage of times a drawn value was closer to the target value than to the mean on the feature dimension. Returning to Metzger's (1941) definition of prägnant Gestalts: the essence of a Gestalt is context-dependent, and this will influence whether leveling or sharpening of a feature will lead to the best organization in the specific context.

6.5 Open and reproducible practices statement

This manuscript was written in R Markdown using the papaja package (Aust & Barth, 2022) with code for data analysis integrated into the text. The data, materials, and analysis and manuscript code for the experiment are available at https://doi.org/10.17605/osf.io/hqcja.

Part VI

Order and complexity

Complexity can contribute positively to aesthetic appreciation, but only when it is made easier to handle by a high level of *order*.

Chapter 7

The Order & Complexity Toolbox for Aesthetics (OCTA): A systematic approach to study the relations between order, complexity, and aesthetic appreciation

Do individuals prefer stimuli that are ordered or disordered, simple or complex, or that strike the right balance of order and complexity? Earlier research mainly focused on the separate influence of order and complexity on aesthetic appreciation. When order and complexity were studied in combination, stimulus manipulations were often not parametrically controlled, only rather specific types of order (i.e., balance or symmetry) were usually studied, and/or the multidimensionality of order and complexity was largely ignored. Progress has also been limited by the lack of an easy way to create reproducible and expandible stimulus sets, including both order and complexity manipulations. The Order & Complexity Toolbox for Aesthetics (OCTA), a Python toolbox that is also available as a point-and-click Shiny application, aims to fill this gap. OCTA provides researchers with a free and easy way to create multi-element displays varying qualitatively (i.e., different types) and quantitatively (i.e., different levels) in order and complexity, based on regularity and variety along multiple element features (e.g., shape, size, color, orientation). The standard vector-based output is ideal for experiments on the web and the creation of dynamic interfaces and stimuli. OCTA will not only facilitate reproducible stimulus construction and experimental design in research on order, complexity, and aesthetics. In addition, OCTA can be a very useful tool in any type of research using visual stimuli, or even to create digital art. To illustrate OCTA's potential, we propose several possible applications and diverse questions that can be addressed using OCTA.

Van Geert, E., Bossens, C., & Wagemans, J. (2022). The Order & Complexity Toolbox for Aesthetics (OCTA): A systematic approach to study the relations between order, complexity, and aesthetic appreciation. ***Behavior Research Methods.*** **https://doi.org/10.3758/s13428-022-01900-w**

Paper | Postprint | Figshare | Python toolbox | Shiny application | Additional resources

The Order & Complexity Toolbox for Aesthetics (OCTA): A systematic approach to study the relations between order, complexity, and aesthetic appreciation

7.1 Introduction

Order, complexity, and the balance between order and complexity have long been considered important factors influencing aesthetic appreciation (for a review, see Van Geert & Wagemans, 2020). Although most empirical work has confirmed the importance of order and complexity, concrete results on the extent and direction of order and complexity influences, as well as the relationship between order and complexity, vary widely. Previous research has some limitations concerning the stimuli and manipulations used, however, and leaves some important research options uncovered (cf. The need for OCTA below). The Order & Complexity Toolbox for Aesthetics (OCTA) fills these gaps by providing researchers an easy, open, and reproducible way to create stimuli varying in order and complexity on multiple element features (see Figure 7.1) for some OCTA example stimuli). Being available as both a Python package and a Shiny application, OCTA[1] is accessible to both researchers with and without programming experience.

7

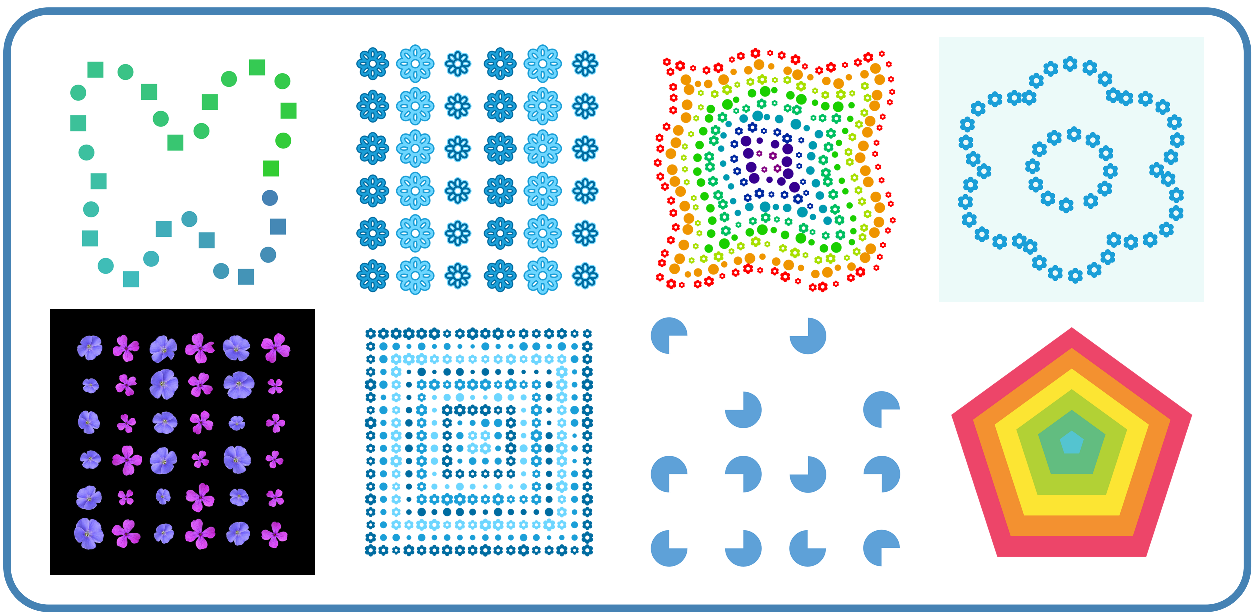


Figure 7.1: Example stimuli created in OCTA. In the PDF version, clicking a stimulus leads to the octa code used to generate it. To view dynamic and interactive versions of some of these stimuli, visit this webpage. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.17708702. The bottom left example stimulus contains natural flower images from Hůla and Flegr (2016).

We will first discuss how we define order and complexity, and explain how we view the relation between order, complexity, and appreciation. Next, we list existing stimulus sets and stimulus generation tools (to be) used in research on the perception and appreciation of order and complexity, point out the limitations to the existing research and tools, and indicate how OCTA can provide a fulfilling answer to the existing gaps. Then, we introduce

[1]We will use octa in small letters when referring specifically to the Python package and OCTA in capital letters when referring to the toolbox in general, regardless of whether it concerns the Python implementation or the Shiny app.

the newly created stimulus generation toolbox and discuss its steps, concepts, and options in more detail. Afterwards, we treat more advanced uses of OCTA, provide a range of potential applications for OCTA, list some possible future extensions of OCTA, and give advice on how to start using OCTA.

7.1.1 Defining order and complexity

In this work, we define (stimulus) order as all aspects related to the *structure* and *organization* of information (in a stimulus) and (stimulus) complexity as all aspects related to the *quantity* and *variety* of information (in a stimulus; Van Geert & Wagemans, 2020). For example, Figure 7.2a and b have the same number of distinguishable elements and the same dissimilarity between those elements (i.e., complexity; Berlyne, 1960) but differ in the way the elements are arranged (i.e., order). Similarly, also Figure 7.2c and d share the same level of complexity, but differ in the way the elements are organized. In Figure 7.2a and b, the difference in order is a *qualitative* difference, a difference in the *type* of order applied. In Figure 7.2c and d, it concerns a *quantitative* difference, a difference in the *level* of order present. Figure 7.2a and c have the same overall organization (i.e., level and type of order), but differ in the quantity and variety of information present (i.e., different level of complexity).

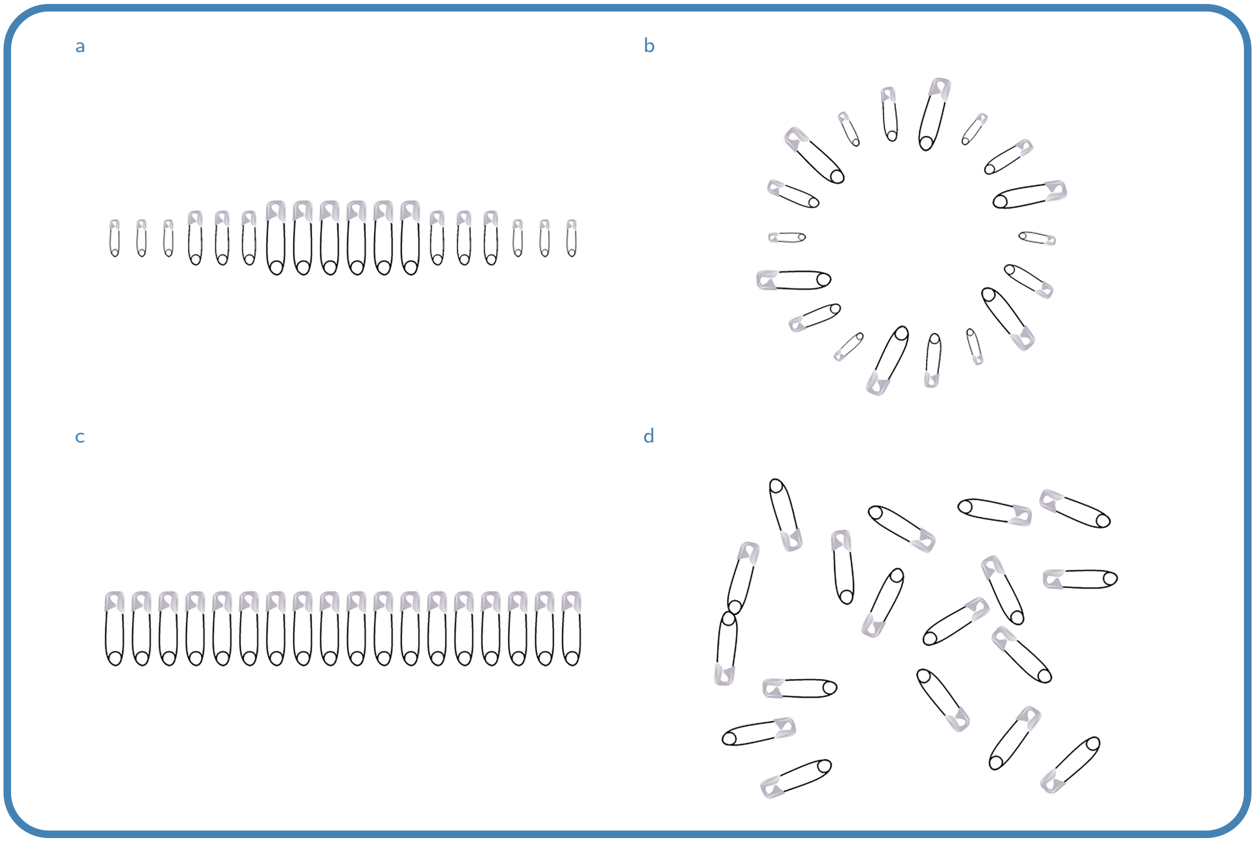


Figure 7.2: Order and complexity on different stimulus dimensions, illustrated with stimuli created in OCTA. In the PDF version, clicking a stimulus leads to the octa code used to generate it. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.19634649. Vector image used to generate the OCTA stimuli : a safety pin created by Clker-Free-Vector-Images available on Pixabay.

In this example, we referred to the multidimensional concepts of *element* order and *element* complexity as referring to a combination of order and complexity on the color, shape, and size dimensions: we focused on the order and complexity of the distinguishable elements present in the stimulus on the different feature dimensions in combination. It is however also possible to investigate other forms of order and complexity in the stimulus, for example position order and complexity, specific feature order and complexity, or the complexity of individual feature values used. Qualitatively different position patterns can be used to place the elements in the stimulus (i.e., type of position order), and these positions can also show more or less diversity (e.g., equally-spaced positions in rows and columns versus completely random positions; position complexity). Feature order and complexity are part of element order and complexity, but focus on one specific element feature at a time. For example, Figure 7.2a has a higher level of size complexity than Figure 7.2c, but Figure 7.2b and d have a higher level of orientation complexity. Furthermore, the individual feature values of the elements present in the stimulus, or the features of the stimulus as a whole, may be more or less complex. For example, a safety pin may be a more complex shape than a circle, or placing the elements on a circular outline may form a less complex overall organization than placing them on a rectangular outline. Acknowledging the multidimensionality of order and complexity may be important in the search for more consistent relationships between order, complexity, and aesthetic appreciation (Nadal et al., 2010; Van Geert & Wagemans, 2020).

Important to understand is that order and complexity are no direct opposites, and can be concurrently present, either on the same or on different feature dimensions. We distinguish order from simplicity (i.e., opposite of complexity) and complexity from disorder (i.e., opposite of order). Complexity on a feature dimension allows for a broader range of order levels and a larger set of order types to be present than is the case for simplicity. For example, the safety pins in Figure 7.2a can be arranged in more qualitatively and quantitatively different ways on the size dimension than the safety pins in Figure 7.2c. However, simplicity on the size dimension does not prohibit disorder on other feature dimensions (e.g., orientation of the elements in Figure 7.2d). As explained in Van Geert & Wagemans (2020), we view order and complexity as partial complements and partial opposites. On the one hand, complexity allows order to show its potential and order helps to make highly complex stimuli ‘digestible’ and aesthetically appreciated. Order and complexity thus need each other’s presence to optimize appreciation. On the other hand, order and complexity can also relate negatively: uniformity for example indicates a low level of complexity, but also implies a high level of order on that feature dimension. Similarly, different types of symmetry may be seen as different types of order, but typically also reduce the range of complexity present in the display. Given this complex relationship between order and complexity, it is important to study their relation to aesthetic appreciation in combination, not separately (Van Geert & Wagemans, 2020).

One additional distinction that is relevant here, is the distinction between objective and subjective definitions and measures of order and complexity. Whereas *objective* order and complexity refer to the level and type of order or complexity that is physically present in the stimulus, *subjective* order and complexity entail the individuals’ perception of the order or complexity present in the stimulus (Van Geert & Wagemans, 2020). It is important to make this distinction, as research results may differ depending on how order and complexity are defined and/or measured: the concepts of objective order and complexity may for example be more easily separable from each other than the concepts of subjective order and complexity. Although many objective measures for complexity exist (for overviews, see Donderi, 2006; Van Geert & Wagemans, 2020), the set of objective measures available for order is rather limited (Van Geert & Wagemans, 2020). Since Arnheim's (1971) seminal essay on "Entropy and art" (which was effectively about disorder and order in art), several scientists have proposed measures of order (and disorder) derived from information theory and dynamical systems theory. For instance, Redies and colleagues have defined and computed edge-orientation entropy or anisotropy in relation to preference for patterns and images of paintings (e.g., Grebenkina et al., 2018; Redies et al., 2017). However, a discussion of these would be beyond the scope of the present paper. The relative lack of objective order measures may be due to the difficulty of quantifying aspects of order other than perceptual balance or symmetry in an objective or automatized way. It is our intuition that the degree of order can be quantified for specific types of order (such as balance and symmetry) but that further research is needed to compared the degree of perceived order for different types of order. Our aim is to facilitate this research by providing a toolbox to systematically generate different types of order (e.g., different pattern types, different types of repetitions vs. changes) to compare their overall degree of perceived order (and how these interact with the number of elements, their degree of similarity and variety, etc.). Below we give an overview of different stimulus sets used in earlier research on the perception and appreciation of order and complexity that attempted to manipulate order or complexity in a parametric way.

7.1.2 Previous parametric stimulus sets in research on the perception and appreciation of order and complexity

Studies investigating the perception or appreciation of *order* often focused on different types of symmetry or on different measures of perceptual balance (see Figure 7.3). Chipman & Mendelson (1979) chose black and white square patterns with different types of order: unstructured, horizontal and/or vertical symmetry, diagonal symmetry, checkerboard organization, and rotational organization. Locher et al. (1998) asked participants to create designs consisting of a set of either nine circles, squares, rectangles, or leaves varying in size (3 large, 3 medium, 3 small forms per set). They investigated the changes in the center of balance as the design was completed. Wilson & Chatterjee (2005) created stimuli using the proprietary software Adobe Photoshop 7.0 and then selected a subset of these stimuli for their experiments based on a calculated balance score (i.e., the average score of eight measures of symmetry). These stimuli were later also used by Hübner & Fillinger (2016). Gollwitzer et al. (2017) generated unbroken and broken patterns to study pattern deviancy aversion in the context of social psychological research on prejudice. Although many studies in empirical aesthetics compared symmetric and asymmetric stimuli (Bertamini & Rampone, 2020), only some recent studies investigated symmetry perception or appreciation using the different systematic characterizations of symmetry in mathematics (e.g., Alp et al., 2018; Clarke et al., 2011; Kohler et al., 2016; Martin et al., 2020). Four basic symmetrical transformations are defined for two-dimensional shapes (translation, rotation, reflection, glide reflection; Grünbaum & Shephard, 1989). With these transformations, seven distinct border patterns or frieze groups (i.e., two-dimensional designs with translations in only one direction) and seventeen distinct wallpaper pattern types (i.e., two-dimensional designs with translations in two independent directions) can be specified (Grünbaum & Shephard, 1989; Thomas, 2012). Martin et al. (2020) compared beauty ratings for each of the seven frieze patterns using a comma, flag, and a filled texture as elements. Clarke et al. (2011) studied the perceptual similarity of the seventeen wallpaper pattern types. They created exemplars starting from a white noise (triangular, rectangular, or square) fundamental region and combined them into a rectangular repeating tile to produce the patterns (see also Alp et al., 2018; Kohler et al., 2016).

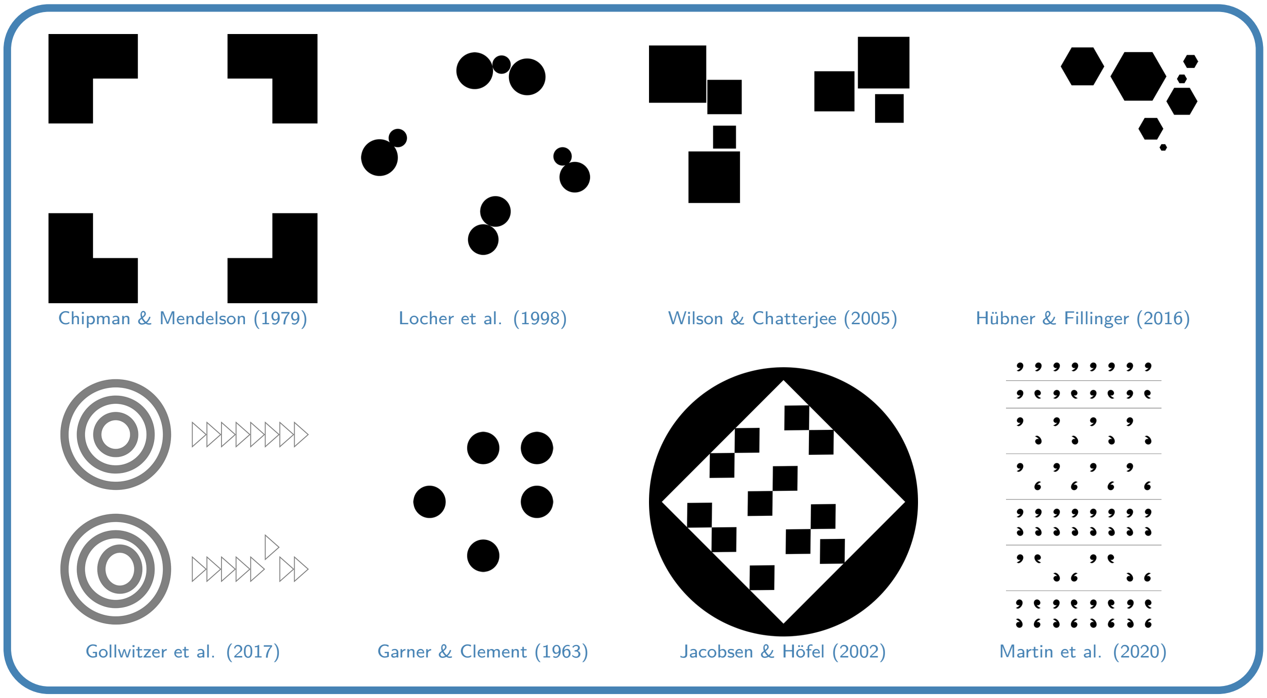


Figure 7.3: Example stimuli from previous research on order and complexity recreated in OCTA. In the PDF version, clicking a stimulus leads to the octa code used to generate it. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.17739860.

Studies focusing on the perception or appreciation of *complexity* have used a broad range of different stimulus types. Some studies used ecologically valid images and calculated statistical image properties to indicate complexity (e.g., Braun et al., 2013). Other studies used different types of fractal images (e.g., Bies et al., 2016; Spehar et al., 2016). Sun & Firestone (2021) used a measure of structural surprisal (i.e., the level of surprise associated with a shape's skeletal structure) to categorize abstract shapes as simple and complex while controlling for the number of sides. The Matlab code to generate the shapes for that study was shared publically on the Open Science Framework.

When both *order and complexity* were manipulated parametrically, researchers often used random polygons (e.g., Arnoult, 1960; Attneave, 1957), black and white square patterns (e.g., Chipman, 1977; Smets, 1973), or dot patterns (e.g., Garner & Clement, 1963; Hamada & Ishihara, 1988). Attneave (1957) constructed random polygon shapes based on a varying number of turns (i.e., complexity manipulation) and manipulated them to be symmetrical or asymmetrical (i.e., order manipulation). Attneave & Arnoult (1956) provided a series of methods for the construction of random shapes. This method was later used, for example by Vanderplas & Garvin (1959), to generate random shapes of six levels of complexity (i.e., varying number of points). Smets (1973) created patterns of black and white squares in which both subjective redundancy (i.e., the percentage of correctly selected black and white elements when reproducing the pattern, related to order) and maximal information (i.e., the number of independent elements, related to complexity) were varied. These patterns were also used in studies by Berlyne (Berlyne, 1974; Cupchik &

Berlyne, 1979). Chipman (1977) used similar but smaller patterns and varied structure and complexity by selecting different stimulus sets by hand. Garner & Clement (1963) produced 90 patterns by placing five dots in an imaginary 3-by-3 square matrix, with the restriction of at least one dot in each row and column, and investigated the influence of the number of patterns in the equivalence group for each pattern (i.e., the number of different patterns that can be created by rotation or reflection of the pattern) as well as symmetry. Hamada & Ishihara (1988) created rotation and reflection invariant dot patterns using imaginary rectangular and hexagonal frameworks. The dot patterns varied in the number of dots they consisted of, the order of the symmetry groups as well as the symmetry group they belonged to (i.e., cyclic and dihedral groups). In more recent years, new stimulus sets focused on symmetry and complexity manipulations. Jacobsen & Höfel (2002) constructed stimuli consisting of a solid black circle showing a centered, quadratic, rhombic cut-out and an arrangement of 86 to 88 small black triangles. Half of the patterns created were symmetric (four possible symmetry axes), the other half were not symmetric. Gartus & Leder (2013) created a new set of abstract black and white patterns using a simulated annealing stochastic optimization algorithm. They varied the number of objects (i.e., complexity manipulation) as well as the type of symmetry: no symmetry, and full and broken symmetry patterns with one, two, and four symmetry axes.

7.1.3 Existing software to generate aesthetic stimuli

7.1.3.1 FlexTiles

Westphal-Fitch and colleagues (2012) created FlexTiles, a custom-written image manipulation program, in which each of 36 tiles can be rotated to four possible orientations (0°, 90°, 180°, 270°). Participants were asked to rotate the tiles as much or as little as they liked. Patterns that emerged showed rotation, translational symmetry, bilateral symmetry along vertical axis, rotational symmetry, symmetry along diagonal axis, or local linear groupings with no overall symmetry. In later experiments, they created flawed and unflawed versions of patterns, with violations determined by color and/or orientation. Muth et al. (2019) also made use of the FlexTiles software and asked participants to produce patterns that would be liked or would be interesting. A separate group of participants rated the produced stimuli on liking, interestingness, complexity, and order. Although Muth et al. (2019) shared the stimuli used in their experiments, the FlexTiles software to recreate or adapt the stimuli has not been made openly available for use by other researchers. For OCTA stimuli created in the style of FlexTiles, see Figure 7.4.

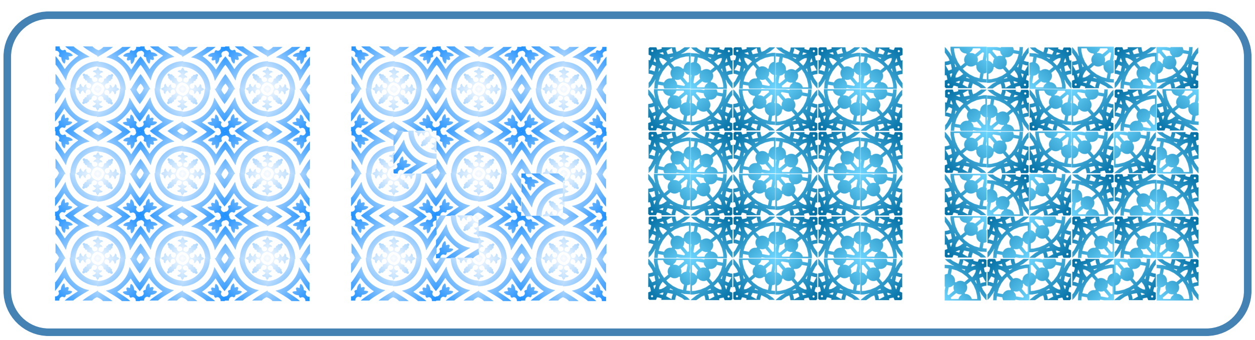


Figure 7.4: Example stimuli in the style of FlexTiles recreated in OCTA. In the PDF version, clicking a stimulus leads to the octa code used to generate it. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.17740640. Vector images used to generate the OCTA stimuli : a tile created by Jorge Carrillo and a tile created by Lluisa Iborra available on the Noun Project.

7.1.3.2 Statistical Geometry Sampler

Güçlütürk et al. (2016) generated statistical geometric patterns using a space filling algorithm that places non-overlapping geometric shapes that monotonically decrease in size in a random fashion on a canvas (Shier, 2011; Shier & Bourke, 2013). A parameter c determining the size of the first shape element and the speed with which the size decreased was manipulated as well as the geometric shape used (i.e., circle, hexagon, square, and triangle). Although the algorithm, documented C code, and a lot of example stimuli are openly available on the website of the authors of the algorithm, and the authors mention to be open to support people interested to get started on their own images, no easy interface to create new stimuli is provided.

7.1.3.3 Aesthetic Abstract Textures Generator

One tool that does generate reproducible stimuli for use in aesthetics is the random abstract texture generator of Alvarez et al. (2021). Jean-Michel Morel, Luis Alvarez, and colleagues did design a point-and-click online tool to generate aesthetic stimuli based on composition principles and random sampling (Alvarez et al., 2021, 2015).[2] This tool is meant for artists and designers to explore and test new styles, and does not allow users to manually specify the location, color, or shape of any element in the display. The shape elements are placed automatically based on the chosen composition principles and random sampling. Although this tool is very useful for the generation of reproducible aesthetic stimuli in general, it does not allow for individually controlled parametric order and complexity manipulations on different element features.

7.1.4 The need for OCTA (see Figure 7.5)

As may be clear from the examples given, the existing research on the perception and appreciation of order and complexity leaves important gaps. First, most empirical work investigated the relation of order and complexity to aesthetic appreciation separately rather than in combination. Second, the relation of appreciation with complexity has been studied much more frequently than the relation with order and studies usually focused on specific aspects of order (i.e., symmetry and perceptual balance). However, many other types of organization exist (e.g., similarity grouping, proximity grouping, alternation/iteration, systematic alteration/gradient) and their influence on aesthetic appreciation has not yet been systematically studied (Van Geert & Wagemans, 2020). Furthermore, the multidimensionality (e.g., order and complexity of element colors, shapes, and sizes) of both order and complexity has often been) neglected (Nadal et al., 2010; Van Geert & Wagemans, 2020). To our knowledge, no single existing stimulus set combines systematic manipulations of order and complexity on different element features (e.g., shape, size, color, orientation). Findings based on earlier research not distinguishing between different element features may be tied to the specific aspect of order or complexity that was investigated, and may not generalize to other order or complexity dimensions.[3] Third, when order and complexity were studied in combination, researchers often focused on a binary classification in high and low order and/or complexity rather than more fine-grained order or complexity manipulations[4], or more ecologically valid stimuli were used but then parametric control over both

[2] The same authors also created an abstract shape generator which may be of interest to the readers.

[3] This will always be the case to some extent but experimental control and generalization are made much easier in OCTA because it provides easy handles to create and test systematically many more variations.

[4] Whether a more continuous manipulation of order levels also leads to a continuous change in perceived order is an empirical question that can be addressed using the OCTA toolbox.

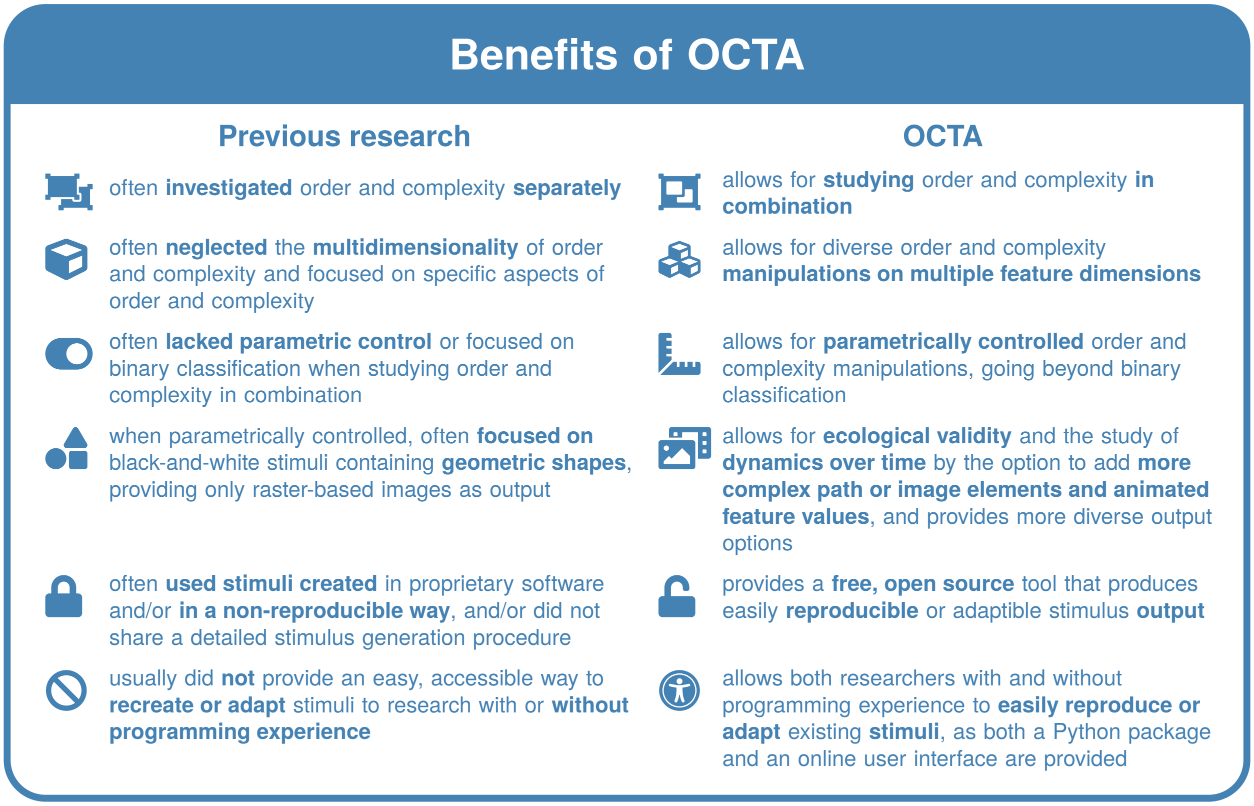


Figure 7.5: Benefits of OCTA. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.17740904.

order and complexity dimensions was lost (e.g., Nadal et al., 2010; Van Geert & Wagemans, 2021). Findings based on this last type of research, using less controlled stimuli, require replication with a more parametrically controlled stimulus set where both order and complexity of the stimulus can be manipulated as independently as possible.[5] Fourth, when parametrically varied stimulus sets were used, the focus was often on black-and-white stimuli containing geometric shapes. Fifth, the stimuli used in aesthetics research were often created using proprietary software, in a non-reproducible and non-adaptable way, or not openly available to the research community, or the researchers did not share enough details concerning how the stimuli were generated. Sixth, in case a detailed stimulus generation procedure was provided, stimuli could not be recreated or adapted easily without programming experience. Therefore, research on order and complexity would benefit from an easy way to create standardized stimuli varying qualitatively and quantitatively in many different aspects of order and complexity, with options for increased ecological validity but without losing parametric control.

The Order & Complexity Toolbox for Aesthetics (OCTA) is such a free, openly available tool and creates many opportunities to investigate order and complexity in a standardized and multidimensional manner, with a focus on multi-element displays. With OCTA, one can manipulate and measure order and complexity in a systematic way, and study their effects in combination. OCTA acknowledges the multidimensionality of order and complexity

[5]Although it is important to study order and complexity *in combination* (i.e., in a factorial design taking into account both order and complexity), we do believe it is important to study order and complexity *as independently as possible* (i.e., in a non-confounded way, separating both factors as much as possible within the same experiment).

by allowing separate manipulations of different element features (e.g., shape, size, color, orientation). It allows for parametric control on many different stimulus and element features. It also acknowledges the potential of research on the dynamics of order and complexity by including animation options for several element features including color, size, and orientation. To provide more ecological validity, OCTA makes it possible to use of images or path elements[6] in a stimulus. OCTA is freely available, open source, and enables users to create reproducible, code-based stimulus sets that are easily adaptable or extendible. The standard vector-based output is ideal for experiments on the web and the creation of dynamic interfaces and stimuli, while raster-based output is possible as well. Furthermore, OCTA is accessible to both researchers with and without programming experience, as both a Python package and an online user interface are provided. With OCTA being fast, flexible, and transparent, we strongly believe that OCTA will facilitate reproducible stimulus construction and experimental design in research on order, complexity, and aesthetics. In addition, OCTA can be a very useful tool to investigate visual perceptual organization, to create visual stimuli for any type of experimental or cognitive task, or even to create digital art.

7.2 The OCTA toolbox as a Python package and a Shiny application

The octa package is a Python package developed in Python 3.8 (Van Rossum & Drake Jr, 2009) and is available (on the Python Package Index or on GitHub) as open-source software under the GNU Lesser General Public License (Version 3) as published by the Free Software Foundation.[7] We chose to create this toolbox in Python as Python is a powerful and popular open source programming language for which a large online community is available. Elaborate documentation on the octa Python package is available, offering a step-by-step introduction to the full functionality of the OCTA toolbox. To get users started, a website with example stimuli (see dynamic stimuli here) and the corresponding code is provided. For users who want to dig further into the toolbox, detailed function documentation is available. Besides developing the Python package, we also developed a Graphical User Interface for the OCTA toolbox, in the form of a Shiny application.[8] The procedure described below will be for using the Python package, but most aspects are similar in the Shiny application (we will indicate the most important differences in text). Additional instructions for the use of the Shiny application are provided in the app itself.

Figure 7.6 gives an overview of the terminology used in OCTA.[9] In OCTA, you can create *stimuli* consisting of multiple *elements*. First, the user specifies the *type* of stimulus they want to create (i.e., Grid, Outline, Concentric; see Figure 7.8) as well as the general *stimulus features* they want to adapt (e.g., stimulus size, background color or shape, stimulus orientation). Second, they can replace the default *position pattern* for the stimulus type by a custom one (e.g., sine grid, custom shape, custom positions). Third, the *element feature patterns* can be specified

[6] Path SVG elements can be used to draw any type of lines, curves, or custom shapes based on a combination of straight or curved lines. For more information on how to define paths, have a look at these online tutorials by W3Schools and Mozilla.

[7] To make full use of octa, Python 3.6 or higher is required, and the toolbox depends on the Python packages svgwrite (Moitzi, 2021), svg.path (Regebro, 2021), svgpathtools (Port, 2021), svgutils (Telenczuk, 2021), jsonpickle (Aguilar, 2021), html2image (vgalin, 2021), svglib (Gherman, 2021), reportlab (Robinson et al., 2021), colour (Lab, 2021), and IPython (Pérez & Granger, 2007). In addition, Google Chrome (Windows, MacOS) or Chromium Browser (Linux) needs to be installed to be able to generate PNG, JPG, or TIFF raster images based on the stimulus.

[8] When using either the octa Python package or the OCTA Shiny app in your (academic) work, please cite this paper to acknowledge the authors.

[9] We realize that it is quite difficult to understand and remember this terminology. This difficulty follows from the wide range of stimulus variations we want OCTA to be able to generate. First-time users might be put off by this difficulty. However, we suggest you try working with the toolbox to familiarize yourself with the options, and then Figure 7.6 and the descriptions in the main text become a handy reference in case you want to develop a more refined understanding all the available options or gain more experience with them.

by defining *pattern type*, *pattern direction*, and *pattern values* per element feature. Fourth, *position, element, and feature deviations* can be added. Finally, the user can save the resulting stimulus in the preferred formats: as a vector-based image (SVG), as a raster-based image (PNG, JPG, PDF, or TIFF), and/or as a computer-readable file (JSON). The SVG format is recommended for online use, as it has the same quality at all viewing sizes and keeps the possibility for animated element and stimulus features. The JSON output can be used to recreate the stimulus in Python using the OCTA toolbox without the original code (using the LoadFromJSON function).[10] In the Shiny app, the user can also view or download the Python code needed to reproduce the current stimulus with the octa package in Python.

[10] Keep in mind that to reproduce the exact same stimulus, one needs to set a seed before running (each line of) the OCTA Python code and also keep this seed (these seeds) along with the JSON file in case any randomization procedures are used in the stimulus generation. In addition, the safest way to ensure reproducibility is by keeping track of the code, seed, and OCTA toolbox version you used to create your stimuli.

7

Terminology used in OCTA

order	aspects related to the structure and organization of elements in a stimulus
complexity	aspects related to the quantity and variety of elements in a stimulus
stimulus	overall display, total configuration including one or more elements
element	object, part of stimulus display which is usually repeated and spatially separated from other parts
stimulus feature	certain feature of the stimulus as a whole (e.g., size, orientation, background color, background size)
element feature	certain feature of the elements in the stimulus (e.g., shape, size, color, orientation)
(element) position pattern	overall configuration of the element positions in the stimulus (e.g., rectangular grid, sinegrid, circle, shape)
(element) feature pattern	overall configuration of the element features in the stimulus (combination of pattern values, pattern type, and pattern direction)
pattern values	the values of an element feature that will be used in the stimulus
pattern type	type of structure or organization present for a certain feature across the elements in the stimulus (e.g., pattern repeat, element repeat, mirror symmetry, gradient)
pattern direction	the direction in which the pattern values are applied according to the pattern type across the stimulus (e.g., across elements, across rows, across columns)
position deviation	deviation from the position pattern, either by the addition of random jitter to the element positions or by the adaptation or removal of specific element positions
element deviation	deviation from the element feature patterns present in the stimulus, by removing elements from the stimulus, swapping element positions, or randomizing element positions
feature deviation	deviation from at least one element feature pattern, by removing elements from the stimulus, swapping one or more features between elements in the stimulus, randomizing element or feature positions, changing the feature value of one or more random or specified elements in the stimulus, or by jittering one or more numeric element features

Figure 7.6: Terminology used in OCTA. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.18102443.

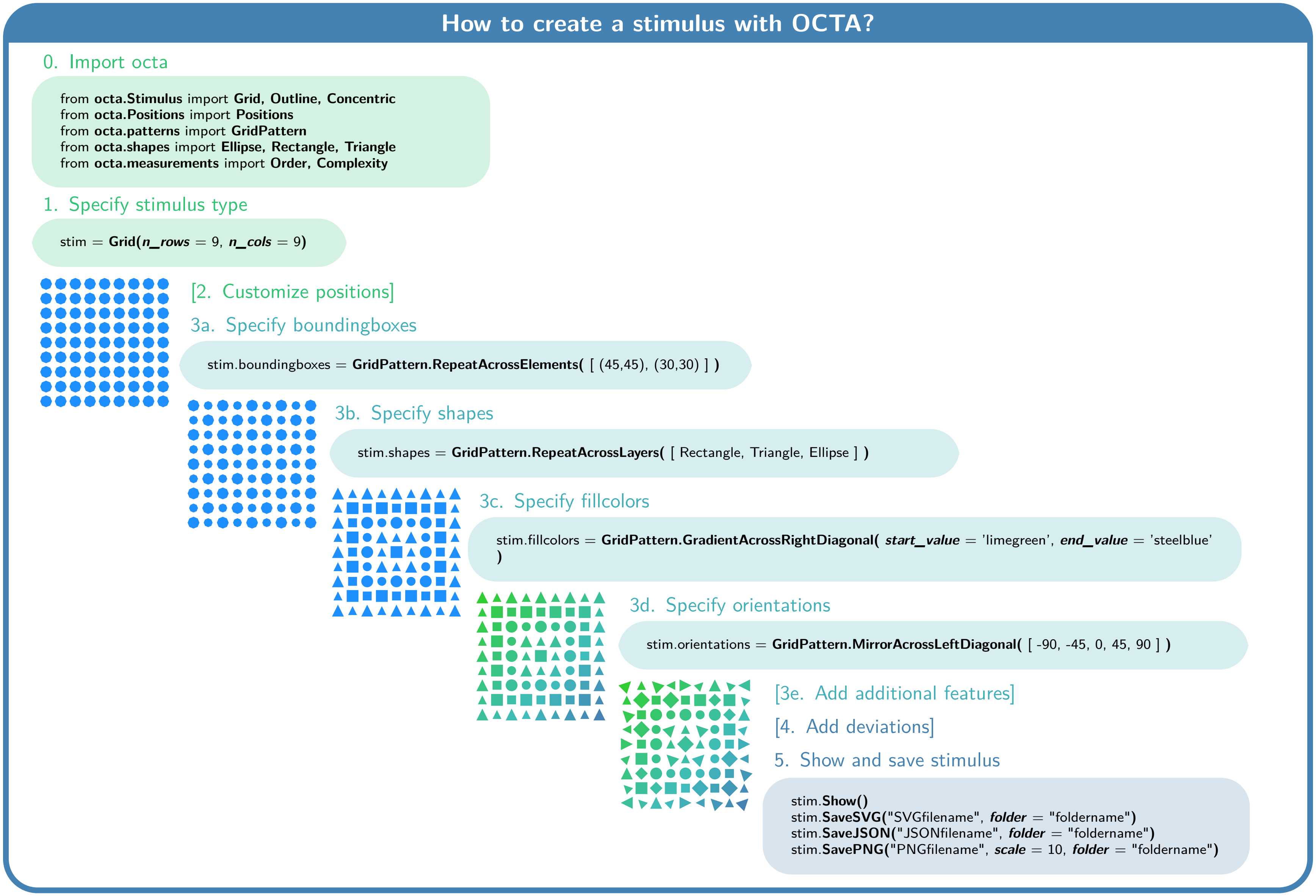


Figure 7.7: Tutorial on how to create a stimulus in OCTA. In the PDF version, clicking a stimulus leads to the octa code used to generate it. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.17743082.

7.2.1 Step-by-step guide to creating a stimulus using OCTA (see Figure 7.7)

The following section gives a more detailed overview of the steps of creating a stimulus (set) in OCTA.[11] For a visual summary of the different steps needed to create a first stimulus in OCTA, consult Figure 7.7.

7.2.1.1 Step 0. Load the octa package in Python

To start using the octa Python package, install octa as well as its dependencies (svgwrite, svg.path, svgpathtools, svgutils, jsonpickle, html2image, svglib, reportlab, colour, and IPython). Once everything is installed, it can be helpful to import specific functions from the octa package. In the Shiny app, this step is unnecessary as all required packages are loaded automatically. Furthermore, the user can download the Python code from the Shiny app and copy it to Python to import the functions from the octa package necessary for a specific stimulus.

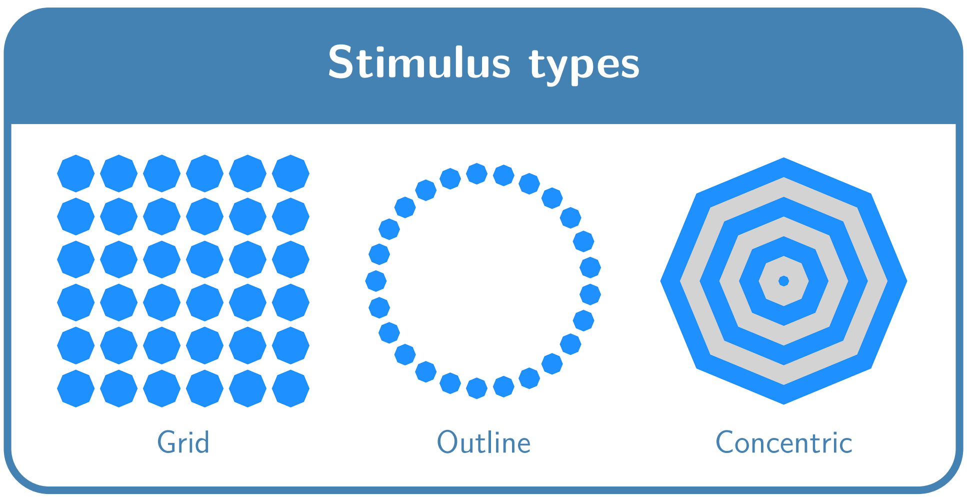


Figure 7.8: Stimulus types available in OCTA. In the PDF version, clicking a stimulus leads to the octa code used to generate it. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.17748872.

7.2.1.2 Step 1. Specify a stimulus type

The first step when creating a stimulus is to define the type of stimulus one wants to create: a Grid stimulus, an Outline stimulus, or a Concentric stimulus (see Figure 7.8).[12] These stimulus types correspond to frequently used stimulus types in the literature. For instance, Grid stimuli have been used by Garner & Clement (1963) and Chipman & Mendelson (1979), and have formed the basis for FlexTiles and many stimuli in visual search tasks. Outline stimuli could form the basis for hierarchical stimuli like those used by Navon (1977), Poirel et al. (2006), and Krakowski et al. (2016). For Concentric stimuli, see for instance Gollwitzer et al. (2017). Whereas a Grid stimulus requires a specified number of rows and columns, Outline or Concentric stimuli require a specified number of elements.[13] If desired, the user can specify additional arguments to customize the stimulus. The following features of the stimulus

[11] In line with footnote 10, the following description of the available options (especially in steps 2-4) will probably be overwhelming for most first-time users. The best way to come to grips with this is to first check the figures (especially Figures 7.7-7.15) for a quick visual impression of the available options and then to read the accompanying text for more details about specific possibilities. Again, this description will become more understandable and useful after having gained some initial experience with OCTA.

[12] These options were chosen to allow the user to create a broad range of stimuli with only a limited number of specified stimulus types. The set of specified stimulus types is partially arbitrary and necessarily incomplete, and can be expanded in later generations of the toolbox based on user feedback or requests.

[13] The stimulus types Outline and Concentric are defined as Grid stimuli with the number of rows equal to one and the number of columns equal to the number of elements.

as a whole can be specified: spacing between rows and columns (in case of a Grid stimulus), shape and shape bounding box (in case of an Outline stimulus), horizontal and vertical margin (when automatic sizing method is used), stimulus size (when fixed size is needed), background color, background shape, stimulus orientation, mirror value for the stimulus as a whole, mask to apply for the stimulus as a whole, and link, class label, and id label for the stimulus. By default, stimuli will be autocentered with a horizontal and vertical margin of 20 units in the current user coordinate system (i.e., user units), have a white background color, and an orientation of 0°. The Grid category has a default row and column spacing of 50 user units, and the Outline category by default has a circular shape with a bounding box of 150 by 150 user units. For more information on all stimulus features, please consult the online documentation. In the Shiny app, additional stimulus features can be specified under the tab '0. Add stimulus features'. The general stimulus type has been merged with the other position pattern options however, and can be changed under the tab '1. Add position pattern'.

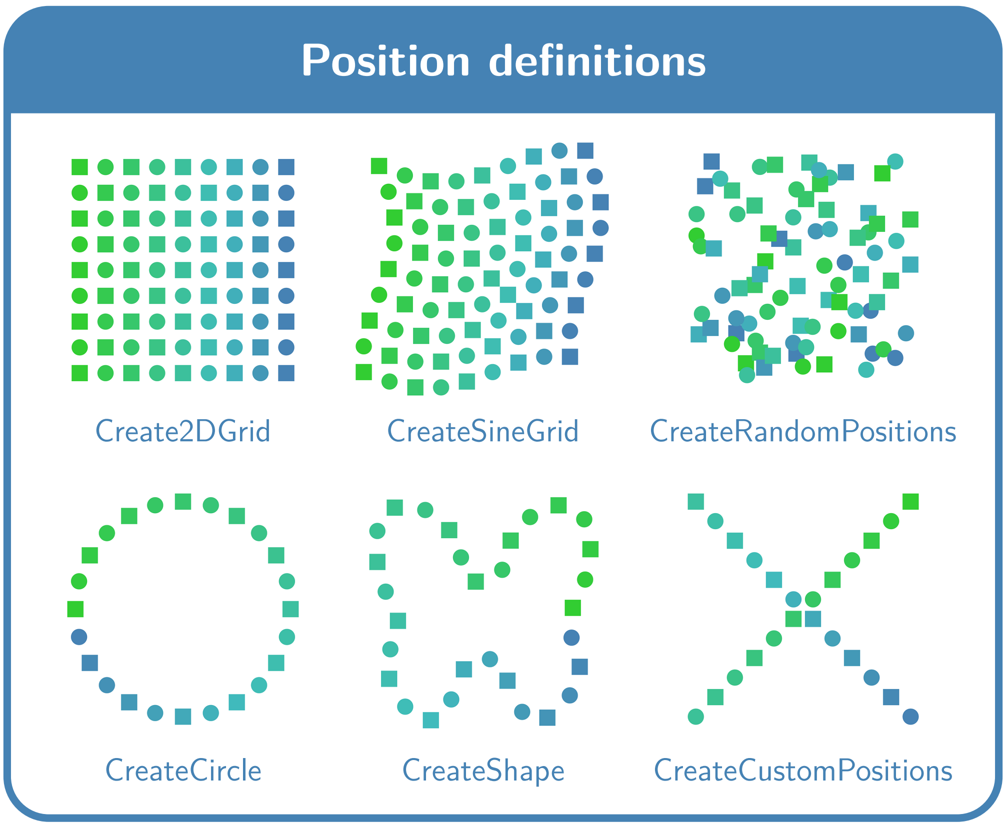


Figure 7.9: Position definition options available in OCTA. In the PDF version, clicking a stimulus leads to the octa code used to generate it. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.17749139.

7.2.1.3 Step 2. Specify element positions

Having created the stimulus, the resulting x and y coordinates for the element positions can be requested via `stimulus.positions.GetPositions()`. The position of an element is determined by the center of the element's bounding box (i.e., the rectangle 'bounding' the size of the element shape). The default position patterns used for Grid, Outline, and Concentric stimuli are a rectangular grid with a row and column spacing of 50 user units, a circle outline with radius 150, and identical (0,0) positions, respectively. The user can replace the default position pattern for the stimulus type by a custom one (e.g., a sinewave-shaped grid, a custom shape outline, random positions within a specified rectangle, manually specified positions). Figure 7.9 gives an overview of all currently available position pattern functions.[14] Important to note is that any custom set of positions can be defined using the CreateCustomPositions function. For more information on all position pattern definitions, please consult the online documentation. In the Shiny app, the element positions can be specified under the tab '1. Add position pattern'.

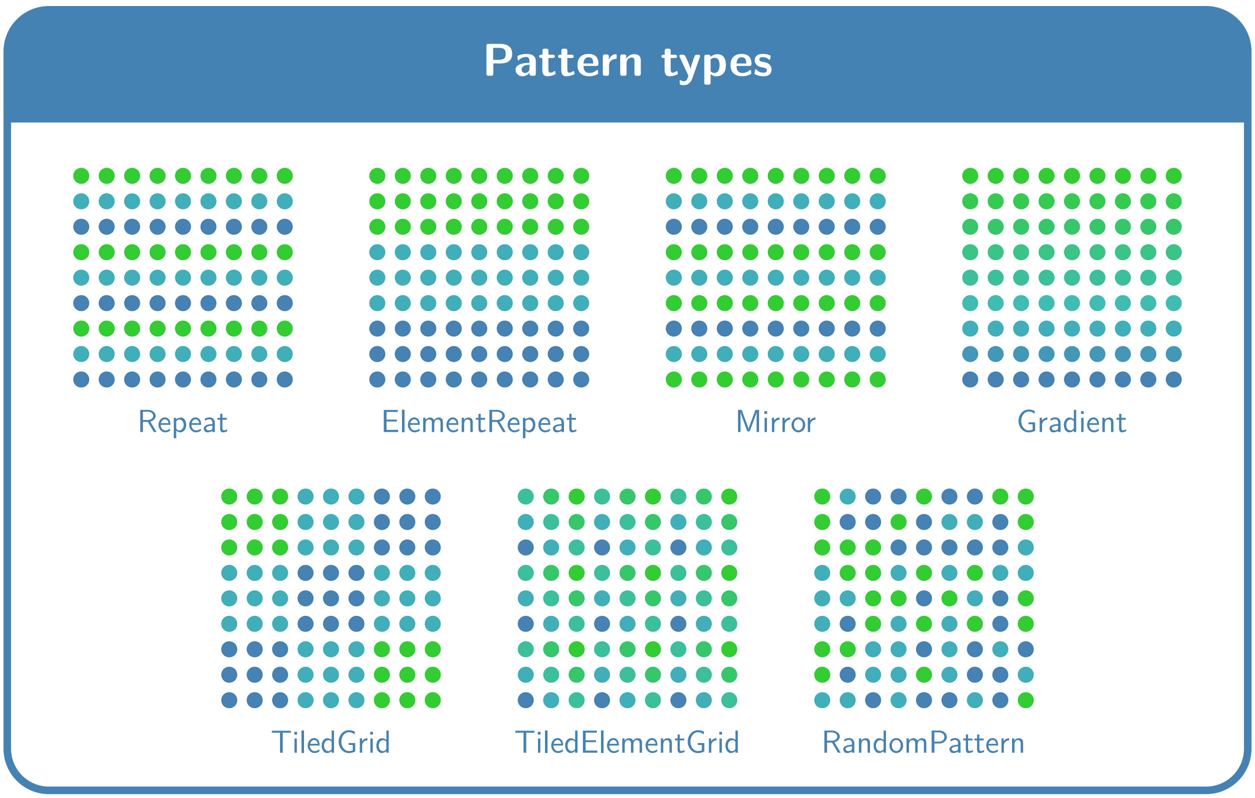


Figure 7.10: Pattern types available in OCTA. In the PDF version, clicking a stimulus leads to the octa code used to generate it. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.17749340.

7.2.1.4 Step 3. Specify patterns and pattern values for different element features

Once the stimulus features and element positions are specified, the pattern types, pattern directions, and pattern values for the different element features can be adjusted. The patterns can be applied to the following element features: shapes, boundingboxes, fillcolors, orientations, borderwidths, bordercolors, opacities, mirrorvalues, links, classlabels, and idlabels (see Element features and Figure 7.12). In Grid and Outline stimuli, one value per element feature is repeated across all elements in the stimulus by default. In Concentric stimuli, a pattern with two fillcolors is repeated across all elements in the stimulus by default, and the boundingbox sizes follows a decreasing

[14]These options were chosen to allow the user to create a broad range of stimuli with only a limited number of specified position definitions. The set of specified position definition functions is partially arbitrary and necessarily incomplete, but can be expanded in later generations of the toolbox based on user feedback or requests.

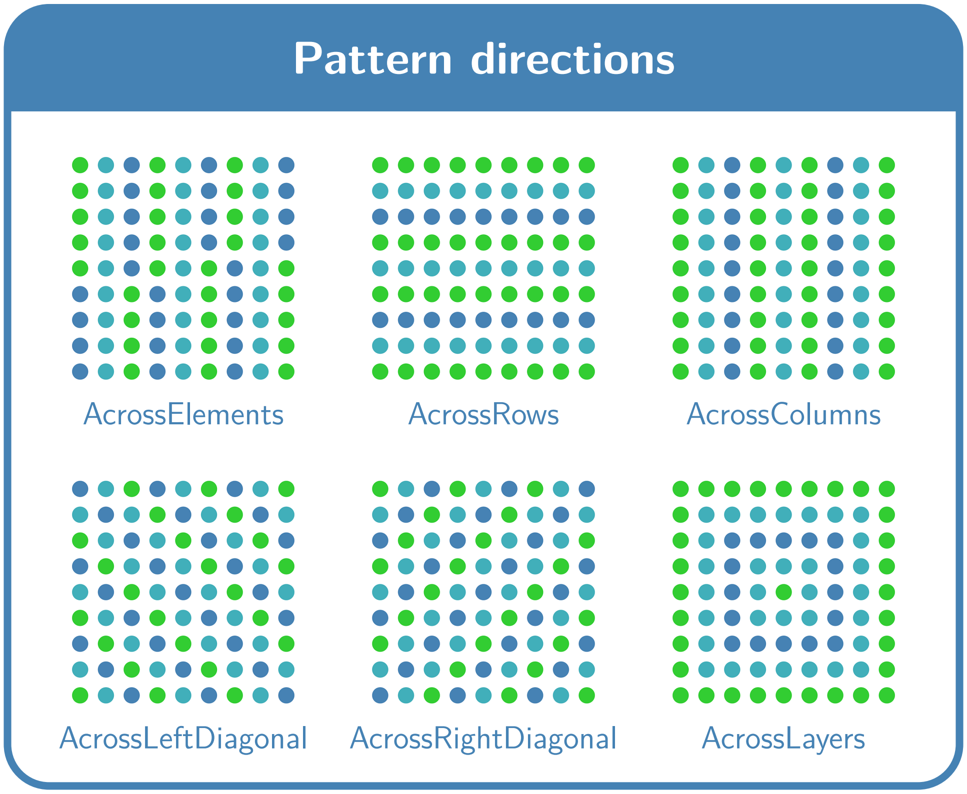


Figure 7.11: Pattern directions available in OCTA. In the PDF version, clicking a stimulus leads to the octa code used to generate it. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.17749634.

7

gradient across elements. Available *pattern types* (see Figure 7.10)[15] include pattern repetition (Repeat), element repetition (ElementRepeat), mirror symmetry (Mirror), and a gradient from a start value to an end value (Gradient). In Grid stimuli, these patterns can be applied according to the following *pattern directions*: across elements, across rows, across columns, across the left diagonal, across the right diagonal, and across layers (see Figure 7.11). In addition, more complex patterns can be constructed using the TiledGrid and TiledElementGrid options. Whereas a TiledGrid copies the feature values in a source grid a specified number of times in the row and column directions, a TiledElementGrid copies each element in the source grid a specified number of times in the row and column directions. Finally, a random application of the values for the element feature across the elements is possible too (RandomPattern). By default, the pattern values in the RandomPattern are repeated until the length is equal to the number of elements in the stimulus. Optionally, a list of frequencies can be provided to determine how many times each pattern value has to be present. In Outline and Concentric stimuli, it is strongly advised to apply patterns across elements (as other pattern directions will not be distinguishable in Outline and Concentric stimuli). In the Shiny app, the feature patterns can be specified using the tabs under ‘2. Add feature patterns’. Some predefined pattern values are provided for each element feature, but custom pattern values can be specified too.

7.2.1.4.1 Element features (see Figure 7.12)

[15] These options were chosen based on the currently available literature using different pattern types, e.g., Chipman & Mendelson (1979), and to allow the user to create a broad range of stimuli with only a limited number of specified pattern types. The set of specified pattern types is necessarily incomplete, but can be expanded in later generations of the toolbox based on user feedback or requests.

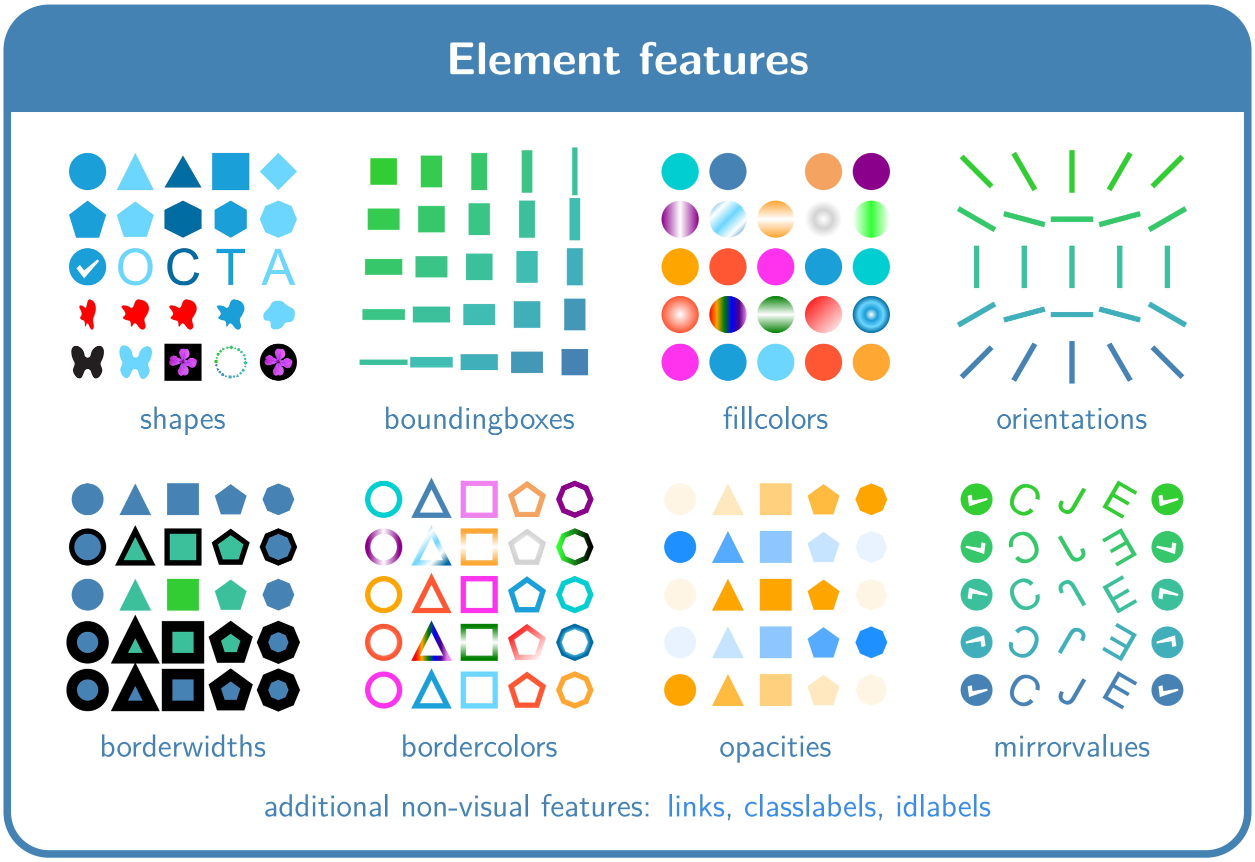


Figure 7.12: Element features available in OCTA. In the PDF version, clicking a stimulus leads to the octa code used to generate it. To view dynamic and interactive versions of some of these stimuli, visit this webpage. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.17749229. The shapes example stimulus contains a butterfly image from the Auckland Optotypes (Hamm et al., 2018) and a flower image from Hůla and Flegr (2016).

7.2.1.4.1.1 Shapes

Available element shape types are geometric shapes, i.e., Ellipse, Rectangle, Triangle, Polygon(n_sides), RegularPolygon(n_sides); more complex path elements[16] based on a provided path string or an existing SVG file, i.e., Path(path, xsize, ysize), PathSvg(source); images based on a provided filename or source url, i.e., Image(source), FitImage(source); and text elements[17], i.e., Text(text). Keep in mind that for the order and complexity measurements, the shapes argument only takes the shape *type* into account and does not distinguish between different polygons, different paths, different images, or different texts.[18] Currently no built-in animation options are available for the shape feature, but it is possible to include dynamic image files (e.g., dynamic SVG or gif file) as shape values in the stimulus. Also in the Shiny app, it is possible to specify custom shapes to use in the stimulus, including Image and PathSvg shapes that are publicly accessible online. It is however not possible to use local files when working with the online Shiny app.

7.2.1.4.1.2 Boundingboxes

Boundingboxes are always rectangular but do not have to be squared. Boundingbox values are defined in user units and can be provided as follows: (xsize, ysize). Although currently no built-in animation options are available for the boundingbox feature, it is possible to either generate separate stimuli and combine them in time afterwards, or to include dynamic image files (e.g., dynamic SVG or GIF file with shape changing in size) as shape values in the stimulus. Shapes will take their maximal size possible within the specified boundingbox, taking into account shape definitions (e.g., the Image shape type will retain original aspect ratio of the image, whereas the FitImage shape type will fit the image to the boundingbox without taking original aspect ratio into account).[19]

7.2.1.4.1.3 Fillcolors

Besides static uniform colors defined based on their hexadecimal color code or color name, the user can specify a radial or linear color gradient with multiple colorvalues (i.e., radial, horizontal, vertical, diagonal) or dynamic fillcolors by setting a new color (i.e., 'set') when a certain action occurs (e.g., when the element is clicked) or by animating the color value (i.e., 'animate'). Fillcolors are not visible for image elements.

7.2.1.4.1.4 Orientations

Orientations are defined in degrees. Besides static orientation values, dynamic definition of orientations is possible (i.e., 'set' or 'animate').

7.2.1.4.1.5 Borderwidths

Borderwidths are defined in user units. Besides static borderwidth values, dynamic definition of borderwidths is possible (i.e., 'set' or 'animate'). Borderwidths are not taken into account when a shape is fit to a particular

[16]When providing a path definition or SVG file, keep in mind that the path or file may contain margins, making the visible result smaller than the defined boundingbox size. Correctly fetching a path from an existing SVG file may fail in case the SVG file contains multiple paths, no paths, or paths are specified differently than expected (e.g., including margins).

[17]Text elements are an experimental feature of the OCTA toolbox. The sizing of the text elements is optimized for a single capital letter to fit in the boundingbox provided.

[18]For example, in an OCTA stimulus containing two different images as image elements, the number of shape *types* used will be one. To take into account the number of different images, polygons, paths, or texts used, one can add the 'data' argument to the element features used in the order and complexity measurements.

[19]As specified in footnote 12, when providing a path definition or SVG file, keep in mind that the path or file may contain margins, making the visible result smaller than the defined boundingbox size. Furthermore, the sizing of the text elements is optimized for a single capital letter to fit in the boundingbox provided.

boundingbox value, which means that half of the borderwidth will fall outside of the specified shape and thus potentially outside of the specified boundingbox (as the borderwidth is centered around the shape border). Keep in mind that the default bordercolor is transparent, so bordercolor needs to be set for the borderwidth value to have a visible effect. In addition, borderwidths are not visible for image elements and may behave unexpectedly for path elements.[20]

7.2.1.4.1.6 Bordercolors

Besides static uniform bordercolors defined based on their hexadecimal color code or color name, the user can specify a radial or linear color gradient with multiple bordercolor values (i.e., radial, horizontal, vertical, diagonal) or dynamic bordercolors by setting a new color (i.e., 'set') when a certain action occurs (e.g., when the element is clicked) or by animating the color value (i.e., 'animate'). Bordercolors are not visible for image elements. Keep in mind that the default borderwidth is zero, so borderwidth needs to be set for the bordercolor values to have a visible effect.

7.2.1.4.1.7 Opacities

Opacities are defined between 0 (no opacity) and 1 (full opacity). Besides static values, dynamic definition of opacities is possible (i.e., 'set' or 'animate').

7.2.1.4.1.8 Mirrorvalues

As element shapes can be mirrored along the horizontal and/or vertical axis, or not be mirrored, this element feature can take any of four different values: 'none', 'horizontal', 'vertical', 'horizontalvertical'. Currently no built-in animation options are available for the mirrorvalue feature, but it is possible to either generate separate stimuli and combine them in time afterwards, or to include dynamic image files (e.g., dynamic SVG or GIF file with shape changing in mirrorvalue) as shape values in the stimulus.

7.2.1.4.1.9 Links

A custom hyperlink can be added to every element in the stimulus. This entails a hand cursor shown when the viewer hovers over that specific element, and the opening of a hyperlink when the element is clicked.

7.2.1.4.1.10 Classlabels and idlabels

Although classlabels and idlabels do not have a directly visible effect on the resulting element, they can be used to add additional JavaScript actions or CSS style changes to individual elements (using the idlabel) or to a group of elements (i.e., all elements with the same classlabel). This method can, for example, be used to add sounds when hovering over an element or when an element is clicked.

7.2.1.4.1.11 Data

Data is a hidden element feature not meant for direct user interaction. It stores the additional arguments given to any of the shapes (e.g., the number of sides for Polygon objects, the source argument for Image objects). This element feature may become important when calculating order and complexity measures within the OCTA toolbox,

[20] Related to footnote 12, when providing a path definition or an SVG file containing a path definition, the path definition may contain margins, making the visible result smaller than the defined boundingbox size. In addition, the path may be defined on another size than it is scaled to, and the borderwidth will be applied to the original size, making the visible borderwidth often smaller than intended. A correct path definition and a borderwidth value custom to the original path boundingbox is thus key to get a correct borderwidth for path elements.

to distinguish between polygons with a different number of sides, different path definitions, different text elements, or images with different sources.

7.2.1.5 Step 4. Add deviations and calculate measures

Once all element feature patterns are specified as desired, the user can add position, element, or feature deviations to decrease the order and/or increase the complexity of the stimulus. Figure 7.13 gives examples of the position, element, and feature deviations that can be added. Although most of these deviation types affect both objective order and objective complexity, some manipulations specifically target either order or complexity. Position, element, and feature deviations are saved separately from the original patterns used to create the stimulus. If multiple deviations are added, later deviations could overwrite earlier deviations if they concern the same position, element or element feature. In the Shiny app, deviations can be added under the tab '3. Add deviations'. Within this tab, position and element deviations are bundled under 'Position deviations' and element feature deviations are displayed under 'Feature deviations'.

7.2.1.5.1 Position deviations

Random jitter or specified deviations can be added to the element positions. Although the impact of position deviations is dependent on the specific deviations added as well as the position pattern used as starting point, they generally increase objective complexity in the positions used in the stimulus (i.e., they increase the variety of positions present in the stimulus). Element or feature complexity stays unchanged when deviating position only. Whether the position deviations influence objective order in positions, elements, or features depends on the specific deviation.[21] For normally distributed or uniformly distributed ***position jitter***, one can specify whether the jitter needs to be applied to the x coordinates, y coordinates, both axes equally, or both axes independently using the axis argument. In the case of uniformly distributed jitter, the user specifies a minimum and a maximum value (min_val and max_val). In the case of normally distributed jitter, the user specifies mean (mu) and standard deviation (std). Default values for axis, distribution, min_val, max_val, mu, and std are 'xy', 'normal', -1, 1, 0, and 1, respectively. To add ***specific position deviations***, the user specifies the element ids (starting from 0 until n_elements - 1), x offsets and/or y offsets for each of the elements to which a deviation relative to the predetermined position needs to be added. Element id needs to be given an integer value or a list of integer values. X and y offsets can be numeric values or a list of numeric values.

7.2.1.5.2 Element deviations

To add element deviations, it is possible to ***remove*** a set of random or specified ***elements*** from the display, to ***swap*** the positions of distinct ***elements***, or to ***randomize*** the order of all ***elements*** in a particular direction. Bringing an additional element into the pattern is possible as well, but requires the user to add an additional element in the stimulus definition and in the definition of the stimulus positions. Swapping or randomizing the position of distinct elements in the display decreases objective element order, but leaves objective position, element, or feature complexity unchanged. If also non-distinct elements would be swapped or randomized, the objective element order

[21] One may argue that position deviations decrease the objective order in the positions, but this may not always be the case. For example, when adding symmetric position deviations, these positions deviations may only alter—or in some cases even increase—the structure and organization of the positions present in the stimulus.

Figure 7.13: Position, element, and feature deviation options in OCTA. In the PDF version, clicking a stimulus leads to the octa code used to generate it. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.17749763.

may stay unchanged. Removing elements does complicate the position pattern and potentially reduces element or feature complexity, but more generally also decreases element and feature order.

7.2.1.5.3 Feature deviations

To add feature deviations in the stimulus, it is possible to ***change*** a ***feature value*** for a number of random or specified elements, to ***swap*** the ***feature values*** for a number of random pairs of (distinct) elements in the display, to ***randomize*** the order of all ***feature values*** in a particular direction, or to ***jitter*** any of the ***numeric feature values*** across all elements. Swapping or randomizing the position of distinct feature values in the display decreases objective order for the feature dimensions involved, but also increases objective element complexity. Changing a feature value to a value that is not yet in the pattern values for that feature or jittering numeric feature values will additionally increase objective feature complexity. An advantage of adding feature deviations is that order can be distorted on one feature dimension specifically but preserved for other feature dimensions (contrary to what is the case with element deviations).

7.2.1.5.4 Order and complexity measures and manipulations

Although deviations are one way to increase or decrease different types of order and complexity in the stimulus, other approaches are possible too. Figures 7.14 and 7.15 give an overview of some order and complexity manipulations that are possible in OCTA. Figure 7.16 lists the order and complexity measures available in OCTA.

7.2.1.5.4.1 Manipulating order

Position order can be changed qualitatively by changing the type of position pattern. Element (and feature) order can be changed qualitatively by changing the different element feature pattern types and directions.[22] To induce quantitative changes in order, the user can swap the positions of distinct elements or randomize the elements in the stimulus (which keep element complexity level constant), or add any other element or feature deviations (but these other deviations may influence element or feature complexity as well; cf. Step 4. Add deviations). The user can also make stimulus features and element features (in)congruent to impact the order level of the stimulus. In addition, the congruency of patterns, pattern types, or pattern directions across feature dimensions can be adapted.[23]

7.2.1.5.4.2 Manipulating complexity

Qualitative changes in complexity can be achieved by changing the feature dimension on which the complexity is present (e.g., shape, color, or size complexity). Quantitative element complexity changes can include (a) changing the number of visible elements present in the stimulus (i.e., by removing elements or by changing the position pattern of the stimulus), (b) manipulating the variety of elements (i.e., by including more pattern values on a feature dimension, by choosing more diverse pattern values, by adding feature deviations, or by changing the congruency of patterns across feature dimensions), (c) changing the complexity (familiarity, unintelligibility, etc.) of individual feature pattern values (e.g., use complex path shape instead of rectangles), or (d) changing the complexity of individual stimulus features. Position complexity can be changed quantitatively by adding random position jitter or structured position deviations (cf. Position deviations).

[22] Be aware that some feature pattern changes may also impact element complexity because of emerging (non-)congruency between different feature patterns.

[23] Keep in mind that changes in pattern congruency may simultaneously influence element complexity.

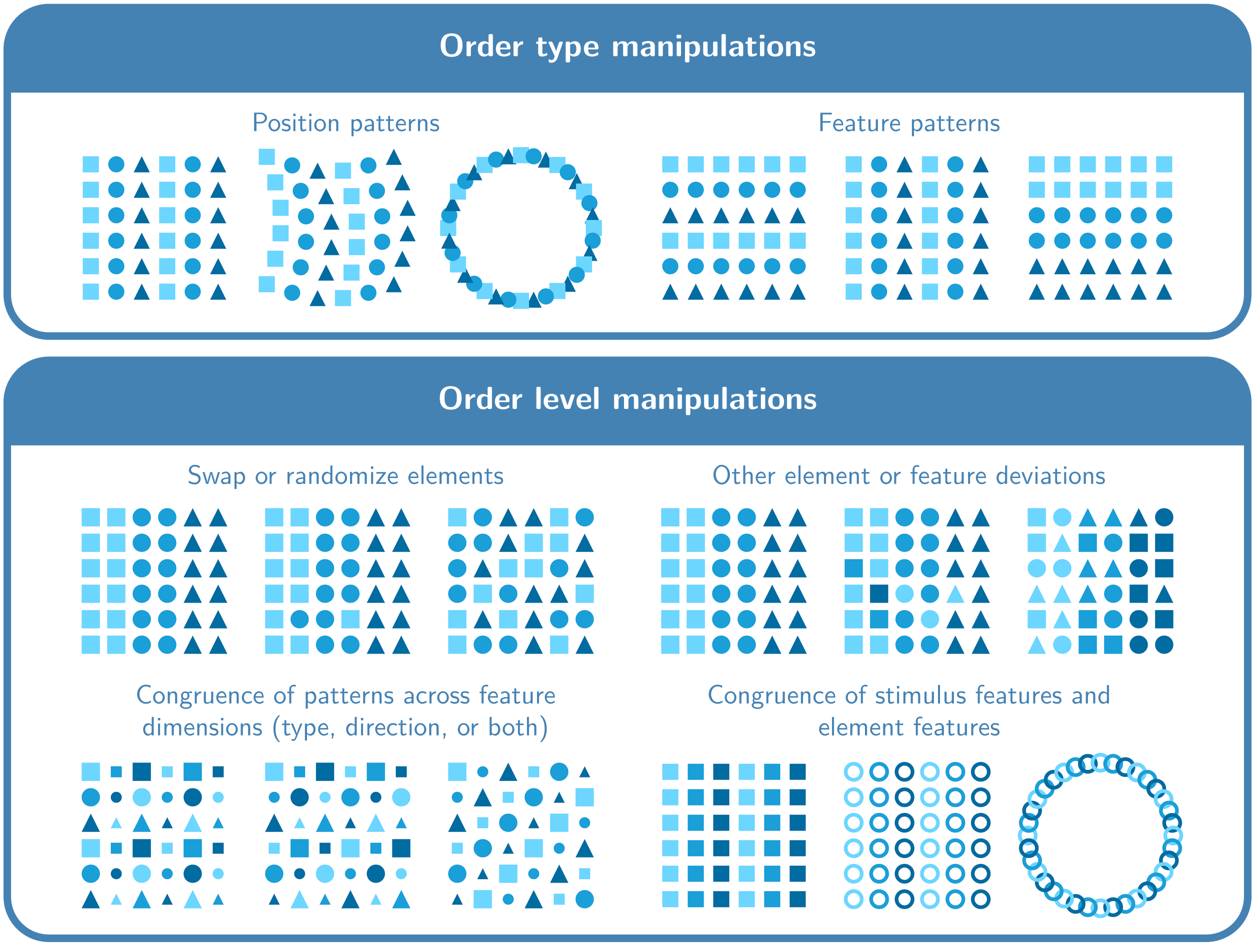


Figure 7.14: Order manipulations in OCTA. In the PDF version, clicking a stimulus leads to the octa code used to generate it. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.17749853.

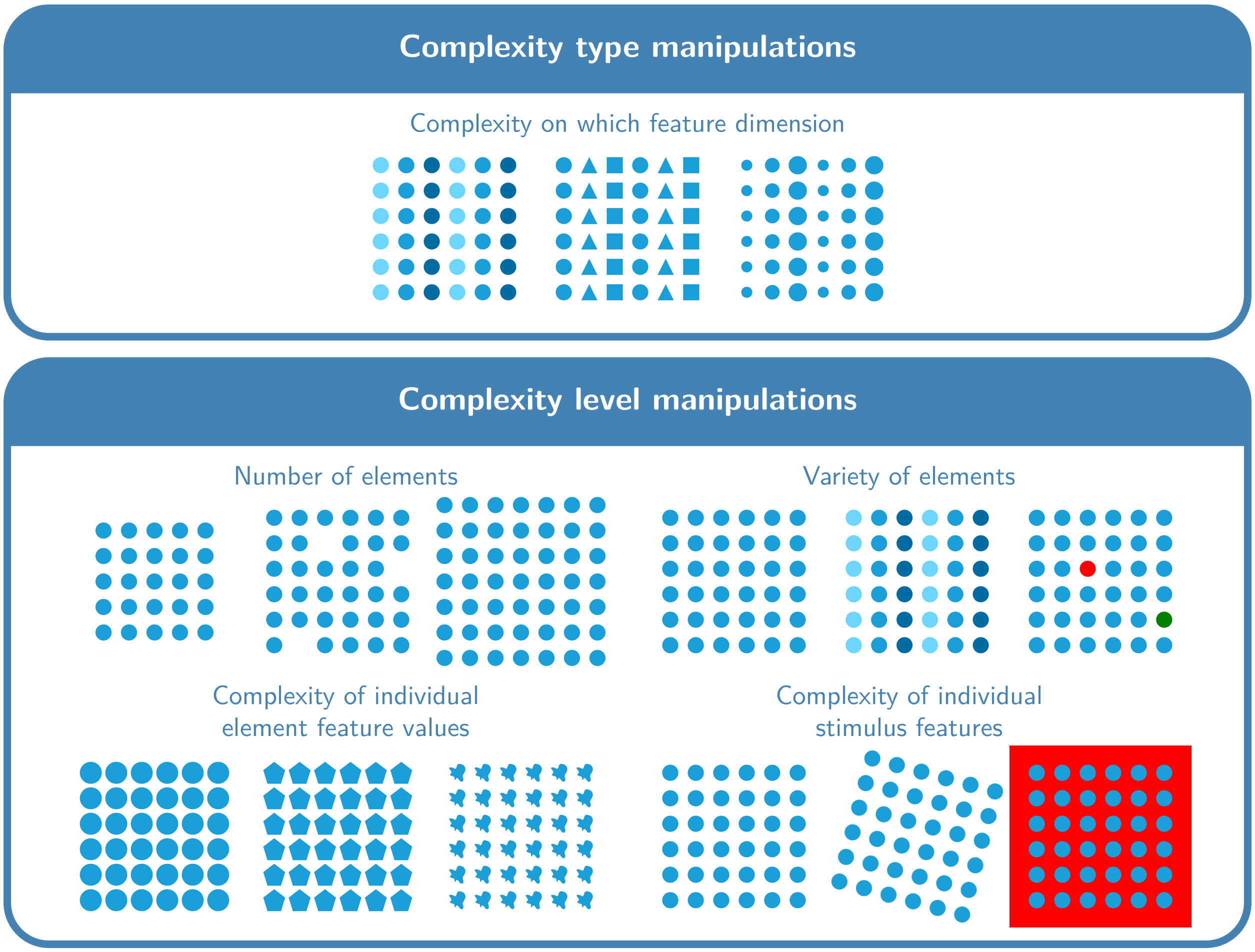


Figure 7.15: Complexity manipulations in OCTA. In the PDF version, clicking a stimulus leads to the octa code used to generate it. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.17749958.

7

7.2.1.5.4.3 Measuring order and complexity

OCTA provides some basic functionality to measure aspects of order and complexity in the created stimulus. When it comes to complexity measures, it is possible to calculate (a) the ***number of elements*** present in the display (Number or N), (b) ***how many different types of elements*** are present in the display based on the feature dimensions specified (Level of Complexity based on Elements or LOCE), (c) ***how many different features*** are present across all feature dimensions (Level of Complexity or LOC), and (d) ***how many different feature dimensions*** have ***more than one feature value*** (i.e., have non-identical values; Level of Complexity based on Identity or LOCI). For order, the user can ***request*** the applied ***patterns***, pattern types, and pattern directions across all feature dimensions; ***check*** whether all specified feature dimensions have ***congruent patterns***, pattern types, or pattern directions; ***calculate how many*** specified feature dimensions have ***congruent patterns***, pattern types, or pattern directions; calculate the ***number of deviant elements*** that are present given the specified feature dimensions (e.g., by added element or feature deviations); and calculate the ***number of deviant positions*** that are present in the stimulus (i.e., by added position jitter or specified position deviations). In the Shiny app, order and complexity measures can be calculated under the tab '4. Calculate measures'. Also in the Shiny app the user has the option to specify which element features to take into account when calculating the measures.

7.2.1.6 Step 5. Generate the stimulus, show it on screen, and save to the desired output format

7

Finally, the user can save the resulting stimulus in the preferred format: as a vector-based image (SVG), as a raster-based image (PNG, JPG, PDF, or TIFF), or as a computer-readable file (JSON). The SVG format is recommended for online use, as it has the same quality at all viewing sizes and keeps the possibility for animated element and stimulus features. For raster-based image output, a scale value can be added to increase the quality when starting from a very small stimulus that would otherwise have pixel artefacts.[24] The JSON output can be used to recreate the stimulus in Python using the octa package without the original code (using the `LoadFromJSON` function). In the Shiny app, the user can download the stimulus as a vector-based SVG image or as a raster-based PNG or PDF image, the JSON output, or the Python code needed to reproduce the current stimulus with the octa package in Python. Besides, the Shiny app user can view the Python code (without the octa import statements), the raw SVG code, and the JSON code in the app itself.

7.3 Discussion and conclusion

With the OCTA toolbox, it is possible to study order and complexity in combination. It acknowledges the multidimensionality of order and complexity and provides the tools to manipulate and measure order and complexity in several ways and on different feature dimensions.[25] The possibility to add more complex shapes and images in the stimulus enables the user to find their own balance between ecological validity and experimental

[24] It is also possible to add this scale value to the SaveSVG function, to scale the vector-based image to the preferred size.

[25] It is a conscious choice to not provide a full set of commonly used order and complexity measures within OCTA and only keep it to order and complexity measures directly deriving from the stimulus construction procedure. Calculating additional complexity measures can be useful, for which other tools already exist, e.g., imagefluency, Mayer (2021), described in Mayer & Landwehr (2018b) and Mayer & Landwehr (2018a); image spectral slope, fractal dimension and Shannon entropy as described in Mather (2018) and Mather (2020); PHOG measures described in Braun et al. (2013); and edge-orientation entropy as described in Redies et al. (2017). The user can use the vector- or raster-based image output to calculate these additional measures.

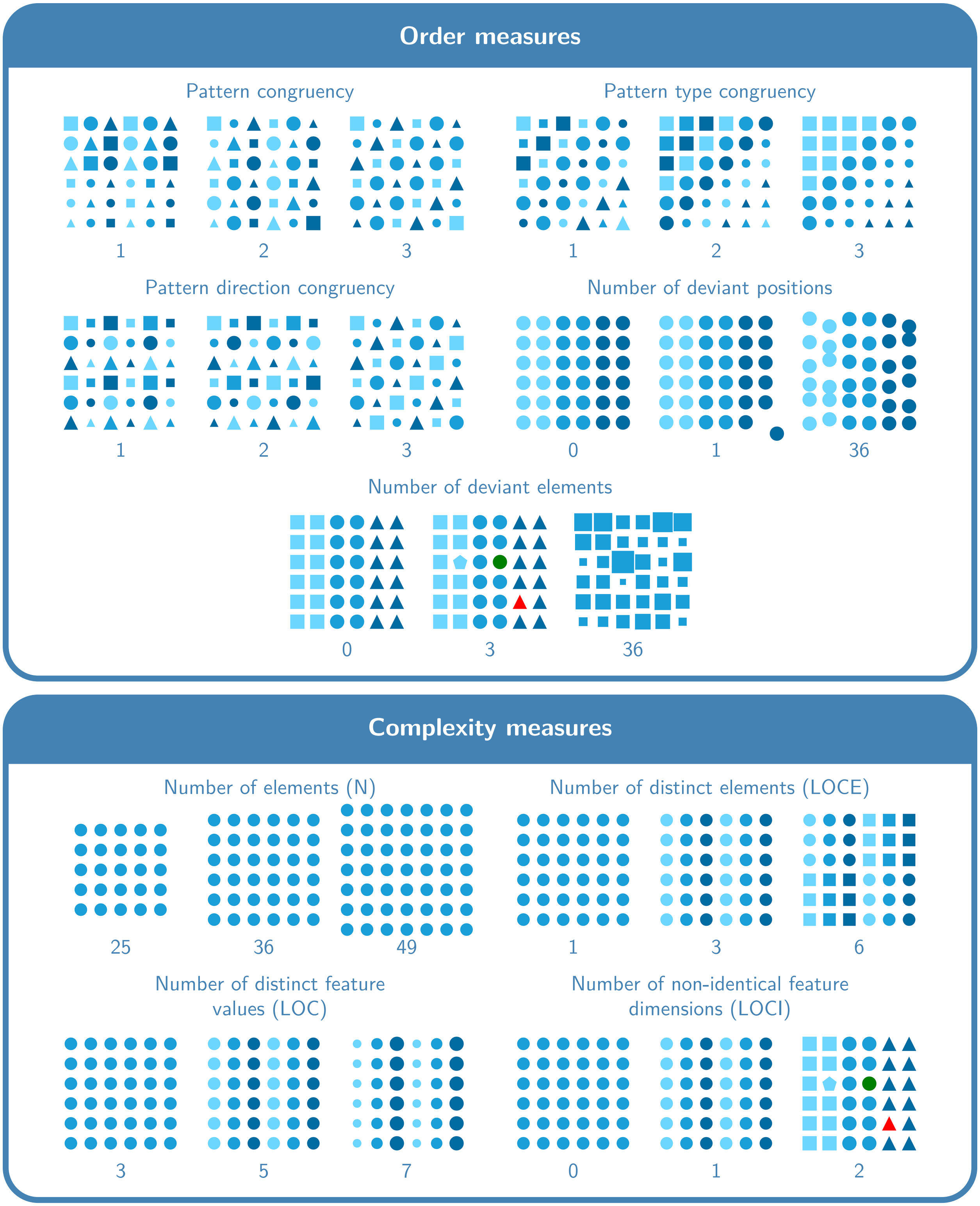


Figure 7.16: Order and complexity measures in OCTA. Examples include boundingboxes, fillcolors, and shapes as (distinction) features. In the PDF version, clicking a stimulus leads to the octa code used to generate it. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.17750069.

7

control. As online testing becomes ever more common, vector-based OCTA stimuli are ideal for online use and allow room for dynamical features. The OCTA toolbox however is not constrained to vector-based output as it still allows for saving the stimuli as raster-based image formats. Importantly, OCTA provides a reproducible way of creating stimuli, as all OCTA code is open source, and tools are made available to reproduce the stimuli, either based on the original seed used and the original code or with the original seed used and a JSON file. Furthermore, as both a Python package and an online app are provided and a multitude of documentation and example stimuli are provided, using OCTA is possible for both researchers with and without programming experience, making it a widely accessible tool. Below we discuss some more advanced uses of OCTA, potential applications, and give advice on how to start using OCTA.

7.3.1 Using animated stimulus and element feature values

One of the big advantages of the OCTA toolbox being vector-based, is the option to animate a diverse set of stimulus and element features either directly within the OCTA toolbox (i.e., using SVG animation to animate stimulus orientation and element fillcolors, orientations, borderwidths, bordercolors, and opacities) or after the OCTA stimulus has been created (i.e., using some additional CSS or JavaScript code, making use of the class and id labels that can be added to specific elements within OCTA stimuli and to OCTA stimuli as a whole).

Within the OCTA toolbox, a wide range of animation options is available: for example, animations initiated by clicking or by a specified starting time, animations including discrete steps or continuous change in feature values, one-time or indefinitely repeating animations. For use of animated feature values in OCTA, have a look at some dynamic stimulus examples or try out some of the animated example feature values in the OCTA Shiny app. For a more detailed reference on animation options in the SVG language, consult the SVG documentation on animation by The World Wide Web Consortium or Mozilla, or the documentation of the svgwrite Python package that is used within the octa Python package.

For more information on how to add CSS or JavaScript animations to SVG images based on class or id labels once the OCTA stimulus has been generated, consult the general documentation on CSS or JavaScript animations by W3Schools or search the internet for more specific tutorials on SVG animation using CSS or JavaScript. For a very simple demo using OCTA stimuli, consult the part on adding stimulus class and id labels in the OCTA documentation (or go directly to the example).

7.3.2 Creating sets of stimuli

Depending on the research question at hand, one can create differently controlled stimulus sets using OCTA. For example, a researcher interested in investigating the influence of the level of order on perception and appreciation of an image under different levels of complexity may create stimuli varying in order level but keeping order type, complexity level and complexity type constant at a smaller number of values. A researcher interested in investigating the generalizability of particular findings concerning order, complexity, and aesthetic appreciation may create stimuli keeping order and complexity levels and types constant but varying the pattern values used in each of the feature dimensions (e.g., fillcolors, boundingboxes, shapes). Furthermore, a researcher interested in studying to what extent and in which way order and complexity on different feature dimensions influence the perception and appreciation of an image, may create stimuli varying the type of complexity present (e.g., number and variety of fillcolors, boundingboxes, or shapes) and keeping order type, order level, and complexity level constant.[26] Researchers investigating perceptual grouping principles including proximity and different types of similarity may use the OCTA toolbox to create congruent and incongruent stimuli, vary row and column spacing in Grid stimuli, or vary the absolute feature values used to investigate the generalizability of the grouping strengths beyond typically used feature values (e.g., black-colored circular elements in the commonly used dot lattices). Moreover, researchers concerned with ecological validity of earlier findings may create equivalent stimulus sets with more abstract and more concrete shapes relevant to daily life (e.g., comparing a standard circle and a circle looking like a button, or comparing a standard triangle and a triangle looking like a tent). These are only examples of stimulus sets that could be created, as the OCTA toolbox gives researchers a very elaborate range of options that can be combined in any way preferred.

When creating sets of stimuli, it may be advisable to use the octa Python package rather than the app, as using Python directly will give you more opportunities to use loops and create multiple stimuli at once. This does not hold users back to first create one of the stimuli in the Shiny app, and then copy the code to Python for creating the complete set of stimuli. For researchers more familiar with R than with Python, the reticulate package (Allaire et al., 2017) can be very useful to interact with the octa Python code when creating and saving the combinations of parameter values for the stimuli in R rather than in Python directly.

7.3.3 Applications

Although OCTA has originally been created to study order and complexity in the context of empirical aesthetics, the toolbox can be used for generation of static or dynamic stimuli in a much broader field of research using visual stimuli as well as in non-academic use contexts. Examples of some research fields that could benefit from the use of OCTA are perceptual organization (e.g., grouping principles), symmetry detection, local-global processing and part-whole relationships, texture perception, visual search, visual illusions, perceptual averaging, ensemble perception, other types of visual perception research (for a review of many of these topics, see Wagemans, 2018), creativity, joy of ordering, aesthetic appreciation in general, and many more (see Figure 7.17). For example, although OCTA is not built with a focus on different types of symmetry specifically, all four basic two-dimensional symmetry operations (rotation, translation, reflection, and glide reflection) are at least in some form automatically available in OCTA[27], and the seven frieze patterns as well as the 17 wallpaper patterns can be created using OCTA.[28] In sum, OCTA can be a useful tool in any type of research using visual stimuli, and even to generate digital art (for some more examples, see Figure 7.1). In addition, OCTA stimuli can be used in many different kinds of tasks

[26] One prerequisite for studying the multidimensionality of order and complexity may be finding equivalent levels of change on different feature dimensions.

[27] Rotation can be achieved using the orientation feature dimension, translation in horizontal and vertical directions is possible using the col_spacing and row_spacing parameters in the Grid stimulus type and translation in any direction is possible using custom position definitions; reflection can be achieved using the mirrorvalue feature dimension; and glide reflection can be achieved using a combination of translation and reflection manipulations.

[28] As OCTA currently builds mostly from rectangular grids, generating wallpaper patterns starting from these requires less customization (e.g., custom position definitions) than creating wallpaper patterns starting from other grid types. One potential avenue for further development of the OCTA toolbox is adding additional grids and related pattern types, to make the creation of some of these wallpaper patterns more straightforward in OCTA.

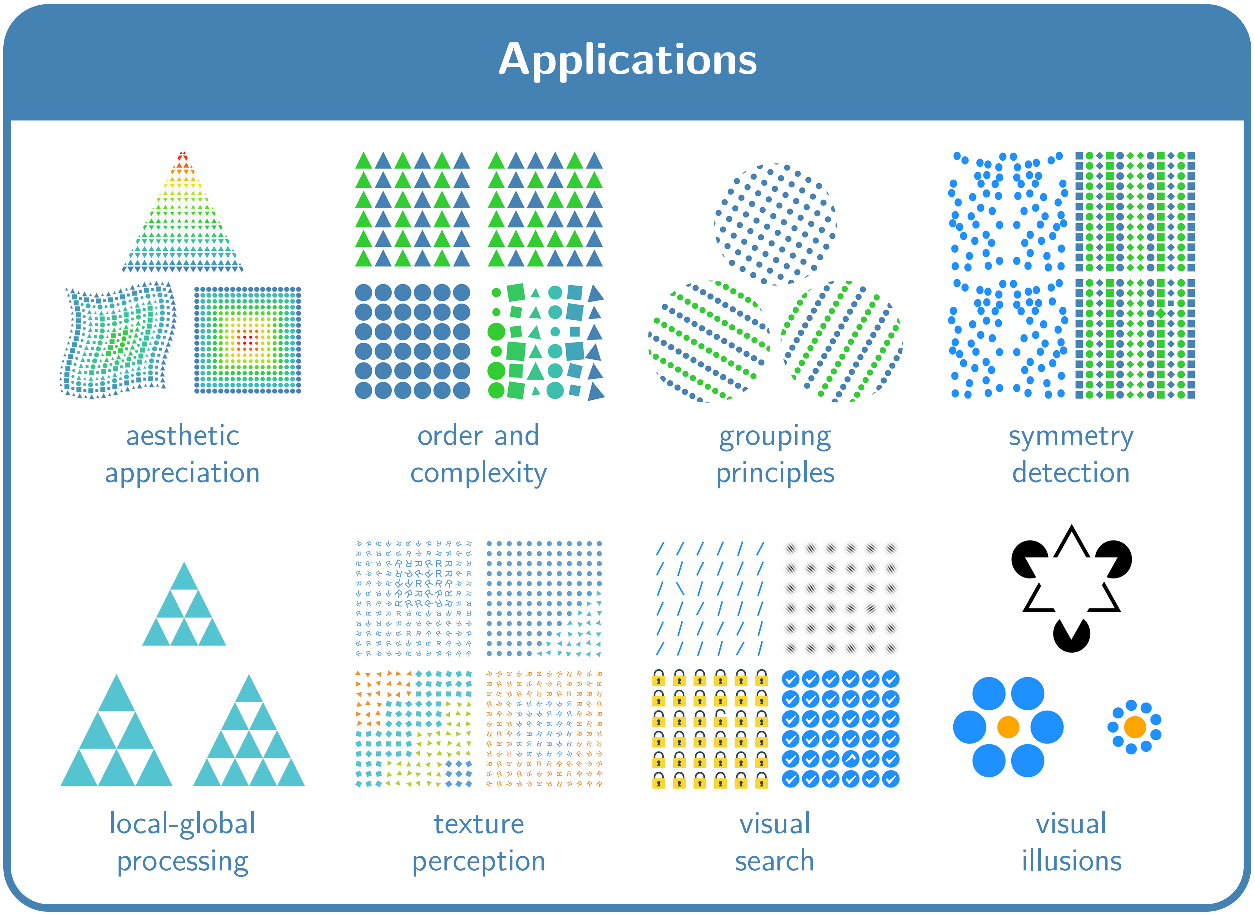


Figure 7.17: Applications of OCTA. In the PDF version, clicking a stimulus leads to the octa code used to generate it. Figure licensed under CC BY 4.0 by the authors. Retrieved from doi.org/10.6084/m9.figshare.17750123. The OCTA stimuli related to local-global processing are based on stimuli used in Kimchi and Palmer (1982, Exp. 1). The OCTA stimuli related to texture perception are inspired by Julesz (1981). The OCTA stimuli related to visual illusions represent a Kanizsa triangle and the Ebbinghaus illusion.

(rating, pairwise comparison, ranking, sorting, detection, construction, adjustment, etc., for a review of many of these tasks, see Palmer et al., 2013). For some experiment demos in jspsych (de Leeuw, 2015), consult the linked webpages (demos, code). Preliminary versions of OCTA are currently already being used both in research on the perception and appreciation of order and complexity (including order and complexity variations in shapes, boundingboxes, fillcolors, and number of elements, Van Geert, Warny, et al., in preparation; Van Geert, Hofmann, et al., in preparation) and in research on the proximity principle in perceptual organization (including manipulations of row spacing and column spacing, Van der Hulst et al., in preparation).

7.3.4 Advice on starting to use OCTA

Although learning to work with a new tool can lead to some cold feet, the online application as well as all the additional resources (e.g., manual, example stimuli, detailed function documentation; for an overview, visit elinevg.github.io/OCTA/) help users along the way. As OCTA is a new tool, feedback is welcomed and there are many opportunities for further development (e.g., additional position patterns and deviation options, additional animation options, option to create multiple stimuli at once). We do believe however that the current functionality already provides an immense array of options unexplored in aesthetics and visual perception research so far, and invite readers to explore the options[29] in the app (for starting users) or the Python package (for researchers with some prior Python programming experience; on the Python Package Index or on GitHub).[30]

7.4 Open practices statement

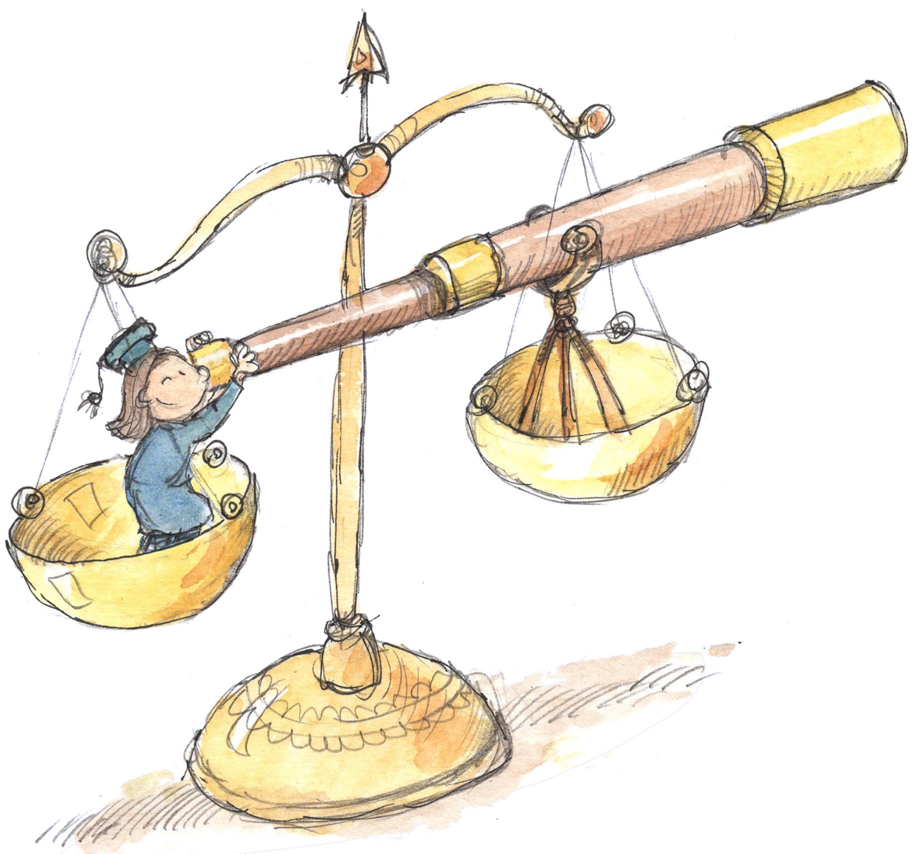
The octa Python package is available (on the Python Package Index or on GitHub) as open-source software under the GNU Lesser General Public License (Version 3) as published by the Free Software Foundation. The following webpage collects all additional resources concerning the OCTA toolbox: https://elinevg.github.io/OCTA/, including the Shiny app, the OCTA manual, the octa function documentation, example stimuli created in OCTA, demo experiments using OCTA stimuli, and more. When using either the octa Python package or the OCTA Shiny app in your (academic) work, please cite this paper to acknowledge the authors.

[29]With some additional JavaScript, it is even possible to add sounds to specific elements or stimuli as a whole and thus to create multimodal stimuli (cf. class labels and id labels in the online documentation).

[30]To start working with the octa Python package, it can be helpful to copy Python code from the Shiny app, the manual or the example stimuli in Python and start from there.

Part VII

Discussion



© Joris Snaet

To come to better psychological organizations, we find a *balance* between what we already know and the new input we receive.

Chapter 8
General discussion

Koffka (1935) posited the Prägnanz principle as the most important principle to guide research on perceptual organization. To advance research on perceptual organization, a more complete view on the meaning of Prägnanz is thus desirable. The central aim of this dissertation was to develop a more fine-grained understanding of Prägnanz and its added value for current theories and research on human visual perception and aesthetic appreciation. Importantly, the Prägnanz principle does not specify what will happen exactly in each and every specific situation (Arnheim, 1987; Rausch, 1952). Gestalt psychology, and as a consequence the Prägnanz principle, can be seen as the outcome of concrete research results, but should also be seen as a means towards further discoveries (Wertheimer, 1924/1999).

Using Prägnanz and the Prägnanz principle as a generative framework for further research is exactly what I have tried to do in this dissertation. Rather than aiming to 'test' the Prägnanz principle (which would assume Prägnanz to be a magical one-fits-all solution), I took the Prägnanz principle as a starting point for concrete research on perceptual organization (and aesthetic appreciation) in the visual modality.

In this General discussion, I first give an overview of the main findings reported in this dissertation. Afterwards, I dedicate some space to specifying how the Gestalt view connects to other theories on perception, with a focus on perception as Bayesian inference. By doing so, I aim to clarify the added value of a Gestalt perspective for current theories and research on human visual perception, aesthetic appreciation, and beyond.

8.1 Summary of the main findings

How do we perceptually and cognitively organize incoming stimulation? The Gestalt psychologists answered this question by providing evidence for several specific principles of organization, with the law of Prägnanz as an overarching principle: psychological organization will always be as 'good' as the prevailing conditions allow (Koffka, 1935). In Chapter 2, I reviewed how the Gestalt school further specified this law, by providing answers to (a) what a 'good' psychological organization entails, (b) how the Prägnanz tendency can be realized, and (c) which prevailing conditions need to be taken into account. Furthermore, I discussed four main uses of Prägnanz in more detail: (a) Prägnanz as a tendency present in each organizational process, (b) Prägnanz as a property of a Gestalt, (c) Prägnanz steps as internal reference points, and (d) Prägnanz in the context of aesthetic appreciation and artistic practice.

Importantly, Prägnanz is a multifaceted concept indicating the psychological clarity of an experienced overall organization. Which factors contribute to this clarity of the experienced organization? First of all, at least some unity or regularity needs to be experienced in the overall organization. The experienced organization thus needs to be a Gestalt, different from a pure sum of its sensory elements. In a Gestalt, the different elements mutually support and determine each other: in experience, they are parts of a whole. In addition to the necessary requirement of experiencing at least some (1) unity or regularity in the organization, the clarity of an organization increases if it is experienced as (2) autonomous or independent rather than derived, (3) integer or complete rather than disrupted, (4) simple of structure rather than complicated of structure, (5) element rich rather than meager, (6) expressive rather than lacking expressiveness, and/or (7) meaningful rather meaningless. From this list of seven Prägnanz aspects, it becomes clear that Prägnanz can increase not only based on figural qualities, but also based on how the form of the organization relates to a semantic content (cf. 6th aspect) and how well the organization can be embedded in the organism's knowledge system (cf. 7th aspect). Furthermore, it also shows that Prägnanz can increase both with increasing order (i.e., structure and organization of elements part of a stimulus) and increasing complexity (i.e., quantity and diversity of elements part of a stimulus).

How can a perceiving organism then clarify the incoming stimulation, i.e., arrive at the most prägnant overall organization possible given the prevailing conditions? Gestalt psychology posits two main ways to clarify the incoming stimulation: either make the input more similar to what we already know (i.e., attraction), or make it more different (i.e., repulsion). Both tendencies are concurrently present in every perceptual process, although depending on the prevailing (internal and external) conditions they may be applied (a) in comparison to different reference distributions (i.e., internal or local reference), (b) to different aspects of the organization (e.g., color, shape, length, width, orientation), (c) with different strengths, and (d) from the stimulus to the percept (i.e., primary Prägnanz) or on the phenomenal level (i.e., secondary Prägnanz)[1].

8.1.1 Robustness and sensitivity

The part on robustness and sensitivity dealt with how long-term internal representations of good Gestalts can influence how the incoming stimulation will be perceptually and psychologically experienced. While the repulsive 'anchor' effect has been found mostly in direct perceptual comparisons, the attractive 'magnet' effect showed in tasks with a larger memory component and more limited perceptual input (Quinn, 2000). Hence, under strong external conditions sensitivity will be larger close to Prägnanz steps, under weak external conditions Prägnanz steps will mainly show their robustness (i.e., the visual input will be perceived as more similar to the Prägnanz steps than it is in reality).

In Chapter 3, we explained how an attractive tendency towards a Prägnanz step can lead to a category boundary effect. The category boundary effect indicates that differences between stimuli belonging to the same category are perceived as smaller than differences between stimuli belonging to different categories, even when the objective similarity between the stimuli is the same.

In an empirical study, we compared two types of morph stimuli on three different tasks: a categorization, a successive discrimination, and a similarity judgment task. For stimuli part of recognizable morph series, strong pre-existing internal reference regions were expected, leading to strong perceptual attraction towards these reference regions under limited viewing conditions (in this case, limited presentation duration). For stimuli part of non-recognizable morph series, we did not expect such strong reference points and hence less attraction.

[1]The latter is probably related to the localization of the tendencies in the 'hierarchy' of the brain (i.e., 'lower-level' processes influenced more heavily by the input vs. 'higher-level' processes influenced more heavily by the perceiver).

In the recognizable morph series, this attraction towards the internal reference points can explain the category boundary effect as follows. In within-category pairs, both stimuli are attracted towards the same internal reference, making the perceived difference between them smaller than the actual difference. In between-category pairs, stimuli are attracted towards different internal reference points, making the perceived difference between them larger than the actual difference. As a consequence, stimuli will be perceived as more similar in within-category pairs than in between-category pairs. How we view it this is not an effect of the category *boundary*, however, rather a consequence of attraction towards the internal reference regions on a dimension.

The results of our study are congruent with the explanations above. We drew two main conclusions. Firstly, the overall category boundary effect in discrimination performance and similarity judgment was much stronger for recognizable than for non-recognizable morph series. As expected, recognizable morph series also showed stronger categorization than non-recognizable morph series. Hence, the stronger the categorization for a morph series, the stronger the overall category boundary effect in discrimination performance and similarity judgment.

Secondly, for the recognizable morph series we saw a clear gradient in the results for the discrimination task, with discrimination performance gradually increasing the further the stimuli in the pair were from a reference point. The same gradual pattern was visible in the similarity judgment task, with stimuli being perceived as increasingly similar the closer they were to a reference point. Hence, for objectively equidistant stimulus pairs belonging to morph series that show evidence for categorization, the stimuli in the pair will be perceived as more similar and will be more difficult to discriminate the closer the stimuli in the pair are to one of the reference points. To accurately predict discrimination performance and similarity judgments, the distance from the reference points for each of the individual stimuli in the pair thus needs to be taken into account. In this study, 'distance from the reference points' can be equated to categorization strength, goodness of the experienced organization, or level of Prägnanz.

8

In Prägnanz terms, this study thus investigated attractive Prägnanz tendencies from the stimulation to the percept (i.e., primary Prägnanz) and on the phenomenal level (i.e., secondary Prägnanz)[2], both in conditions under which a non-uniform internal reference distribution with Prägnanz steps is available (i.e., recognizable morph series) and in conditions under which a more uniform internal reference distribution is present (i.e., non-recognizable morph series). Both the presence of strong internal Prägnanz steps (i.e., type of morph series) and the Prägnanz level of the currently experienced organization (i.e., based on morph level, indicating closeness to Prägnanz step) were varied. In Rausch's (1966) terminology, it considers the Prägnanz aspects of meaningfulness (7) and integrity (3), respectively. As stimuli were only presented for 300 milliseconds, the external conditions were weak, and primary Prägnanz tendencies could lead to deviations from the stimulus to the percept that were not directly noticeable by the observer.

Related to this theme on robustness and sensitivity, I conclude that *we use long-term internal representations of good Gestalts (i.e., Prägnanz steps) as reference points to clarify the incoming visual stimulation.*

[2]As we asked for categorization, discrimination, and similarity judgment, and not for direct reports of their percept, we leave open the option for secondary Prägnanz tendencies on the phenomenal level. That is, in addition to deviations from the stimulus to the percept, also a — potentially conscious — evaluation of the experienced closeness of the percept to the reference may influence the judgments asked for.

8.1.2 Hysteresis and adaptation

The part on hysteresis and adaptation dealt with how the immediate temporal context of a stimulus can influence how that stimulus will be perceived. While repulsive effects have been found for the immediate stimulus history, attractive effects have been found for the immediate perceptual history (e.g., Gepshtein & Kubovy, 2005; Schwiedrzik et al., 2014).

In Chapter 4, we investigated whether individuals differ consistently in how they combine previous visual input and experience with current visual input in their perception, using a multistable dot lattice paradigm.

The results of this Registered Report indicated that although individuals differed considerably in the size of their hysteresis and adaptation effects, almost everyone showed both effects in the expected direction. Furthermore, these differences in how strongly individuals were influenced by their stimulus history and perceptual history showed very stable across one to two weeks' time. Individual differences in hysteresis and adaptation were also strongly positively correlated. That is, individuals with a strong attractive effect of the previous percept, in general also showed a stronger repulsive effect of the previous stimulus evidence. Although not the main interest of this study, individuals also differed considerably in the direct effect of grouping by proximity (i.e., one of the traditional Gestalt laws of perceptual organization; Wertheimer, 1923). In addition, the relation between short-term and long-term context effects was not straightforward. This was potentially due to the low stability of absolute orientation bias *strengths* for at least some participants. Individual differences in the *direction* of the absolute orientation biases (i.e., the absolute orientation region with the highest Prägnanz) stayed relatively consistent across time, however. Furthermore, as effects in the control task were present but considerably smaller than in the main experimental task, we could conclude that the attractive temporal context effect was partially perceptual and partially decisional. Also, the results hinted at potential individual differences in the nature of the attractive temporal context effect.

In Prägnanz terms, this study thus investigated attractive and repulsive Prägnanz tendencies, mainly from the stimulation to the percept (i.e., primary Prägnanz). The current stimulus was fixed, but the Prägnanz level of the previously experienced organization varied, both by manipulating the aspect ratio of the previous stimulus and by differences in how that stimulus was experienced. In Rausch's (1966) terminology, the aspect ratio manipulation considers the Prägnanz aspect of integrity (3). The manipulation of perceptual and stimulus history in general may be seen as part of the by Wertheimer (1923) described Gestalt principle of set [*Einstellung*]. For the current stimulus, the external conditions were weak (presented for 300 milliseconds), but the immediately preceding stimulus was more clearly visible (presented for 800 milliseconds). Given the low visibility of the current stimulus, primary Prägnanz tendencies could lead to deviations from the stimulus to the percept that were not directly noticeable by the observer.

To explain the co-occurring attractive and repulsive temporal context effects in multistable dot lattice perception, we developed an efficient Bayesian observer model that included a likelihood on the stimulus level and a prior on the perceptual level (cf. Chapter 5). While likelihood repulsion on the *stimulus* level could explain the repulsive context effect, *perceptual* prior attraction could explain the attractive temporal context effect. Furthermore, we were able to reproduce the strong positive correlation between individual differences in hysteresis and adaptation effects by assuming a strong correlation between updating speeds at the stimulus and the perceptual level.

In a follow-up analysis (not part of this dissertation) using the categorization data from the study in Chapter 3, we found evidence for attractive and repulsive temporal context effects in the data for the non-recognizable morph series, but not in the data for the recognizable morph series (Van Geert & Wagemans, in preparation a; for an early presentation of this analysis at ECVP 2019, see this poster). Hence, short-term attractive and repulsive temporal context effects seem to have a larger influence on perception when no strong internal reference points are available.

In a follow-up study (not part of this dissertation), I zoomed in on attractive and repulsive temporal context effects in a categorization task using non-recognizable morph shapes, and explored potential individual differences in the size and direction of these temporal context effects (Van Geert & Wagemans, in preparation b; for an early presentation of this study at VSS 2021, see this poster and corresponding poster tour). The non-recognizable morph series used were set 1 and set 2 from Chapter 3, and the same categorization task procedure was used as in that study. Different from the dot lattice study reported in Chapter 4, not only the previous stimulus evidence and the previous percept varied, but also the current stimulus evidence. In this follow-up study, we found evidence for a clear attractive effect of the previous percept (i.e., hysteresis): keeping the current stimulus equal, this stimulus was more often perceived as part of category B when the previous stimulus was also perceived as part of category B. On the other hand, there were also indications of a repulsive effect of the previous stimulus evidence, but only when the current stimulus evidence was incongruent with the previous percept. When the current stimulus evidence was congruent with the previous percept, there was an attractive effect of the previous stimulus evidence. Put differently, sensitivity to the current stimulus evidence was higher when the previous stimulus was in accordance with the previous percept, and sensitivity was lower when the previous percept was a 'mistake'. Hence, when also the strength of the current stimulus evidence comes into play, the pattern of results may become more complex than was the case in the dot lattice study in Chapter 4, in which the current stimulus was always equally ambiguous (cf. also Gallagher & Benton, 2022). As a sidenote, Mao & Stocker (2022) proposed a hierarchical efficient Bayesian observer model precisely to be able to correctly predict the effects of different noise manipulations on previous and current stimulus. In the future, it may thus be worthwhile to also develop a hierarchical efficient Bayesian observer model that can explain the results structure for this follow-up study using non-recognizable morph shapes.

With these additional studies in mind, it becomes clear that the attractive and repulsive temporal context effects, as reported in Chapter 4, may have special importance for perception (a) when no strong long-term internal reference points are available, and (b) when the current stimulus is ambiguous.

Related to this theme on hysteresis and adaptation, I conclude that *the immediate temporal context* in which a perceiver receives the incoming stimulation *can have a strong influence on how that (ambiguous) stimulation is perceptually organized*, and individuals consistently differ in the extent to which they use this temporal context to clarify their percepts. Furthermore, an efficient Bayesian observer model can predict these co-occurring attractive and repulsive immediate temporal context effects using perceptual prior attraction and stimulus likelihood repulsion.

8.1.3 Simplification and complication

The part on simplification and complication dealt with how the immediate spatial context of a stimulus can influence how that stimulus will be experienced and communicated. When a difference between a stimulus and its spatial context is visible and experienced as important, the organization of that stimulus will be repelled from the spatial context. When a difference between a stimulus and its spatial context is barely noticeable and/or experienced as clutter, the organization of that stimulus will be attracted to the spatial context.

In Chapter 6, we asked participants to draw a target figure in such a way that another participant would be able to recognize it from the presented alternative figures. We investigated whether the importance of a feature for discrimination of the target figure among the alternative figures influenced whether the feature was repelled from the spatial context (i.e., complication, sharpening) or rather attracted to the spatial context (i.e., simplification, leveling). Importance of a feature for discrimination was operationalized in different ways: (a) whether the alternative figures were qualitatively or only quantitatively different from the target (i.e., far or close context); (b) whether the range of variability for a feature across the alternatives was more narrow or wide; and (c) whether the target value was an extreme value on the feature dimension or fell in-between the values for the alternative figures.

The study's results indicated that sharpening occurred more often for extreme feature values, for features exhibiting more variability, and for features of figures that were presented amongst alternative figures that were visually quite similar, than for non-extreme feature values, features exhibiting less variability, or features of figures presented amongst alternative figures that were totally different. In line with one of Metzger's (1941) definitions of prägnant Gestalts (i.e., good Gestalts as those structures that most purely and compellingly represent an essence), the *essence* of a Gestalt may be context-dependent, and this will influence whether leveling or sharpening of a feature will lead to the best organization in the specific context.

In Prägnanz terms, this study thus investigated attractive and repulsive Prägnanz tendencies, mainly on the phenomenal level (i.e., secondary Prägnanz), but potentially also from stimulus to percept (i.e., primary Prägnanz). The current stimulus was fixed, but the Prägnanz level of the currently experienced organization varied based on a manipulation of the spatial context in which the stimulus was presented. The manipulation of spatial context may be seen as part of the by Wertheimer (1923) described Gestalt principle of set [*Einstellung*]. In this study, the external conditions were rather strong (presented with high contrast for unlimited time, although at relatively small size), leaving less room for primary Prägnanz tendencies and deviations from the stimulus to the percept that were not directly noticeable by the observer, but leaving the possibility of (potentially conscious) secondary Prägnanz tendencies to play an important role.

Related to this theme on simplification and complication, I conclude that *the immediate spatial context* in which a perceiver receives the incoming stimulation *can have a strong influence on whether attractive or repulsive Prägnanz tendencies will occur*, and consequently how that stimulation is psychologically experienced.

8.1.4 Order and complexity

The part on order and complexity considered the relation between good perceptual and psychological organization and aesthetic appreciation. Order and complexity are not only important contributors to a better, more prägnant organization, they also relate to our experience of aesthetic appreciation (Van Geert & Wagemans, 2020). My earlier research (Van Geert et al., 2021; Van Geert & Wagemans, 2021) indicated that while order is almost always appreciated, appreciation of complexity is more variable across individuals and cultures. This research used non-parametrically controlled stimulus sets, however. Also other existing research had its limitations: it mainly focused on the separate influence of order and complexity on aesthetic appreciation, considered rather specific types of order (i.e., balance or symmetry), and/or ignored that order and complexity can show on multiple feature dimensions (e.g., color, shape, size, orientation).

In Chapter 7, I presented my answer to these limitations. With the Order & Complexity Toolbox for Aesthetics (OCTA), we provide an easy way to create reproducible and expandible stimulus sets, including both order and complexity manipulations on multiple feature dimensions.

Two first empirical studies using OCTA (not part of this dissertation) brought additional evidence for the conclusion drawn above, but now with parametrically controlled stimulus sets: order is a more consistent and situation-independent factor in appreciation, while the level of complexity that is appreciated varies more heavily between individuals, feature dimensions, and experienced levels of order (Van Geert, Hofmann, et al., in preparation; Van Geert, Warny, et al., in preparation, for an early presentation of these studies, see this oral presentation at TeaP 2022).

For Prägnanz, some form of regularity is always required, whereas the level and type of complexity a viewer can handle differs between contexts and individuals. The same seems to be true for aesthetic appreciation: while order is always a positive contributor, the impact of complexity is more condition-dependent.

Related to this theme on order and complexity, I conclude that *order and complexity are antagonistic but complementary factors that can both contribute to perceptual organization as well as aesthetic appreciation.* Furthermore, aesthetic appreciation may be related to the absolute level of Prägnanz height, but also to a relative increase in Prägnanz (i.e., the strength of the experienced Prägnanz tendency).

8.2 General considerations

8.2.1 Antagonistic but complementary attractive and repulsive Prägnanz tendencies

In the parts on robustness and sensitivity, hysteresis and adaptation, simplification and complication, and order and complexity, every time a pair of concepts is introduced that are antagonistic but also complementary. On the one hand, they work against each other: in the first three pairs listed, the first concept represents attraction to a reference, and the second concerns repulsion from a reference. In a univariate situation, they are thus necessarily contradictory. In most situations, however, stimulation is multivariate, leaving room for both tendencies to work together towards the best psychological organization possible under the prevailing conditions. And especially because these contradictory tendencies can complement each other to reach a goal, it is important to study them together.

The same is true for order and complexity. Although this concept pair is not directly linked to attractive and repulsive tendencies from a reference, it also concerns a balance between (a) organizing the input in such a way

8

that we can better understand and (b) embracing the unique complexities of the newly experienced organization from which we can learn. Just like in the case of attractive and repulsive tendencies, it concerns partial opposites that can both contribute positively to a common goal — regardless of whether it concerns perceptual organization or aesthetic appreciation.

This focus on interacting tendencies indicates how important dynamics are in the Gestalt approach, and may remind readers of the approach of nonlinear dynamical systems theory. The dynamical systems approach describes how a stable balance is reached when arriving at attractor positions on a dimension (Wagemans, Feldman, et al., 2012). The stronger these attractors, the higher their Prägnanz (van Leeuwen, 1990).

8.2.2 Dependence of primary Prägnanz tendencies on the incoming stimulation

Both in the part on robustness and sensitivity and in the part on hysteresis and adaptation, viewing conditions for the currently presented stimulus were limited (because of brief presentation duration). In Chapter 3 the ambiguity of the current stimulus was manipulated, in Chapter 4 the current stimulus was always ambiguous and the ambiguity of the preceding one was manipulated. Because of the limited visibility of the current stimulus, primary Prägnanz tendencies (i.e., deviations from stimulus to percept) had more leeway. In Chapter 3, attraction to an internal (not currently visible) reference dominated. In Chapter 4, attractive and repulsive tendencies could be distinguished as being related to the previous percept and the previous stimulus (which was clearly visible just before the current stimulus, as presentation duration was not limited), respectively.

These results hint at a general pattern: primary *repulsive* tendencies may be more prominent for direct perceptual comparisons, in which the reference to which the input is compared is more visibly present (or has been shortly before). Primary *attractive* tendencies may be less dependent on a strong visibility of the reference, and also occur for references for which there is less perceptual basis (e.g., internal reference, or less visible preceding stimulus). This dissociation between attraction and repulsion may indicate that both act in separate cortical networks in the brain. Some evidence has already been gathered for this idea. Schwiedrzik et al. (2014) for example related hysteresis to a widespread network of higher-order visual and fronto-parietal areas, whereas repulsion was confined to 'early' visual areas. Also other researchers have found indications of this dissociation between attractive and repulsive context effects, with attractive effects being stronger in perceptually ambiguous stimuli, non-attended stimuli (Kiyonaga et al., 2017), for stable attributes (Taubert et al., 2016), and repulsive effects being stronger for non-ambiguous stimuli, attended stimuli, and for changeable attributes.

I do not agree with this idea of 'lower-level' processing happening earlier in time than 'higher-level' processing, however (cf. also Hochstein & Ahissar, 2002), neither would the Gestaltists agree in my opinion. What it can indicate in my opinion, is different cortical networks in the brain, and future research can further explore this distinction.

When it comes to secondary Prägnanz tendencies, these can occur even under very strong visibility of the current stimulus, as they do indicate deviations on the phenomenal level and not necessarily deviations from the stimulus to the percept. This is potentially why both attractive and repulsive tendencies could be found in Chapter 6.

8

8.2.3 Distinguishing low visibility, uncertainty in feature, and ambiguity in feature

The research projects I undertook for my PhD show the importance of distinguishing reduced *visual* strength (i.e., Koffka's (1935) weak external conditions, due to limited viewing conditions), reduced *featural* strength (i.e., related to uncertainty in a feature part of the stimulus), and *ambiguity* in how to organize incoming stimulation (i.e., multistability; cf. also 'The Prägnanz framework and perception as Bayesian inference'). In an efficient Bayesian observer model, reduced visual strength is related to high internal sensory noise (i.e., leading to an asymmetric likelihood distribution in stimulus space), while reduced featural strength is related to high external stimulus noise (i.e., wider likelihood peaks in stimulus space). Ambiguity in a feature would show as multiple peaks in the likelihood, not wider peaks, hence not necessarily increasing external stimulus noise. Given that these different uncertainties can impact psychological organization differently, making these distinctions clear is strongly advisable in future research manipulating aspects of the stimulus.

8.2.4 An incomplete exploration

The research I present in this dissertation does not provide a complete exploration of the ways in which Prägnanz may be used as an inspiration for current research. Importantly, that was also not the purpose of this project. What I do provide is some case studies, that illustrate some of the options, but certainly not all. I only looked at some stimulus manipulations, some Prägnanz aspects, some limitations on the viewing conditions. Prägnanz provides a framework that pays special attention to any constraints on generality: what the best overall organization will be depends on input, person, context, and all their interactions. The studies in this dissertation provide first steps in specifying some of these constraints on generality, but none of the specified constraints on generality is complete. By showing some examples across a wide range of different research topics, I hope to inspire future research that takes a more systematic approach in exploring several Prägnanz aspects, several options within each Prägnanz aspect, primary and secondary Prägnanz tendencies, and manipulations of internal and external conditions, as well as a further specification of these constraints on generality that make the Prägnanz framework worthwhile. Importantly, the effects of these constraints are in no sense random, we only need to organize all findings in a coherent whole to come to the knowledge of a system (cf. also Koffka, 1935).

Given that determining the best organization is dependent on input, person, context, and their interactions, one could try to get a better idea of which of these factors should be given most weight (i.e., input, person, context, or interactions). How strong each of these factors will be depends on the total combination of factors, however. As one example, in a series of studies on spontaneous perceptual dynamics when perceiving an ambiguous motion quartet, Boeykens et al. (2021) found that individual differences in perceptual bias can be strongly constrained by reducing the temporal distance between displays. When temporal distance between the displays was longer, individual differences became more influential. As another example, the follow-up analysis on attractive and repulsive immediate temporal context effects using morph stimuli (Van Geert & Wagemans, in preparation a) showed these context effects to be present only when no strong internal reference points were available. Based on the same reasoning, Gestalt psychologists described their Gestalt principles as *ceteris paribus* principles: a Gestalt principle is supposed to hold within the constraints of the prevailing (internal and external) conditions (Wagemans, 2018). When the prevailing conditions change, other Gestalt principles could come into play and overrule the Gestalt principle in question. Hence, systematic investigations of the *interactions* between principles are needed.

8.2.5 Veridicality and Prägnanz

Whereas some researchers believed perception to be aimed directly at achieving a sufficiently veridical representation of the physical world — and this is actually still the mainstream view in current vision science (Koenderink, 2015) — the Gestalt theorists did not view veridicality as a direct aim of perception. There is a correspondence between experienced and physical world, but this correspondence results from the presence of similar natural, autochthonous principles guiding spontaneous self-organization in both the experienced and the physical world (Bischof, 1966; Hüppe, 1984). As the course of spontaneous self-organization depends on the properties of the organism in which the organizational process takes place, it is impossible to predict phenomenal experience from the stimulus conditions only (Hüppe, 1984; Koffka, 1935). This dependence of the natural laws on the organism, however, does not make perceptual organization of the visual field a matter of personal choice: a particular grouping and segregation is perceived beyond personal control (Wertheimer, 1923; Wertheimer et al., 2012).

Koenderink (2015) summarized the mainstream view as follows: they indicate that the world contains all structure, that science's task is to uncover that structure, and that perception is veridical to the extent that it represents that structure. Nevertheless, actually every individual adds structure to construct his/her own experienced reality ('esse est percipi'), and the higher goal of the organism is biological fitness, not veridicality ('verum factum est,' Koenderink, 2015). Koenderink's view is similar in some respects to the view of Gestalt theory, denying veridicality as a direct aim of perception. Koenderink (2019) describes perception as an idiosyncratic user interface, which is — through evolutionary tendencies — optimized for utility, not veridicality. This utility perspective comes back in Koenderink's description of the two main aspects of Prägnanz (Koenderink et al., 2018), one related to the organism's capacity limit (i.e., 'structural complexity bottleneck'), and a second one related to biological fitness. That is, as long as new complexities can be processed and organized, they may contribute to Prägnanz and biological relevance. When they exceed the organism's capacity limit, they are not behaviorally useful in that context.

8.2.6 Behavioral relevance of Prägnanz

Although the Prägnanz tendency indicates that we will always organize the incoming stimulation in the best, clearest way possible given the prevailing conditions, the best Gestalt will not always be the most behaviorally relevant way of organizing the stimulation without keeping these prevailing conditions in mind. What the 'best' organization is behaviorally, strongly depends on the task for example. I do believe that Prägnanz tendencies are useful in the long term, however.

Koenderink et al. (2018) suggest a strong link between Prägnanz and biological relevance. More specifically, they posit the (potential) relevance to the organism's biological fitness (or the affinity to its optical user interface) as one of two main aspects part of Prägnanz[3]. They also refer to releasers (i.e., stimulus constellations that trigger a

[3] The other main aspect of Prägnanz in Koenderink et al. (2018) is the structural complexity bottleneck of visual systems.

fixed behavioral pattern in a particular species) in ethology as extreme examples of good Gestalts (Koenderink et al., 2018): these stimuli have a structure that represents their essence in the most pure and compelling way. Or more correctly, maybe not in the *most* pure and compelling way, but in a pure and compelling *enough* way. In ethology, also supernormal stimuli are discussed (Tinbergen, 1951): non-natural stimuli that can have an even stronger effect than the natural releasers as they represent the essence even more purely and compellingly. In the specific case of supernormal stimuli, Prägnanz may thus show to be destructive in cases where it leads to unproductive behavior that may cause the death of the individual or extinction of the species. In the more common case of releasers, the effect of Prägnanz is clearly beneficial, however.

The same idea, Prägnanz as a double-edged sword, being mostly beneficial but sometimes also destructive, comes back in other situations. More specific Prägnanz steps may lead to faster change detection (e.g., Archambault et al., 1999) and more specific target templates improve performance in visual search (e.g., Hout & Goldinger, 2015). Congruency between typical and perceptual size differences can decrease reaction times in visual search (e.g., Riou et al., 2011). Furthermore, representations can be highly accurate, even when local measurements are very noisy (Alvarez & Oliva, 2008), and can increase the accuracy with which items are stored in visual working memory (Brady & Alvarez, 2011). But on the other hand, the temporal context may bias memory for individual items (Brady & Alvarez, 2011), Kanizsa (1979) discussed Prägnanz as a potential obstacle to problem solving (Legrenzi, 1994), and inflexible precise priors in individuals with Autism Spectrum Disorder may impede generalization and learning of broader high-level abstractions (HIPPEA, Van de Cruys et al., 2014).

So, 'what good is goodness' (Garner, 1974)? Although Garner (1974) had a very specific view on Prägnanz (cf. Chapter 2) — with good patterns being invariant to as many rotations and reflections as possible, leading to a small subset size and high internal redundancy — he specified more clearly in which ways goodness can influence perceptual and cognitive processing. Regardless of the task, he found an effect of goodness on both encoding and memory generation, while there was no conclusive evidence for an effect on speed of comparison (Garner, 1974). In addition, Goetschalckx et al. (2019) found that memorability was higher for distinctive images (i.e., more clearly different from an internal reference; better memory generation) and that images that could be categorized more accurately (i.e., more similar to an internal reference; easier encoding) when distinctiveness was controlled for. In that sense, both attractive and repulsive Prägnanz tendencies seemed to improve performance in a memory task in different ways.

8.2.7 The Prägnanz framework and perception as Bayesian inference

Here I dig deeper in the connections I see between Gestalt and Bayesian views on perception, specifically focusing on the efficient Bayesian observer model as presented by Wei & Stocker (2015). Typically, 'simplicity' and 'likelihood' have been presented as contrasting principles, while overlap in ideas and predictions following from these principles seems to have been largely ignored. Therefore, I find it important to stress that there are important parallels and similarities to draw, and that both views on perception can support each other in many ways. Given that both Bayesian and Gestalt psychological views posit a connection to regularities in the real world — be it directly by proposing veridicality or indirectly by proposing parallellism — it is understandable that both views will lead to similar results in many cases.

The core of a Bayesian view on perception is the combination of current sensory evidence (resulting in a *likelihood*) and prior expectations about its frequency of occurrence (i.e., a *prior* probability). Also in Prägnanz, both external (stimulus-dependent) and internal (viewer-dependent) conditions codetermine the organization that will be experienced.

A core idea of the efficient Bayesian observer model is that neural resources are limited, and therefore *efficient coding* will occur, optimizing sensory representations relative to the stimulus statistics of the natural environment (Wei & Stocker, 2015). In other words, encoding precision of feature values will be variable in line with their frequency of occurrence in the natural environment. For this idea, Wei & Stocker (2015) referred back to Attneave (1954) and his work on redundancy in visual perception. In this paper, Attneave (1954) defined a good Gestalt as a figure with a high degree of internal redundancy. Although I do not agree with how Attneave (1954) equated Prägnanz with a minimum principle, the efficient coding idea in general certainly represents an important aspect of Prägnanz. As mentioned in Chapter 2, Köhler (1920) focused on the equivalence of the Prägnanz tendency to the physical tendency towards minimal structural energy, with minimal structural energy being attained when in a stable, stationary state. Koffka (1935) also described the effect of the energy level of an organism on which Prägnanz tendencies will occur, and Metzger (1941) discussed how individual differences in for example comprehension capacity and energy level may influence whether conflicting Gestalt factors will lead to an unclear percept or a richer organization. The idea of a constraint on encoding resources also comes back as one aspect of Koenderink's (2018) description of Prägnanz (i.e., a structural complexity bottleneck). This upper boundary on the level of structural complexity that can be encoded is somehow similar to the upper boundary put on encoding capacity in the efficient Bayesian observer model (Wei & Stocker, 2015). Noel et al. (2021) also found empirical evidence for differences in visual encoding capacity, operationalized using Fisher information. Whereas Gestalt theory thus specified the tendency towards minimal structural energy, the efficient Bayesian observer model provides a more specific approach to implement this tendency. However, in the Bayesian view it does not necessarily concern *structural* complexity, and Gestalt theory would probably say the coding is optimized in relation to the organism rather than to the stimulus statistics in the natural environment (see below).

In the efficient Bayesian observer model (Wei & Stocker, 2015), two types of noise are distinguished, with differential consequences for perception. High *internal* sensory noise Huang (2022) leads to an asymmetric likelihood distribution in the stimulus space and potentially biases away from the peak of the prior distribution. When Koffka (1935) referred to weak external conditions (i.e., low visibility, due to, e.g., brief presentation, low contrast, small size), he referred to conditions that are similar to those mentioned to lead to high internal sensory noise. High *external* stimulus noise Huang (2022) only widens the likelihood distribution and leads to a stronger attractive influence of the prior distribution.

From the efficient Bayesian observer model proposed by Wei & Stocker (2015), Wei & Stocker (2017) derived a lawful relation between perceptual bias and discriminability: perceptual bias is proportional to the slope of the square of the discrimination threshold. This relation implies that the perceptual bias will be zero for locations on the dimensions for which discrimination sensitivity is maximal or minimal, while the bias will be largest for locations on the dimensions where the discrimination sensitivity changes most quickly (Wei & Stocker, 2017). In Prägnanz terms: Prägnanz tendencies will be absent at internal reference points (i.e., where discrimination is maximal) and exactly in between two reference points (i.e., where discrimination sensitivity is minimal), while the largest Prägnanz tendencies will happen in between those two extremes (i.e., closer to one of the reference points than the other).

Nevertheless, the presented Prägnanz framework is not just a rephrasing of perception as Bayesian inference; important differences remain. One of the main distinctions I see is the way in which Prägnanz steps and Prägnanz functions — or priors in Bayesian terms — are formed. Bayesian inference bets strongly on 'previous experience'. Gestalt theory is not in contrast to any influence from learning or previous experience, rather to the contrary. For example, Wertheimer (1923) explicitly described the Gestalt factor of objective set [*Einstellung*], referring to the influence of the immediate temporal of spatial context on perception, as a strong one that needs to be considered very carefully in experiments. Furthermore, also familiarity or past experience — here meant as only based on arbitrary habit or drill, not related to the content or any aspect of the specific configuration — is mentioned as a separate Gestalt principle (Wertheimer, 1923). In addition, Koffka (1935) makes note of the fact that the structure of the human nervous system is not only influenced by innate factors but also by previous experience. The main difference here between Gestalt theory and Bayesian inference is that in Gestalt theory, 'previous experience' is not viewed as a satisfying explanation for all questions about human perceptual organization, and emphasizes the importance of more general (biologically useful, utility-related) principles of organization next to the influences of previous experience. If one would want to attribute all Gestalt principles to prior experience, one must be able to show concretely based on which prior experience these principles are based, and also that there is no basis in prior experience for alternative principles and organizations (Wertheimer, 1923). Put differently, previous experience should not be used as a magical one-fits-all solution either. In any case, how exactly internal representations are formed and used as a reference to compare to the incoming stimulation, is an interesting area for further research, regardless of whether one wants to study it from a Gestalt or a Bayesian perspective.

In my opinion, Gestalt theory thus provides a broader range of influencing factors than Bayesian inference. In the end you need to come to a 'prior distribution', which is easy when only one dimension is varied, but more difficult when multiple interacting factors are at play. Gestalt theory gives a broader insight in different ways in which reference regions and prior probabilities could arise. Experienced organizations have many facets, and in many cases the 'prior' or 'internal reference' does not concern a single one-dimensional probability distribution. Gestalt psychology raised many factors (i.e., different Gestalt laws, different Prägnanz aspects related to both structure and meaning) that could influence this multivariate prior, with the potential effects of previous experience as one of them. I believe that it is not about how many times a specific stimulus value has been presented, but more about the extent to which a value is probable given all presented stimulus values on that dimension (cf. ensemble and distribution learning, e.g., Corbett et al., 2023; Khayat & Hochstein, 2019). Also, new Prägnanz steps are formed when they are no longer seen as a disruption of an existing Prägnanz step and transform in a derived one, with its own right of existence (when it becomes *useful* to form a new category, which is related to the function of a percept in the organism's life world, behavioral context and action tendencies of the organism). In other words, does a certain organization differ enough from the pre-existing reference point that it requires differential action/experience of the organism? Also utility could play a role here: for instance, perceiving a wolf as a dog has more detrimental consequences than perceiving a dog as a wolf, and therefore, perception can take these utility functions into account in its perceptual organizational principles. Even previous experience itself can entail many factors (e.g., longer- and shorter-term temporal context, concurrent spatial context, also different features of the context can be taken into account), and it is not specified how the Bayesian perspective disentangles those or how they are weighted.

It has to be said that when it comes to quantifying their vision, models of Bayesian inference have been further elaborated than those of Gestalt. On the other hand, Prägnanz does not necessarily concern probabilities in the real world, although the subjective probabilities that Prägnanz does take into account may be related to them. Bayesian perspectives often start from real-world probabilities, but as Chapter 4 indicates, also perceptual interpretations of the previously presented stimulation need to be taken into account. Furthermore, the subjective probability may be different based on other internal conditions, like the interest in the occurrence of a certain experienced organization. Therefore, I believe that the Bayesian modeling perspective could become a more fruitful approach to perception, by taking a broader perspective on factors influencing probability (i.e., not only aspects the stimulus) based on the Prägnanz framework.

Compared to the typical Bayesian framework, Gestalt psychology is also different in that not only attractive tendencies (i.e., *towards* a reference), but also repulsive tendencies (i.e., *away from* a reference) can occur. This difference is resolved when using an efficient Bayesian observer model (Wei & Stocker, 2015), which does allow for both attraction towards and repulsion from a prior peak (Wei & Stocker, 2015; cf. also Chapter 5 and Hahn & Wei, 2022).

8.3 Future directions

8.3.1 Further specifying the prevailing conditions

The Prägnanz principle entails that we will always organize the incoming stimulation in the best way possible given the prevailing conditions. Furthermore, both simplification and complication tendencies can occur, and also here, which tendency will occur will depend on the prevailing conditions. Therefore, one important goal for future research is further specifying these prevailing conditions. Which factors certainly need to be taken into account to determine which tendencies can be expected? In my own work, I investigated some specific combinations of factors, and touched upon the importance of the visibility of the external stimulus, the strength of the internal reference points, the immediate temporal and spatial context, and the energy level of the organism. Nevertheless, further research should aim for a more systematic analysis of the factors that may influence which tendencies will occur or how they will show. In addition to further investigating and specifying the prevailing conditions to take into account, also relating these conditions to different types of uncertainty — internal sensory noise (related to limited visibility), external stimulus noise (related to uncertainty in a stimulus feature), and multistability in a feature — seems a promising avenue to obtain a more complete understanding of the impact of different factors on diverse Prägnanz tendencies. Relevant in this context is also the work of Huang (2015, 2022), who next to featural strength and visual strength also defines spatial strength as a general factor in visual processing (although ambiguity is not taken into account in Huang's framework). One way to approach this systematic specification, is to conduct studies with a large set of manipulations and tasks for one and the same stimulus type, and later aim to generalize the findings for this stimulus type to other stimuli. This generalization will not always occur, however.

8.3.2 Cross-fertilization across fields through conceptual integration

In research on perception and cognition, many different terms are used for rather similar effects. It is not always clear whether it concerns effects or processes that are actually identical, or only similar on a superficial level. For example, this is a non-exhaustive list of terms used to indicate attractive changes in perception: attractive effects, serial dependencies, serial effects, sequential effects, history effects, priming, integration, recency bias, perceptual stability, facilitation, stabilization, attraction, assimilation, contraction, magnet effect, robustness, simplification, leveling, tension-reducing tendencies, minimum simplicity, central tendency bias, categorization. For repulsive effects on perception, the following terms are in use: contrastive effects, repulsion, adaptation, habituation, contrast, perceptual warping, differentiation, sensitivity, anchor effect, complication, articulation, sharpening. I think a rationalization in the number of useful theoretical concepts, and thereby clarifying the relationships between these terms, will largely benefit the research community, as in that way also communication and cross-fertilization across research fields using different terminologies becomes more feasible.

8.3.3 Enrich interaction across fields through quantitative modeling

Next to qualitative clarification in the concepts we employ as one way to enrich communication across fields, another avenue I see is the development of general quantitative modeling approaches than can be applied across diverse sets of tasks and paradigms. One such promising modeling approach is the category of efficient Bayesian observer models (Hahn & Wei, 2022; Mao & Stocker, 2022; Ni & Stocker, 2023; Wei & Stocker, 2015; Wei & Stocker, 2017), provided that a broader view is taken on the factors influencing probability than is typically done (cf. the section on 'The Prägnanz framework and perception as Bayesian inference'). Recent extensions of the original efficient Bayesian observer model formulation have already broadened its original field of application extensively (Hahn & Wei, 2022; Mao & Stocker, 2022; Ni & Stocker, 2023). For the future, it is important to provide accessible introductions to the modeling approach, to use the model as a starting point for better understanding similarities and differences in psychological processes underlying task behavior across research fields, and to show the model's value across an even wider variety of tasks, paradigms, and research fields. Especially multidimensional extensions (e.g., for tasks in which more than one stimulus feature is varied) will prove important to further extend the model's field of application. As part of my postdoctoral project, I will aim to develop a hierarchical efficient Bayesian observer model, with the goal of advancing theoretical understanding of universality and diversity across conditions (i.e., stimuli, individuals, contexts) in perception and appreciation, after first establishing robust similarities and differences across conditions that the model should be able to explain[4].

[4] For more information, please contact Eline Van Geert.

8

8.3.4 How are Prägnanz steps formed?

One important point for further research is clarifying how Prägnanz steps (i.e., internal reference regions), and by extension Prägnanz functions — or prior distributions in Bayesian terms — are formed. Can the formation of these Prägnanz distributions be reduced to past experience alone, or is there more to it? Given that this aspect is one of the crucial points of disagreement between Gestalt and Bayesian views, concrete research on how these templates are formed, and how they temporally change, is valuable for better understanding the perceptual process in general.

8.3.5 Relation between Prägnanz and appreciation

This dissertation suggests two different ways in which Prägnanz and appreciation can be related. On the one hand, appreciation can be associated with the *absolute* level of Prägnanz (i.e., Prägnanz height), on the other hand it can be associated with the *relative* increase in Prägnanz experienced (i.e., strength of the experienced Prägnanz tendency). Both routes towards appreciation may co-exist, however. In their pleasure-interest model of aesthetic liking, Graf and Landwehr (2015, 2017) proposed such a view encompassing both perspectives. Further uncovering the relative importance of both routes under diverse conditions, and also investigating different Prägnanz aspects and their relation to aesthetics, seem interesting ways forward for the future.

8.4 Reflections on my approach

In my doctoral research, I aimed to combine a strong theoretical basis and a wish for theoretical integration with conducting open and reproducible research. On the open, reproducible research practices side, I created open source software (i.e., OCTA); practiced reproducible reporting for all manuscripts included in this dissertation as well as the dissertation as a whole (using papaja, R Markdown, and Quarto); provided open data, materials, and code (on the Open Science Framework); used Bayesian statistics; wrote one manuscript as a Registered Report; provided interactive graphs where deemed useful; and posted preprints to allow early and open access to my manuscripts. Although I found this way of working sometimes challenging, and I am still learning and improving my workflow, I strongly believe that these practices are worthwhile to enable other researchers to better and more completely understand and evaluate my work, and to use my data, materials, and/or analysis approach in future research projects. Furthermore, I hope that by applying these practices, I can inspire other researchers to try out some of these practices themselves. Moreover, not only others can benefit from this approach: when data requests or other questions related to my earlier research come, or when I want to reuse my own data, materials, or code, this openness will also make it easier for my future self.

In addition to this focus on openness and reproducibility, I also find it important to start my work from a theoretical basis when possible. In this dissertation, the Gestalt theoretical concept of Prägnanz provides this overall framework. Rather than focusing on theory testing, I focused on using this theoretical framework as a tool for developing new empirical research. Theory can thus inspire which questions are worth answering, and can also provide a tool for integrating research findings that seem diverse at first sight. Especially this theoretical integration is often lacking in current research, although the problem was already known in earlier decades. Newell (1973, p. 16) put it this way: “We never seem in the experimental literature to put the results of all the experiments together.”

In this dissertation, I aimed to provide some first steps in the direction of theoretical integration, in diverse ways. Firstly, I aimed to bring the study of attractive and repulsive effects together, as well as the study of order and complexity. Also, I aimed to bring together effects studied in different fields, like longer-term effects based on categorization and more short-term immediate context effects. Secondly, I tested the generalizability of at least some findings in the dissertation across tasks, across stimuli, across individuals, and/or across time. Thirdly, I aimed to provide a contribution to extending the efficient Bayesian observer model into a processing model that can include and predict many different effects. One other way to stimulate theoretical integration that was not part of this dissertation, is systematically testing different manipulations on one single task and stimulus type, and only then generalizing to other tasks and stimulus types (Newell, 1973, cf. 'Future directions').

8.5 Concluding remarks

I hope that, with this doctoral dissertation, I was able to show that Gestalt psychology, and in my case the concept of Prägnanz in particular, can support and guide the generation of new research questions in the field of perceptual organization, aesthetic appreciation, and beyond. Prägnanz was really meant as a general concept, not specifically tied to simple stimuli or human visual perception. This generality does not necessarily imply vagueness, rather it implies multifacetedness. And similarly, multifacetedness does not imply randomness, but rather a strong dependence on internal (organism-dependent) and external (input-dependent) conditions. To come to clearer as well as aesthetically appreciated psychological organizations, tendencies are at work that are both antagonistic and complementary: although they tend to decrease each other's influence, they can also work together towards a better psychological organization. What the optimal balance of both tendencies entails exactly will depend on the input the individual receives, the individual in question, as well as the context in which the individual receives the input. Further specifying how input, person, context, and all their interactions combine in influencing which Prägnanz tendencies will occur requires further research. This dissertation at least provides some case studies of how the Prägnanz framework can inspire concrete research into the influences of these conditions on human visual perception and aesthetic appreciation. Moreover, it aims to serve as an example of how Prägnanz can support the integration of research findings across diverse fields of study into one coherent framework, and thereby bring us one step closer to understanding the human psychological system (cf. also Koffka, 1935).

Not only in our perception, but also in our role as researchers, we are balancing new results and findings with our already existing worldview. As a consequence, some may find this dissertation too qualitative and dreamy because of its reference to Prägnanz, others may find the methods too simplistic and linear to match complex Gestalt theoretical ideas. Different researchers may have different ideas of how science should be conducted, but at least I hope that this dissertation may have provided some new external stimulation, and hopefully it can be well integrated into the readers' knowledge system and lead to a prägnant Gestalt.

References

Abrahamyan, A., Silva, L. L., Dakin, S. C., Carandini, M., & Gardner, J. L. (2016). Adaptable history biases in human perceptual decisions. *Proceedings of the National Academy of Sciences*, *113*(25), E3548–E3557. https://doi.org/10.1073/pnas.1518786113

Acker, B. E., Pastore, R. E., & Hall, M. D. (1995). Within-category discrimination of musical chords: Perceptual magnet or anchor? *Perception & Psychophysics*, *57*(6), 863–874. https://doi.org/10.3758/bf03206801

Aguilar, D. (2021). *Jsonpickle* (Version 2.0.0) [Computer software]. https://github.com/jsonpickle/jsonpickle

Allaire, J., Ushey, K., Tang, Y., & Eddelbuettel, D. (2017). *Reticulate: R interface to Python*. https://github.com/rstudio/reticulate

Allport, G. W. (1930). Change and decay in the visual memory image. *British Journal of Psychology*, *21*(2), 133–148. https://doi.org/10.1111/j.2044-8295.1930.tb00581.x

Alp, N., Kohler, P. J., Kogo, N., Wagemans, J., & Norcia, A. M. (2018). Measuring integration processes in visual symmetry with frequency-tagged EEG. *Scientific Reports*, *8*(1), 6969. https://doi.org/10.1038/s41598-018-24513-w

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*(4), 392–398. https://doi.org/10.1111/j.1467-9280.2008.02098.x

Alvarez, L., Gousseau, Y., Morel, J.-M., & Salgado, A. (2015). Exploring the space of abstract textures by principles and random sampling. *Journal of Mathematical Imaging and Vision*, *53*(3), 332–345. https://doi.org/10.1007/s10851-015-0582-z

Alvarez, L., Monzón, N., & Morel, J.-M. (2021). Interactive design of random aesthetic abstract textures by composition principles. *Leonardo*, *54*(2), 179–184. https://doi.org/10.1162/leon_a_01768

Analytics, R., & Weston, S. (2020). *Iterators: Provides iterator construct*. https://CRAN.R-project.org/package=iterators

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. *Psychological Bulletin*, *78*(4), 266–278. https://doi.org/10.1037/h0033117

Archambault, A., O'Donnell, C., & Schyns, P. G. (1999). Blind to object changes: When learning the same object at different levels of categorization modifies its perception. *Psychological Science*, *10*(3), 249–255. https://doi.org/10.1111/1467-9280.00145

Arnheim, R. (1971). *Entropy and art: An essay on disorder and order*. University of California Press. https://books.google.be/books?id=gpTpAAAAMAAJ

Arnheim, R. (1974). *Art and visual perception*. University of California Press.

Arnheim, R. (1975). *Anwendungen gestalttheoretischer Prinzipien auf die Kunst* [Applications of Gestalt theoretical principles to art]. In S. Ertel, L. Kemmler, & M. Stadler (Eds.), *Gestalttheorie in der modernen Psychologie [Gestalt theory in modern psychology]* (pp. 278–284). Steinkopff. https://doi.org/10.1007/978-3-642-72312-4_28

Arnheim, R. (1986). The two faces of Gestalt psychology. *American Psychologist*, *41*(7), 820–824. https://doi.org/10.1037/0003-066X.41.7.820

Arnheim, R. (1987). Prägnanz and its discontents. *Gestalt Theory*, *9*(2), 102–107.

Arnold, J. B. (2021). *Ggthemes: Extra themes, scales and geoms for 'ggplot2'*. https://CRAN.R-project.org/package=ggthemes

Arnoult, M. D. (1960). Prediction of perceptual responses from structural characteristics of the stimulus. *Perceptual and Motor Skills*, *11*(3), 261–268. https://doi.org/10.2466/pms.1960.11.3.261

Ash, M. G. (1995). *Gestalt psychology in German culture, 1890–1967: Holism and the quest for objectivity*. Cambridge University Press.

Ashourian, P., & Loewenstein, Y. (2011). Bayesian inference underlies the contraction bias in delayed comparison tasks. *PLoS ONE*, *6*(5), e19551. https://doi.org/10.1371/journal.pone.0019551

Attneave, F. (1955). Symmetry, information, and memory for patterns. *The American Journal of Psychology*, *68*(2), 209–222. https://doi.org/10.2307/1418892

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*(3), 183–193. https://doi.org/10.1037/h0054663

Attneave, F. (1957). Physical determinants of the judged complexity of shapes. *Journal of Experimental Psychology*, *53*(4), 221–227. https://doi.org/10.1037/h0043921

Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin*, *53*(6), 452–471. https://doi.org/10.1037/h0044049

Aust, F., & Barth, M. (2022). *papaja: Prepare reproducible APA journal articles with R Markdown*. https://github.com/crsh/papaja

Barth, M. (2022). *tinylabels: Lightweight variable labels*. https://cran.r-project.org/package=tinylabels

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Bates, D., & Maechler, M. (2021). *Matrix: Sparse and dense matrix classes and methods*. https://CRAN.R-project.org/package=Matrix

Berlyne, D. E. (Ed.). (1960). *Conflict, arousal and curiosity*. McGraw-Hill. https://doi.org/10.1037/11164-000

Berlyne, D. E. (Ed.). (1974). *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation*. Hemisphere.

Bertamini, M., & Rampone, G. (2020). The study of symmetry in empirical aesthetics. In M. Nadal & O. Vartanian (Eds.), *The Oxford Handbook of Empirical Aesthetics*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198824350.013.23

Best, R. M., & Goldstone, R. L. (2019). Bias to (and away from) the extreme: Comparing two models of categorical perception effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(7), 1166–1176. https://doi.org/10.1037/xlm0000609

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115–147. https://doi.org/10.1037/0033-295X.94.2.115

Bies, A. J., Blanc-Goldhammer, D. R., Boydston, C. R., Taylor, R. P., & Sereno, M. E. (2016). Aesthetic responses to exact fractals driven by physical complexity. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00210

Bischof, N. (1966). *Erkenntnistheoretische Grundlagenprobleme der Wahrnemungspsychologie* [Basic epistemological problems of the psychology of perception]. In W. Metzger, R. Bergius, & H. Thomae (Eds.), *Allgemeine Psychologie [General psychology]*. Hogrefe.

Blake, A. B., Nazarian, M., & Castel, A. D. (2015). The Apple of the mind's eye: Everyday attention, metamemory, and reconstructive memory for the Apple logo. *The Quarterly Journal of Experimental Psychology*, *68*(5), 858–865. https://doi.org/10.1080/17470218.2014.1002798

Bock, H., & Pfeiffer, T. (1987). *Prototypikalität von Bedeutungsvarianten des Verbs "überholen" im Lichte der gestalttheoretischen Bezugssystemlehre* [Prototypicality of meaning variants of the verb "to overtake" in the light of Gestalt theoretical frame of reference theory]. *Gestalt Theory*, *9*(1), 3–16.

Boeykens, C., Wagemans, J., & Moors, P. (2021). Perception of the ambiguous motion quartet: A stimulus-observer interaction approach. *Journal of Vision*, *21*(13), 12. https://doi.org/10.1167/jov.21.13.12

Bosch, E., Fritsche, M., Ehinger, B. V., & de Lange, F. P. (2020). Opposite effects of choice history and evidence history resolve a paradox of sequential choice bias. *Journal of Vision*, *20*(12), 9–9. https://doi.org/10.1167/jov.20.12.9

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, *22*(3), 384–392. https://doi.org/10.1177/0956797610397956

Brascamp, J. W., Knapen, T. H. J., Kanai, R., Noest, A. J., van Ee, R., & van den Berg, A. V. (2008). Multi-timescale perceptual history resolves visual ambiguity. *PLOS ONE*, *3*(1), e1497. https://doi.org/10.1371/journal.pone.0001497

Braun, J., Amirshahi, S. A., Denzler, J., & Redies, C. (2013). Statistical image properties of print advertisements, visual artworks and images of architecture. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00808

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, *13*(2), 207–230. https://doi.org/10.1016/0010-0285(81)90008-6

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software*, *100*(5), 1–54. https://doi.org/10.18637/jss.v100.i05

Burnett, H. G., & Jellema, T. (2013). (Re-)conceptualisation in Asperger's syndrome and typical individuals with varying degrees of autistic-like traits. *Journal of Autism and Developmental Disorders*, *43*(1), 211–223. https://doi.org/10.1007/s10803-012-1567-z

Caddigan, E., Choo, H., Fei-Fei, L., & Beck, D. M. (2017). Categorization influences detection: A perceptual advantage for representative exemplars of natural scene categories. *Journal of Vision*, *17*(1), 21. https://doi.org/10.1167/17.1.21

Campitelli, E. (2022). *Ggnewscale: Multiple fill and colour scales in 'ggplot2'*. https://CRAN.R-project.org/package=ggnewscale

Carmichael, L., Hogan, H. P., & Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology*, *15*(1), 73–86. https://doi.org/10.1037/h0072671

Carter, O., Snyder, J. S., Fung, S., & Rubin, N. (2014). Using ambiguous plaid stimuli to investigate the influence of immediate prior experience on perception. *Attention, Perception, & Psychophysics*, *76*(1), 133–147. https://doi.org/10.3758/s13414-013-0547-5

Chamberlain, R., & Wagemans, J. (2016). The genesis of errors in drawing. *Neuroscience & Biobehavioral Reviews*, *65*, 195–207. https://doi.org/10.1016/j.neubiorev.2016.04.002

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review*, *103*(3), 566–581.

Chaussé, P. (2010). Computing generalized method of moments and generalized empirical likelihood with R. *Journal of Statistical Software*, *34*(11), 1–35. https://doi.org/10.18637/jss.v034.i11

Checkosky, S. F., & Whitlock, D. (1973). Effects of pattern goodness on recognition time in a memory search task. *Journal of Experimental Psychology*, *100*(2), 341–348. https://doi.org/10.1037/h0035692

Cheng, J., Sievert, C., Schloerke, B., Chang, W., Xie, Y., & Allen, J. (2021). *Htmltools: Tools for HTML*. https://CRAN.R-project.org/package=htmltools

Chetverikov, A., & Kristjánsson, Á. (2016). On the joys of perceiving: Affect as feedback for perceptual predictions. *Acta Psychologica*, *169*, 1–10. https://doi.org/10.1016/j.actpsy.2016.05.005

Chipman, S. F. (1977). Complexity and structure in visual patterns. *Journal of Experimental Psychology: General*, *106*(3), 296–301.

Chipman, S. F., & Mendelson, M. J. (1979). Influence of six types of visual structure on complexity judgments in children and adults. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(2), 365–378.

Cicchini, G. M., Mikellidou, K., & Burr, D. (2017). Serial dependencies act directly on perception. *Journal of Vision*, *17*(14), 6. https://doi.org/10.1167/17.14.6

Claessens, P. M. E., & Wagemans, J. (2008). A Bayesian framework for cue integration in multistable grouping: Proximity, collinearity, and orientation priors in zigzag lattices. *Journal of Vision*, *8*(7), 33. https://doi.org/10.1167/8.7.33

Clarke, A. D. F., Green, P. R., Halley, F., & Chantler, M. J. (2011). Similar symmetries: The role of wallpaper groups in perceptual texture similarity. *Symmetry*, *3*(2), 246–264. https://doi.org/10.3390/sym3020246

Clement, D. E. (1964). Uncertainty and latency of verbal naming responses as correlates of pattern goodness. *Journal of Verbal Learning and Verbal Behavior*, *3*(2), 150–157. https://doi.org/10.1016/S0022-5371(64)80033-5

Clement, D. E., & Varnadoe, K. W. (1967). Pattern uncertainty and the discrimination of visual patterns. *Perception & Psychophysics*, *2*(9), 427–431. https://doi.org/10.3758/BF03208782

Cohen, D. J. (2005). Look little, look often: The influence of gaze frequency on drawing accuracy. *Perception & Psychophysics*, *67*(6), 997–1009. https://doi.org/10.3758/BF03193626

Cohen, D. J., & Bennett, S. (1997). Why can't most people draw what they see? *Journal of Experimental Psychology. Human Perception and Performance*, *23*(3), 609–621. https://doi.org/10.1037//0096-1523.23.3.609

Coppola, D. M., Purves, H. R., McCoy, A. N., & Purves, D. (1998). The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(7), 4002–4006. https://doi.org/10.1073/pnas.95.7.4002

Corbett, J. E., Utochkin, I., & Hochstein, S. (2023). The pervasiveness of ensemble perception: Not just your average review. *Elements in Perception*. https://doi.org/10.1017/9781009222716

Corporation, M., & Weston, S. (2020). *doParallel: Foreach parallel adaptor for the 'parallel' package*. https://CRAN.R-project.org/package=doParallel

Cupchik, G. C., & Berlyne, D. E. (1979). The perception of collative properties in visual stimuli. *Scandinavian Journal of Psychology*, *20*(1), 93–104. https://doi.org/10.1111/j.1467-9450.1979.tb00688.x

Curray, J. R. (1956). The analysis of two-dimensional orientation data. *The Journal of Geology*, *64*(2), 117–131.
de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y
Donderi, D. C. (2006). Visual complexity: A review. *Psychological Bulletin*, *132*(1), 73–97. https://doi.org/10.1037/0033-2909.132.1.73
Dunn, J. C. (1983). Spatial metrics of integral and separable dimensions. *Journal of Experimental Psychology*, *9*(2), 242–257. https://doi.org/10.1037/0096-1523.9.2.242
Eddelbuettel, D., & Balamuta, J. J. (2018). Extending extitR with extitC++: A Brief Introduction to extitRcpp. *The American Statistician*, *72*(1), 28–36. https://doi.org/10.1080/00031305.2017.1375990
Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, *40*(8), 1–18. https://doi.org/10.18637/jss.v040.i08
Ellis, W. D. (1938). *A source book of Gestalt psychology*. Routledge & Kegan Paul.
Eysenck, H. J. (1942). The experimental study of the 'good Gestalt'–a new approach. *Psychological Review*, *49*(4), 344–364. https://doi.org/10.1037/h0057013
Fan, J. E., Hawkins, R. D., Wu, M., & Goodman, N. D. (2020). Pragmatic inference and visual abstraction enable contextual flexibility during visual communication. *Computational Brain & Behavior*, *3*(1), 86–101. https://doi.org/10.1007/s42113-019-00058-7
Fan, J. E., Yamins, D. L. K., & Turk-Browne, N. B. (2018). Common object representations for visual production and recognition. *Cognitive Science*, *42*(8), 2670–2698. https://doi.org/10.1111/cogs.12676
Fechner, G. T. (1860). Elemente der Psychophysik *[Elements of psychophysics]*. Leipzig: Breitkopf und Härtel.
Fehrer, E. V. (1935). An investigation of the learning of visually perceived forms. *The American Journal of Psychology*, *47*(2), 187–221. https://doi.org/10.2307/1415826
Feldman, J. (2003). What is a visual object? *Trends in Cognitive Sciences*, *7*(6), 252–256. https://doi.org/10.1016/S1364-6613(03)00111-6
Feldman, J. (2000). Bias toward regular form in mental shape spaces. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 152–165. https://doi.org/10.1037/0096-1523.26.1.152
Feldman, J. (2021). Mutual information and categorical perception. *Psychological Science*, *32*(8), 1298–1310. https://doi.org/10.1177/0956797621996663
Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, *116*(4), 752–782. https://doi.org/10.1037/a0017196
Fritsche, M., Mostert, P., & de Lange, F. P. (2017). Opposite effects of recent history on perception and decision. *Current Biology: CB*, *27*(4), 590–595. https://doi.org/10.1016/j.cub.2017.01.006
Fritsche, M., Spaak, E., & de Lange, F. P. (2020). A Bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception. *eLife*, *9*, e55389. https://doi.org/10.7554/eLife.55389
Froyen, V., Feldman, J., & Singh, M. (2015). Bayesian hierarchical grouping: Perceptual grouping as mixture estimation. *Psychological Review*, *122*(4), 575–597. https://doi.org/10.1037/a0039540
Gallagher, G. K., & Benton, C. P. (2022). Stimulus uncertainty predicts serial dependence in orientation judgements. *Journal of Vision*, *22*(1), 6. https://doi.org/10.1167/jov.22.1.6
Garner, W. R. (1974). *The processing of information and structure*. Erlbaum.
Garner, W. R., & Clement, D. E. (1963). Goodness of pattern and pattern uncertainty. *Journal of Verbal Learning and Verbal Behavior*, *2*(5-6), 446–452. https://doi.org/10.1016/S0022-5371(63)80046-8
Gartus, A., & Leder, H. (2013). The small step toward asymmetry: Aesthetic judgment of broken symmetries. *I-Perception*, *4*(5), 361–364. https://doi.org/10.1068/i0588sas
Gati, I., & Tversky, A. (1982). Representations of qualitative and quantitative dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 325–340. https://doi.org/10.1037//0096-1523.8.2.325
Genz, A., & Bretz, F. (2009). *Computation of multivariate normal and t probabilities*. Springer-Verlag.
Gepshtein, S., & Kubovy, M. (2005). Stability and change in perception: Spatial organization in temporal context. *Experimental Brain Research*, *160*(4), 487–495. https://doi.org/10.1007/s00221-004-2038-3
Gherman, D. (2021). *Svglib* (Version 1.1.0) [Computer software]. https://github.com/deeplook/svglib
Gibson, J. J. (1929). The reproduction of visually perceived forms. *Journal of Experimental Psychology*, *12*(1), 1–39. https://doi.org/10.1037/h0072470
Gillebert, C. R., Op de Beeck, H. P., & Wagemans, J. (2009). The influence of categorisation on the perceived shape similarity of everyday objects. *Psychologica Belgica*, *48*(4), 261. https://doi.org/10.5334/pb-48-4-261

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*(7), 926–932. https://doi.org/10.1038/nn.2831

Glanzer, M., & Clark, W. H. (1963). Accuracy of perceptual recall: An analysis of organization. *Journal of Verbal Learning & Verbal Behavior*, *1*, 289–299. https://doi.org/10.1016/S0022-5371(63)80008-0

Goetschalckx, L., Moors, P., Vanmarcke, S., & Wagemans, J. (2019). Get the Picture? Goodness of Image Organization Contributes to Image Memorability. *Journal of Cognition*, *2*(1), 22. https://doi.org/10.5334/joc.80

Gohel, D., & Skintzos, P. (2021). *Ggiraph: Make 'ggplot2' graphics interactive*. https://CRAN.R-project.org/package=ggiraph

Goldmeier, E. (1937). *Über Ähnlichkeit bei gesehenen Figuren* [About similarity in seen figures]. *Psychologische Forschung*, *21*, 146–208.

Goldmeier, E. (1972). Similarity in visually perceived forms. *Psychological Issues*, *8*(1).

Goldmeier, E. (1982). *The memory trace: Its formation and its fate*. Erlbaum.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(1), 69–78. https://doi.org/10.1002/wcs.26

Gollwitzer, A., Marshall, J., Wang, Y., & Bargh, J. A. (2017). Relating pattern deviancy aversion to stigma and prejudice. *Nature Human Behaviour*, *1*(12), 920–927. https://doi.org/10.1038/s41562-017-0243-x

Graf, L. K. M., & Landwehr, J. R. (2015). A dual-process perspective on fluency-based aesthetics: The Pleasure-Interest Model of Aesthetic Liking. *Personality and Social Psychology Review*, *19*(4), 395–410. https://doi.org/10.1177/1088868315574978

Graf, L. K. M., & Landwehr, J. R. (2017). Aesthetic pleasure versus aesthetic interest: The two routes to aesthetic liking. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00015

Granit, A. R. (1922). A study on the perception of form. *British Journal of Psychology. General Section*, *12*(3), 223–247. https://doi.org/10.1111/j.2044-8295.1922.tb00055.x

Grebenkina, M., Brachmann, A., Bertamini, M., Kaduhm, A., & Redies, C. (2018). Edge-orientation entropy predicts preference for diverse types of man-made images. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00678

Grinband, J., Hirsch, J., & Ferrera, V. P. (2006). A neural representation of categorization uncertainty in the human brain. *Neuron*, *49*(5), 757–763. https://doi.org/10.1016/j.neuron.2006.01.032

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*(3), 1–25. https://www.jstatsoft.org/v40/i03/

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D. S., Forster, J. J., Wagenmakers, E.-J., & Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, *81*, 80–97. https://doi.org/10.1016/j.jmp.2017.09.005

Grünbaum, B., & Shephard, G. C. (1989). *Tilings and patterns*. W. H. Freeman; company.

Güçlütürk, Y., Jacobs, R. H. A. H., & van Lier, R. (2016). Liking versus complexity: Decomposing the inverted u-curve. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00112

Haaf, J. M., & Rouder, J. N. (2019). Some do and some don't? Accounting for variability of individual difference structures. *Psychonomic Bulletin & Review*, *26*(3), 772–789. https://doi.org/10.3758/s13423-018-1522-x

Hahn, M., & Wei, X.-X. (2022). *A unifying theory explains seemingly contradicting biases in perceptual estimation* [Preprint]. bioRxiv. https://doi.org/10.1101/2022.12.12.519538

Hamada, J., & Ishihara, T. (1988). Complexity and goodness of dot patterns varying in symmetry. *Psychological Research*, *50*(3), 155–161. https://doi.org/10.1007/BF00310176

Hammad, S., Juricevic, I., Rajani, S., & Kennedy, J. (2008). Angle illusion on a picture's surface. *Spatial Vision*, *21*(3-5), 451–462. https://doi.org/10.1163/156856808784532554

Hanley, J. R., & Roberson, D. (2011). Categorical perception effects reflect differences in typicality on within-category trials. *Psychonomic Bulletin & Review*, *18*(2), 355–363. https://doi.org/10.3758/s13423-010-0043-z

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In *Categorical perception: The groundwork of cognition*. Cambridge University Press.

Harnad, S. (2003). Categorical perception. *Encyclopedia of Cognitive Science*.

Hartendorp, M. O., Van der Stigchel, S., Burnett, H. G., Jellema, T., Eilers, P. H. C., & Postma, A. (2010). Categorical perception of morphed objects using a free-naming experiment. *Visual Cognition*, *18*(9), 1320–1347. https://doi.org/10.1080/13506285.2010.482774

Hellström, Å. (2007). Temporal asymmetry and "magnet effect"" in similarity and discrimination of prototypical and nonprototypical stimuli: Consequences of differential sensation weighting. *Fechner Day 2007: Proceedings of the 23rd Annual Meeting of the International Society for Psychophysics*, 283–288.
Hendrickx, M., & Wagemans, J. (1999). A critique of Leyton's theory of perception and cognition. Review of *Symmetry, Causality, Mind*, by Michael Leyton. *Journal of Mathematical Psychology*, *43*(2), 314–345. https://doi.org/10.1006/jmps.1998.1232
Henle, M. (1987). On breaking out of dichotomies. *Gestalt Theory*, *9*(3/4), 140–149.
Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools*. https://CRAN.R-project.org/package=purrr
Henry, L., Wickham, H., & Chang, W. (2020). *Ggstance: Horizontal 'ggplot2' components*. https://CRAN.R-project.org/package=ggstance
Hester, J., & Bryan, J. (2022). *Glue: Interpreted string literals*. https://CRAN.R-project.org/package=glue
Hochberg, J. (2003). Acts of perceptual inquiry: Problems for any stimulus-based simplicity theory. *Acta Psychologica*, *114*(3), 215–228. https://doi.org/10.1016/j.actpsy.2003.07.002
Hochberg, J. (1968). *Perception*. Prentice Hall.
Hochberg, J., & McAlister, E. (1953). A quantitative approach, to figural "goodness". *Journal of Experimental Psychology*, *46*(5), 361–364. https://doi.org/10.1037/h0055809
Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804. https://doi.org/10.1016/S0896-6273(02)01091-7
Hoffman, D. D. (2009). The interface theory of perception: Natural selection drives true perception to swift extinction. In S. J. Dickinson, A. Leonardis, B. Schiele, & M. J. Tarr (Eds.), *Object Categorization* (pp. 148–166). Cambridge University Press. https://doi.org/10.1017/CBO9780511635465.009
Hoffman, D. D., Singh, M., & Prakash, C. (2015). The interface theory of perception. *Psychonomic Bulletin & Review*, *22*(6), 1480–1506. https://doi.org/10.3758/s13423-015-0890-8
Holzman, P. S., & Gardner, R. W. (1960). Leveling-sharpening and memory organization. *Journal of Abnormal and Social Psychology*, *61*(2), 176–180. https://doi.org/10.1037/h0041295
Hout, M. C., & Goldinger, S. D. (2015). Target templates: The precision of mental representations affects attentional guidance and decision-making in visual search. *Attention, Perception, & Psychophysics*, *77*(1), 128–149. https://doi.org/10.3758/s13414-014-0764-6
Huang, L. (2015). Visual features: Featural strength and visual strength are two dissociable dimensions. *Scientific Reports*, *5*(1), 13769. https://doi.org/10.1038/srep13769
Huang, L. (2022). FVS 2.0: A unifying framework for understanding the factors of visual-attentional processing. *Psychological Review*, *129*(4), 696–731. https://doi.org/10.1037/rev0000314
Hubbell, M. B. (1940). Configurational properties considered 'good' by naive subjects. *The American Journal of Psychology*, *53*(1), 46. https://doi.org/10.2307/1415960
Hübner, R., & Fillinger, M. G. (2016). Comparison of objective measures for predicting perceptual balance and visual aesthetic preference. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00335
Hüppe, A. (1984). *Prägnanz - ein gestalttheoretischer Grundbegriff: Experimentelle Untersuchungen [Prägnanz - a basic concept in gestalt theory: Experimental investigations]*. Profil-Verlag.
Irwin, D. E., & Pachella, R. G. (1985). Effects of stimulus probability and visual similarity on stimulus encoding. *The American Journal of Psychology*, *98*(1), 85. https://doi.org/10.2307/1422769
Jacobsen, T., & Höfel, L. (2002). Aesthetic judgments of novel graphic patterns: Analyses of individual judgments. *Perceptual and Motor Skills*, *95*(3), 755–766. https://doi.org/10.2466/pms.2002.95.3.755
Jäkel, F., Singh, M., Wichmann, F. A., & Herzog, M. H. (2016). An overview of quantitative approaches in Gestalt perception. *Vision Research*, *126*, 3–8. https://doi.org/10.1016/j.visres.2016.06.004
Jastrow, J. (1899). The mind's eye. *Popular Science Monthly*, *54*, 299–312.
Kanai, R., & Rees, G. (2011). The structural basis of inter-individual differences in human behaviour and cognition. *Nature Reviews Neuroscience*, *12*(4), 231–242. https://doi.org/10.1038/nrn3000
Kanizsa, G. (1975). "Pragnanz" as an obstacle to problem-solving. *Giornale Italiano Di Psicologia*, *2*, 417–425.
Kanizsa, G. (1979). *Organization in vision: Essays on Gestalt perception*. Praeger.
Kanizsa, G., & Luccio, R. (1986). *Die Doppeldeutigkeiten der Prägnanz* [The ambiguities of Prägnanz]. *Gestalt Theory*, *8*, 99–135.
Kay, M. (2021a). *ggdist: Visualizations of distributions and uncertainty*. https://doi.org/10.5281/zenodo.3879620
Kay, M. (2021b). *tidybayes: Tidy data and geoms for Bayesian models*. https://doi.org/10.5281/zenodo.1308151

Kayaert, G., Op de Beeck, H. P., & Wagemans, J. (2011). Dynamic prototypicality effects in visual search. *Journal of Experimental Psychology. General*, *140*(3), 506–519. https://doi.org/10.1037/a0023494

Kennedy, G. J., Orbach, H. S., & Loffler, G. (2008). Global shape versus local feature: An angle illusion. *Vision Research*, *48*(11), 1281–1289. https://doi.org/10.1016/j.visres.2008.03.003

Khayat, N., & Hochstein, S. (2019). Relating categorization to set summary statistics perception. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-019-01792-7

Kiyonaga, A., Scimeca, J. M., Bliss, D. P., & Whitney, D. (2017). Serial dependence across perception, attention, and memory. *Trends in Cognitive Sciences*, *21*(7), 493–497. https://doi.org/10.1016/j.tics.2017.04.011

Koenderink, J. (2014). The All Seeing Eye? *Perception*, *43*(1), 1–6. https://doi.org/10.1068/p4301ed

Koenderink, J. (2015). Esse est percipi & verum factum est. *Psychonomic Bulletin & Review*, *22*(6), 1530–1534. https://doi.org/10.3758/s13423-014-0754-7

Koenderink, J. (2019). Vision, an optical user interface. *Perception*, 0301006619853758. https://doi.org/10.1177/0301006619853758

Koenderink, J., van Doorn, A., & Pinna, B. (2018). Measures of Prägnanz? *Gestalt Theory*, *40*, 7–28. https://doi.org/10.2478/gth-2018-0002

Koffka, K. (1935). *Principles of Gestalt psychology*. Harcourt, Brace.

Koffka, K. (1940). Problems in the psychology of art. In R. Bernheimer (Ed.), *Art: A Bryn Mawr symposium* (pp. 180–273). New York: Oriole Editions.

Kohler, P. J., Clarke, A., Yakovleva, A., Liu, Y., & Norcia, A. M. (2016). Representation of maximally regular textures in human visual cortex. *The Journal of Neuroscience*, *36*(3), 714–729. https://doi.org/10.1523/JNEUROSCI.2962-15.2016

Köhler, W. (1920). *Die physischen Gestalten in Ruhe und im stationären Zustand [The physical Gestalten at rest and in stationary state]*. Friedr. Vieweg & Sohn.

Köhler, W. (1940). *Dynamics in psychology*. Liveright.

Köhler, W. (1993). Letter to Abraham S. Luchins (December 6, 1951). "... The principle of Prägnanz is probably in need of a revised formulation...". *Gestalt Theory*, *15*(3–4), 297–298. (Original work published 1951)

Kondo, A., Murai, Y., & Whitney, D. (2022). The test-retest reliability and spatial tuning of serial dependence in orientation perception. *Journal of Vision*, *22*(4), 5. https://doi.org/10.1167/jov.22.4.5

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language*, *37*(4), 555–583. https://doi.org/10.1006/jmla.1997.2529

Krakowski, C.-S., Poirel, N., Vidal, J., Roëll, M., Pineau, A., Borst, G., & Houdé, O. (2016). The forest, the trees, and the leaves: Differences of processing across development. *Developmental Psychology*, *52*(8), 1262–1272. https://doi.org/10.1037/dev0000138

Kruse, P. (1986). *Wie unabhängig ist das Wahrnehmungsobjekt vom Prozeß der Identifikation: Ein Kommentar zu G. Kanizsa und R. Luccio* [How independent is the perceptual object from the process of identification: A comment on G. Kanizsa and R. Luccio]. *Gestalt Theory*, *8*(2), 141–143.

Kubilius, J., Sleurs, C., & Wagemans, J. (2017). Sensitivity to nonaccidental configurations of two-line stimuli. *I-Perception*, *8*(2). https://doi.org/10.1177/2041669517699628

Kubilius, J., Wagemans, J., & Op de Beeck, H. P. (2014). Encoding of configural regularity in the human visual system. *Journal of Vision*, *14*(9), 11. https://doi.org/10.1167/14.9.11

Kubovy, M., & Berg, M. (2002). Oblique effects in grouping: Surprising individual differences. *Journal of Vision*, *2*(7), 480–480. https://doi.org/10.1167/2.7.480

Kubovy, M., Holcombe, A. O., & Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology*, *35*(1), 71–98. https://doi.org/10.1006/cogp.1997.0673

Kubovy, M., & van den Berg, M. (2008). The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychological Review*, *115*(1), 131–154. https://doi.org/10.1037/0033-295X.115.1.131

Kubovy, M., & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative Gestalt theory. *Psychological Science*, *6*(4), 225–234. https://doi.org/10.1111/j.1467-9280.1995.tb00597.x

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93–107. https://doi.org/10.3758/BF03212211

Lab, V. (2021). *Colour* (Version 0.1.5) [Computer software]. http://github.com/vaab/colour

Langlois, T. A., Jacoby, N., Suchow, J. W., & Griffiths, T. L. (2021). Serial reproduction reveals the geometry of visuospatial representations. *Proceedings of the National Academy of Sciences*, *118*(13), e2012938118. https://doi.org/10.1073/pnas.2012938118

Lee, K., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: Testing the face-space framework. *Psychological Science*, *11*(5), 379–385. https://doi.org/10.1111/1467-9280.00274

Leeuwenberg, E. L. J., & Boselie, F. (1988). Against the likelihood principle in visual form perception. *Psychological Review*, *95*(4), 485–491. https://doi.org/10.1037/0033-295x.95.4.485

Leeuwenberg, E. L. J., & van der Helm, P. A. (2012). *Structural Information Theory: The simplicity of visual form*. Cambridge University Press. https://doi.org/10.1017/CBO9781139342223

Leeuwenberg, E. L. J., & van der Helm, P. A. (1991). Unity and variety in visual form. *Perception*, *20*(5), 595–622. https://doi.org/10.1068/p200595

Legrenzi, P. (1994). Kanizsa's analysis of "Prägnanz" as an obstacle to problem solving and the theory of mental models. *Japanese Psychological Research*, *36*(3), 121–125.

Leyton, M. (1992). *Symmetry, causality, mind*. MIT Press.

Lieder, I., Adam, V., Frenkel, O., Jaffe-Dax, S., Sahani, M., & Ahissar, M. (2019). Perceptual bias reveals slow-updating in autism and fast-forgetting in dyslexia. *Nature Neuroscience*, *22*(2), 256–264. https://doi.org/10.1038/s41593-018-0308-9

Locher, P. J., Stappers, P. J., & Overbeeke, K. (1998). The role of balance as an organizing design principle underlying adults' compositional strategies for creating visual displays. *Acta Psychologica*, *99*(2), 141–161. https://doi.org/10.1016/S0001-6918(98)00008-0

Long, B., Fan, J. E., Huey, H., Chai, Z., & Frank, M. C. (2021). *Parallel developmental changes in children's production and recognition of line drawings of visual concepts* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/5yv7x

Long, B., Wang, Y., Christie, S., Frank, M. C., & Fan, J. E. (2022). *Developmental consistency in children's drawings of object categories* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/fpebs

Luccio, R. (2019). Perceptual simplicity: The true role of Prägnanz and Occam. *Gestalt Theory*, *41*(3), 263–276. https://doi.org/10.2478/gth-2019-0024

Luchins, A. S., & Luchins, E. H. (1998). Commentary on Vicario's "On Wertheimer's principles of organization". *Gestalt Theory*, *20*(4), 270–282.

Manassi, M., Liberman, A., Kosovicheva, A., Zhang, K., & Whitney, D. (2018). Serial dependence in position occurs at the time of perception. *Psychonomic Bulletin & Review*, *25*(6), 2245–2253. https://doi.org/10.3758/s13423-018-1454-5

Mao, J., & Stocker, A. A. (2022). *Holistic inference explains human perception of stimulus orientation* [Preprint]. bioRxiv. https://doi.org/10.1101/2022.06.24.497534

Mardia, K. V., & Jupp, P. E. (2000). *Directional Statistics*. Wiley.

Marković, S., & Gvozdenovi, V. (2001). Symmetry, complexity and perceptual economy: Effects of minimum and maximum simplicity conditions. *Visual Cognition*, *8*(3-5), 305–327. https://doi.org/10.1080/13506280143000025

Martin, P., Uy, N., Kvapil, M., & Friedenberg, J. (2020). *The aesthetics of frieze patterns A preference for emergent features* [Poster]. https://doi.org/10.13140/RG.2.2.34413.74721

Mather, G. (2018). Visual image statistics in the history of Western art. *Art and Perception*, *6*(2-3), 97–115. https://doi.org/10.1163/22134913-20181092

Mather, G. (2020). Aesthetic image statistics vary with artistic genre. *Vision*, *4*(1), 10. https://doi.org/10.3390/vision4010010

Mattar, M. G., Carter, M. V., Zebrowitz, M. S., Thompson-Schill, S. L., & Aguirre, G. K. (2018). Individual differences in response precision correlate with adaptation bias. *Journal of Vision*, *18*(13). https://doi.org/10.1167/18.13.18

Mattar, M. G., Kahn, D. A., Thompson-Schill, S. L., & Aguirre, G. K. (2016). Varying timescales of stimulus integration unite neural adaptation and prototype formation. *Current Biology*, *26*(13), 1669–1676. https://doi.org/10.1016/j.cub.2016.04.065

Matthews, W. J., & Adams, A. (2008). Another reason why adults find it hard to draw accurately. *Perception*, *37*(4), 628–630. https://doi.org/10.1068/p5895

Mauro, R., & Kubovy, M. (1992). Caricature and face recognition. *Memory & Cognition*, *20*(4), 433–440. https://doi.org/10.3758/BF03210927

Maus, G. W., Chaney, W., Liberman, A., & Whitney, D. (2013). The challenge of measuring long-term positive aftereffects. *Current Biology : CB*, *23*(10). https://doi.org/10.1016/j.cub.2013.03.024

Mayer, S. (2021). *Imagefluency: Image statistics based on processing fluency*. Zenodo. https://doi.org/10.5281/zenodo.5614666

Mayer, S., & Landwehr, J. R. (2018a). Objective measures of design typicality. *Design Studies*, *54*, 146–161. https://doi.org/10.1016/j.destud.2017.09.004

Mayer, S., & Landwehr, J. R. (2018b). Quantifying visual aesthetics based on processing fluency theory: Four algorithmic measures for antecedents of aesthetic preferences. *Psychology of Aesthetics, Creativity, and the Arts*, *12*(4), 399–431. https://doi.org/10.1037/aca0000187

McCloud, S. (1993). *Understanding comics: The invisible art*. HarperCollins Publishers.

McGovern, D. P., Walsh, K. S., Bell, J., & Newell, F. N. (2017). Individual differences in context-dependent effects reveal common mechanisms underlying the direction aftereffect and direction repulsion. *Vision Research*, *141*, 109–116. https://doi.org/10.1016/j.visres.2016.08.009

McMurray, B. (2022). *The myth of categorical perception* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/dq7ej

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist*, *44*(12), 1469–1481. https://doi.org/10.1037/0003-066X.44.12.1469

Medin, D. L., & Barsalou, L. W. (1987). Categorization processes and categorical perception. In *Categorical perception: The groundwork of cognition* (pp. 455–490). Cambridge University Press.

Mersmann, O., Trautmann, H., Steuer, D., & Bornkamp, B. (2018). *Truncnorm: Truncated normal distribution*. https://CRAN.R-project.org/package=truncnorm

Metzger, W. (1941). *Psychologie: Die Entwicklung ihrer Grundannahmen seit der Einführung des Experiments [Psychology: The development of its basic assumptions since the introduction of the experiment.]*. Springer-Verlag. https://link.springer.com/book/10.1007/978-3-642-53395-2

Metzger, W. (1954). Grundbegriffe der Gestaltpsychologie. In *Aktuelle Probleme der Gestalttheorie* (Ajuriaguerra, Juan de).

Metzger, W. (1966). *Figural-wahrnemung* [Figural perception]. In W. Metzger, R. Bergius, & H. Thomae (Eds.), *Allgemeine Psychologie [General psychology]* (pp. 693–744). Hogrefe.

Metzger, W. (1975). *Gesetze des Sehens [Laws of seeing]* (Third edition). Kramer.

Metzger, W. (2006). *Laws of seeing*. MIT Press. (Original work published 1936)

Microsoft, & Weston, S. (2020). *Foreach: Provides foreach looping construct*. https://CRAN.R-project.org/package=foreach

Miller, J., & Schwarz, W. (2018). Implications of individual differences in on-average null effects. *Journal of Experimental Psychology: General*, *147*(3), 377–397. https://doi.org/10.1037/xge0000367

Miller, M. B., & Gazzaniga, M. S. (1998). Creating false memories for visual scenes. *Neuropsychologia*, *36*(6), 513–520. https://doi.org/10.1016/S0028-3932(97)00148-6

Mitchell, P., Ropar, D., Ackroyd, K., & Rajendran, G. (2005). How perception impacts on drawings. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(5), 996–1003. https://doi.org/10.1037/0096-1523.31.5.996

Moitzi, M. (2021). *Svgwrite* (Version 1.4.1) [Computer software]. http://github.com/mozman/svgwrite.git

Mollon, J. D., Bosten, J. M., Peterzell, D. H., & Webster, M. A. (2017). Individual differences in visual science: What can be learned and what is good experimental practice? *Vision Research*, *141*, 4–15. https://doi.org/10.1016/j.visres.2017.11.001

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of bayes factors for common designs*. https://CRAN.R-project.org/package=BayesFactor

Müller, K. (2020). *Here: A simpler way to find your files*. https://CRAN.R-project.org/package=here

Müller, K., & Wickham, H. (2022). *Tibble: Simple data frames*. https://CRAN.R-project.org/package=tibble

Murdoch, D., & Chow, E. D. (2020). *Ellipse: Functions for drawing ellipses and ellipse-like confidence regions*. https://CRAN.R-project.org/package=ellipse

Muth, C., & Carbon, C.-C. (2013). The Aesthetic Aha: On the pleasure of having insights into Gestalt. *Acta Psychologica*, *144*(1), 25–30. https://doi.org/10.1016/j.actpsy.2013.05.001

Muth, C., & Carbon, C.-C. (2016). SeIns: Semantic instability in art. *Art and Perception*, *4*(1-2), 145–184. https://doi.org/10.1163/22134913-00002049

Muth, C., Pepperell, R., & Carbon, C.-C. (2013). Give me Gestalt! Preference for cubist artworks revealing high detectability of objects. *Leonardo*, *46*(5), 488–489. https://doi.org/10.1162/LEON_a_00649

Muth, C., Westphal-Fitch, G., & Carbon, C.-C. (2019). Seeking (dis)order: Ordering appeals but slight disorder and complex order trigger interest. *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000284

Nadal, M., Munar, E., Marty, G., & Cela-Conde, C. J. (2010). Visual complexity and beauty appreciation: Explaining the divergence of results. *Empirical Studies of the Arts*, *28*(2), 173–191. https://doi.org/10.2190/EM.28.2.d

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383. https://doi.org/10.1016/0010-0285(77)90012-3

Necker, L. A. (1832). Observations on some remarkable optical phaenomena seen in Switzerland; and on an optical phaenomenon which occurs on viewing a figure of a crystal or geometrical solid. *London and Edinburgh Philosophical Magazine and Journal of Science. Third Series*, *1*, 329–337.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In *Visual Information Processing* (pp. 283–308). Elsevier. https://doi.org/10.1016/B978-0-12-170150-5.50012-3

Newell, F. N., & Bulthoff, H. H. (2002). Categorical perception of familiar objects. *Cognition*, *85*, 113–143. https://doi.org/10.1016/S0010-0277(02)00104-X

Ni, L., & Stocker, A. A. (2023). Efficient sensory encoding predicts robust averaging. *Cognition*, *232*, 105334. https://doi.org/10.1016/j.cognition.2022.105334

Noel, J.-P., Zhang, L.-Q., Stocker, A. A., & Angelaki, D. E. (2021). Individuals with autism spectrum disorder have altered visual encoding capacity. *PLOS Biology*, *19*(5), e3001215. https://doi.org/10.1371/journal.pbio.3001215

Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, *23*(1), 94–140. https://doi.org/10.1016/0010-0285(91)90004-8

Ooms, J. (2021). *Magick: Advanced graphics and image-processing in r*. https://CRAN.R-project.org/package=magick

Op de Beeck, H., Wagemans, J., & Vogels, R. (2003a). Asymmetries in stimulus comparisons by monkey and man. *Current Biology*, *13*(20), 1803–1808. https://doi.org/10.1016/j.cub.2003.09.036

Op de Beeck, H., Wagemans, J., & Vogels, R. (2003b). The effect of category learning on the representation of shape: Dimensions can be biased but not differentiated. *Journal of Experimental Psychology: General*, *132*(4), 491–511. https://doi.org/10.1037/0096-3445.132.4.491

Ostrofsky, J., Kozbelt, A., & Cohen, D. J. (2015). Observational drawing biases are predicted by biases in perception: Empirical support of the misperception hypothesis of drawing accuracy with respect to two angle illusions. *Quarterly Journal of Experimental Psychology*, *68*(5), 1007–1025. https://doi.org/10.1080/17470218.2014.973889

Ostrofsky, J., Kozbelt, A., & Seidel, A. (2012). Perceptual constancies and visual selection as predictors of realistic drawing skill. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(2), 124–136. https://doi.org/10.1037/a0026384

Palmer, S. E. (1982). Symmetry, transformation, and the structure of perceptual systems. In J. Beck (Ed.), *Organization and representation in perception* (pp. 95–144). Lawrence Erlbaum.

Palmer, S. E. (1991). Goodness, Gestalt, groups, and Garner: Local symmetry subgroups as a theory of figural goodness. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure: Essays in honor of Wendell R. Garner* (pp. 23–39). American Psychological Association. https://doi.org/10.1037/10101-001

Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human preference. *Annual Review of Psychology*, *64*(1), 77–107. https://doi.org/10.1146/annurev-psych-120710-100504

Panis, S., Wagemans, J., & Op de Beeck, H. P. (2011). Dynamic norm-based encoding for unfamiliar shapes in human visual cortex. *Journal of Cognitive Neuroscience*, *23*(7), 1829–1843. https://doi.org/10.1162/jocn.2010.21559

Pascucci, D., Mancuso, G., Santandrea, E., Libera, C. D., Plomp, G., & Chelazzi, L. (2019). Laws of concatenated perception: Vision goes for novelty, decisions for perseverance. *PLOS Biology*, *17*(3), e3000144. https://doi.org/10.1371/journal.pbio.3000144

Pascucci, D., Tanrikulu, Ö. D., Ozkirli, A., Houborg, C., Ceylan, G., Zerr, P., Rafiei, M., & Kristjánsson, Á. (2023). Serial dependence in visual perception: A review. *Journal of Vision*, *23*(1), 9. https://doi.org/10.1167/jov.23.1.9

Pastore, R. E. (1987). Categorical perception: Some psychophysical models. In *Categorical perception: The groundwork of cognition* (pp. 29–52). Cambridge University Press.

Patching, G. R., Englund, M. P., & Hellström, Å. (2012). Time- and space-order effects in timed discrimination of brightness and size of paired visual stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 915–940. https://doi.org/10.1037/a0027593

Pedersen, T. L. (2021). *Ggforce: Accelerating 'ggplot2'*. https://CRAN.R-project.org/package=ggforce

Pedersen, T. L. (2022). *Patchwork: The composer of plots*. https://CRAN.R-project.org/package=patchwork

Peirce, J. W. (2007). PsychoPy–Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pepperell, R. (2018). Art, energy, and the brain. In J. F. Christensen & A. Gomila (Eds.), *Progress in Brain Research* (Vol. 237, pp. 417–435). Elsevier. https://doi.org/10.1016/bs.pbr.2018.03.022

Pérez, F., & Granger, B. E. (2007). IPython: A system for interactive scientific computing. *Computing in Science & Engineering*, *9*(3). https://doi.org/10.1109/MCSE.2007.53

Petermann, B. (1931). *Das Gestaltproblem in der Psychologie im Lichte analytischer Besinnung: Ein Versuch zu grundsätzlicher Orientierung [The Gestalt problem in psychology in the light of analytical reflection: An attempt at fundamental orientation]*. Verlag von Johann Ambrosius.

Peterson, M. A., & Gibson, B. S. (1994). Object recognition contributions to figure-ground organization: Operations on outlines and subjective contours. *Perception & Psychophysics*, *56*(5), 551–564. https://doi.org/10.3758/BF03206951

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, *6*(1), 7–11. https://journal.r-project.org/archive/

Poirel, N., Pineau, A., & Mellet, E. (2006). Implicit identification of irrelevant local objects interacts with global/local processing of hierarchical stimuli. *Acta Psychologica*, *122*(3), 321–336. https://doi.org/10.1016/j.actpsy.2005.12.010

Polk, T. A., Behensky, C., Gonzalez, R., & Smith, E. E. (2002). Rating the similarity of simple perceptual stimuli: Asymmetries induced by manipulating exposure frequency. *Cognition*, *82*(3), B75–B88. https://doi.org/10.1016/S0010-0277(01)00151-2

Pomerantz, J. R. (1977). Pattern goodness and speed of encoding. *Memory & Cognition*, *5*(2), 235–241. https://doi.org/10.3758/BF03197367

Pomerantz, J. R., & Garner, W. R. (1973). The role of configuration and target discriminability in a visual search task. *Memory & Cognition*, *1*(1), 64–68. https://doi.org/10.3758/BF03198070

Pomerantz, J. R., & Kubovy, M. (1986). Theoretical approaches to perceptual organization: Simplicity and likelihood principles. In *Handbook of perception and human performance, Vol. 2: Cognitive processes and performance.* (pp. 1–46). John Wiley & Sons.

Port, A. (2021). *Svgpathtools* (Version 1.4.1) [Computer software]. https://github.com/mathandy/svgpathtools

Post, R. A. G., Blijlevens, J., & Hekkert, P. (2016). "To preserve unity while almost allowing for chaos": Testing the aesthetic principle of unity-in-variety in product design. *Acta Psychologica*, *163*, 142–152. https://doi.org/10.1016/j.actpsy.2015.11.013

Prasad, D., & Bainbridge, W. A. (2022). The visual Mandela effect as evidence for shared and specific false memories across people. *Psychological Science*, *33*(12), 1971–1988. https://doi.org/10.1177/09567976221108944

Quinlan, P. T., & Wilton, R. N. (1998). Grouping by proximity or similarity? Competition between the Gestalt principles in vision. *Perception*, *27*(4), 417–430. https://doi.org/10.1068/p270417

Quinn, P. C. (2000). Perceptual reference points for form and orientation in young infants: Anchors or magnets? *Perception & Psychophysics*, *62*(8), 1625–1633. https://doi.org/10.3758/BF03212160

R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Rausch, E. (1952). *Struktur und Metrik figural-optischer Wahrnehmung [Structure and metrics of figural-optical perception]*. Verlag Dr. Waldemar Kramer.

Rausch, E. (1966). *Das Eigenschaftsproblem in der Gestalttheorie der Wahrnemung* [The property problem in the Gestalt theory of perception]. In W. Metzger, R. Bergius, & H. Thomae (Eds.), *Allgemeine Psychologie [General psychology]* (pp. 866–953). Hogrefe.

Rausch, E. (1979/1992). *Neun Wünsche an die Zukunft der Psychologie (Auszugsweiser Nachdruck eines 1979 erschienenen Gesprächs mit E. Rausch)* [Nine wishes for the future of psychology (Excerpt reprint of a 1979 reprint of a conversation with E. Rausch published in 1979)]. *Gestalt Theory*, *14*(2), 143–144.

Rauschenberger, R., & Yantis, S. (2006). Perceptual encoding efficiency in visual search. *Journal of Experimental Psychology: General*, *135*(1), 116–131. https://doi.org/10.1037/0096-3445.135.1.116

Redies, C., Brachmann, A., & Wagemans, J. (2017). High entropy of edge orientations characterizes visual artworks from diverse cultural backgrounds. *Vision Research*, *133*, 130–144. https://doi.org/10.1016/j.visres.2017.02.004

Regebro, L. (2021). *Svg.path* (Version 4.1) [Computer software]. https://github.com/regebro/svg.path

Repp, B. H., & Liberman, A. M. (1987). Phonetic category boundaries are flexible. In *Categorical perception: The groundwork of cognition* (pp. 89–112). Cambridge University Press.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology*, *19*(4), 473–497. https://doi.org/10.1016/0010-0285(87)90016-8

Rhodes, G., & McLean, I. G. (1990). Distinctiveness and expertise effects with homogeneous stimuli: Towards a model of configural coding. *Perception*, *19*(6), 773–794. https://doi.org/10.1068/p190773

Riou, B., Lesourd, M., Brunel, L., & Versace, R. (2011). Visual memory and visual perception: When memory improves visual search. *Memory & Cognition*, *39*(6), 1094–1102. https://doi.org/10.3758/s13421-011-0075-2

Roberson, D., Damjanovic, L., & Pilling, M. (2007). Categorical perception of facial expressions: Evidence for a "category adjustment" model. *Memory & Cognition*, *35*(7), 1814–1829. https://doi.org/10.3758/BF03193512

Roberson, D., Hanley, J. R., & Pak, H. (2009). Thresholds for color discrimination in English and Korean speakers. *Cognition*, *112*(3), 482–487. https://doi.org/10.1016/j.cognition.2009.06.008

Robert, M. B. L. (1999). A unified account of the effects of caricaturing faces. *Visual Cognition*, *6*(1), 1–42. https://doi.org/10.1080/713756800

Robinson, A., Becker, R., the ReportLab team, & the community. (2021). *Reportlab: The Reportlab Toolkit* (Version 3.6.1) [Computer software]. http://www.reportlab.com/

Rodríguez, J., Bortfeld, H., Rudomín, I., Hernández, B., & Gutiérrez-Osuna, R. (2009). The reverse-caricature effect revisited: Familiarization with frontal facial caricatures improves veridical face recognition. *Applied Cognitive Psychology*, *23*(5), 733–742. https://doi.org/10.1002/acp.1539

Roediger III, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Rogers, B. (2014). Delusions about Illusions. *Perception*, *43*(9), 840–845. https://doi.org/10.1068/p7731

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and acquisition of language* (pp. 111–144). Academic Press. https://doi.org/10.1016/B978-0-12-505850-6.50010-4

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, *7*(4), 532–547. https://doi.org/10.1016/0010-0285(75)90021-3

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Lawrence Erlbaum.

Rosielle, L. J., & Hite, L. A. (2009). The caricature effect in drawing: Evidence for the use of categorical relations when drawing abstract pictures. *Perception*, *38*(3), 357–375. https://doi.org/10.1068/p5831

Rouder, J. N. (2019). *On the interpretation of Bayes Factors: A reply to de Heide and Grunwald* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/m6dhw

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, *21*(2), 301–308. https://doi.org/10.3758/s13423-014-0595-4

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, *26*(2), 452–467. https://doi.org/10.3758/s13423-018-1558-y

Sablé-Meyer, M., Fagot, J., Caparos, S., van Kerkoerle, T., Amalric, M., & Dehaene, S. (2021). Sensitivity to geometric shape regularity in humans and baboons: A putative signature of human singularity. *Proceedings of the National Academy of Sciences*, *118*(16), e2023123118. https://doi.org/10.1073/pnas.2023123118

Sadil, P., Cowell, R., & Huber, D. E. (2021). *The push-pull of serial dependence effects: Attraction to the prior response and repulsion from the prior stimulus* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/f52yz

Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics*, *31*(4), 307–314. https://doi.org/10.3758/BF03202653

Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., Elberg, A., & Crowley, J. (2021). *GGally: Extension to 'ggplot2'*. https://CRAN.R-project.org/package=GGally

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, *25*(1), 128–142. https://doi.org/10.3758/s13423-017-1230-y

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322–339. https://doi.org/10.1037/met0000061

Schumann, F. (1914). *Bericht über den VI. Kongreß für experimentelle Psychologie in Göttingen vom 15. Bis 18. April 1914 [Report on the VI Congress of Experimental Psychology in Göttingen from April 15 to 18, 1914]*.

Schurger, A., Sarigiannidis, I., Naccache, L., Sitt, J. D., & Dehaene, S. (2015). Cortical activity is more stable when sensory stimuli are consciously perceived. *Proceedings of the National Academy of Sciences*, *112*(16), E2083–E2092. https://doi.org/10.1073/pnas.1418730112

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. *Nature Human Behaviour*, 1–17. https://doi.org/10.1038/s41562-020-00938-0

Schwiedrzik, C. M., Ruff, C. C., Lazar, A., Leitner, F. C., Singer, W., & Melloni, L. (2014). Untangling perceptual memory: Hysteresis and adaptation map into separate cortical networks. *Cerebral Cortex*, *24*(5), 1152–1164. https://doi.org/10.1093/cercor/bhs396

Schwiedrzik, C. M., Sudmann, S. S., Thesen, T., Wang, X., Groppe, D. M., Mégevand, P., Doyle, W., Mehta, A. D., Devinsky, O., & Melloni, L. (2018). Medial prefrontal cortex supports perceptual memory. *Current Biology*, *28*(18), R1094–R1095. https://doi.org/10.1016/j.cub.2018.07.066

Seamon, J. G., Luo, C. R., Schlegel, S. E., Greene, S. E., & Goldenberg, A. B. (2000). False memory for categorized pictures and words: The category associates procedure for studying memory errors in children and adults. *Journal of Memory and Language*, *42*(1), 120–146. https://doi.org/10.1006/jmla.1999.2676

Sheehan, T. C., & Serences, J. T. (2023). *Distinguishing response from stimulus driven history biases* [Preprint]. bioRxiv. https://doi.org/10.1101/2023.01.11.523637

Shepard, R. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323. https://doi.org/10.1126/science.3629243

Shier, J. (2011). *Filling space with random fractal non-overlapping simple shapes*. 10.

Shier, J., & Bourke, P. (2013). An algorithm for random fractal filling of space: An algorithm for random fractal filling of space. *Computer Graphics Forum*, *32*(8), 89–97. https://doi.org/10.1111/cgf.12163

Sims, C. R. (2018). Efficient coding explains the universal law of generalization in human perception. *Science*, *360*(6389), 652–656. https://doi.org/10.1126/science.aaq1118

Smets, G. (1973). *Aesthetic judgment and arousal: An experimental contribution to psycho-aesthetics*. Leuven University Press.

Smith, B. (Ed.). (1988). *Foundations of Gestalt theory*. Philosophia Verlag.

Snyder, H. K., Rafferty, S. M., Haaf, J. M., & Rouder, J. N. (2019). Common or distinct attention mechanisms for contrast and assimilation? *Attention, Perception, & Psychophysics*, *81*(6), 1944–1950. https://doi.org/10.3758/s13414-019-01713-8

Snyder, J. S., Schwiedrzik, C. M., Vitela, A. D., & Melloni, L. (2015). How previous experience shapes perception in different sensory modalities. *Frontiers in Human Neuroscience*, *9*, 594. https://doi.org/10.3389/fnhum.2015.00594

Song, C., Schwarzkopf, D. S., Lutti, A., Li, B., Kanai, R., & Rees, G. (2013). Effective connectivity within human primary visual cortex predicts interindividual diversity in illusory perception. *Journal of Neuroscience*, *33*(48), 18781–18791. https://doi.org/10.1523/jneurosci.4201-12.2013

Song, C., Schwarzkopf, D. S., & Rees, G. (2013). Variability in visual cortex size reflects tradeoff between local orientation sensitivity and global orientation modulation. *Nature Communications*, *4*(1), 2201. https://doi.org/10.1038/ncomms3201

Sorge, S. (1940). *Neue versuche über die wiedergabe abstrakter optischer gebilde* [New experiments on the reproduction of abstract optical formations]. *Archiv für die gesamte Psychologie*, *106*, 1–88.

Spehar, B., Walker, N., & Taylor, R. P. (2016). Taxonomy of individual variations in aesthetic responses to fractal patterns. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00350

Spröte, P., Schmidt, F., & Fleming, R. W. (2016). Visual perception of shape altered by inferred causal history. *Scientific Reports*, *6*(1), 36245. https://doi.org/10.1038/srep36245

Stadler, M., Stegnano, L., & Trombini, G. (1979). *Quantitative Analyse der Rauschschen Prägnanzaspekte* [quantitative analysis of Rausch' Prägnanz aspects]. *Gestalt Theory*, *1*, 28–40.

Stan Development Team. (2020a). *RStan: The R interface to Stan*. http://mc-stan.org/

Stan Development Team. (2020b). *StanHeaders: Headers for the R interface to Stan*. https://mc-stan.org/

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Stein, H., Barbosa, J., Rosa-Justicia, M., Prades, L., Morató, A., Galan-Gadea, A., Ariño, H., Martinez-Hernandez, E., Castro-Fornieles, J., Dalmau, J., & Compte, A. (2020). Reduced serial dependence suggests deficits in synaptic potentiation in anti-NMDAR encephalitis and schizophrenia. *Nature Communications*, *11*(1), 4250. https://doi.org/10.1038/s41467-020-18033-3

Stevenage, S. V. (1995). Can caricatures really produce distinctiveness effects? *British Journal of Psychology*, *86*(1), 127–146.

Strother, L., & Kubovy, M. (2006). On the surprising salience of curvature in grouping by proximity. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(2), 226–234. https://doi.org/10.1037/0096-1523.32.2.226

Strother, L., & Kubovy, M. (2012). Structural salience and the nonaccidentality of a Gestalt. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 827–832. https://doi.org/10.1037/a0027939

Sun, Z., & Firestone, C. (2021). Curious objects: How visual complexity guides attention and engagement. *Cognitive Science*, *45*(4). https://doi.org/10.1111/cogs.12933

Sundqvist, F. (2003). *Perceptual dynamics: Theoretical foundations and philosophical implications of Gestalt psychology* [PhD thesis]. Göteborg University; Acta Universitatis Gothoburgensis.

Taubert, J., Alais, D., & Burr, D. (2016). Different coding strategies for the perception of stable and changeable facial attributes. *Scientific Reports*, *6*(1), 32239. https://doi.org/10.1038/srep32239

Telenczuk, B. (2021). *Svgutils* (Version 0.3.4) [Computer software]. https://svgutils.readthedocs.io

Thomas, B. G. (2012). 15 - Colour symmetry: The systematic coloration of patterns and tilings. In J. Best (Ed.), *Colour Design* (pp. 381–432). Woodhead Publishing. https://doi.org/10.1533/9780857095534.3.381

Tiedemann, F. (2020). *Gghalves: Compose half-half plots using your favourite geoms*. https://CRAN.R-project.org/package=gghalves

Tinbergen, N. (1951). *The study of instinct*. Clarendon Press.

Tomassini, A., Morgan, M. J., & Solomon, J. A. (2010). Orientation uncertainty reduces perceived obliquity. *Vision Research*, *50*(5), 541–547. https://doi.org/10.1016/j.visres.2009.12.005

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352. https://doi.org/10.1037/0033-295X.84.4.327

Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, *8*(1), 14637. https://doi.org/10.1038/ncomms14637

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, *121*(4), 649–675. https://doi.org/10.1037/a0037665

Van de Cruys, S., & Wagemans, J. (2011). Putting reward in art: A tentative prediction error account of visual art. *I-Perception*, *2*(9), 1035–1062. https://doi.org/10.1068/i0466aap

van der Helm, P. A. (2000). Simplicity versus likelihood in visual perception: From surprisals to precisals. *Psychological Bulletin*, *126*(5), 770–800. https://doi.org/10.1037/0033-2909.126.5.770

van der Helm, P. A. (2017). On Bayesian simplicity in human visual perceptual organization. *Perception*, *46*(11), 1269–1282. https://doi.org/10.1177/0301006617719604

Van der Hulst, E., Van Geert, E., & Wagemans, J. (in preparation). *Shape variation in proximity grouping: An individual differences approach*.

Van der Hulst, E., van Heusden, E., Wagemans, J., & Moors, P. (2022). *Grouping by proximity and luminance similarity is additive for everyone: An analysis of individual differences in grouping sensitivity*. Retrieved from osf.io/p845j.

Van Geert, E., Bossens, C., & Wagemans, J. (2022). The Order & Complexity Toolbox for Aesthetics (OCTA): A systematic approach to study the relations between order, complexity, and aesthetic appreciation. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01900-w

Van Geert, E., Ding, R., & Wagemans, J. (2021). *A cross-cultural comparison of aesthetic preferences for neatly organized compositions: Native Chinese- vs. native Dutch-speaking samples* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/679zm

Van Geert, E., Frérart, L., & Wagemans, J. (2022). *Towards the most prägnant Gestalt: Leveling and sharpening as contextually dependent adaptive strategies* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/t3bzw

Van Geert, E., Hofmann, D., & Wagemans, J. (in preparation). *The perception and appreciation of order and complexity*.

Van Geert, E., Moors, P., Haaf, J., & Wagemans, J. (2022). Same stimulus, same temporal context, different percept? Individual differences in hysteresis and adaptation when perceiving multistable dot lattices. *I-Perception*, *13*(4), 20416695221109300. https://doi.org/10.1177/20416695221109300

Van Geert, E., & Wagemans, J. (2020). Order, complexity, and aesthetic appreciation. *Psychology of Aesthetics, Creativity, and the Arts*, *14*(2), 135–154. https://doi.org/10.1037/aca0000224

Van Geert, E., & Wagemans, J. (2021). Order, complexity, and aesthetic preferences for neatly organized compositions. *Psychology of Aesthetics, Creativity, and the Arts*, *15*(3), 484–504. https://doi.org/10.1037/aca0000276

Van Geert, E., & Wagemans, J. (2022). *What good is goodness? The effects of reference points on discrimination and categorization of shapes* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/6x75c

Van Geert, E., & Wagemans, J. (2023). *Prägnanz in visual perception* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/pxdg4

Van Geert, E., & Wagemans, J. (in preparation a). *Pre-existing categorization diminishes attractive and repulsive temporal history effects on perception.*

Van Geert, E., & Wagemans, J. (in preparation b). *Individual differences in the use of perceptual history in the visual categorization of abstract shapes.*

Van Geert, E., Warny, A., & Wagemans, J. (in preparation). *A systematic approach to study preferences for complexity.*

van Leeuwen, C. (1990). Perceptual-learning systems as conservative structures: Is economy an attractor? *Psychological Research*, *52*(2), 145–152. https://doi.org/10.1007/BF00877522

van Lier, R., van der Helm, P., & Leeuwenberg, E. L. J. (1994). Integrating global and local aspects of visual occlusion. *Perception*, *23*(8), 883–903. https://doi.org/10.1068/p230883

Van Rossum, G., & Drake Jr, F. L. (1995). *Python reference manual.* Centrum voor Wiskunde en Informatica Amsterdam.

Van Rossum, G., & Drake Jr, F. L. (2009). *Python 3 reference manual.* CreateSpace.

Vanderplas, J. M., & Garvin, E. A. (1959). Complexity, association value, and practice as factors in shape recognition following paired-associates training. *Journal of Experimental Psychology*, *57*(3), 155–163. https://doi.org/10.1037/h0042010

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth). Springer. http://www.stats.ox.ac.uk/pub/MASS4/

vgalin. (2021). *html2image* (Version 1.1.2) [Computer software]. https://github.com/vgalin/html2image

von Ehrenfels, C. (1916). *Höhe und Reinheit der Gestalt* [Height and purity of Gestalt]. In *Kosmogonie [Cosmogony]* (pp. 93–96). Diederichs.

von Ehrenfels, C. (1922). *Das Primzahlengesetz, entwickelt und dargestellt auf Grund der Gestalttheorie [The prime number law, developed and presented on the basis of the Gestalt theory].* O. R. Reisland.

von Ehrenfels, C. (1937). *Über Gestaltqualitäten (1932)* [On Gestalt qualities]. *Philosophia (Belgrad)*, *2*, 139–141. (Original work published 1932)

Wagemans, J. (1992). Perceptual use of nonaccidental properties. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *46*(2), 236–279. https://doi.org/10.1037/h0084323

Wagemans, J. (2015). Historical and conceptual background: Gestalt theory. In J. Wagemans (Ed.), *The Oxford Handbook of Perceptual Organization.* Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199686858.013.026

Wagemans, J. (2018). Perceptual organization. In *The Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience: Vol. 2. Sensation, Perception & Attention* (pp. 803–872). John Wiley & Sons, Inc. https://doi.org/https://doi.org/10.1002/9781119170174.epcn218

Wagemans, J., Bossche, P. V., Segers, N., & d'Ydewalle, G. (1994). An affine group model and the perception of orthographically projected planar random polygons. *Journal of Mathematical Psychology*, *38*(1), 59–72. https://doi.org/10.1006/jmps.1994.1003

Wagemans, J., Claessens, P. M. E., & Moors, P. (2018). *Perceptual grouping in dot lattices revisited* [41st European Conference on Visual Perception (ECVP)]. Abstract published in *Perception*, 48(S1) (Supplement). https://doi.org/10.1177/0301006618824879

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & Heydt, R. von der. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172–1217. https://doi.org/10.1037/a0029333

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., & van Leeuwen, C. (2012). A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin*, *138*(6), 1218–1252. https://doi.org/10.1037/a0029334

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. https://doi.org/10.1016/j.cogpsych.2009.12.001

Wei, X.-X., & Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nature Neuroscience*, *18*(10), 1509–1517. https://doi.org/10.1038/nn.4105

Wei, X.-X., & Stocker, A. A. (2017). Lawful relation between perceptual bias and discriminability. *Proceedings of the National Academy of Sciences*, *114*(38), 10244–10249. https://doi.org/10.1073/pnas.1619153114

Wellek, A. (1959). *Das Prägnanzproblem der Gestaltpsychologie und das "Exemplarische" in der Pädagogik* [The problem of Prägnanz in Gestalt psychology and the "exemplary" in pedagogy]. *Zeitschrift für experimentelle und angewandte Psychologie*, *6*, 722–736.

Wertheimer, M. (1912). *Über das Denken der Naturvölker. I. Zahlen und Zahlgebilde* [About the thinking of people who live close to nature. I. Numbers and number formations.]. *Zeitschrift für Psychologie*, *60*, 321–378.

Wertheimer, M. (1922). *Untersuchungen zur Lehre von der Gestalt. I. Prinzipielle Bemerkungen* [Investigations into the teachings of Gestalt. I. Remarks on its principles]. *Psychologische Forschung*, *1*, 47–58. https://doi.org/10.1007/BF00410385

Wertheimer, M. (1923). *Untersuchungen zur Lehre von der Gestalt. II* [Investigations into the teachings of Gestalt. II. Psychological research]. *Psychologische Forschung*, *4*, 301–350. https://doi.org/10.1007/BF00410640

Wertheimer, M. (1959). *Productive thinking*. Harper.

Wertheimer, M. (1999). Gestalt theory. *Gestalt Theory*, *21*, 181–183. (Original work published 1924)

Wertheimer, M., Spillmann, L., Sarris, V., & Sekuler, R. (2012). *On perceived motion and figural organization*. MIT Press.

Westphal-Fitch, G., Huber, L., Gómez, J. C., & Fitch, W. T. (2012). Production and perception rules underlying visual patterns: Effects of symmetry and hierarchy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 2007–2022. https://doi.org/10.1098/rstb.2012.0098

Wexler, M., Duyck, M., & Mamassian, P. (2015). Persistent states in vision break universality and time invariance. *Proceedings of the National Academy of Sciences*, *112*(48), 14990–14995. https://doi.org/10.1073/pnas.1508847112

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string operations*. https://CRAN.R-project.org/package=stringr

Wickham, H. (2022a). *Forcats: Tools for working with categorical variables (factors)*. https://CRAN.R-project.org/package=forcats

Wickham, H. (2022b). *Modelr: Modelling functions that work with the pipe*. https://CRAN.R-project.org/package=modelr

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., & Bryan, J. (2022). *Readxl: Read excel files*. https://CRAN.R-project.org/package=readxl

Wickham, H., Bryan, J., & Barrett, M. (2021). *Usethis: Automate package and project setup*. https://CRAN.R-project.org/package=usethis

Wickham, H., François, R., Henry, L., & Müller, K. (2022). *Dplyr: A grammar of data manipulation*. https://CRAN.R-project.org/package=dplyr

Wickham, H., & Girlich, M. (2022). *Tidyr: Tidy messy data*. https://CRAN.R-project.org/package=tidyr

Wickham, H., Hester, J., & Bryan, J. (2022). *Readr: Read rectangular text data*. https://CRAN.R-project.org/package=readr

Wickham, H., Hester, J., Chang, W., & Bryan, J. (2021). *Devtools: Tools to make developing r packages easier*. https://CRAN.R-project.org/package=devtools

Wilhelm, S., & G, M. B. (2022). *tmvtnorm: Truncated multivariate normal and student t distribution*. https://CRAN.R-project.org/package=tmvtnorm

Wilke, C. O. (2020). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'*. https://CRAN.R-project.org/package=cowplot

Wilson, A., & Chatterjee, A. (2005). The assessment of preference for balance: Introducing a new test. *Empirical Studies of the Arts*, *23*(2), 165–180. https://doi.org/10.2190/B1LR-MVF3-F36X-XR64

Wohlfahrt, E. (1932). *Der Auffassungsvorgang an kleinen Gestalten; Ein Beitrag zur Psychologie des Vorgestaltserlebnisses* [the perceptual process of small figures; A contribution to the psychology of pre-Gestalt experience. *Neue Psychologische Studien*, *4*, 347–414.

Wulf, F. (1922). *Beiträge zur Psychologie der Gestalt. VI. Über die Veränderung yon Vorstellungen (Gedächtnis und Gestalt)* [Contributions to the Psychology of Gestalt. VI. On the change of ideas (Memory and Gestalt)]. *Psychologische Forschung*, *1*(1), 333–373. https://doi.org/10.1007/BF00410394

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Chapman; Hall/CRC. https://yihui.org/knitr/

Yang, J., & Fan, J. (2021). Visual communication of object concepts at different levels of abstraction. *Journal of Vision*, *21*(9), 2951. https://doi.org/10.1167/jov.21.9.2951

Yu, G. (2022). *Ggimage: Use image in 'ggplot2'*. https://CRAN.R-project.org/package=ggimage

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, *11*(10), 1–17. https://doi.org/10.18637/jss.v011.i10

Zeileis, A. (2006). Object-oriented computation of sandwich estimators. *Journal of Statistical Software*, *16*(9), 1–16. https://doi.org/10.18637/jss.v016.i09

Zeileis, A., Köll, S., & Graham, N. (2020). Various versatile variances: An object-oriented implementation of clustered covariances in R. *Journal of Statistical Software*, *95*(1), 1–36. https://doi.org/10.18637/jss.v095.i01

Zhang, H., & Alais, D. (2020). Individual difference in serial dependence results from opposite influences of perceptual choices and motor responses. *Journal of Vision*, *20*(8), 2. https://doi.org/10.1167/jov.20.8.2

Zhu, H. (2021). *kableExtra: Construct complex table with 'kable' and pipe syntax*. https://CRAN.R-project.org/package=kableExtra

Zimmer, A. C. (1991). The complementarity of singularity and stability. A comment on Kanizsa & Luccio's "Analysis of the concept of Prägnanz" (1986). *Gestalt Theory*, *13*(4), 276–282.

Appendix A
Reference points, categorization, and discrimination

A.1 R packages used

As indicated in the Methods section, for all our analyses we used R (Version 4.0.4; R Core Team, 2021) and the R-packages *brms* (Version 2.16.1; Bürkner, 2017, 2018), *cowplot* (Version 1.1.1; Wilke, 2020), *dplyr* (Version 1.0.10; Wickham, François, et al., 2022), *forcats* (Version 0.5.2; Wickham, 2022a), *ggimage* (Version 0.3.1; Yu, 2022), *ggiraph* (Version 0.7.10; Gohel & Skintzos, 2021), *ggplot2* (Version 3.3.6; Wickham, 2016), *ggthemes* (Version 4.2.4; Arnold, 2021), *glue* (Version 1.6.2; Hester & Bryan, 2022), *here* (Version 1.0.1; Müller, 2020), *lubridate* (Version 1.8.0; Grolemund & Wickham, 2011), *magick* (Version 2.7.3; Ooms, 2021), *modelr* (Version 0.1.9; Wickham, 2022b), *papaja* (Version 0.1.1; Aust & Barth, 2022), *patchwork* (Version 1.1.2; Pedersen, 2022), *purrr* (Version 0.3.4; Henry & Wickham, 2020), *Rcpp* (Eddelbuettel & Balamuta, 2018; Version 1.0.9; Eddelbuettel & François, 2011), *readr* (Version 2.1.2; Wickham, Hester, et al., 2022), *readxl* (Version 1.4.1; Wickham & Bryan, 2022), *rstan* (Version 2.21.2; Stan Development Team, 2020a), *StanHeaders* (Version 2.21.0.7; Stan Development Team, 2020b), *stringr* (Version 1.4.0; Wickham, 2019), *tibble* (Version 3.1.8; Müller & Wickham, 2022), *tidybayes* (Version 3.0.1; Kay, 2021b), *tidyr* (Version 1.2.1; Wickham & Girlich, 2022), *tidyverse* (Version 1.3.2; Wickham et al., 2019), and *tinylabels* (Version 0.2.3; Barth, 2022).

A

A.2 Bayesian hierarchical model implementation details

A.2.1 Categorization responses

We fitted a hierarchical Bayesian binomial logistic regression model to the categorization response data with morph level as fixed effect, for each morph series separately, and participant ID within each morph series as random effect for both intercept and slope:

$$freqB \mid trials(n) \sim a + b * morph\_level$$

$$a \sim 0 + morph\_series + (1 \mid p \mid morph\_series : pp\_id)$$

$$b \sim 0 + morph\_series + (1 \mid p \mid morph\_series : pp\_id)$$

As priors, we specified a normal distribution with a mean of zero and a standard deviation of 1.5 for the intercept and a normal distribution with a mean of zero and a standard deviation of 0.5 for the slope. For the other model parameters, default priors were used. We used 4 chains consisting of 8000 iterations, with 4000 warmup iterations per chain.

A.2.2 Categorization response times

We fitted a hierarchical Bayesian lognormal regression model to the categorization response time data with morph level and morph level squared as fixed effects, for each morph series separately, and participant ID within each morph series as random effect for both intercept and slopes:

$$catRT \sim a + b * morph\_level + c * morph\_level^2$$

$$a \sim 0 + morph\_series + (1 \mid p \mid morph\_series : pp\_id)$$

$$b \sim 0 + morph\_series + (1 \mid p \mid morph\_series : pp\_id)$$

$$c \sim 0 + morph\_series + (1 \mid p \mid morph\_series : pp\_id)$$

As priors, we specified a normal distribution with a mean of -1 and a standard deviation of 0.5 for the intercept and a normal distribution with a mean of zero and a standard deviation of 0.3 for the slopes. The prior for sigma was specified as a normal distribution with mean 0.4 and standard deviation 0.3, and the priors for the standard deviations of intercept and slopes were specified as a normal distribution with mean 0.3 and standard deviation 0.1. We used 4 chains consisting of 8000 iterations, with 4000 warmup iterations per chain.

A.2.3 Discrimination responses

To investigate the presence of differences in discrimination sensitivity across stimulus pairs, we fitted a hierarchical Bayesian binomial logistic regression model to the discrimination response data with stepsize as fixed effect, for each morph series separately, and with trial stimuli and participant ID within each morph series as random effects for intercept and participant ID within each morph series as random effect for the slope:

$$freqdiff \mid trials(n) \sim a + b * stepsize$$

$$\begin{aligned} a \sim 0 + morph\_series + (1 \mid q \mid morph\_series : trial\_stimuli) \\ + (1 \mid p \mid morph\_series : pp\_id) \end{aligned}$$

$$b \sim 0 + morph\_series + (1 \mid p \mid morph\_series : pp\_id)$$

As priors, we specified a normal distribution with a mean of zero and a standard deviation of 1.5 for the intercept and a normal distribution with a mean of zero and a standard deviation of 0.5 for the slope. We used 4 chains consisting of 8000 iterations, with 4000 warmup iterations per chain.

Figure A.1 shows the empirical proportions and posterior predictive distributions for responding ‘different’ in the successive discrimination task, per stepsize, trial type, and morph series, averaged across participants. Stepsize indicates the absolute difference in morph level between the two morph stimuli presented in a trial, with a minimum of zero (for same trials) and a maximum of eleven (when both extremes of the morph series are presented). In this supplementary figure, all stepsizes are shown.

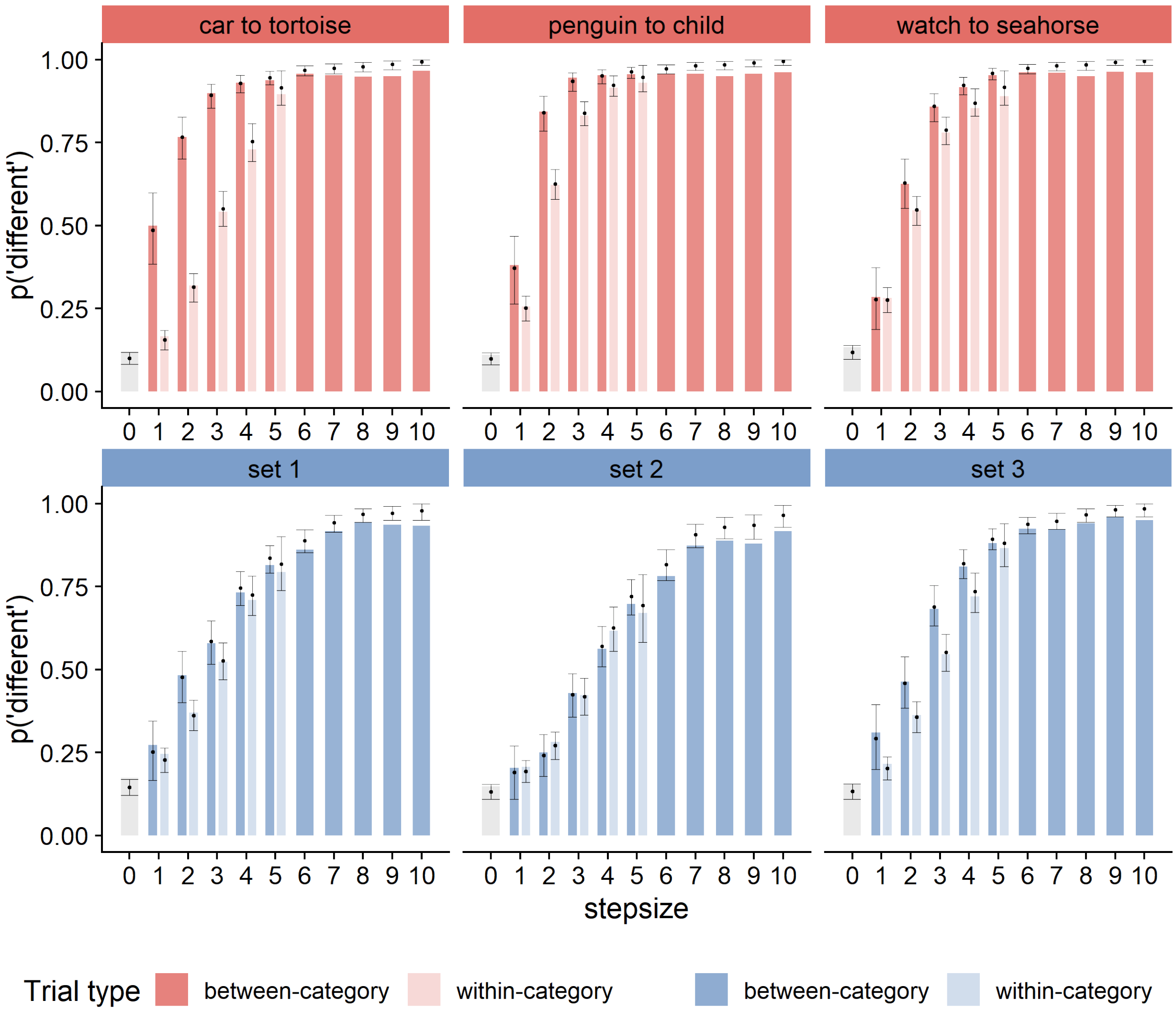


Figure A.1: Proportion of 'different' responses in the successive discrimination task for each stepsize, trial type (i.e., between-category vs. within-category), and morph series separately, averaged across participants. Bars indicate the empirical proportions for responding 'different'. The black dots indicate the mean posterior predictions from the model and the error bars indicate the 95% highest density continuous intervals (HDCI) of the posterior predictive distributions. In this figure, the difference between the darker and the lighter bars (i.e., the category boundary effect: more 'different' responses for between-category compared to within-category pairs, keeping stepsize equal) is on average higher for the recognizable than for the non-recognizable morph series. *Note.* In the interactive version of this figure, you can hover over the colored bars to see the exact percentage different responses, the mean and 95% HDCI of the posterior predictive distributions, and the number of trials related to each bar.

Figure A.2 shows the posterior distributions for the intercept (A) and the effect of stepsize (B) on the probability of responding 'different' in the successive discrimination task, for each morph series separately.

Figure A.3 shows the estimated pairwise differences between the posterior distributions for the effect of stepsize on responding 'different' in the successive discrimination task, for each of the different recognizable and non-recognizable morph series combinations.

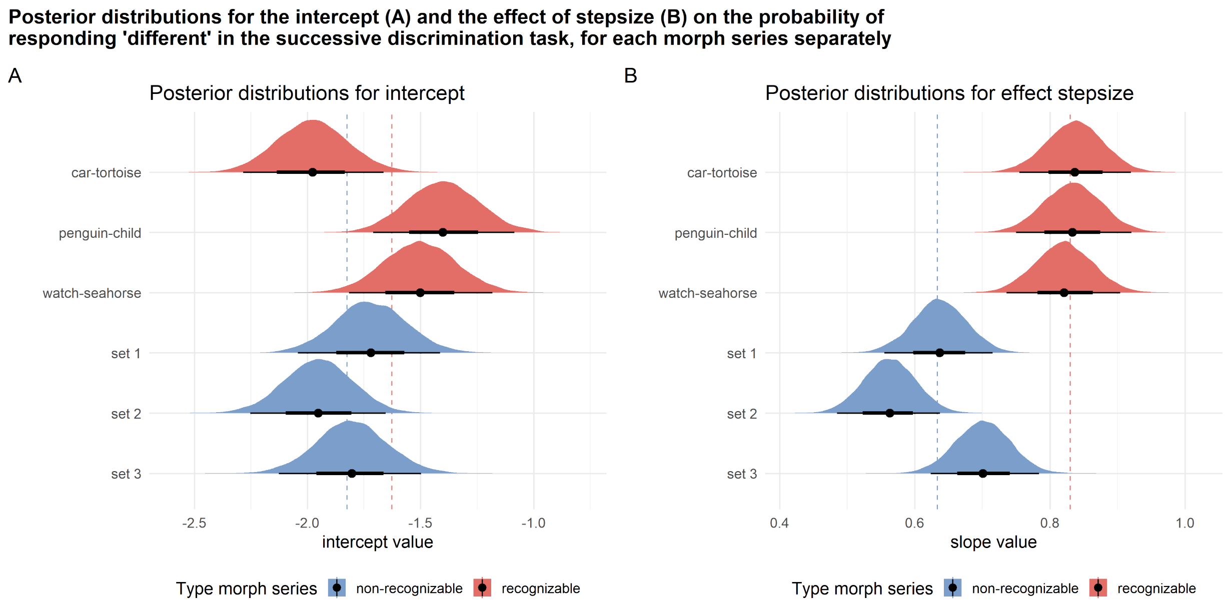


Figure A.2: Posterior distributions for the intercept (A) and the effect of stepsize (B) on the probability of responding 'different' in the successive discrimination task, for each morph series separately, in logodds units. Black dots and intervals indicate the mean, 66%, and 95% highest density continuous interval (HDCI) for each intercept or slope value. The colored dashed vertical lines indicate the estimated mean value per type of morph series (recognizable vs. non-recognizable). In this figure, the estimated effect of stepsize is larger for the recognizable than for the non-recognizable morph series. *Note.* In the interactive version of this figure, you can hover over the intervals to see the related mean and 95% HDCI for each distribution.

To investigate the presence of an overall category boundary effect, we fitted a hierarchical Bayesian binomial logistic regression model to the discrimination response data with stepsize as fixed effect, for each morph series and trial type separately, and with trial stimuli and participant ID within each morph series as random effects for intercept and participant ID within each morph series as random effect for the slope:

$$freqdiff \mid trials(n) \sim a + b * stepsize$$

$$a \sim 0 + morph\_series * between\_category + (1 \mid q \mid morph\_series : trial\_stimuli)$$
$$+ (1 \mid p \mid morph\_series : pp\_id)$$

$$b \sim 0 + morph\_series * between\_category + (1 \mid p \mid morph\_series : pp\_id)$$

As priors, we specified a normal distribution with a mean of zero and a standard deviation of 1.5 for the intercept and a normal distribution with a mean of zero and a standard deviation of 0.5 for the slope. We used 4 chains consisting of 8000 iterations, with 4000 warmup iterations per chain.

Figure A.4 shows the posterior distributions for the intercept (A), effect of stepsize (B), effect of trial type (C), and interaction between stepsize and trial type (D) on the probability of responding 'different' in the successive discrimination task, for each morph series separately.

To investigate the presence of directional asymmetries, we fitted a hierarchical Bayesian binomial logistic regression model to the discrimination response data with stepsize as fixed effect, for each morph series separately,

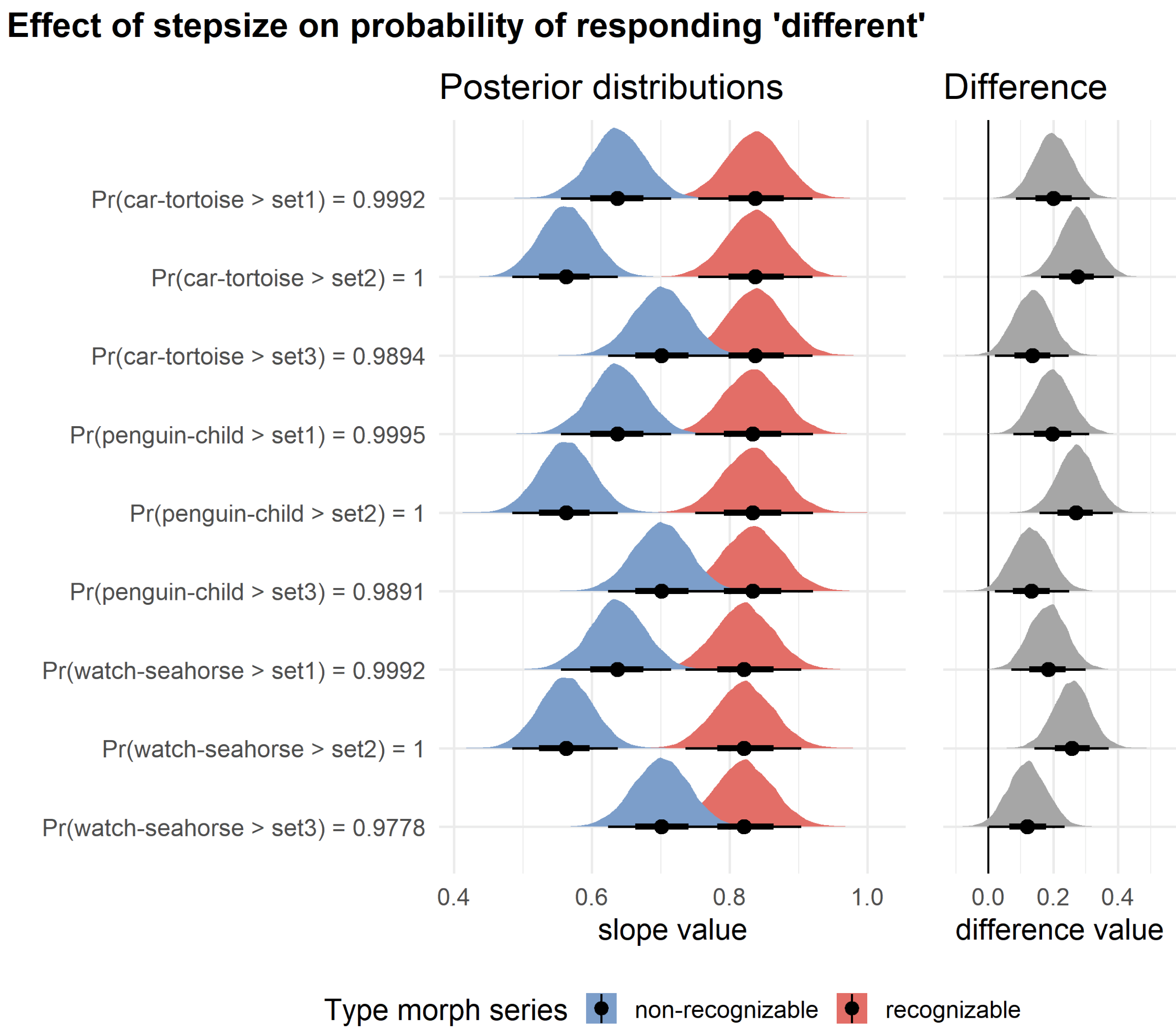


Figure A.3: Estimated pairwise differences between the posterior distributions for the effect of stepsize on the probability of responding 'different' in the successive discrimination task for each of the different recognizable and non-recognizable morph series combinations, in logodds units. Black dots and intervals indicate the mean, 66%, and 95% highest density continuous interval (HDCI) for each slope or difference value. The black vertical line indicates a difference in slope of zero. In this figure, the estimated effect of stepsize is larger for the recognizable than for the non-recognizable morph series. *Note.* In the interactive version of this figure, you can hover over the intervals to see the related mean and 95% HDCI for each distribution.

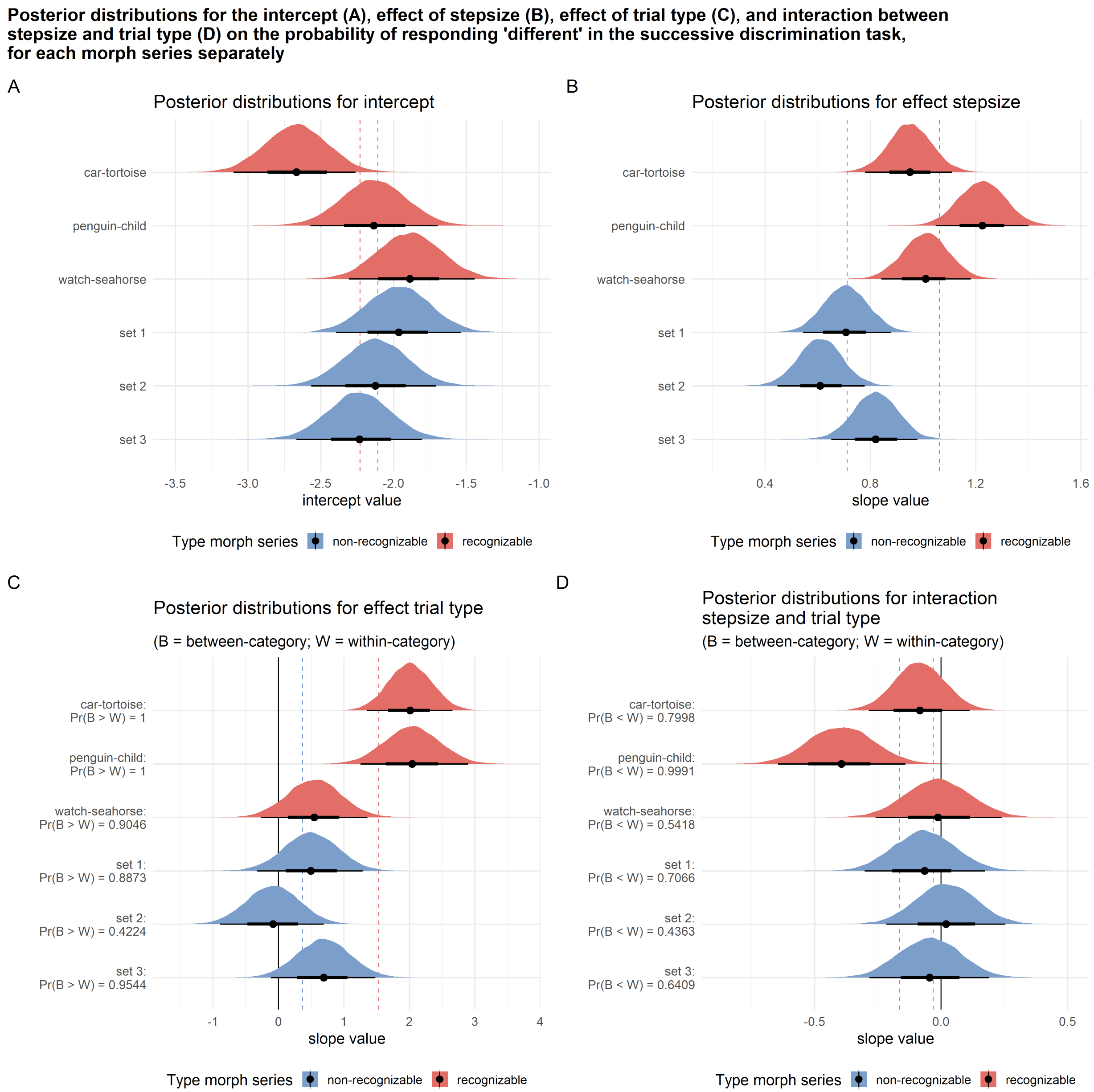


Figure A.4: Posterior distributions for the intercept (A), effect of stepsize (B), effect of trial type (C), and interaction between stepsize and trial type (D) on the probability of responding 'different' in the successive discrimination task, for each morph series separately, in logodds units. Black dots and intervals indicate the mean, 66%, and 95% highest density continuous interval (HDCI) for each intercept or slope value. The colored dashed vertical lines indicate the estimated mean value per type of morph series (recognizable vs. non-recognizable). The black vertical line indicates a difference in slope of zero. In this figure, the estimated effect of stepsize is larger for the recognizable than for the non-recognizable morph series (B). The main effect of trial type is larger for the recognizable series car-tortoise and penguin-child than for all non-recognizable morph series (C). The interaction effect between stepsize and trial type is more negative for the recognizable series penguin-child than for all non-recognizable series (D). *Note.* In the interactive version of this figure, you can hover over the intervals to see the related mean and 95% HDCI for each distribution.

and with ordered trial stimuli and participant ID within each morph series as random effects for intercept and participant ID within each morph series as random effect for the slope:

$$freqdiff \mid trials(n) \sim a + b * stepsize$$

$$\begin{aligned} a \sim 0 + morph\_series + (1 \mid q \mid morph\_series : trial\_stimuli\_ordered) \\ +(1 \mid p \mid morph\_series : pp\_id) \\ b \sim 0 + morph\_series + (1 \mid p \mid morph\_series : pp\_id) \end{aligned}$$

As priors, we specified a normal distribution with a mean of zero and a standard deviation of 1.5 for the intercept and a normal distribution with a mean of zero and a standard deviation of 0.5 for the slope. We used 4 chains consisting of 8000 iterations, with 4000 warmup iterations per chain.

Figure A.5 shows the aggregate behavioral data and posterior predictive distributions for responding "different" in the discrimination task, for each morph series, stepsize, and ordered stimulus pair separately.

A.2.4 Similarity judgments

To investigate the presence of differences in perceived similarity across stimulus pairs, we fitted a hierarchical Bayesian linear regression model to the by-participant-standardized similarity judgments with stepsize as fixed effect, for each morph series separately, and with trial stimuli and participant ID within each morph series as random effects for intercept and participant ID within each morph series as random effect for the slope:

$$z\_response \sim a + b * stepsize$$

$$\begin{aligned} a \sim 0 + morph\_series + (1 \mid q \mid morph\_series : trial\_stimuli) \\ +(1 \mid p \mid morph\_series : pp\_id) \\ b \sim 0 + morph\_series + (1 \mid p \mid morph\_series : pp\_id) \end{aligned}$$

As priors, we specified a normal distribution with a mean of zero and a standard deviation of 1.5 for the intercept and a normal distribution with a mean of zero and a standard deviation of 0.5 for the slope. We used 4 chains consisting of 8000 iterations, with 4000 warmup iterations per chain.

Figure A.6 shows the empirical similarity scores and posterior predictive distributions for the similarity judgments, per stepsize, trial type, and morph series, averaged across participants. Stepsize indicates the absolute difference in morph level between the two morph stimuli presented in a trial, with a minimum of zero (for same trials) and a maximum of eleven (when both extremes of the morph series are presented). In this supplementary figure, all stepsizes are shown.

Figure A.7 shows the posterior distributions for the intercept (A) and the effect of stepsize (B) on the standardized similarity judgments, for each morph series separately.

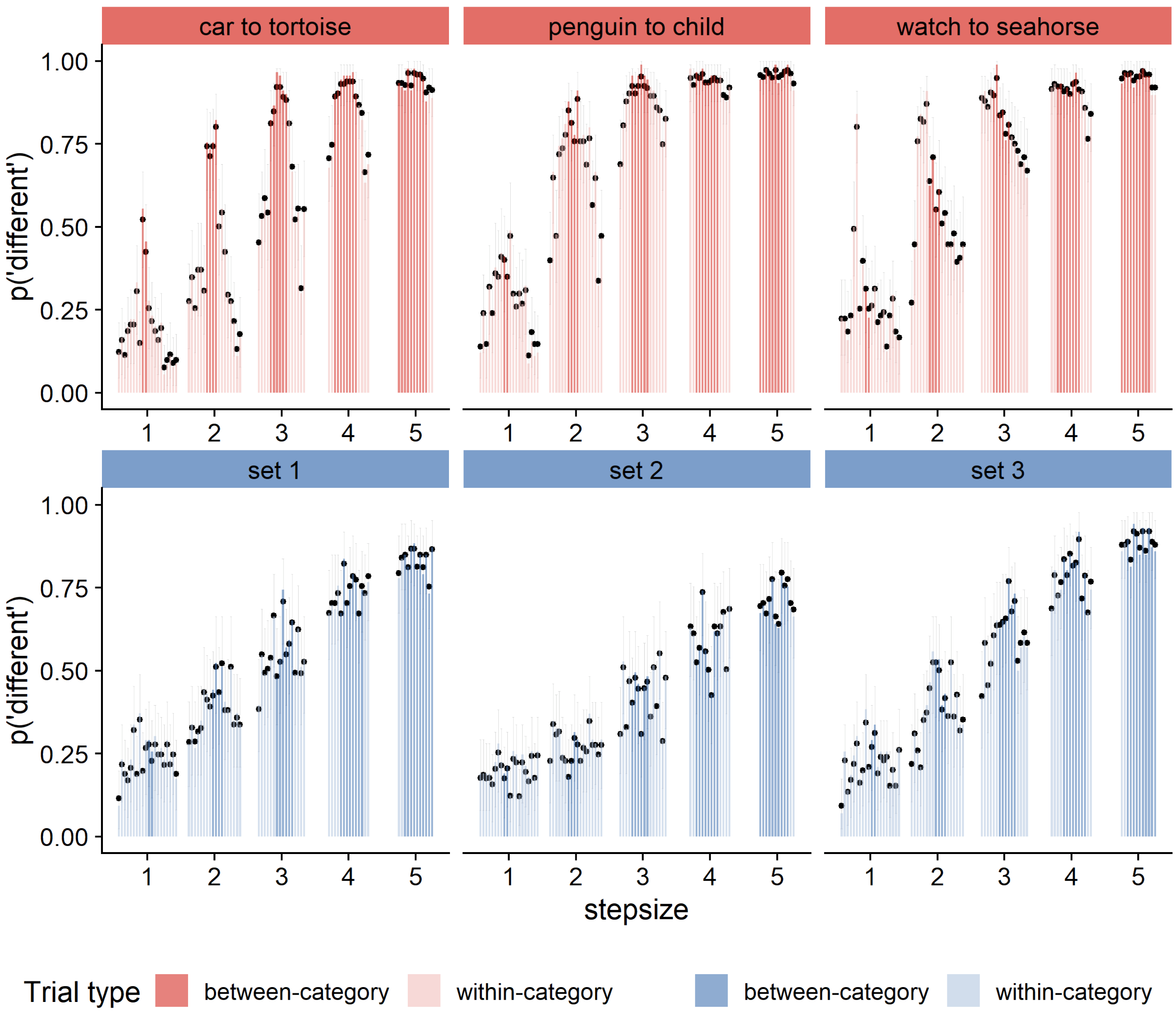


Figure A.5: Proportion of 'different' responses in the successive discrimination task for each stepsize, ordered stimulus pair, and morph series separately, averaged across participants. Stimulus pairs are ordered per stepsize and from left to right in the morph series as presented in Figure 3.3. Bars indicate the empirical proportions for responding 'different'. The black dots indicate the mean posterior predictions from the model and the grey error bars indicate the 95% highest density continuous intervals (HDCI) of the posterior predictive distributions. In this figure, no clear directional asymmetries (i.e., differences in discrimination performance based on the presentation order of the stimuli in the pair) are present. *Note.* In the interactive version of this figure, you can hover over the colored bars to see the stimuli involved in the pair (in the presented order from left to right), the exact percentage different responses, the mean and 95% HDCI of the posterior predictive distributions, and the number of trials related to each bar.

A

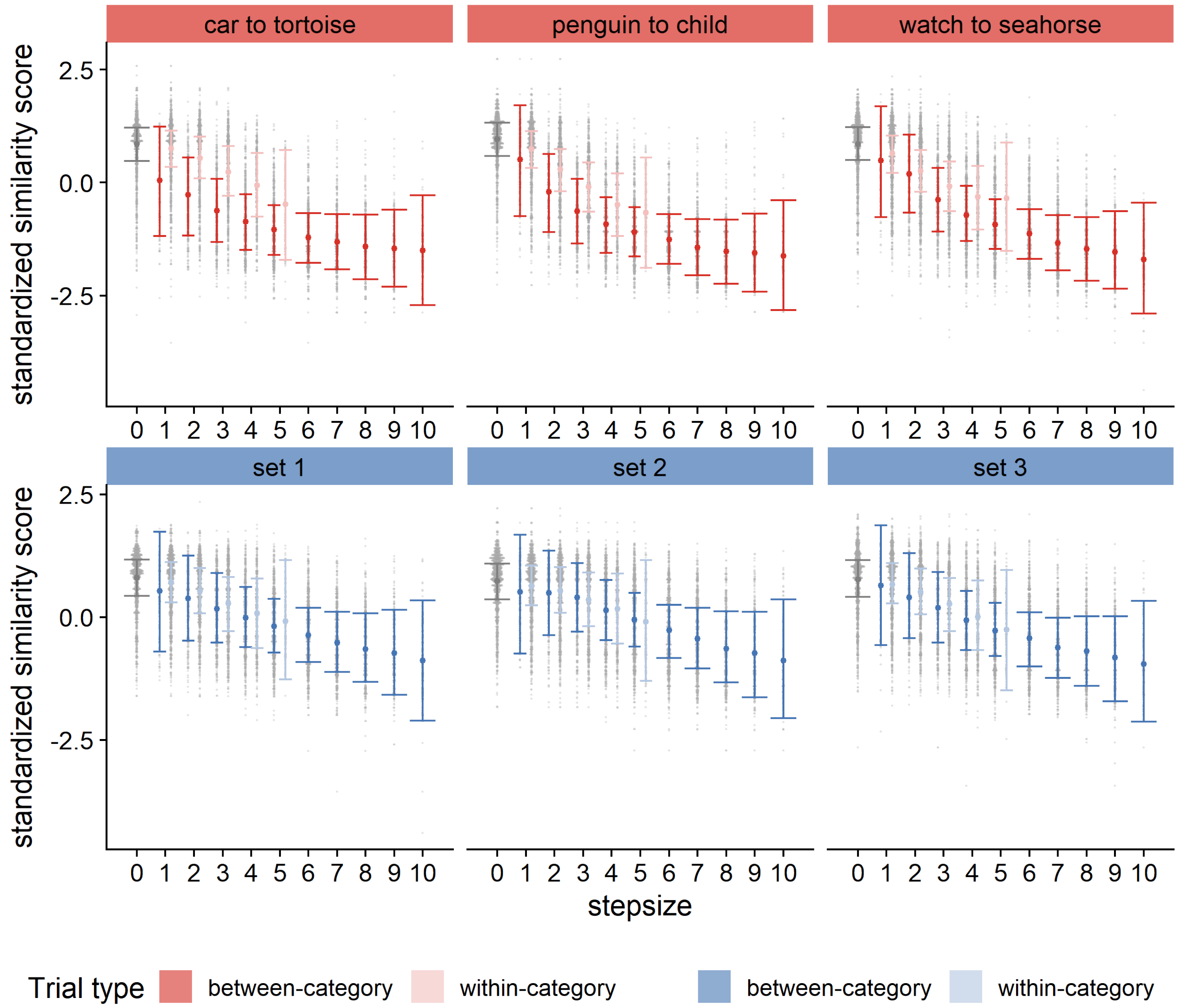


Figure A.6: Standardized similarity scores for each stepsize, trial type (i.e., between-category vs. within-category), and morph series separately, averaged across participants. Grey dots indicate the raw standardized similarity scores. For the conditions that contain less trials, these grey dots are not always clearly visible. The colored dots and error bars indicate the mean posterior predictions from the model and the 95% highest density continuous interval (HDCI) of the posterior predictive distributions. In this figure, the difference between the darker and the lighter intervals (i.e., the category boundary effect: between-category pairs rated as less similar than within-category pairs, keeping stepsize equal) is on average larger for the recognizable than for the non-recognizable morph series. *Note.* In the interactive version of this figure, you can hover over the intervals to see the exact similarity score, the mean and 95% HDCI of the posterior predictive distributions, and the number of trials related to each interval.

A

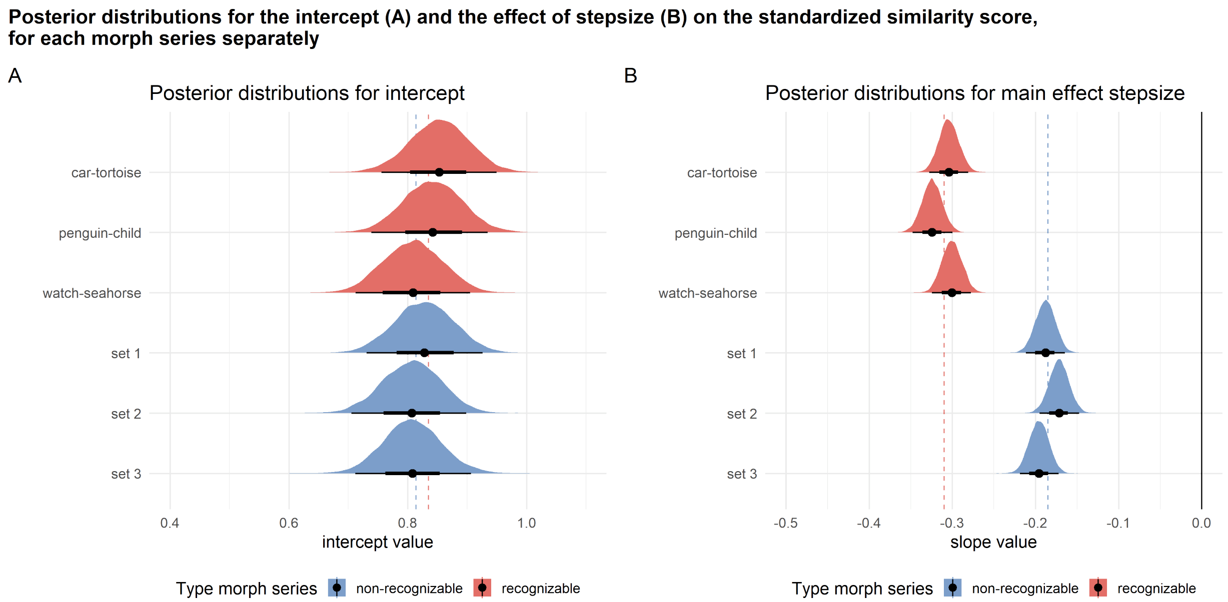


Figure A.7: Posterior distributions for the intercept (A) and the effect of stepsize (B) on the standardized similarity scores, for each morph series separately. Black dots and intervals indicate the mean, 66%, and 95% highest density continuous interval (HDCI) for each intercept or slope value. The colored dashed vertical lines indicate the estimated mean value per type of morph series (recognizable vs. non-recognizable). In this figure, the estimated effect of stepsize is larger (i.e., more different from zero) for the recognizable than for the non-recognizable morph series. *Note.* In the interactive version of this figure, you can hover over the intervals to see the related mean and 95% HDCI for each distribution.

Figure A.8 shows the estimated pairwise differences between the posterior distributions for the effect of stepsize on the similarity judgments, for each of the different recognizable and non-recognizable morph series combinations.

To investigate the presence of an overall category boundary effect, we fitted a hierarchical Bayesian linear regression model to the by-participant-standardized similarity judgments with stepsize as fixed effect, for each morph series and trial type separately, and with trial stimuli and participant ID within each morph series as random effects for intercept and participant ID within each morph series as random effect for the slope:

$$z\_response \sim a + b * stepsize$$

$$a \sim 0 + morph\_series * between\_category + (1 \mid q \mid morph\_series : trial\_stimuli)$$
$$+ (1 \mid p \mid morph\_series : pp\_id)$$

$$b \sim 0 + morph\_series * between\_category + (1 \mid p \mid morph\_series : pp\_id)$$

As priors, we specified a normal distribution with a mean of zero and a standard deviation of 1.5 for the intercept and a normal distribution with a mean of zero and a standard deviation of 0.5 for the slope. We used 4 chains consisting of 8000 iterations, with 4000 warmup iterations per chain.

Figure A.9 shows the posterior distributions for the intercept (A), effect of stepsize (B), effect of trial type (C), and interaction between stepsize and trial type (D) on the standardized similarity judgments, for each morph series separately.

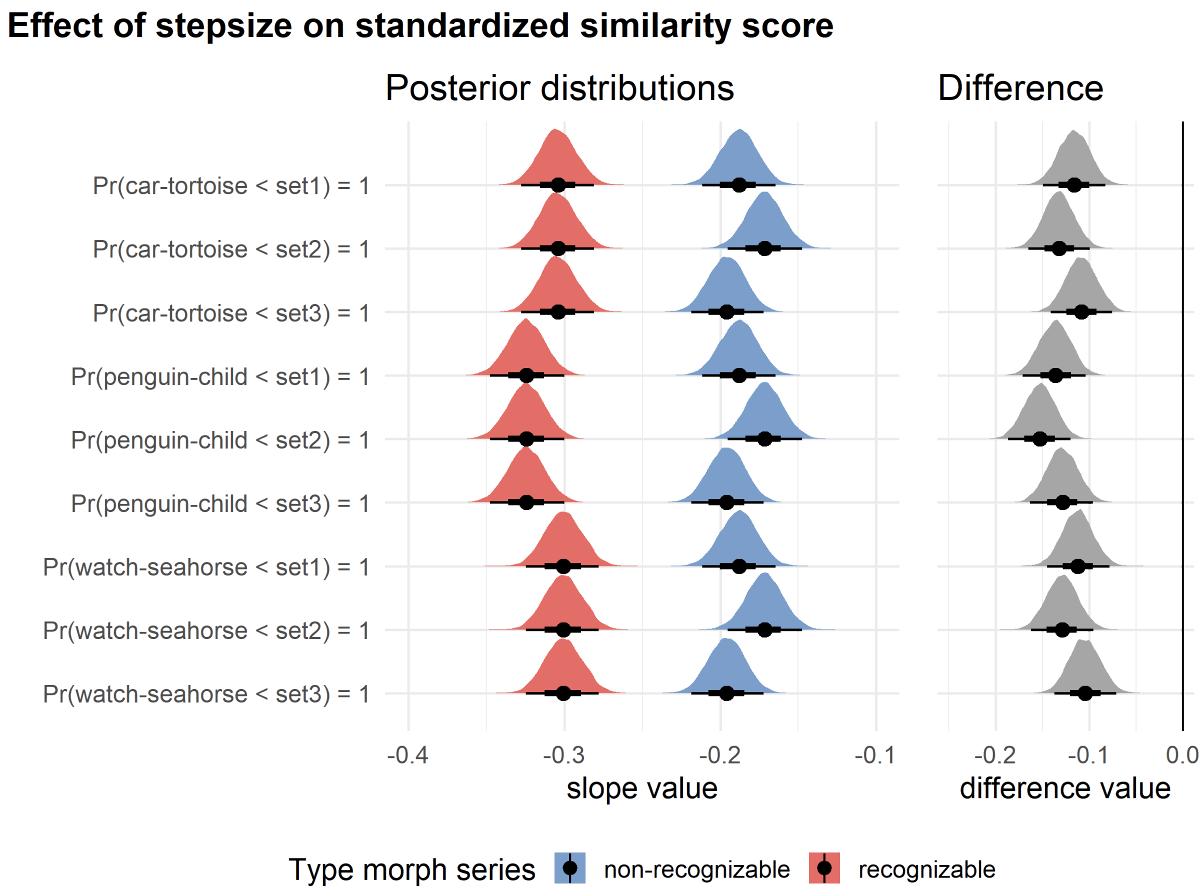


Figure A.8: Estimated pairwise differences between the posterior distributions for the effect of stepsize on the standardized similarity scores for each of the different recognizable and non-recognizable morph series combinations. Black dots and intervals indicate the mean, 66%, and 95% highest density continuous interval (HDCI) for each slope or difference value. The black vertical line indicates a difference in slope of zero. In this figure, the estimated effect of stepsize is larger (i.e., more different from zero) for the recognizable than for the non-recognizable morph series. *Note.* In the interactive version of this figure, you can hover over the intervals to see the related mean and 95% HDCI for each distribution.

A

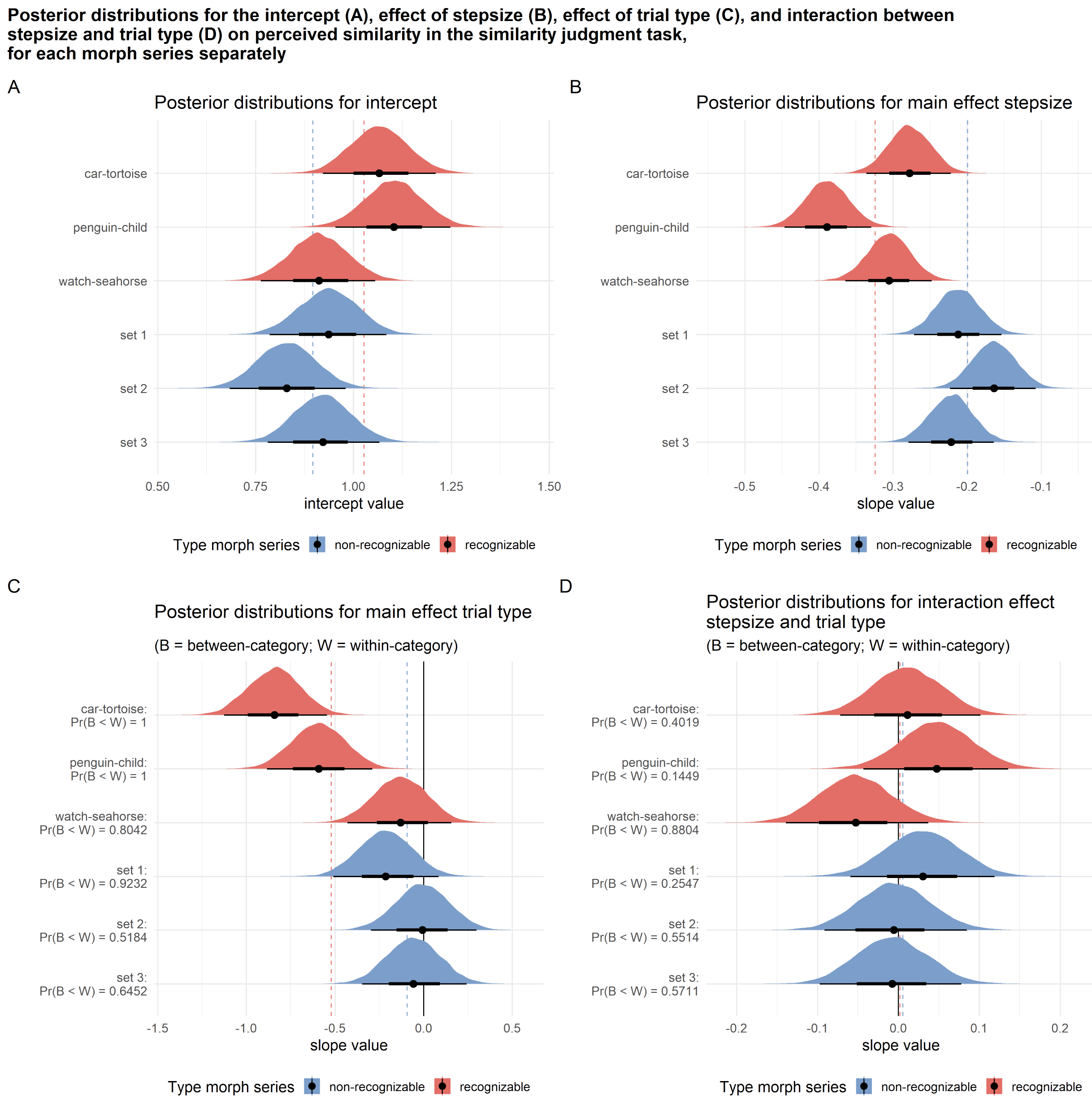


Figure A.9: Posterior distributions for the intercept (A), effect of stepsize (B), effect of trial type (C), and interaction between stepsize and trial type (D) on the standardized similarity scores, for each morph series separately. Black dots and intervals indicate the mean, 66%, and 95% highest density continuous interval (HDCI) for each intercept or slope value. The colored dashed vertical lines indicate the estimated mean value per type of morph series (recognizable vs. non-recognizable). The black vertical line indicates a difference in slope of zero. In this figure, the estimated effect of stepsize is larger (i.e., more different from zero) for the recognizable than for the non-recognizable morph series (B). The main effect of trial type is larger (i.e., more different from zero) for the recognizable series car-tortoise and penguin-child than for all non-recognizable morph series (C). The interaction effect between stepsize and trial type is larger (i.e., more different from zero) for the recognizable series watch-seahorse than for all non-recognizable series (D). *Note.* In the interactive version of this figure, you can hover over the intervals to see the related mean and 95% HDCI for each distribution.

A

To investigate the presence of directional asymmetries, we fitted a hierarchical Bayesian linear regression model to the by-participant-standardized similarity judgments with stepsize as fixed effect, for each morph series separately, and with ordered trial stimuli and participant ID within each morph series as random effects for intercept and participant ID within each morph series as random effect for the slope:

$$z\_response \sim a + b * stepsize$$

$$\begin{aligned} a \sim 0 + morph\_series + (1 \mid q \mid morph\_series : trial\_stimuli\_ordered) \\ + (1 \mid p \mid morph\_series : pp\_id) \end{aligned}$$

$$b \sim 0 + morph\_series + (1 \mid p \mid morph\_series : pp\_id)$$

As priors, we specified a normal distribution with a mean of zero and a standard deviation of 1.5 for the intercept and a normal distribution with a mean of zero and a standard deviation of 0.5 for the slope. We used 4 chains consisting of 8000 iterations, with 4000 warmup iterations per chain.

Figure A.10 shows the posterior predictive distributions for the standardized similarity judgments, for each morph series, step size and ordered stimulus pair separately.

A

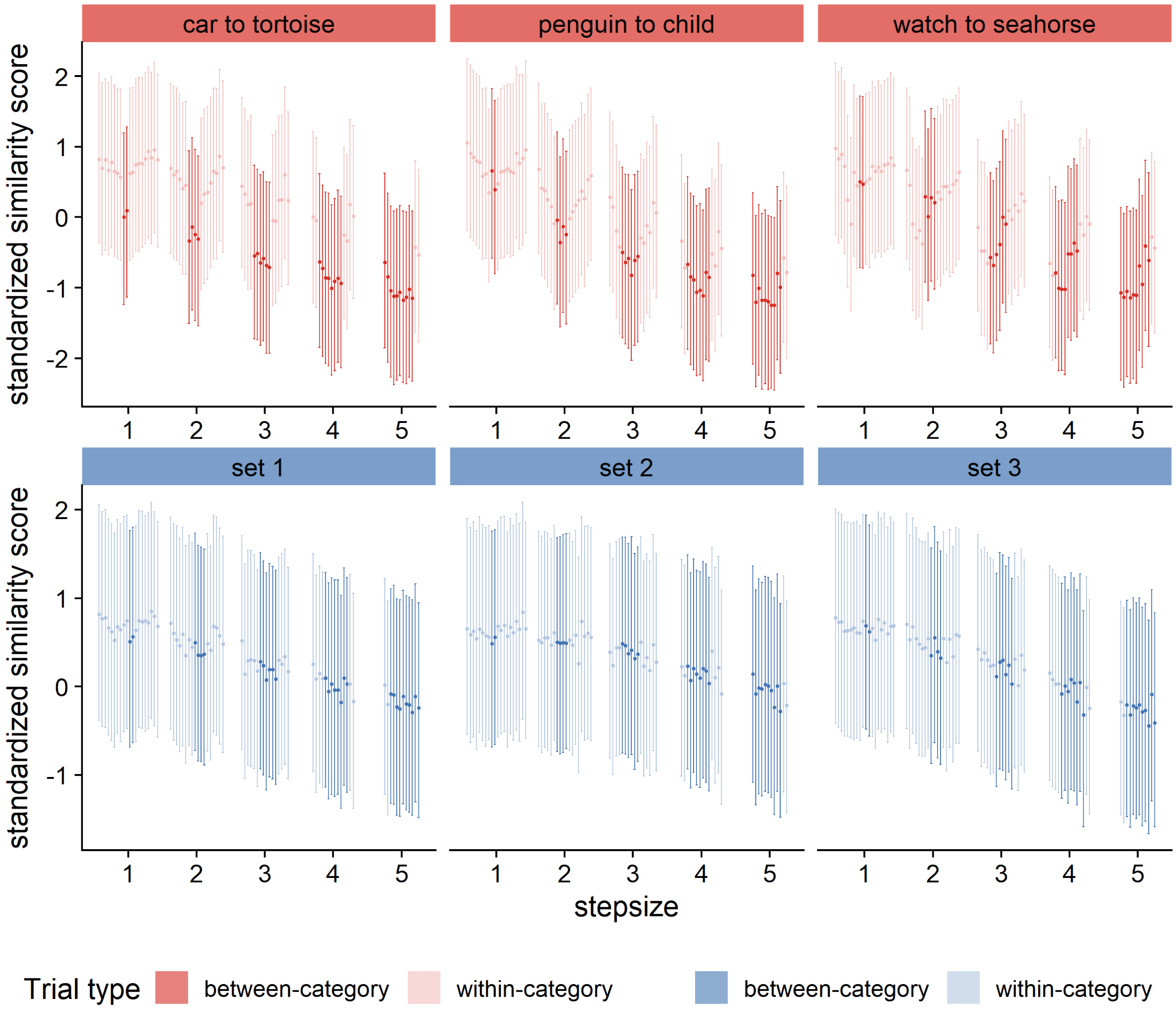


Figure A.10: Posterior predictive distributions for the responses to the successive similarity judgment task, for each morph series, stepsize, and ordered stimulus pair separately. Colored dots and error bars indicate the mean posterior predictions from the model and the 95% highest density continuous intervals of the posterior predictive distributions. In this figure, no clear directional asymmetries (i.e., differences in perceived similarity based on the presentation order of the stimuli in the pair) are present. *Note*. In the interactive version of this figure, you can hover over the mean posterior predictions to see the exact percentage different responses and the number of trials related to each data point.

A

A.3 Supplemental videos

In the HTML version of this Appendix, you can find screen recordings of some trials for each task that was part of this study (cf. Figure A.11).

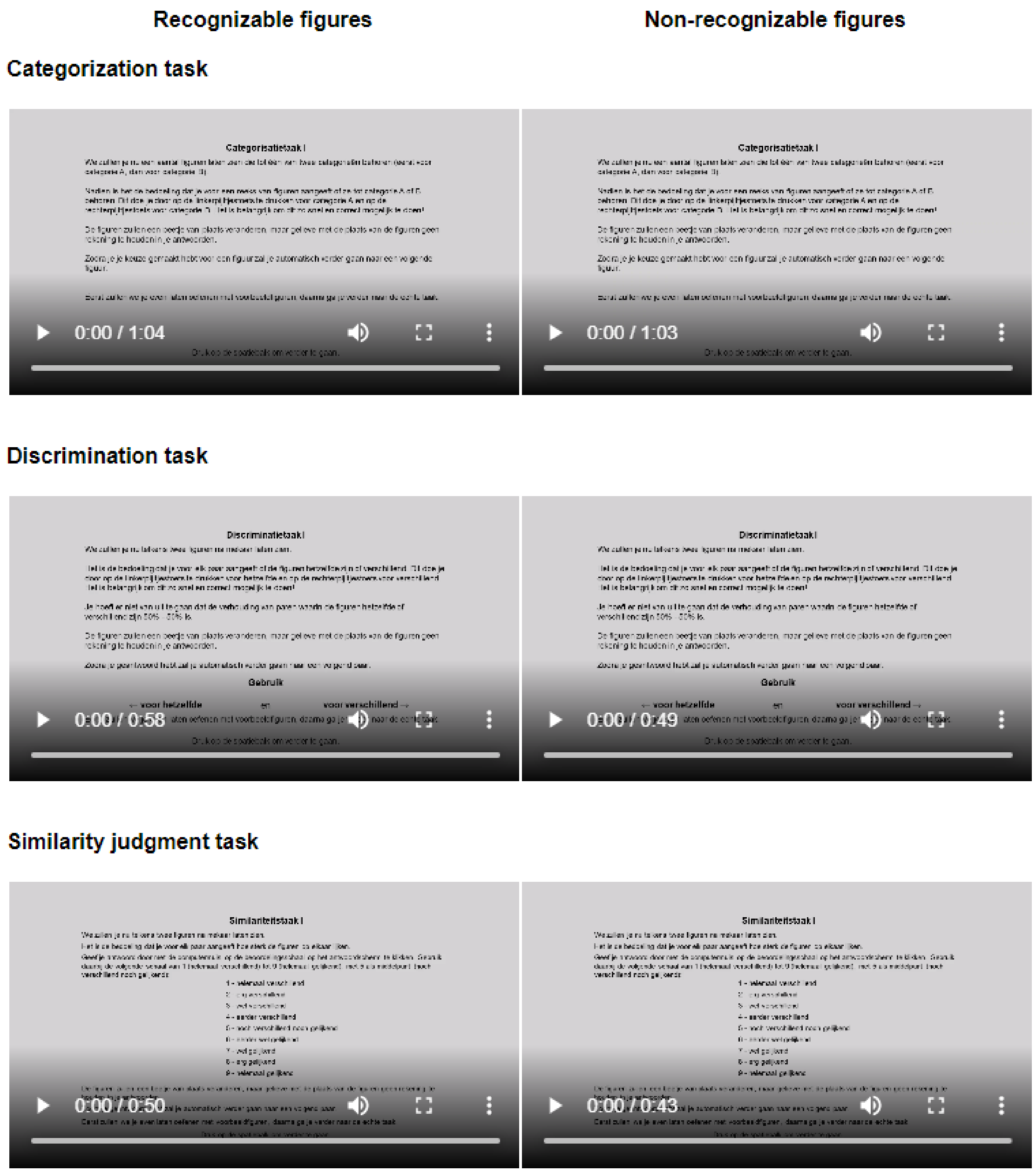


Figure A.11: Screen recordings of some trials for the categorization, discrimination, and similarity judgment tasks that were part of this study.

Appendix B
Individual differences in hysteresis and adaptation

B.1 R packages used

As indicated in the Methods section, for all our analyses we used R (Version 4.0.4; R Core Team, 2021) and the R-packages *BayesFactor* (Version 0.9.12.4.2; Morey & Rouder, 2018), *brms* (Version 2.16.1; Bürkner, 2017, 2018, 2021), *coda* (Version 0.19.4; Plummer et al., 2006), *cowplot* (Version 1.1.1; Wilke, 2020), *devtools* (Version 2.4.4; Wickham, Hester, et al., 2021), *dplyr* (Version 1.0.10; Wickham, François, et al., 2022), *ellipse* (Version 0.4.2; Murdoch & Chow, 2020), *forcats* (Version 0.5.2; Wickham, 2022a), *ggforce* (Version 0.3.3; Pedersen, 2021), *ggimage* (Version 0.3.1; Yu, 2022), *ggiraph* (Version 0.7.10; Gohel & Skintzos, 2021), *ggplot2* (Version 3.3.6; Wickham, 2016), *ggstance* (Version 0.3.5; Henry et al., 2020), *ggthemes* (Version 4.2.4; Arnold, 2021), *glue* (Version 1.6.2; Hester & Bryan, 2022), *here* (Version 1.0.1; Müller, 2020), *kableExtra* (Version 1.3.4; Zhu, 2021), *knitr* (Version 1.39; Xie, 2015), *lubridate* (Version 1.8.0; Grolemund & Wickham, 2011), *Matrix* (Version 1.3.2; Bates & Maechler, 2021), *modelr* (Version 0.1.9; Wickham, 2022b), *papaja* (Version 0.1.1; Aust & Barth, 2022), *patchwork* (Version 1.1.2; Pedersen, 2022), *purrr* (Version 0.3.4; Henry & Wickham, 2020), *Rcpp* (Eddelbuettel & Balamuta, 2018; Version 1.0.9; Eddelbuettel & François, 2011), *readr* (Version 2.1.2; Wickham, Hester, et al., 2022), *readxl* (Version 1.4.1; Wickham & Bryan, 2022), *rstan* (Version 2.21.2; Stan Development Team, 2020a), *StanHeaders* (Version 2.21.0.7; Stan Development Team, 2020b), *stringr* (Version 1.4.0; Wickham, 2019), *tibble* (Version 3.1.8; Müller & Wickham, 2022), *tidybayes* (Version 3.0.1; Kay, 2021b), *tidyr* (Version 1.2.1; Wickham & Girlich, 2022), *tidyverse* (Version 1.3.2; Wickham et al., 2019), *tinylabels* (Version 0.2.3; Barth, 2022), and *usethis* (Version 2.1.6; Wickham, Bryan, et al., 2021).

B.2 Supplementary Figures

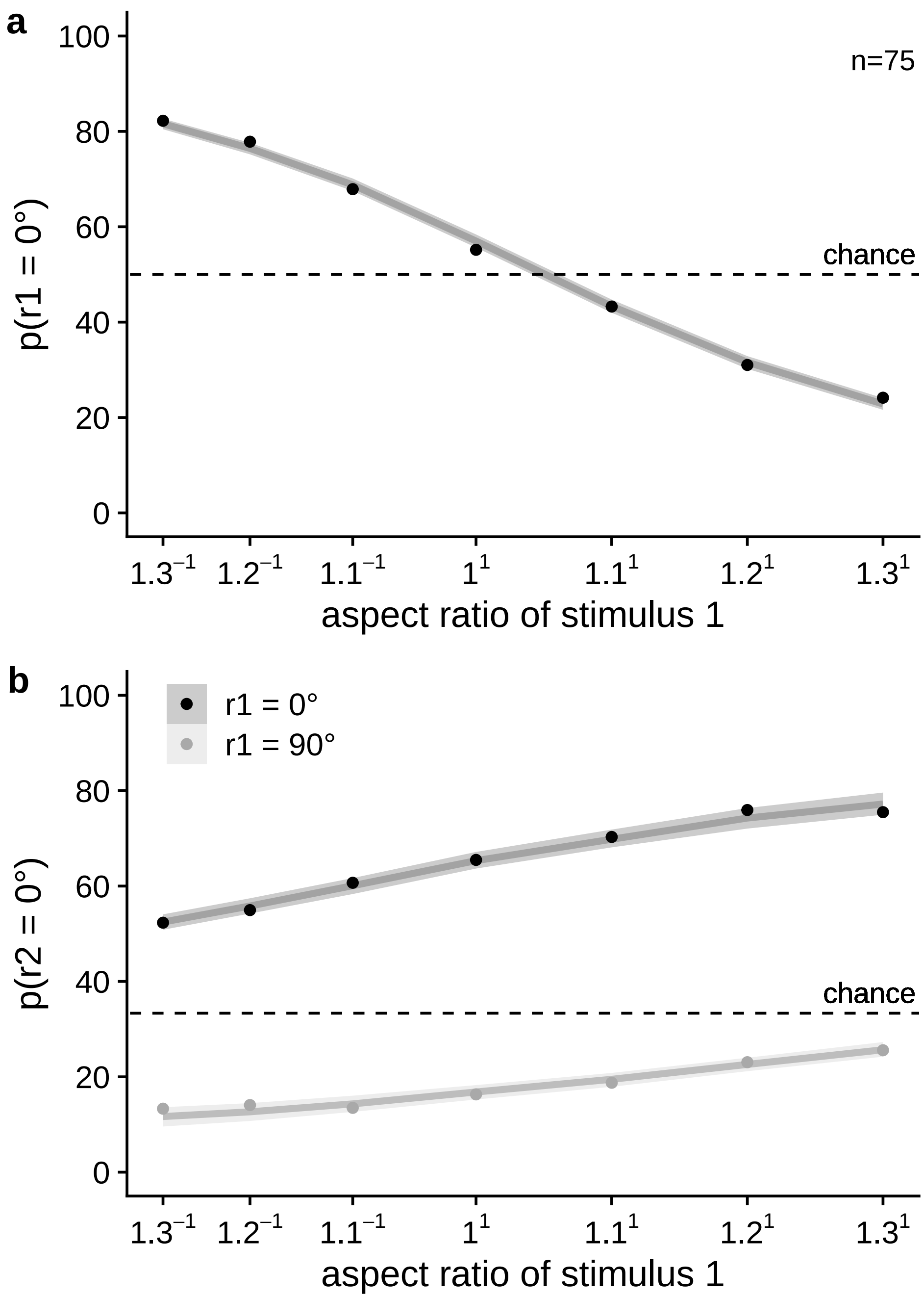


Figure B.1: (a) Mean response to the first stimulus dependent on aspect ratio (percentage). The probability of responding 0° to the first stimulus decreases with aspect ratio (|a|/|b|). The value of aspect ratio increases with increasing distance in the 0°-orientation, leading to more 90° responses. (b) Mean response to the second stimulus dependent on aspect ratio (percentage). The probability of responding 0° to the second stimulus increases with aspect ratio (|a|/|b|; i.e., adaptation effect), and increases when the first stimulus was perceived as 0° rather than 90° (i.e., hysteresis effect). Dots indicate observed values. In addition, mean posterior predictions and their 95% highest density continuous intervals are shown.

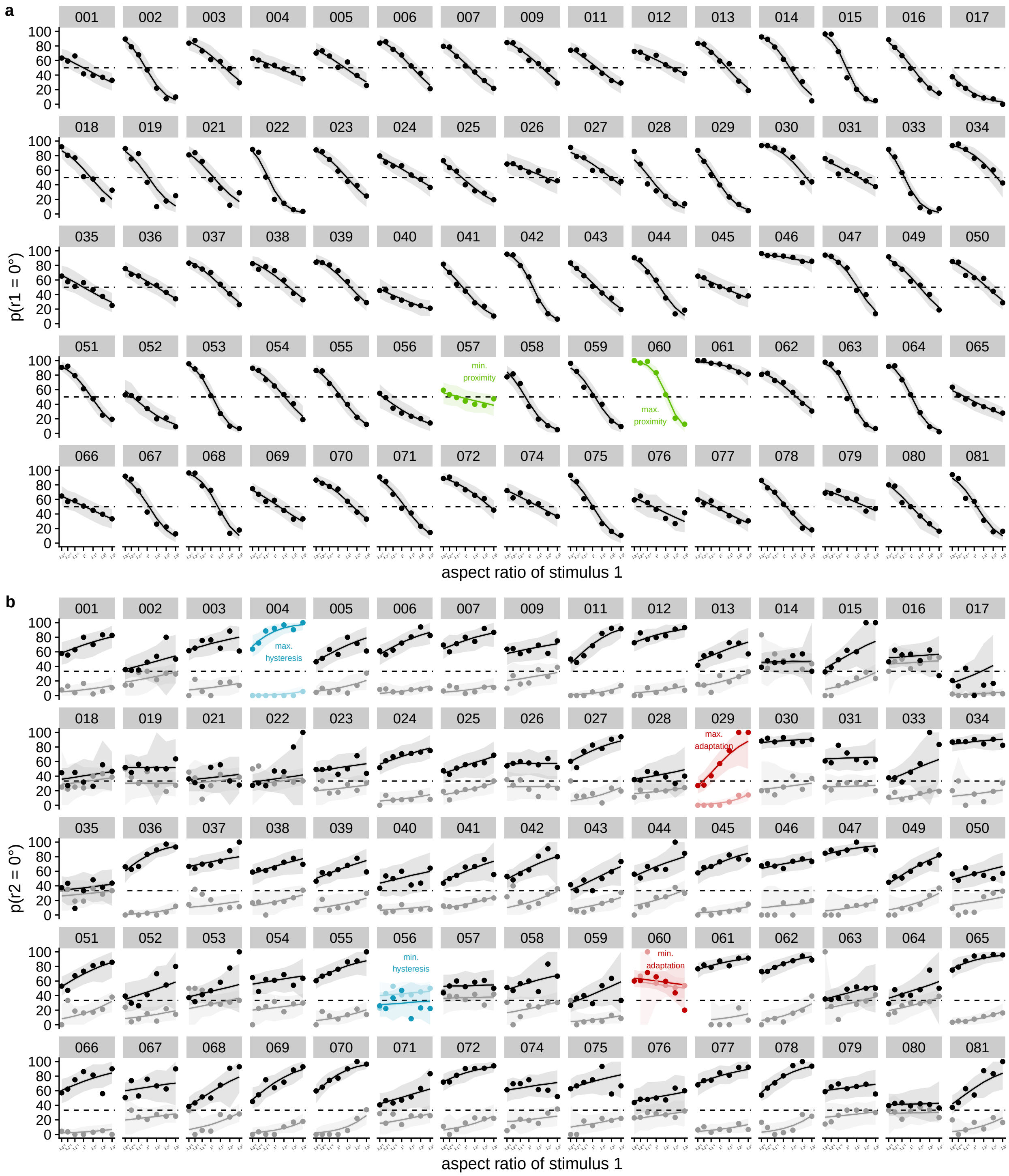


Figure B.2: (a) Mean individual responses to the first stimulus dependent on aspect ratio (percentage). The probability of responding 0° to the first stimulus decreases with aspect ratio (|a|/|b|). The value of aspect ratio increases with increasing distance in the 0°-orientation, leading to more 90° responses. Dots indicate observed values. In addition, mean posterior predictions and their 95% highest density continuous intervals are shown. Plots for participants with the smallest and largest estimated proximity effect are indicated in red. (b) Mean individual responses to the second stimulus dependent on aspect ratio (percentage). The probability of responding 0° to the second stimulus increases with aspect ratio (|a|/|b|; i.e., adaptation effect), and increases when the first stimulus was perceived as 0° rather than 90° (i.e., hysteresis effect). Dots indicate observed values. In addition, mean posterior predictions and their 95% highest density continuous intervals are shown. Plots for participants with the smallest and largest estimated hysteresis effect are indicated in blue, participants with the smallest and largest estimated adaptation effect are indicated in red.

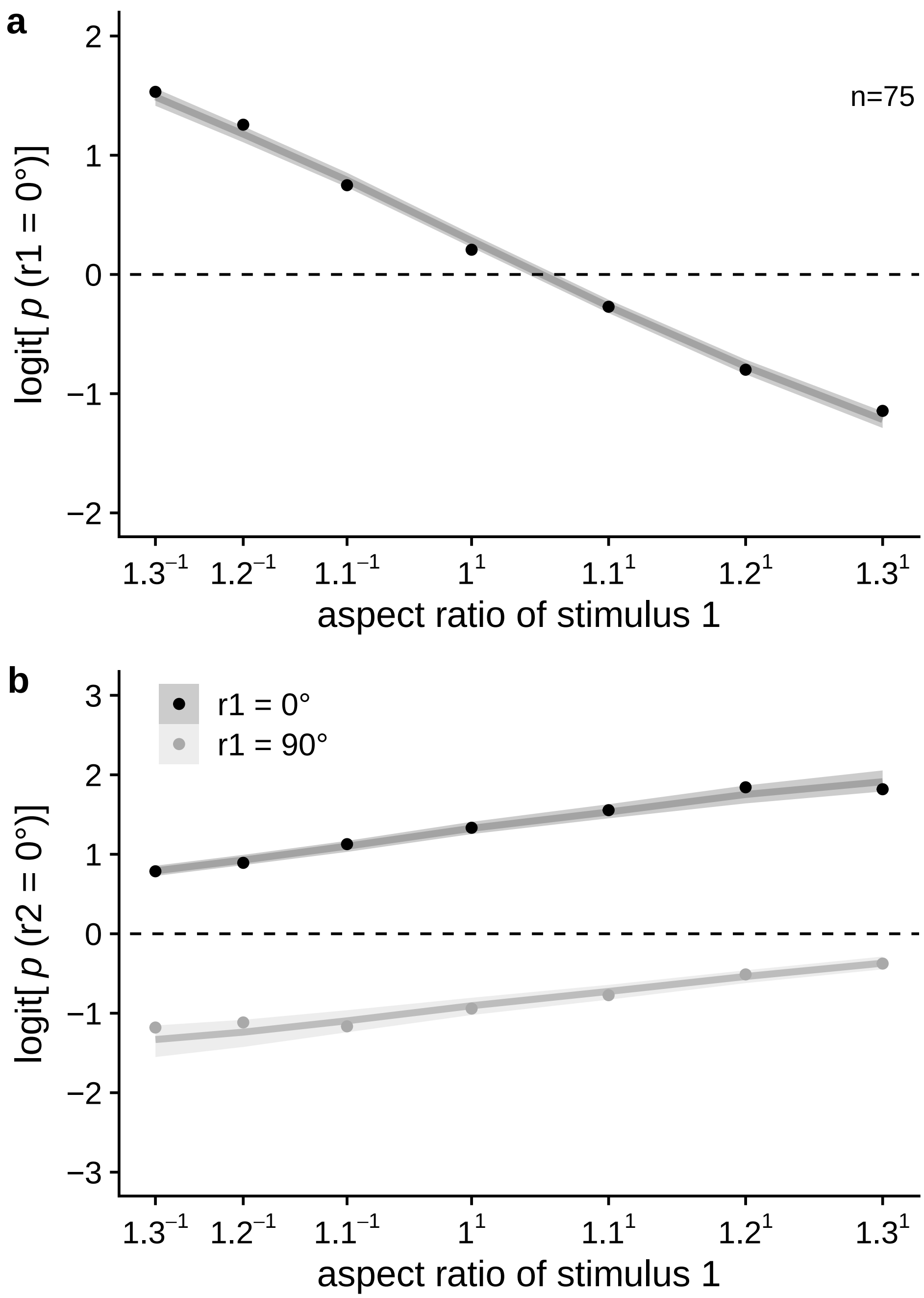


Figure B.3: (a) Mean response to the first stimulus dependent on aspect ratio (logit as used by Gepshtein & Kubovy, 2005, and Schwiedrzik et al., 2014). The probability of responding 0° to the first stimulus decreases with aspect ratio (|a|/|b|). The value of aspect ratio increases with increasing distance in the 0°-orientation, leading to more 90° responses. (b) Mean response to the second stimulus dependent on aspect ratio Schwiedrzik et al. (2014). The probability of responding 0° to the second stimulus increases with aspect ratio (|a|/|b|; i.e., adaptation effect), and increases when the first stimulus was perceived as 0° rather than 90° (i.e., hysteresis effect). Dots indicate observed values. In addition, mean posterior predictions and their 95% highest density continuous intervals are shown.

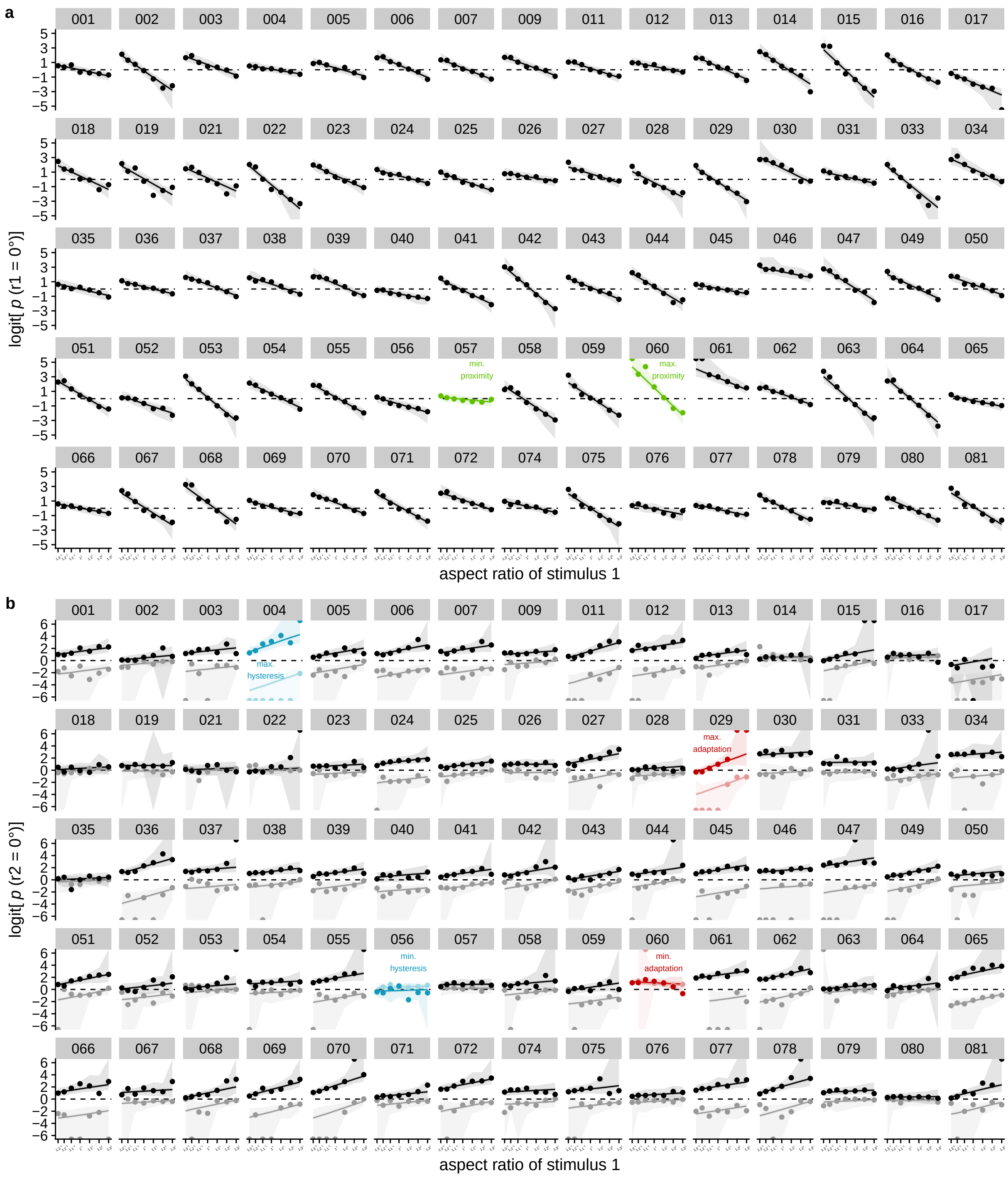


Figure B.4: (a) Mean individual responses to the first stimulus dependent on aspect ratio (logit as used by Gepshtein & Kubovy, 2005, and Schwiedrzik et al., 2014). The probability of responding 0° to the first stimulus decreases with aspect ratio (|a|/|b|). The value of aspect ratio increases with increasing distance in the 0°-orientation, leading to more 90° responses. Dots indicate observed values. In addition, mean posterior predictions and their 95% highest density continuous intervals are shown. Plots for participants with the smallest and largest estimated proximity effect are indicated in red. (b) Mean individual responses to the second stimulus dependent on aspect ratio Schwiedrzik et al. (2014). The probability of responding 0° to the second stimulus increases with aspect ratio (|a|/|b|; i.e., adaptation effect), and increases when the first stimulus was perceived as 0° rather than 90° (i.e., hysteresis effect). Dots indicate observed values. In addition, mean posterior predictions and their 95% highest density continuous intervals are shown. Plots for participants with the smallest and largest estimated hysteresis effect are indicated in blue, participants with the smallest and largest estimated adaptation effect are indicated in red.

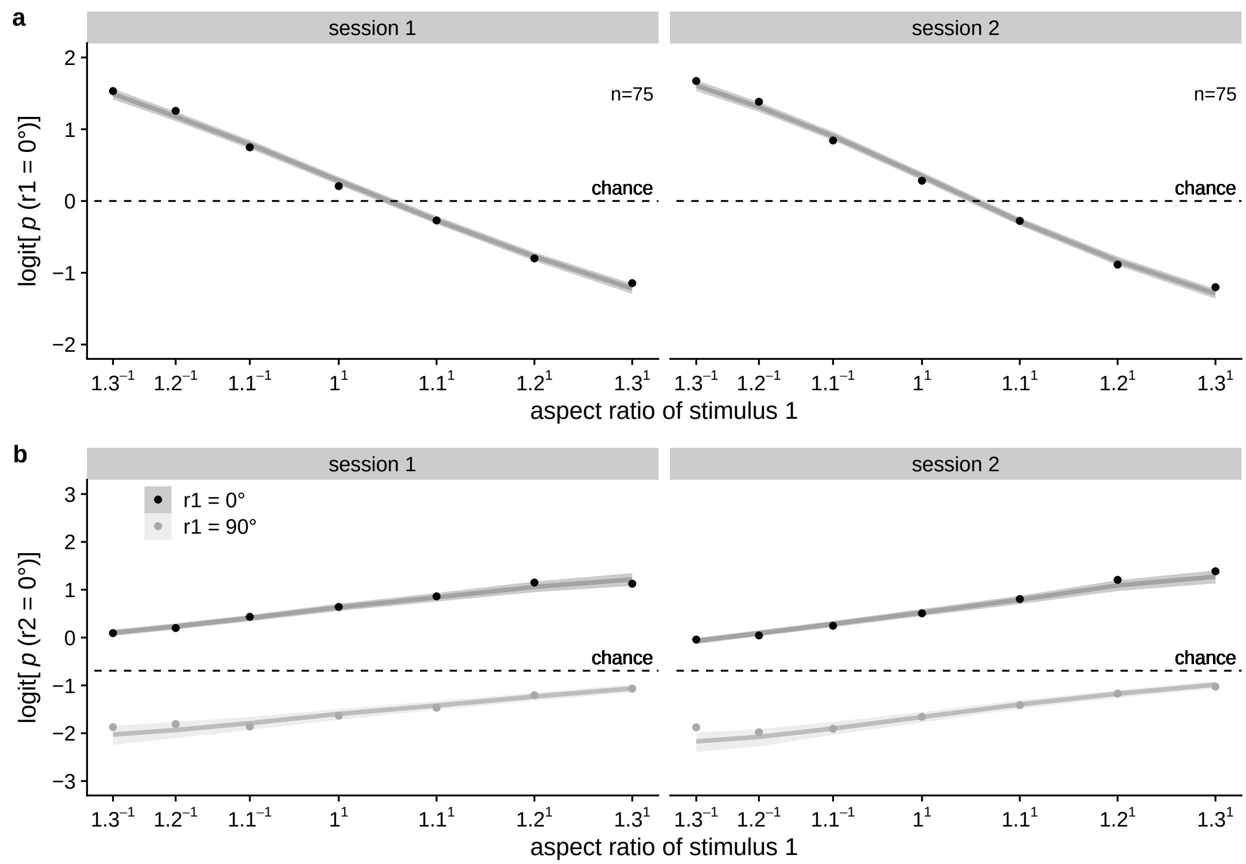


Figure B.5: (a) Mean response to the first stimulus dependent on aspect ratio (logit), for both the first and the second session. The probability of responding 0° to the first stimulus decreases with aspect ratio (|a|/|b|). The value of aspect ratio increases with increasing distance in the 0°-orientation, leading to more 90° responses. (b) Mean response to the second stimulus dependent on aspect ratio (logit), for both the first and the second session. The probability of responding 0° to the second stimulus increases with aspect ratio (|a|/|b|; i.e., adaptation effect), and increases when the first stimulus was perceived as 0° rather than 90° (i.e., hysteresis effect). Dots indicate observed values. In addition, mean posterior predictions and their 95% highest density continuous intervals are shown.

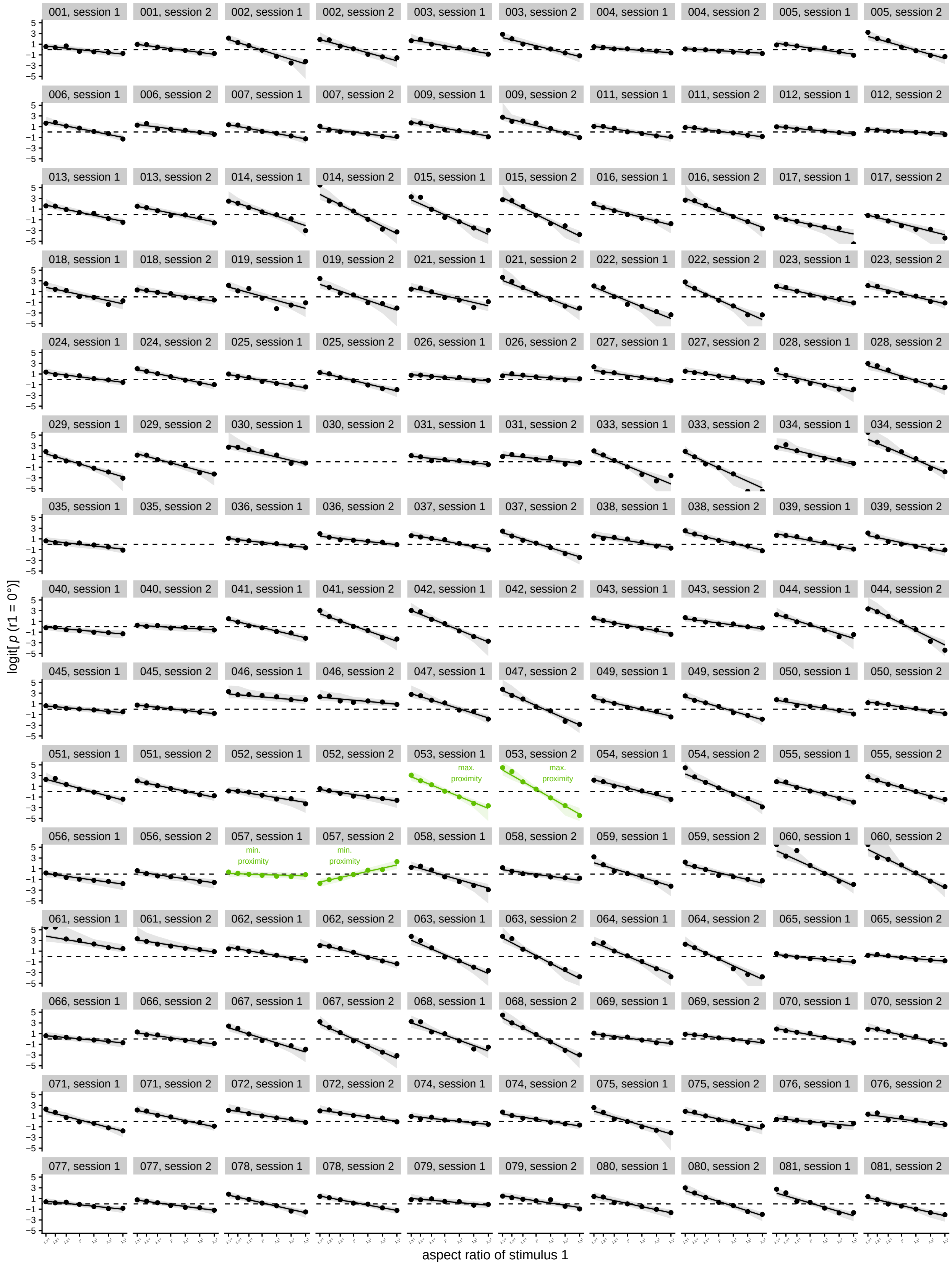


Figure B.6: Mean individual responses to the first stimulus dependent on aspect ratio (logit), for both the first and the second session. The probability of responding 0° to the first stimulus decreases with aspect ratio (|a|/|b|). The value of aspect ratio increases with increasing distance in the 0°-orientation, leading to more 90° responses. Dots indicate observed values. In addition, mean posterior predictions and their 95% highest density continuous intervals are shown. Plots for participants with the smallest and largest estimated proximity effect are indicated in green.

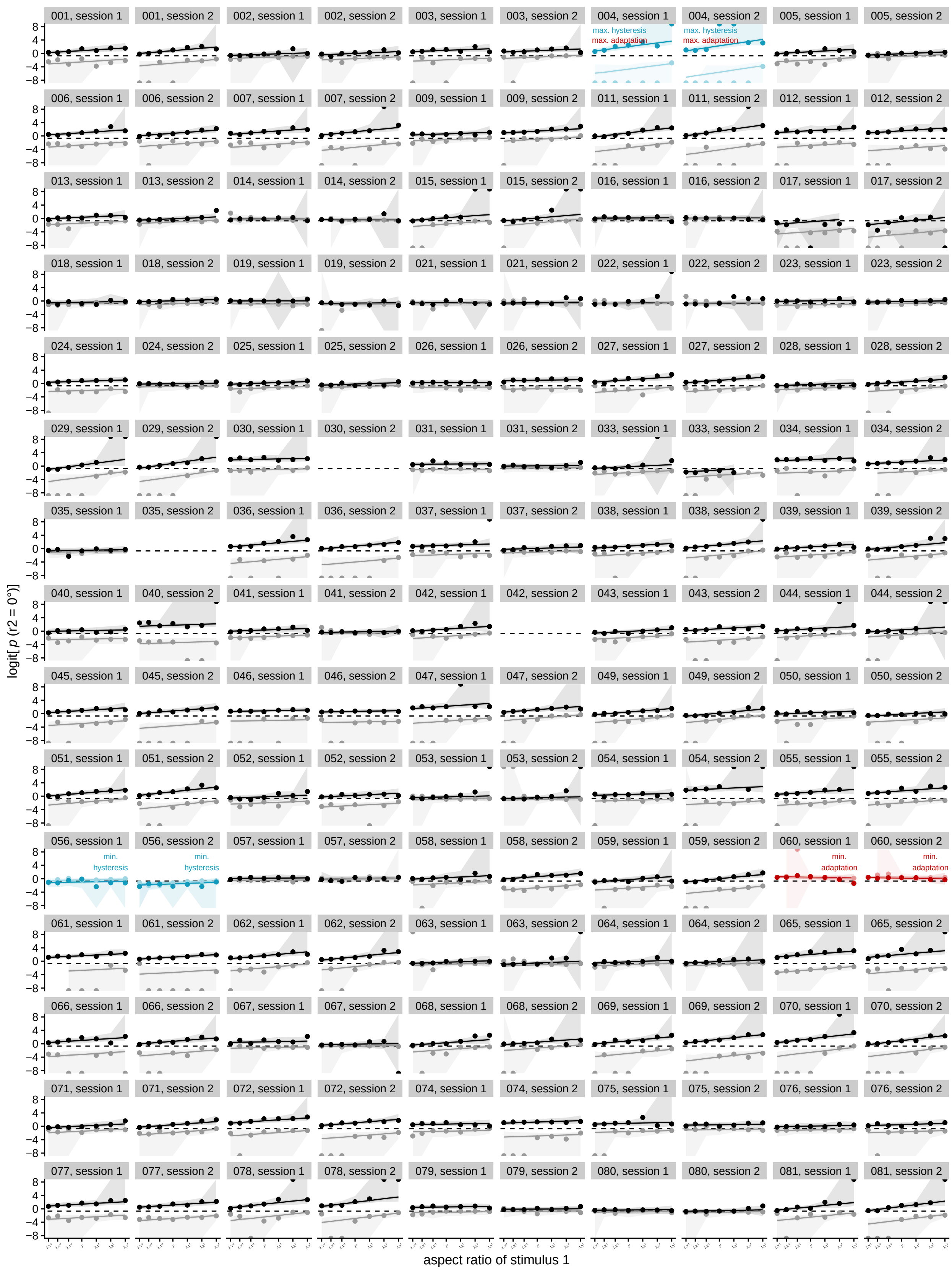


Figure B.7: Mean individual responses to the second stimulus dependent on aspect ratio (logit), for both the first and the second session. The probability of responding 0° to the second stimulus increases with aspect ratio (|a|/|b|; i.e., adaptation effect), and increases when the first stimulus was perceived as 0° rather than 90° (i.e., hysteresis effect). Dots indicate observed values. In addition, mean posterior predictions and their 95% highest density continuous intervals are shown. Plots for participants with the smallest and largest estimated hysteresis effect are indicated in blue, participants with the smallest and largest estimated adaptation effect are indicated in red.

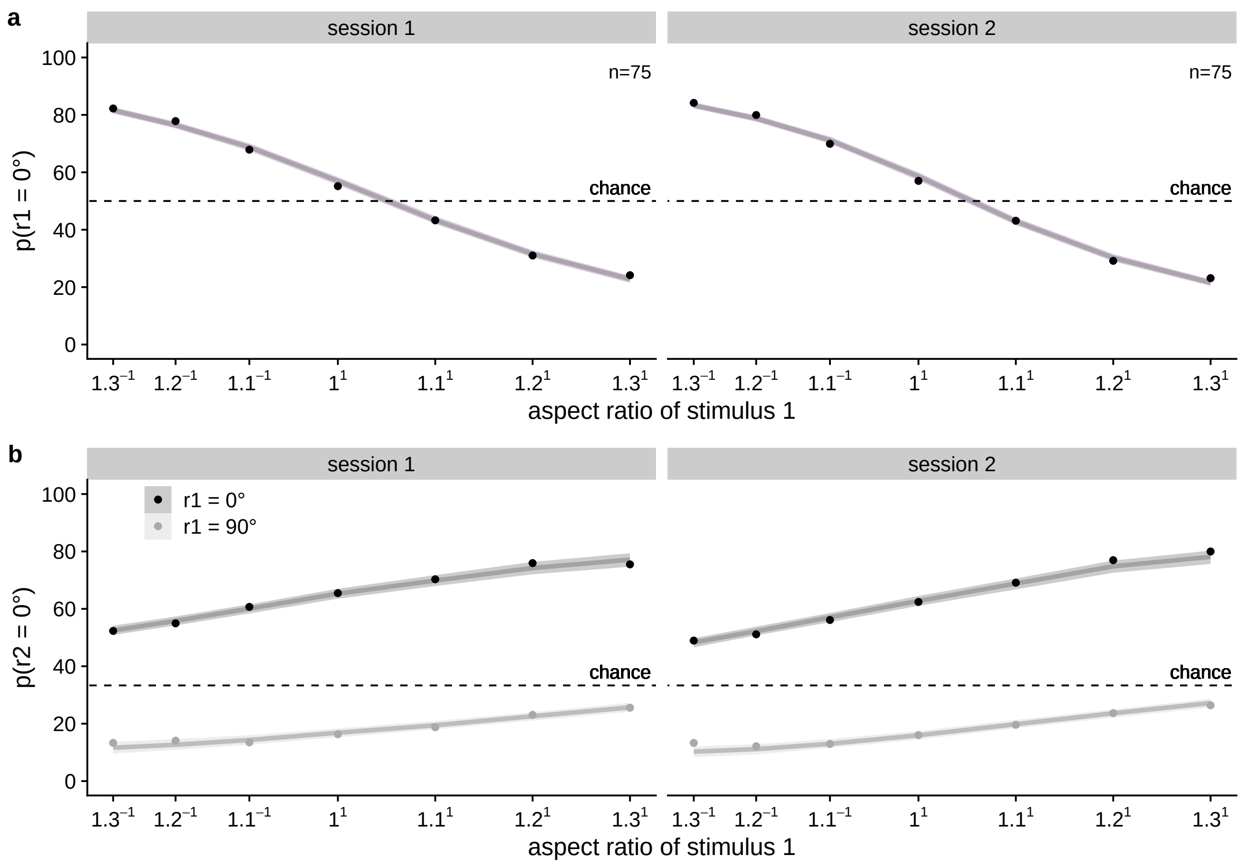


Figure B.8: (a) Mean response to the first stimulus dependent on aspect ratio (percentage), for both the first and the second session. The probability of responding 0° to the first stimulus decreases with aspect ratio (|a|/|b|). The value of aspect ratio increases with increasing distance in the 0°-orientation, leading to more 90° responses. (b) Mean response to the second stimulus dependent on aspect ratio (probability), for both the first and the second session. The probability of responding 0° to the second stimulus increases with aspect ratio (|a|/|b|; i.e., adaptation effect), and increases when the first stimulus was perceived as 0° rather than 90° (i.e., hysteresis effect). Dots indicate observed values. In addition, mean posterior predictions and their 95% highest density continuous intervals are shown.

B

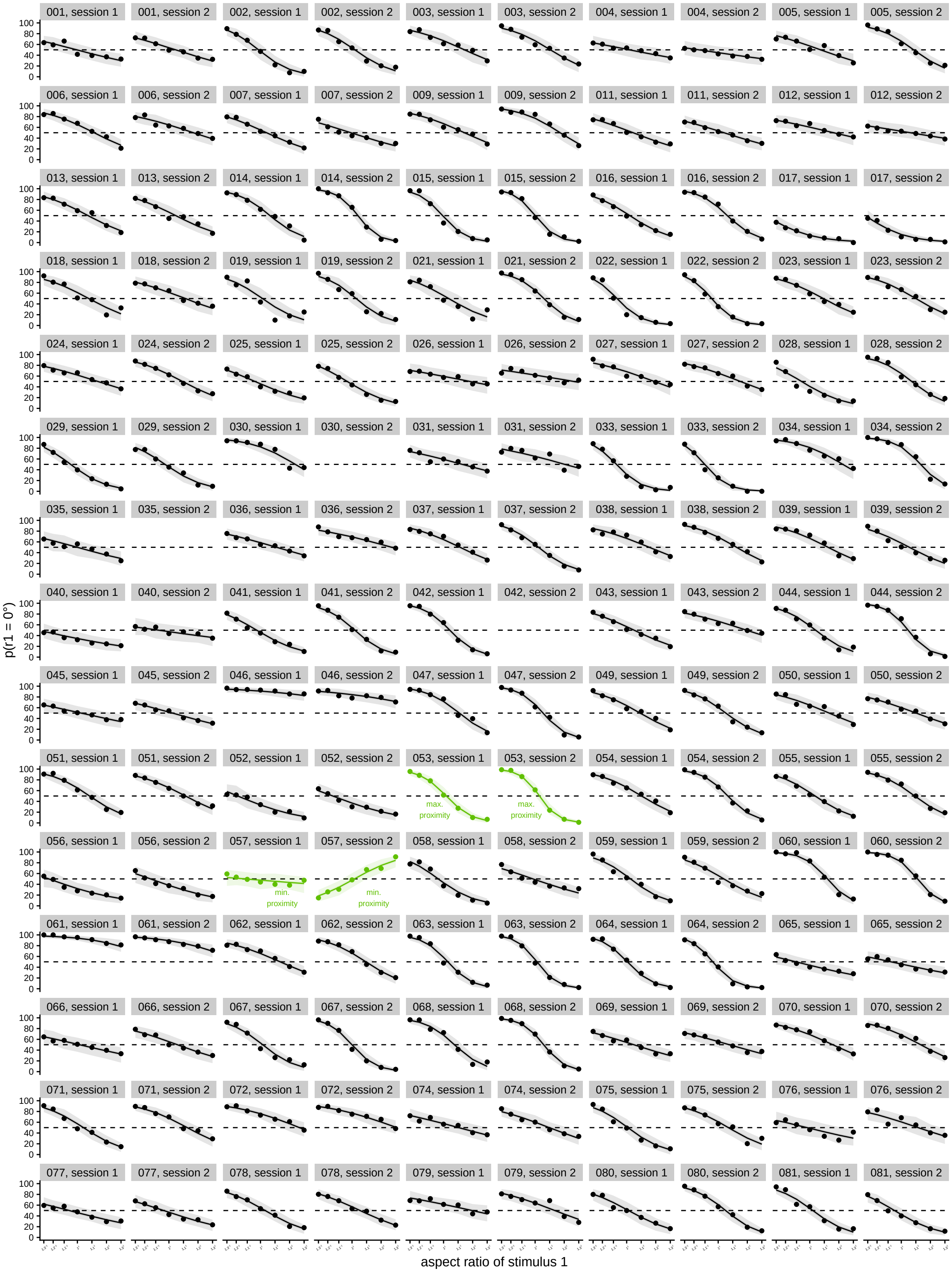


Figure B.9: Mean individual responses to the first stimulus dependent on aspect ratio (percentage), for both the first and the second session. The probability of responding 0° to the first stimulus decreases with aspect ratio (|a|/|b|). The value of aspect ratio increases with increasing distance in the 0°-orientation, leading to more 90° responses. Dots indicate observed values. In addition, mean posterior predictions and their 95% highest density continuous intervals are shown. Plots for participants with the smallest and largest estimated proximity effect are indicated in green.

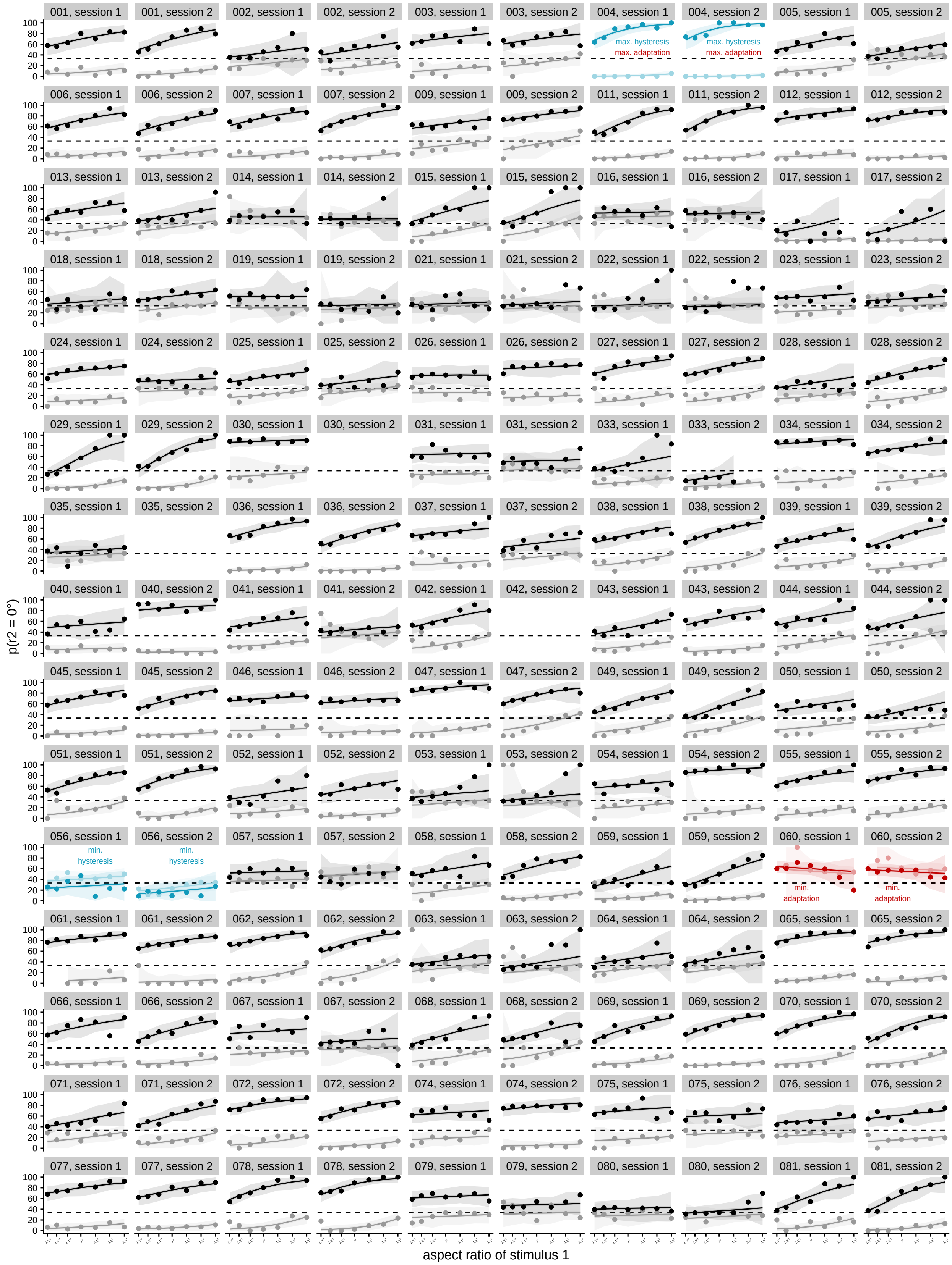


Figure B.10: Mean individual responses to the second stimulus dependent on aspect ratio (percentage), for both the first and the second session. The probability of responding 0° to the second stimulus increases with aspect ratio (|a|/|b|; i.e., adaptation effect), and increases when the first stimulus was perceived as 0° rather than 90° (i.e., hysteresis effect). Dots indicate observed values. In addition, mean posterior predictions and their 95% highest density continuous intervals are shown. Plots for participants with the smallest and largest estimated hysteresis effect are indicated in blue, participants with the smallest and largest estimated adaptation effect are indicated in red.

Figure B.11: Visual representation of the absolute orientation bias of an individual, corrected for implausible 90° responses, separately for the first and the second session. The colored oriented line indicates the mean direction of the absolute orientation bias per individual. Also the numeric values for the mean direction (M) and the magnitude (L) of the absolute orientation bias and the number of included trials (N) per individual are given.

Figure B.12: Visual representation of the absolute orientation bias of an individual per block of 60 trials, corrected for implausible 90° responses, separately for the first and the second session (Part 1/10). The colored oriented line indicates the mean direction of the absolute orientation bias per individual per block. Also the numeric values for the mean direction (M) and the magnitude (L) of the absolute orientation bias and the number of included trials (N) per individual per block are given.

Figure B.13: Visual representation of the absolute orientation bias of an individual per block of 60 trials, corrected for implausible 90° responses, separately for the first and the second session (Part 2/10). The colored oriented line indicates the mean direction of the absolute orientation bias per individual per block. Also the numeric values for the mean direction (M) and the magnitude (L) of the absolute orientation bias and the number of included trials (N) per individual per block are given.

B

Figure B.14: Visual representation of the absolute orientation bias of an individual per block of 60 trials, corrected for implausible 90° responses, separately for the first and the second session (Part 3/10). The colored oriented line indicates the mean direction of the absolute orientation bias per individual per block. Also the numeric values for the mean direction (M) and the magnitude (L) of the absolute orientation bias and the number of included trials (N) per individual per block are given.

B

Figure B.15: Visual representation of the absolute orientation bias of an individual per block of 60 trials, corrected for implausible 90° responses, separately for the first and the second session (Part 4/10). The colored oriented line indicates the mean direction of the absolute orientation bias per individual per block. Also the numeric values for the mean direction (M) and the magnitude (L) of the absolute orientation bias and the number of included trials (N) per individual per block are given.

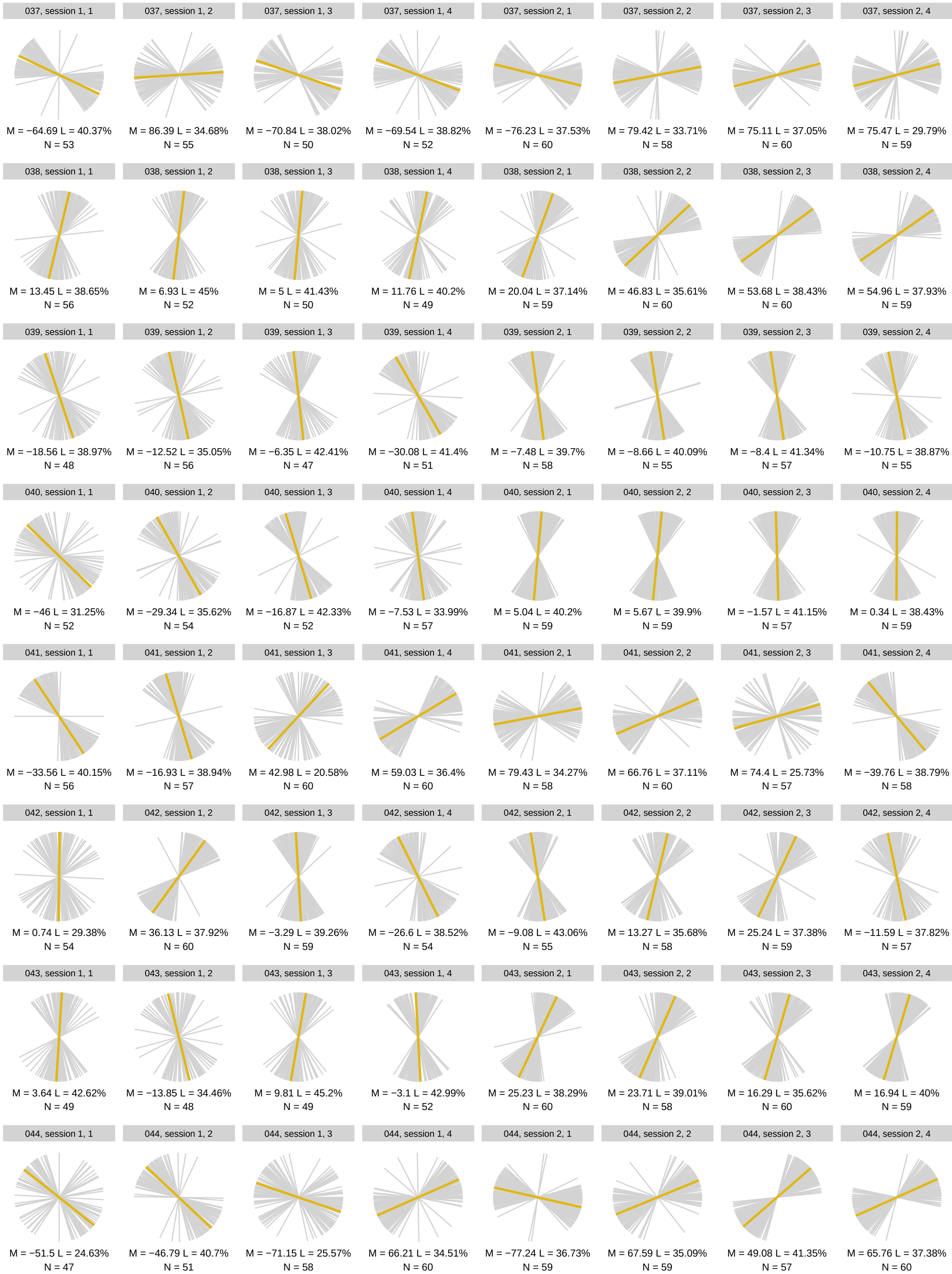


Figure B.16: Visual representation of the absolute orientation bias of an individual per block of 60 trials, corrected for implausible 90° responses, separately for the first and the second session (Part 5/10). The colored oriented line indicates the mean direction of the absolute orientation bias per individual per block. Also the numeric values for the mean direction (M) and the magnitude (L) of the absolute orientation bias and the number of included trials (N) per individual per block are given.

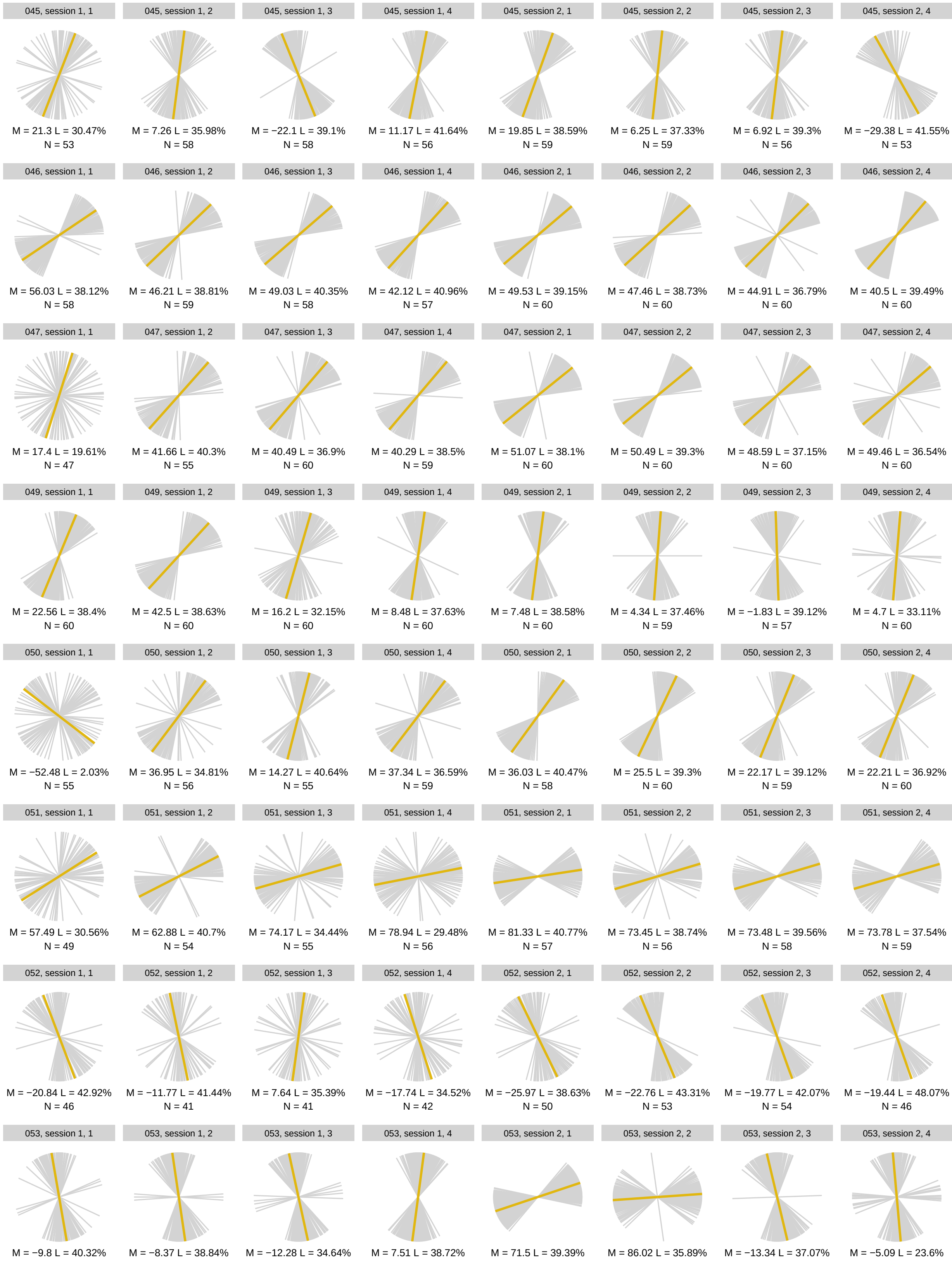


Figure B.17: Visual representation of the absolute orientation bias of an individual per block of 60 trials, corrected for implausible 90° responses, separately for the first and the second session (Part 6/10). The colored oriented line indicates the mean direction of the absolute orientation bias per individual per block. Also the numeric values for the mean direction (M) and the magnitude (L) of the absolute orientation bias and the number of included trials (N) per individual per block are given.

B

Figure B.18: Visual representation of the absolute orientation bias of an individual per block of 60 trials, corrected for implausible 90° responses, separately for the first and the second session (Part 7/10). The colored oriented line indicates the mean direction of the absolute orientation bias per individual per block. Also the numeric values for the mean direction (M) and the magnitude (L) of the absolute orientation bias and the number of included trials (N) per individual per block are given.

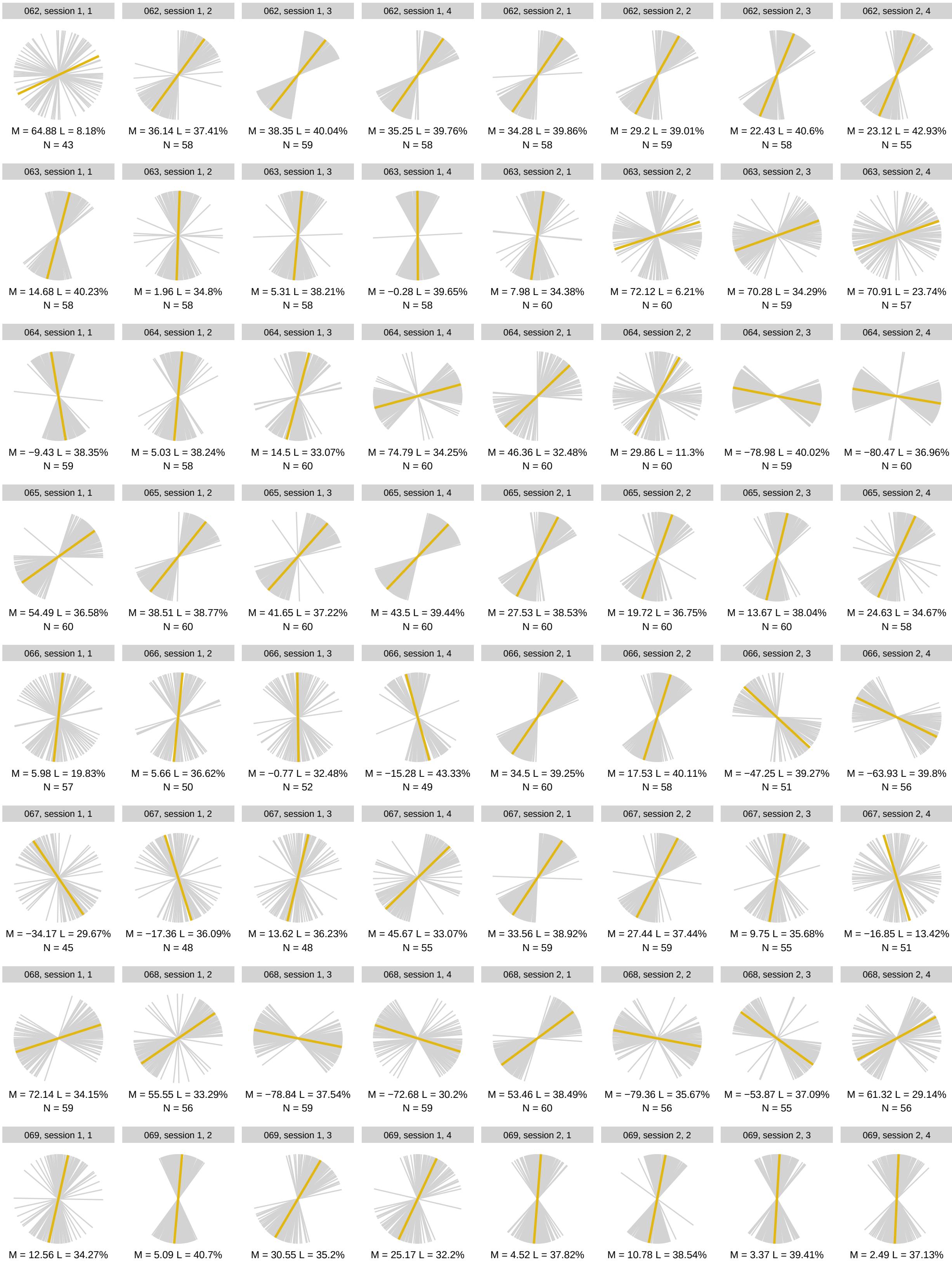


Figure B.19: Visual representation of the absolute orientation bias of an individual per block of 60 trials, corrected for implausible 90° responses, separately for the first and the second session (Part 8/10). The colored oriented line indicates the mean direction of the absolute orientation bias per individual per block. Also the numeric values for the mean direction (M) and the magnitude (L) of the absolute orientation bias and the number of included trials (N) per individual per block are given.

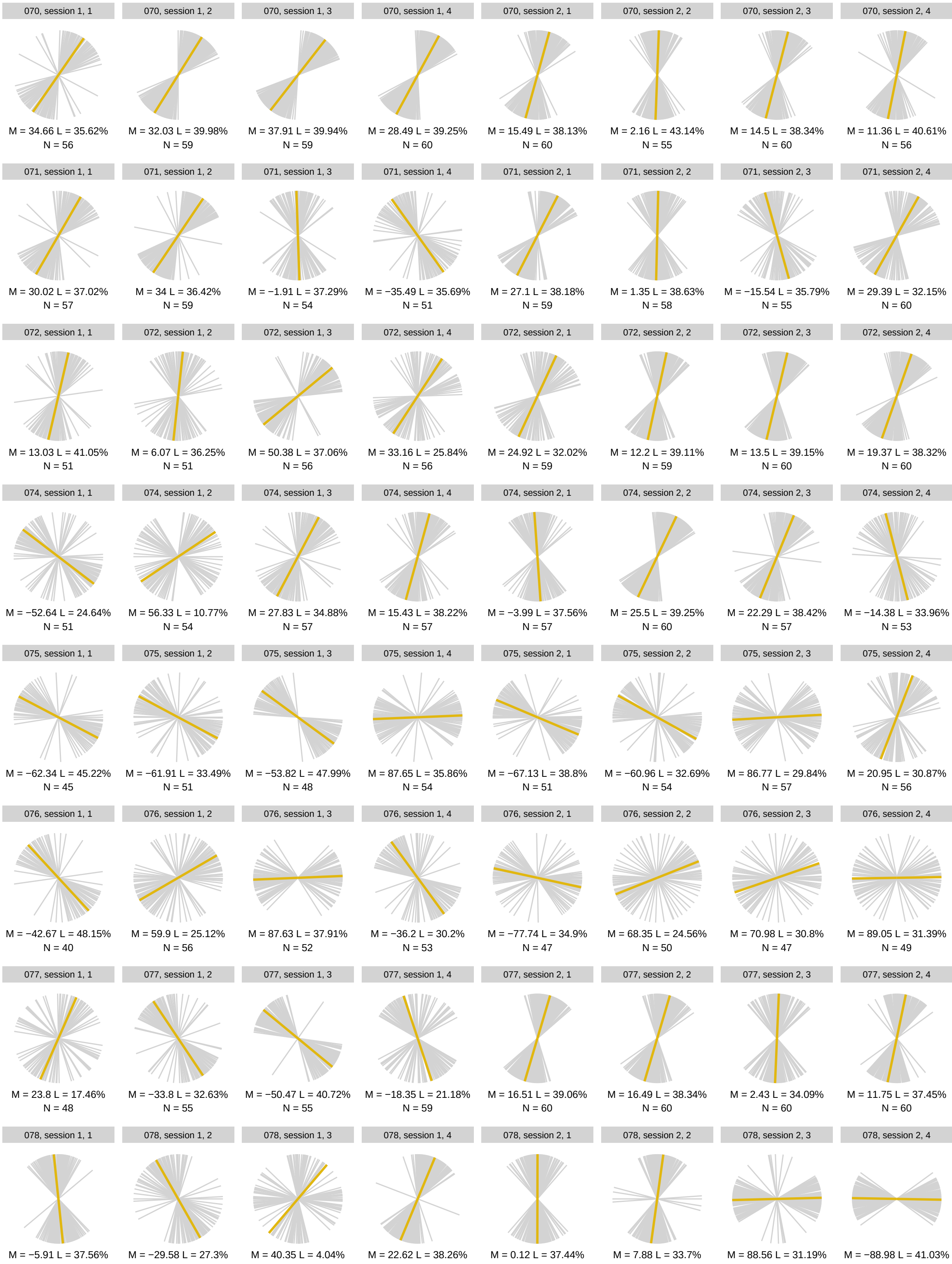


Figure B.20: Visual representation of the absolute orientation bias of an individual per block of 60 trials, corrected for implausible 90° responses, separately for the first and the second session (Part 9/10). The colored oriented line indicates the mean direction of the absolute orientation bias per individual per block. Also the numeric values for the mean direction (M) and the magnitude (L) of the absolute orientation bias and the number of included trials (N) per individual per block are given.

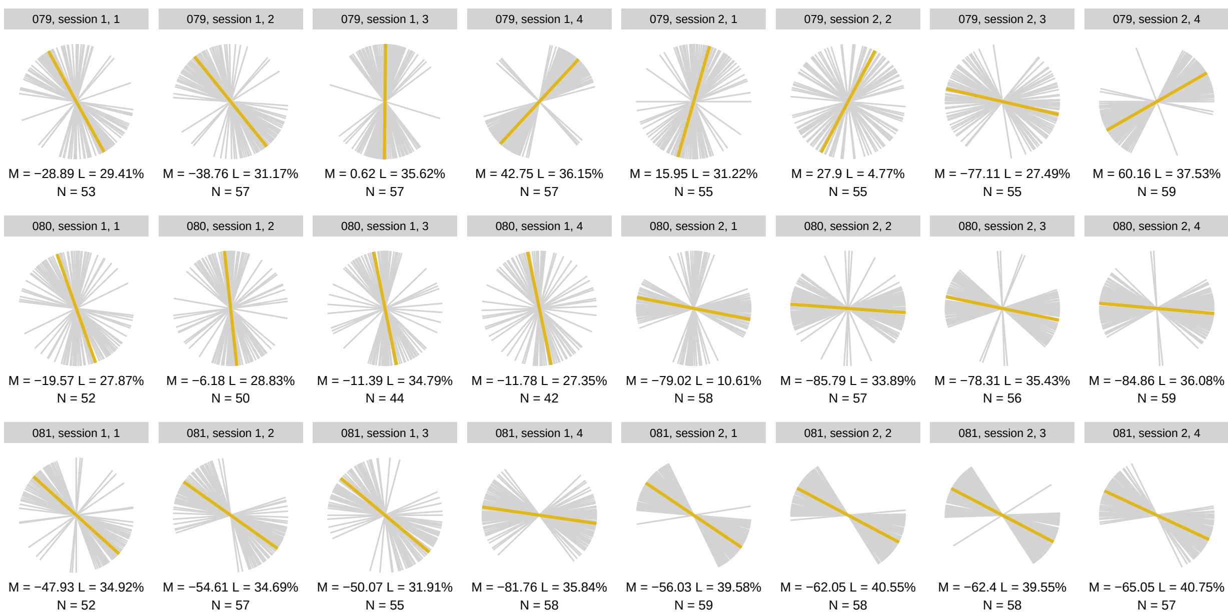


Figure B.21: Visual representation of the absolute orientation bias of an individual per block of 60 trials, corrected for implausible 90° responses, separately for the first and the second session (Part 10/10). The colored oriented line indicates the mean direction of the absolute orientation bias per individual per block. Also the numeric values for the mean direction (M) and the magnitude (L) of the absolute orientation bias and the number of included trials (N) per individual per block are given.

B.3 Supplemental videos

In the HTML version of this Appendix, you can find screen recordings of some trials for each task that was part of this study (cf. Figure B.22).

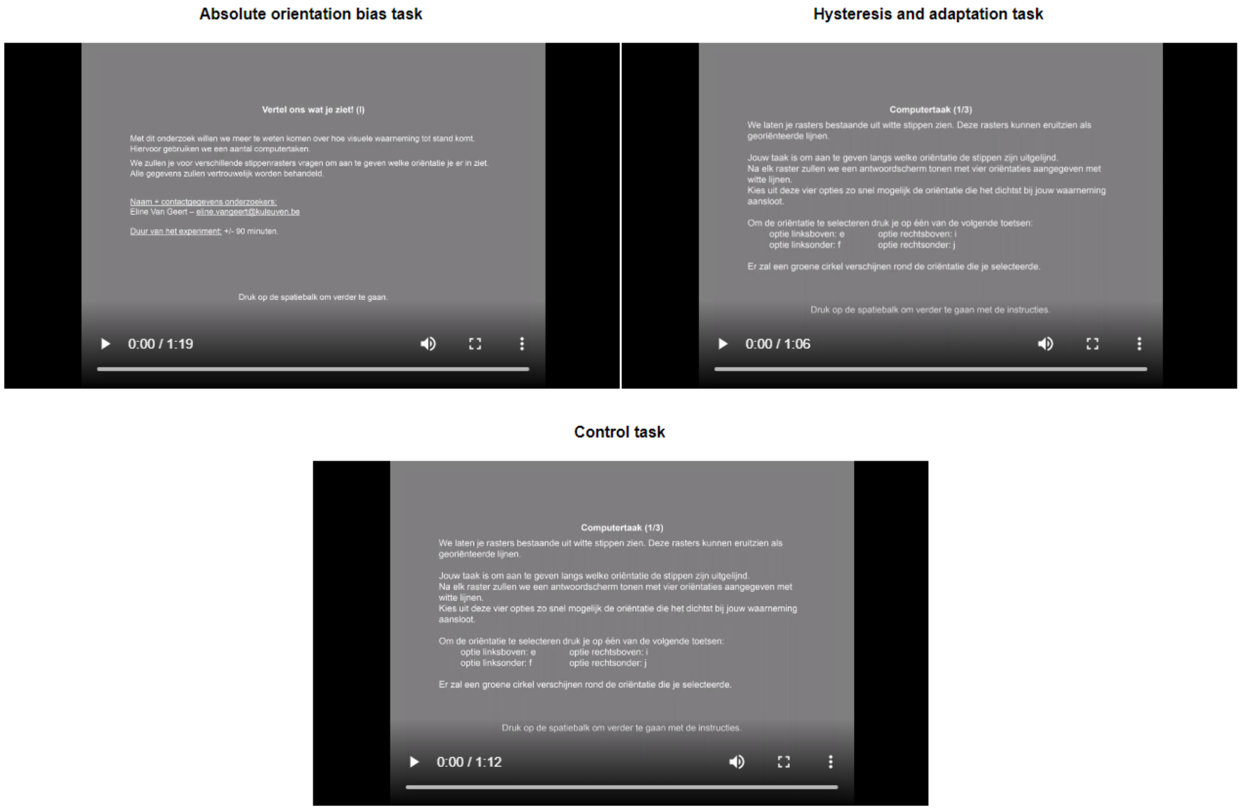


Figure B.22: Screen recordings of some trials for the absolute orientation bias, experimental, and control tasks that were part of this study.

B

Appendix C

Efficient Bayesian observer model of hysteresis and adaptation

C.1 R packages used

As indicated in the Methods section, for all our analyses we used R (Version 4.0.4; R Core Team, 2021) and the R-packages *BayesFactor* (Version 0.9.12.4.2; Morey & Rouder, 2018), *brms* (Version 2.16.1; Bürkner, 2017, 2018, 2021), *coda* (Version 0.19.4; Plummer et al., 2006), *cowplot* (Version 1.1.1; Wilke, 2020), *devtools* (Version 2.4.4; Wickham, Hester, et al., 2021), *doParallel* (Version 1.0.17; Corporation & Weston, 2020), *dplyr* (Version 1.0.10; Wickham, François, et al., 2022), *ellipse* (Version 0.4.2; Murdoch & Chow, 2020), *forcats* (Version 0.5.2; Wickham, 2022a), *foreach* (Version 1.5.2; Microsoft & Weston, 2020), *GGally* (Version 2.1.2; Schloerke et al., 2021), *ggdist* (Version 3.0.0; Kay, 2021a), *ggforce* (Version 0.3.3; Pedersen, 2021), *ggnewscale* (Version 0.4.8; Campitelli, 2022), *ggplot2* (Version 3.3.6; Wickham, 2016), *ggstance* (Version 0.3.5; Henry et al., 2020), *glue* (Version 1.6.2; Hester & Bryan, 2022), *gmm* (Version 1.7; Chaussé, 2010), *here* (Version 1.0.1; Müller, 2020), *iterators* (Version 1.0.14; Analytics & Weston, 2020), *kableExtra* (Version 1.3.4; Zhu, 2021), *knitr* (Version 1.39; Xie, 2015), *lubridate* (Version 1.8.0; Grolemund & Wickham, 2011), *MASS* (Version 7.3.53; Venables & Ripley, 2002), *Matrix* (Version 1.3.2; Bates & Maechler, 2021), *mvtnorm* (Version 1.1.3; Genz & Bretz, 2009; Wilhelm & G, 2022), *papaja* (Version 0.1.1; Aust & Barth, 2022), *patchwork* (Version 1.1.2; Pedersen, 2022), *purrr* (Version 0.3.4; Henry & Wickham, 2020), *Rcpp* (Eddelbuettel & Balamuta, 2018; Version 1.0.9; Eddelbuettel & François, 2011), *readr* (Version 2.1.2; Wickham, Hester, et al., 2022), *rstan* (Version 2.21.2; Stan Development Team, 2020a), *sandwich* (Zeileis, 2004, 2006; Version 3.0.2; Zeileis et al., 2020), *StanHeaders* (Version 2.21.0.7; Stan Development Team, 2020b), *stringr* (Version 1.4.0; Wickham, 2019), *tibble* (Version 3.1.8; Müller & Wickham, 2022), *tidybayes* (Version 3.0.1; Kay, 2021b), *tidyr* (Version 1.2.1; Wickham & Girlich, 2022), *tidyverse* (Version 1.3.2; Wickham et al., 2019), *tinylabels* (Version 0.2.3; Barth, 2022), *tmvtnorm* (Version 1.5; Wilhelm & G, 2022), *truncnorm* (Mersmann et al., 2018), and *usethis* (Version 2.1.6; Wickham, Bryan, et al., 2021).

C.2 Visualization of prior, likelihood, and posterior distributions for one trial

Figure C.1 visualizes the prior, likelihood, and posterior distributions for the first lattice in a single trial of the dot lattices paradigm, given a model with a prior stimulus distribution incorporating the more frequent occurrence of cardinal compared to oblique orientations. Figure C.2 visualizes the prior, likelihood, and posterior distributions for the second lattice in a single trial of the dot lattices paradigm, given the same natural stimulus distribution for the prior of the first lattice.

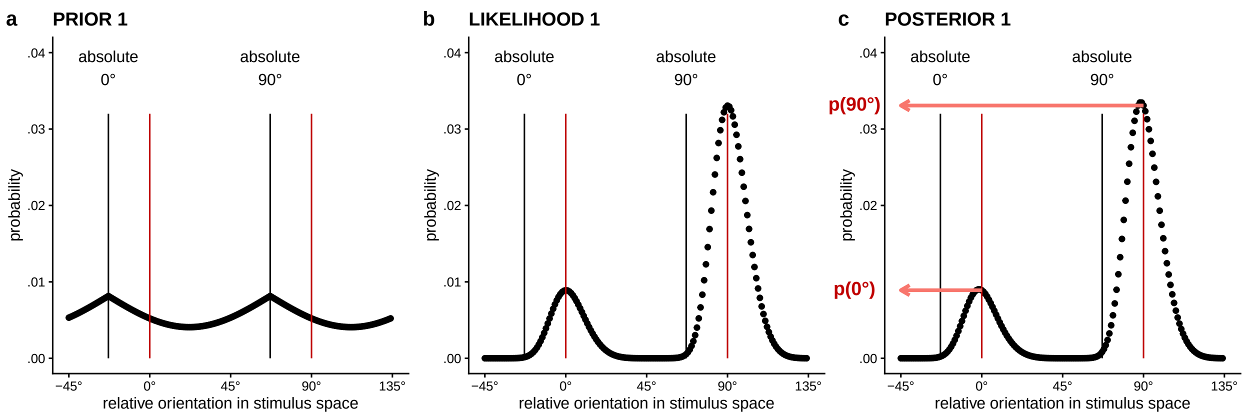


Figure C.1: (a) Natural stimulus prior for the first lattice, defined by the formula: $p(\theta) = c_0(2-|sin(\theta)|)$, in which θ represents the orientation value and c_0 the normalizing constant. (b) Example of likelihood 1 defined in the stimulus space, for a first lattice with an absolute lattice orientation of 23° and AR = 1.3^{-1}, which favors the relative 0° orientation. (c) Posterior distribution for the first lattice. Based on the difference in height of the peaks for the relative 0° and 90° orientation, i.e., p(0°) and p(90°), the probability of a} 0° or 90° response can be determined. *Note.* The red vertical lines in the graph are placed at the two dominant relative 0° and 90° orientations in the lattice. The black vertical lines label the absolute 0° and 90° orientations.

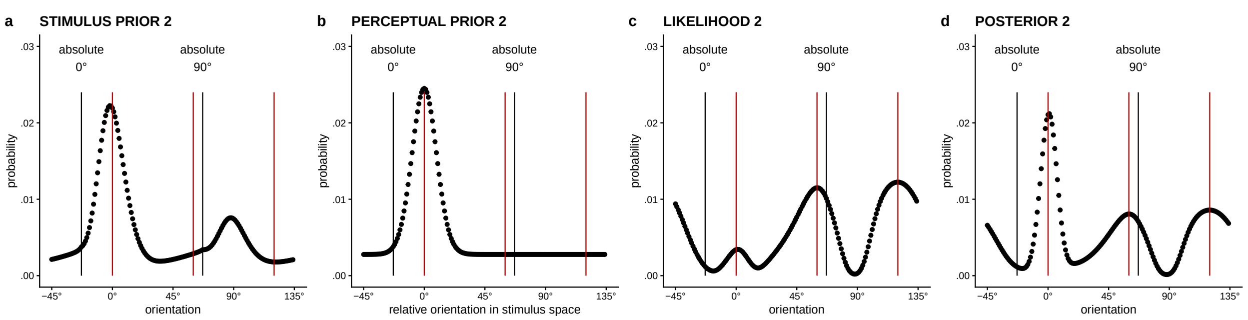


Figure C.2: (a) Stimulus prior for the second lattice given a first lattice with an absolute lattice orientation of 23° and with AR = 1.3^{-1}, which favors the relative 0° orientation. (b) Perceptual prior for the second lattice, given the relative 0° orientation was perceived in the first lattice. (c) Likelihood distribution defined in the stimulus space for the second lattice. This distribution is influenced by the stimulus prior for the second lattice (and hence the aspect ratio of the first lattice) via the stimulus-to-sensory mapping. (d) Posterior distribution for the second lattice, combining perceptual prior and likelihood for the second lattice. Based on the difference in height of the peaks for the relative 0°, 60°, and 120° orientation, the probability of a 0°, 60°, or 120° response can be determined. *Note.* The red vertical lines in the graph are placed at the three dominant relative 0°, 60°, and 120° orientations in the lattice. The black vertical lines label the absolute 0° and 90° orientations.

C.3 Supplementary Figures related to the section "Approximation of average attractive and repulsive temporal context effects"

Figure C.3 visualizes the logit probability of perceiving the relative 0° orientation based on efficient Bayesian observer model with a natural stimulus distribution prior for the first lattice. Figure C.4 visualizes the logit probability of perceiving the relative 0° orientation based on efficient Bayesian observer model without a perceptual prior and with a more extreme weighting of the previous stimulus evidence than in Figure 5.4b.

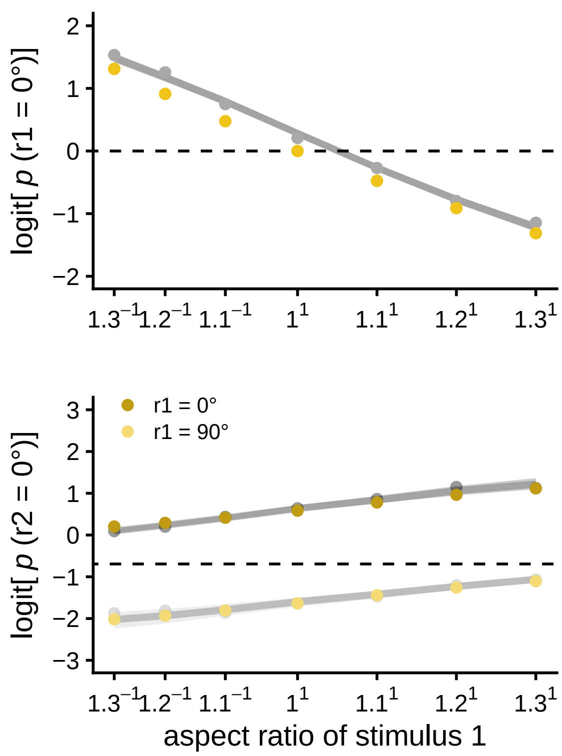


Figure C.3: Visualization of the logit probability to perceive the relative 0° orientation in (a) the first lattice and (b) the second lattice, based on an efficient Bayesian observer model with a natural stimulus distribution prior for the first lattice and the following parameters: $c_{\text{stim}} = 5$, $\kappa_{\text{stimL1}} = 20$, $\kappa_{\text{sensL1}} = 20$, $\kappa_{\text{stimL2}} = 20$, $\kappa_{\text{sensL2}} = 18$, $\kappa_{\text{percL1}} = 10$, w~stimL1 = 0.60, and $w_{\text{percL1}} = 0.50$. The yellow dots indicate the probabilities based on the model. The behavioral results and the estimated effects based on the behavioral results of Van Geert, Moors, et al. (2022), averaged across participants, are indicated in grey.

C.4 Supplementary information concerning Bayesian analyses of interindividual variation in simulation data

We estimated individual hysteresis and adaptation effects using a Bayesian multilevel binomial regression model predicting the percept of the second lattice, with aspect ratio of the first lattice (AR) and the percept of the first lattice ($R10$) as fixed and random effects. To estimate the direct proximity effect (i.e., the direct effect of aspect ratio on perception of the first lattice), we used a Bayesian multilevel binomial regression model predicting the percept of the first lattice, with aspect ratio of the first lattice (AR) as fixed and random effect.

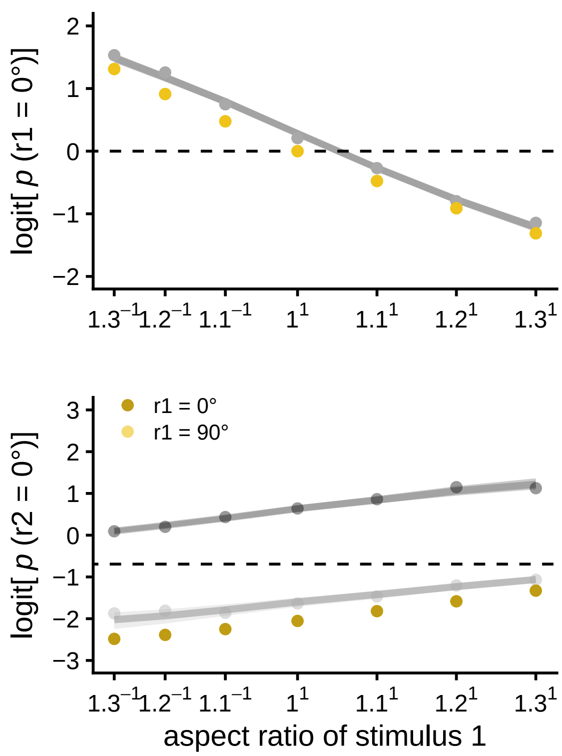


Figure C.4: Visualization of the logit probability to perceive the relative 0° orientation in (a) the first lattice and (b) the second lattice, based on an efficient Bayesian observer model without a perceptual prior for the second lattice, a flat prior distribution for the first lattice, and the following parameters: $c_{\text{stim}} = 5$, $\kappa_{\text{stimL1}} = 20$, $\kappa_{\text{sensL1}} = 20$, $\kappa_{\text{stimL2}} = 20$, $\kappa_{\text{sensL2}} = 18$, $\kappa_{\text{percL1}} = 10$, $w_{\text{stimL1}} = 1$, and $w_{\text{percL1}} = 0$. The blue dots indicate the probabilities based on the model. The behavioral results and the estimated effects based on the behavioral results of Van Geert, Moors, et al. (2022), averaged across participants, are indicated in grey.

C

The model for the first lattice included fixed and individual random effects for aspect ratio *AR* (i.e., proximity effect), and individual random intercepts. This model can be formulated as follows:

$$freq(r1 = 0°)|trials(n) \sim Intercept + AR + (Intercept + AR \mid participant).$$

To determine the size of the proximity effect, we used the individual estimates for the effect of the aspect ratio of the first lattice on the percept of the first lattice.

The model for the second lattice thus included fixed and individual random effects for percept in the first lattice *R10* (i.e., hysteresis effect) as well as aspect ratio in the first lattice *AR* (i.e., adaptation effect), and individual random intercepts. This model can be formulated as follows:

$$freq(r2 = 0°)|trials(n) \sim Intercept + AR + R10 + (Intercept + AR + R10 \mid participant).$$

To determine the size of the hysteresis effect, we used the individual estimates for the effect of the percept of the first lattice on the percept of the second lattice. To determine the size of the adaptation effect, we used the individual estimates for the effect of aspect ratio of the first lattice on the percept of the second lattice. To have an estimate of the strength of the correlation between the size of individuals' hysteresis and adaptation effects, we report the mean and 95% HDCI for the correlation between estimated individual hysteresis and adaptation effects, based on the full model described above.

In both the model for the first and for the second lattice, centered aspect ratio was used, which means that a value of zero corresponds to an aspect ratio of 1, a value of 1.1^{-1} -1 (i.e., ≈ -0.09) corresponds to 1.1^{-1}, and a value of 1.1^{1} -1 (i.e., 0.10) to an aspect ratio of 1.1.

Figure C.5 visualizes the priors we specified for the fixed effects, for the standard deviation of the random effects, and for the correlation matrix.

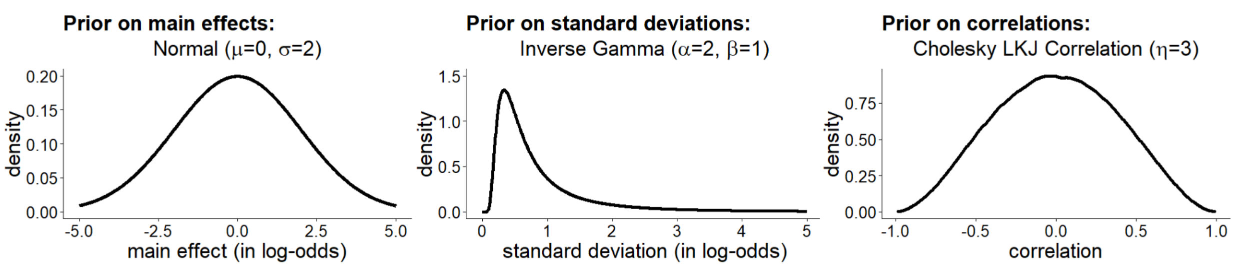


Figure C.5: Illustration of priors used in the model predicting the percept of L1 and L2. Reprinted from Van Geert, Moors, et al. (2022).

We fitted these models of perceived L1 and perceived L2 orientation using brms (Bürkner, 2017, 2018). We used 4 chains with 20000 iterations each with the default number of warmup iterations per chain. For any other sampling specifications we used the default settings. For further details on these analyses and those for calculating the correlation between individuals' proximity, hysteresis, and adaptation estimates, please consult Van Geert, Moors, et al. (2022).

C

C.5 Supplementary Figure concerning interindividual variation in proximity, hysteresis, and adaptation

Figure C.6 visualizes the correlation of estimated individual hysteresis and adaptation effects concerning the second lattice with estimated individual proximity effects concerning the first lattice for the empirical data collected in Van Geert, Moors, et al. (2022) and for the simulation results.

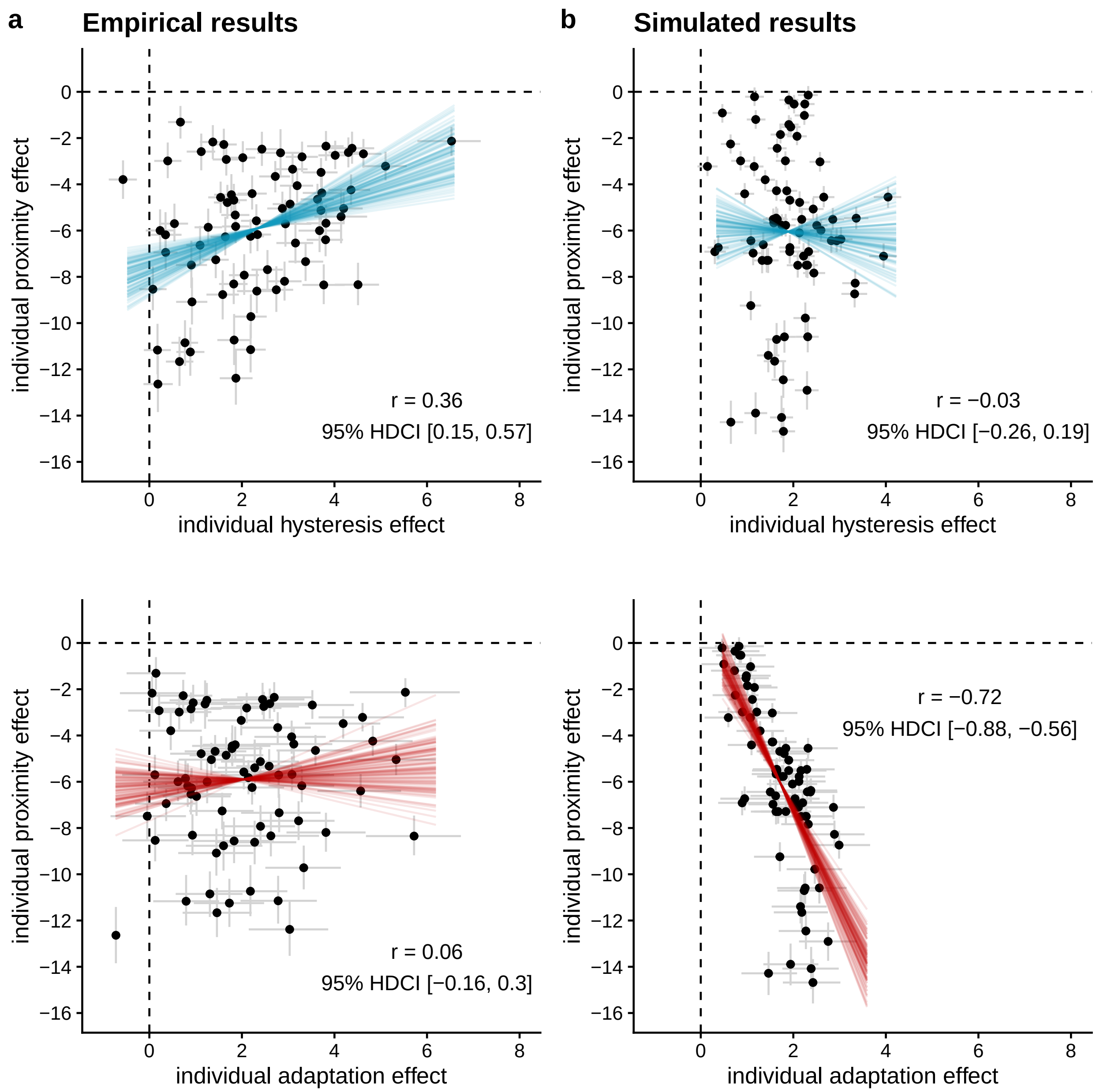


Figure C.6: Correlation of estimated individual hysteresis and adaptation effects concerning the second lattice with estimated individual proximity effects concerning the first lattice for (a) the empirical data collected in Van Geert, Moors, et al. (2022) and (b) the simulation results.

C

C.6 Shiny application

Figure C.7 shows the Shiny application to visualize the prior, likelihood, and posterior distributions for one trial under different parameter settings.

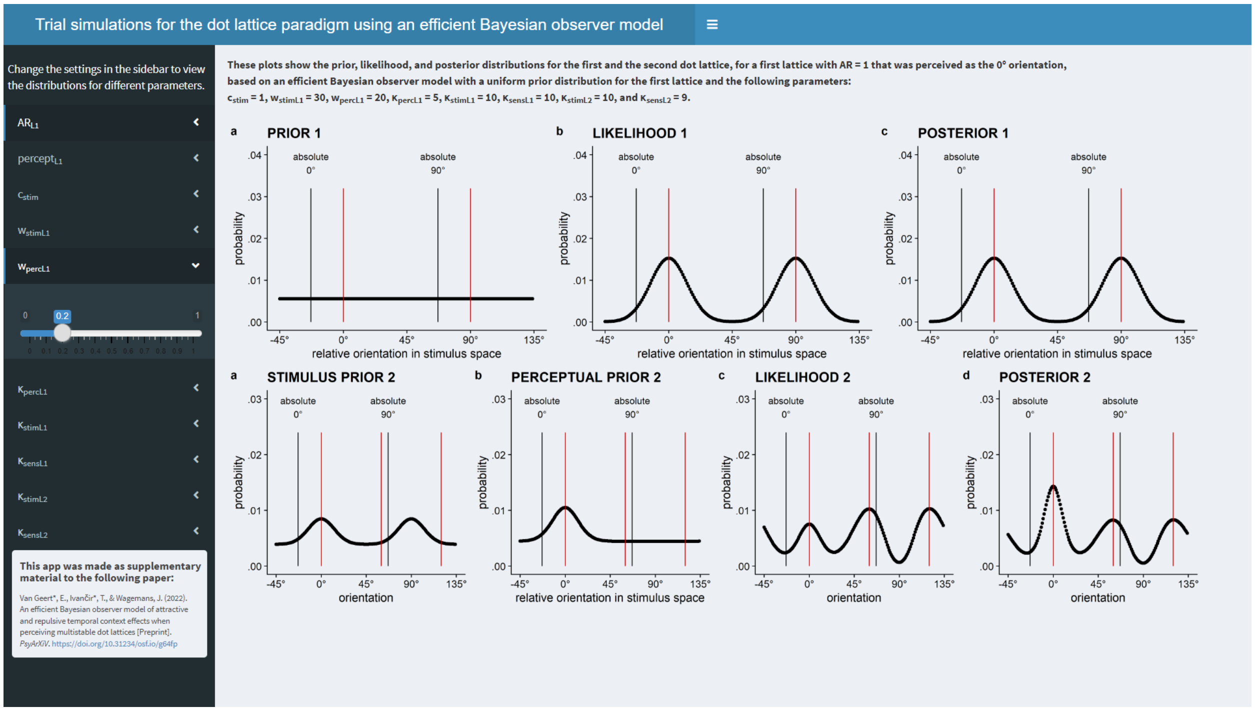


Figure C.7: A Shiny application to visualize the prior, likelihood, and posterior distributions for one trial under different parameter settings. Click the link to try out the application.

Appendix D

Context dependence of leveling and sharpening

D.1 Supplementary Methods

D.1.1 Additional information on stimulus construction

To calculate the specific values on the two feature dimensions, the following steps were taken. First, the length between the minimum and maximum value of the dimension which has the least space to vary relative to the border of the canvas, i.e. the available length, was determined. To compute the available length, a predefined margin was multiplied by two and subtracted from the total length of the dimension which had the least space to vary. This margin was necessary to allow for the sharpening of a dimension. It was defined as one-fifth of the total length. This first step could be mathematically defined as follows:

$$availableLength = totalLength - \frac{2 * totalLength}{5}$$

Second, the size of the steps between the four values of a dimension was computed and centered around their midpoint[1]. The largest steps were defined as one-third of the available length, i.e.,

$$step_{largest} = \frac{availableLength}{3}$$

whereas the smallest steps were defined as one-fourth of the largest steps[2], i.e.,

$$step_{smallest} = \frac{step_{largest}}{4}$$

Third, these steps were randomly paired per series. For instance, a stimulus in series A could have had the second largest value of the largest steps assigned to the first dimension, and the smallest value of the smallest steps assigned to the second dimension. Fourth, the values of the dimensions of the basic shapes that were not

D

[1] Due to rounding errors, the size of the steps between the different values on a feature dimension was not always exactly equal, especially for the feature dimensions with minor variability. In addition, when constructing the stimuli, some errors were made in the minimum or maximum value for the dimension with minor variability for designs 5 (maximum value dimension 1 in series B 28.25 instead of 28.75), 16 (minimum value dimension 1 in series B 30 instead of 31.2), and 22 (minimum value dimension 1 in series B 23.8 instead of 21.2 and maximum value dimension 1 in series B 31.2 instead of 28.8).

[2] For designs 1, 2, and 6, one of the feature dimensions was bidirectional, which required the step size to be added at both sides of the figure. Therefore, in these designs, the smallest steps were one out of two or one out of eight of the largest steps available concurrently for the other feature dimension in the figure. When comparing major and minor variability for the same feature dimension across the A and B series, the ratio of one out of four is correct.

varied, were defined as the mean of the varied dimension of the other basic shape containing a variable dimension. Because of the nature of the task, for both series, only two out of four stimuli could serve as the target, namely those with the maximum and minimum value on the dimension with the largest variability. Because of the random pairing of values on the dimensions with largest and smallest variability, the value of the target on the dimension with minor variability could either have a minimum or maximum value as well (i.e., an extreme value), or could be a value in-between (i.e., a non-extreme value).

D.1.2 R packages used

As indicated in the Methods section, for all our analyses we used R (Version 4.0.4; R Core Team, 2021) and the R-packages *brms* (Version 2.16.1; Bürkner, 2017, 2018, 2021), *cowplot* (Version 1.1.1; Wilke, 2020), *dplyr* (Version 1.0.10; Wickham, François, et al., 2022), *forcats* (Version 0.5.2; Wickham, 2022a), *ggdist* (Version 3.0.0; Kay, 2021a), *gghalves* (Version 0.1.3; Tiedemann, 2020), *ggplot2* (Version 3.3.6; Wickham, 2016), *ggstance* (Version 0.3.5; Henry et al., 2020), *here* (Version 1.0.1; Müller, 2020), *htmltools* (Version 0.5.3; Cheng et al., 2021), *knitr* (Version 1.39; Xie, 2015), *papaja* (Version 0.1.1; Aust & Barth, 2022), *patchwork* (Version 1.1.2; Pedersen, 2022), *purrr* (Version 0.3.4; Henry & Wickham, 2020), *Rcpp* (Eddelbuettel & Balamuta, 2018; Version 1.0.9; Eddelbuettel & François, 2011), *readr* (Version 2.1.2; Wickham, Hester, et al., 2022), *rstan* (Version 2.21.2; Stan Development Team, 2020a), *StanHeaders* (Version 2.21.0.7; Stan Development Team, 2020b), *stringr* (Version 1.4.0; Wickham, 2019), *tibble* (Version 3.1.8; Müller & Wickham, 2022), *tidybayes* (Version 3.0.1; Kay, 2021b), *tidyr* (Version 1.2.1; Wickham & Girlich, 2022), *tidyverse* (Version 1.3.2; Wickham et al., 2019), and *tinylabels* (Version 0.2.3; Barth, 2022).

D.1.3 Preprocessing of drawing data

To be able to analyze the drawing data, rather extensive preprocessing was required. First, shapes that were located outside of the drawing canvas (with an additional margin of 5 units added) were removed from the drawing data. This resulted in the removal of 215 shapes, coming from 171 of the 13032 drawings. We then counted the number of shapes of each type on the canvas and excluded drawings that had an incorrect number of rectangles or triangles (203 of 13032 drawings, 1.56%).

Second, we determined which drawn shape contained which of the relevant feature dimensions, based on the relative horizontal or vertical position (depending on the design in question) of the two drawn shapes on the canvas. If an equal horizontal or vertical position of the two drawn shapes made it impossible to determine which drawn shape contained which feature dimension, the drawing was excluded from analyses (7 of 12829 drawings, 0.05%).

Third, for the 20 out of 24 designs containing background figures, the position of the drawn shapes relative to the background figures was checked (specific exclusion criteria were design-dependent). In this way, 44 additional drawings were excluded (0.34%). This check also identified 115 drawings where the background figure was interpreted as a different shape than was intended, which were excluded from analyses as well (115 out of 12778 drawings, 0.9%).

Fourth, for the four designs without background figure(s), the relative placement of the drawn shapes was checked. Following this check, 11 drawings were excluded (0.09%).

Fifth, the orientation and the point location of the shapes was scrutinized, and the value for the point location was adapted when needed (i.e., because of a combination of mirroring and orientation parameters of the drawn shapes). This step excluded 22 additional drawings (0.17%). In total, 380 out of 13032 drawings were excluded so far (2.92%).

After excluding drawings not surviving the checks above, some of the absolute feature dimension values for the drawings needed rescaling: the values for drawings without background shapes (i.e., drawings for designs 2, 4, 5, and 18) and the values for drawings where the background shape was not necessarily used for scaling (i.e., drawings for designs 1 and 6). More specifically, we rescaled the drawings in such a way that the height or width of the drawn shape for which this feature was not varied served as a reference for the drawn shape in which this dimension was varied, so that the ratio between the two heights or widths was kept constant. This rescaling was needed to be able to compare the drawn feature values with the target values on each dimension.[3]

Although the rescaled absolute drawn values can be used directly for qualitative comparisons across designs (i.e., proportion of times a feature was leveled or sharpened in each context and variability condition), an additional standardization was needed to make quantitative comparisons across designs possible. All feature values were therefore divided by the available range for leveling and sharpening combined. In this relative measure, a value of one indicates use of all available space to draw the feature, with values ranging from zero to one (or larger than one if the feature was drawn larger than the available space).

Due to a programming error in the online experiment for the first wave of data collection, information concerning which figure served as target in a particular trial (i.e., minimum or maximum on the major feature dimension) was not saved, for 6326 out of 12630 drawings (50.09%). We therefore defined the extreme value that was closest to the drawn value on the major feature dimension as the target value on the major dimension, and in that way also determined the target value on the minor feature dimension. This is a conservative assumption that can only diminish the effect size of our results, as for those drawings for which no actual target information was available, it minimized leveling and sharpening for the major feature dimension (i.e., to one side of the mean value on the major feature dimension). To estimate the accuracy of this target assignment method, we compared assumed targets with actual targets for the part of the sample for which target information was saved. In this part of the sample (6304 drawings), an incorrect target was assigned for 40 drawings (0.63%), and no target could be assigned for 58 drawings (0.92%) because the drawn value was equally far from both targets or because target assignment differed when either the absolute or the relative values were taken into account.

Drawings for which no actual target information was available and in which a different target was assumed depending on whether the absolute or relative values were taken into account and/or drawings in which the value on the major feature dimension was equally far from the minimum and maximum target value (making it impossible to retrieve the target value on the minor feature dimension) were excluded from analyses (57 of 12630 drawings, 0.45%). As the wrong figure (with an in-between value rather than an extreme value on the major feature dimension)

[3] One downside of this rescaling procedure is that the height and/or width of the two drawn shapes becomes dependent, potentially indirectly leading to dependent feature values on both dimensions. The rescaling can also lead to feature values relative to the available drawing space that are smaller than -1 or larger than 1. We therefore also conducted the analyses with exclusion of these six designs, which led to slightly smaller but similar qualitative and quantitative results as presented in the main text (see Figure D.5 and Figure D.15).

D

was mistakenly assigned as minimum target for series B of design 5 in the close context, all drawings of design 5 in series B where the minimum target was assigned were excluded, both those presented in close and far context (105 of 12573 drawings, 0.84%).

D.1.4 Bayesian hierarchical model implementation details

D.1.4.1 Qualitative data

We fitted a hierarchical Bayesian logistic regression model to the proportion of times a dimension was sharpened, with context, variability, and their interaction as fixed effects and feature dimension and participant ID as random effects for both intercept and slopes:

$$\begin{aligned} sharpening \sim & Intercept + context + variability + context : variability \\ & + (1 + context + variability + context : variability \,||\, conceptdim) \\ & + (1 + context + variability + context : variability \,||\, pp\_id) \end{aligned}$$

As priors, we specified a normal distribution with a mean of zero and a standard deviation of 2 for the intercept, a normal distribution with a mean of zero and a standard deviation of 2 for the slopes, and an inverse gamma distribution for the standard deviations, with an alpha parameter of 2 and a beta parameter of 1. We used 4 chains consisting of 8000 iterations, with 4000 warmup iterations per chain.

We fitted models with the same specifications for the traditional and the alternative definitions of sharpening, as well as for the definition of being in target range, leveling, and sharpening including a neutral category. Regarding the latter, in the model for sharpening, the neutral category and the leveling responses were collapsed, in the model for leveling the neutral category and the sharpening responses were collapsed, and in the model for being in the target range the leveling and sharpening responses were collapsed.

D.1.4.2 Quantitative data

We fitted a hierarchical Bayesian Gaussian regression model to the signed difference between the drawn and target value (values relative to the available drawing space for the feature dimension in question), with context, variability, and their interaction as fixed effects and feature dimension and participant ID as random effects for both intercept and slopes:

D

$$\begin{aligned} rel\_targetdiff \sim & Intercept + context + variability + context : variability \\ & + (1 + context + variability + context : variability \,||\, conceptdim) \\ & + (1 + context + variability + context : variability \,||\, pp\_id) \end{aligned}$$

As priors, we specified a normal distribution with a mean of zero and a standard deviation of 2 for the intercept, a normal distribution with a mean of zero and a standard deviation of 2 for the slopes, and an inverse gamma distribution for the standard deviations, with an alpha parameter of 2 and a beta parameter of 1. We used 4 chains consisting of 8000 iterations, with 4000 warmup iterations per chain.

D.2 Supplementary Results

D.2.1 Additional qualitative results for the traditional interpretation of leveling and sharpening (binarized; relative values)

When inspecting the standard deviations for the feature dimension random effects (see Figure D.1), it becomes clear that the estimated variability between feature dimensions is larger for the differences between the variability conditions than for the context effect: Whereas for almost all feature dimensions, the estimated sharpening probability was higher in the close than in the far context, not all feature dimensions showed a higher sharpening probability in the major variability condition (see Figure D.2). That the highest density continuous intervals are wider for the effect of context than for the differences between the variability conditions is potentially a consequence of the difference in the number of trials per condition involved in the comparison: for a context comparison, maximally 12 trials per condition per participant, whereas for a variability or extremeness comparison, maximally 24 trials per condition per participant were involved. The estimated variability between feature dimensions was higher than the estimated variability between participants (see also Figure D.3), which is potentially also a consequence of the limited number of data points per participant in comparison to the number of data points per feature dimension.

Figure D.4 shows the results for the equivalent qualitative model of sharpening based on the absolute rather than the relative feature values. Figure D.5 shows the results for the qualitative model of sharpening based on the relative feature values, but with exclusion of concepts 1, 2, 4, 5, 6, and 18.

D.2.2 Qualitative results for the alternative interpretation of leveling and sharpening (binarized; relative values)

In Figure D.6, sharpening is defined as a drawn value closer to the target value than to the feature dimension's mean. Defining sharpening in this manner, increased the differences between the three variability conditions and decreased the size of the context effect. The results stayed in the expected direction, however: sharpening was more likely for the major variability dimension than for the minor variability dimensions, and there was a tendency for sharpening to be more common in the close than in the far context. Furthermore, also using this definition of sharpening, there was much more variability in the percentage of sharpening across feature dimensions in the minor variability conditions than in the major variability condition.

Figure D.7, Figure D.8, and Figure D.9 give the posterior estimates from the model per context and variability condition, and compare the slope strengths across conditions for the effects of context, variability, and extremeness, respectively. In the major variability condition, the posterior probability of the proportion of sharpening in the close context to be larger than in the far context was 93%. Even though the direction of the effect was highly probable in the expected direction, given this alternative definition of leveling and sharpening, the context effect was very limited in size (see Figure D.7).

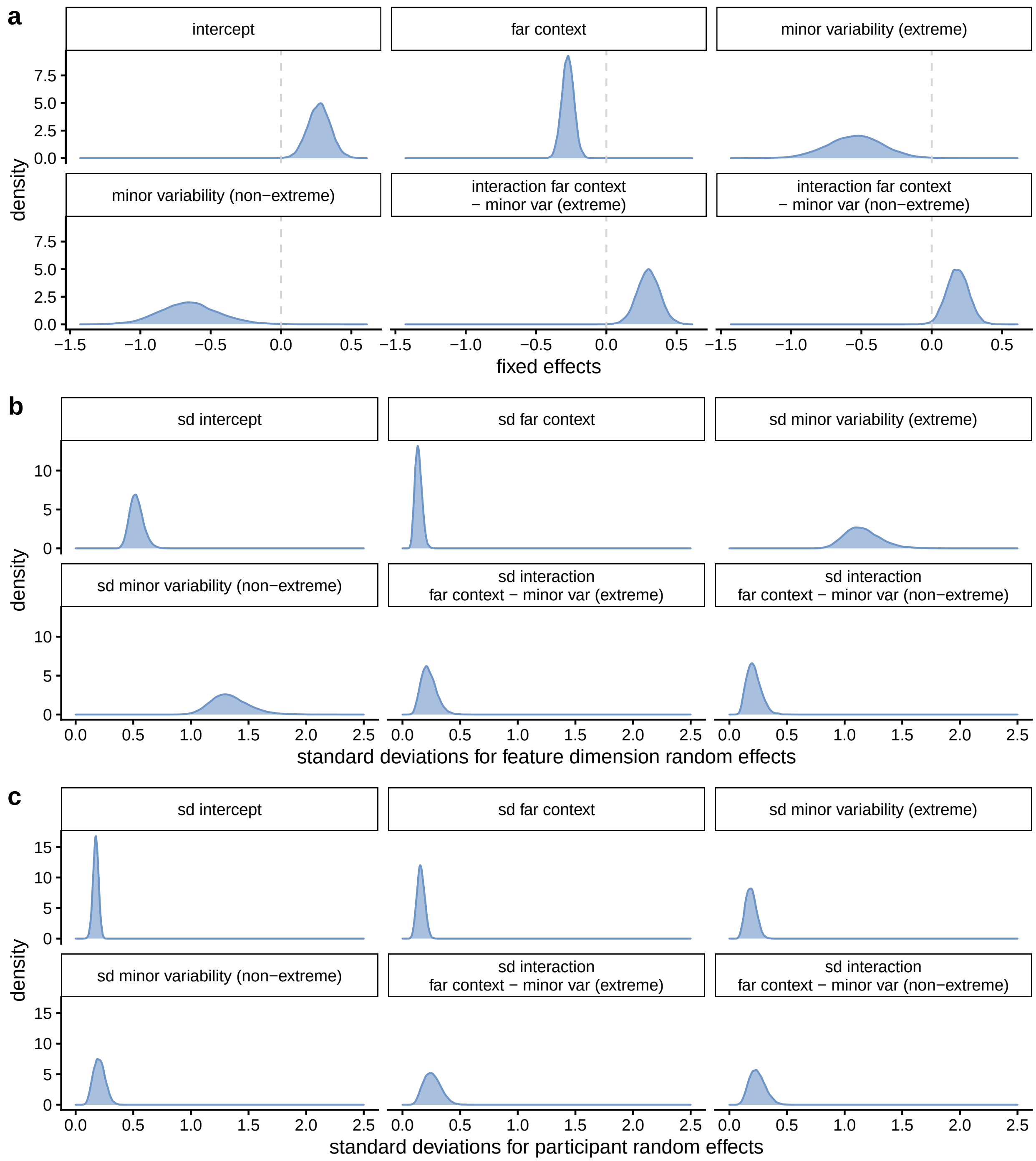


Figure D.1: Posterior distributions of fixed effects and standard deviations of random effects for concept and dimension combined, for the qualitative model (binarized; based on relative feature values).

D

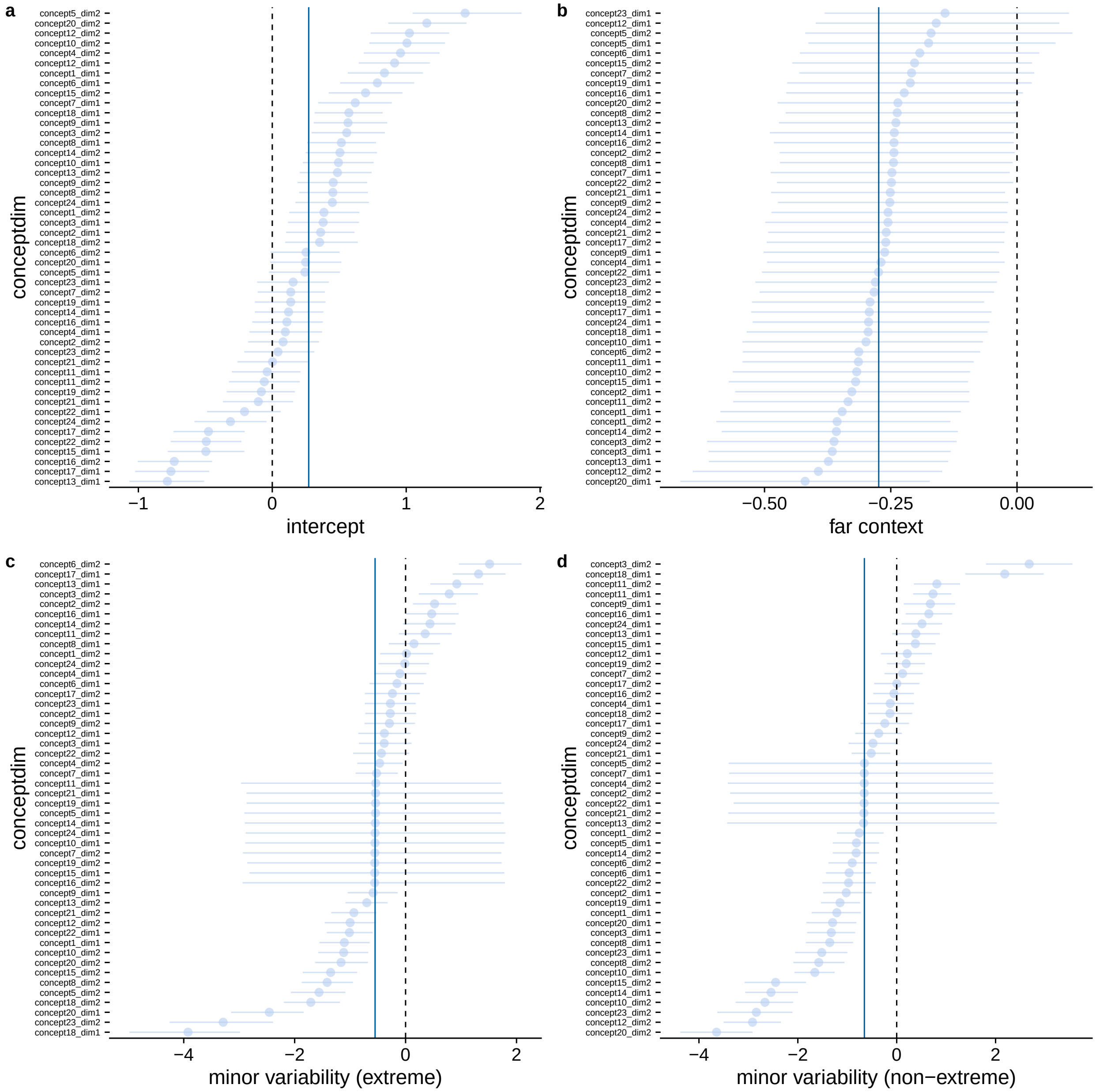


Figure D.2: Slopes for the intercept, the effect of context, the effect of a minor extreme feature value compared to a major extreme feature value, and the effect of a minor non-extreme feature value compared to a major extreme feature value on the proportion of times a feature was sharpened, with sharpening defined as a drawn value equal to the target value or more extreme, per feature dimension. Mean and 95% highest density continuous intervals are shown. The blue line indicates the average mean value across feature dimensions. The dashed black line indicates a value of zero. *Note.* The very wide intervals for some feature dimensions are the consequence of those feature dimensions not being present in the relevant condition, e.g., no minor variability (extreme) condition, for that feature dimension.

D

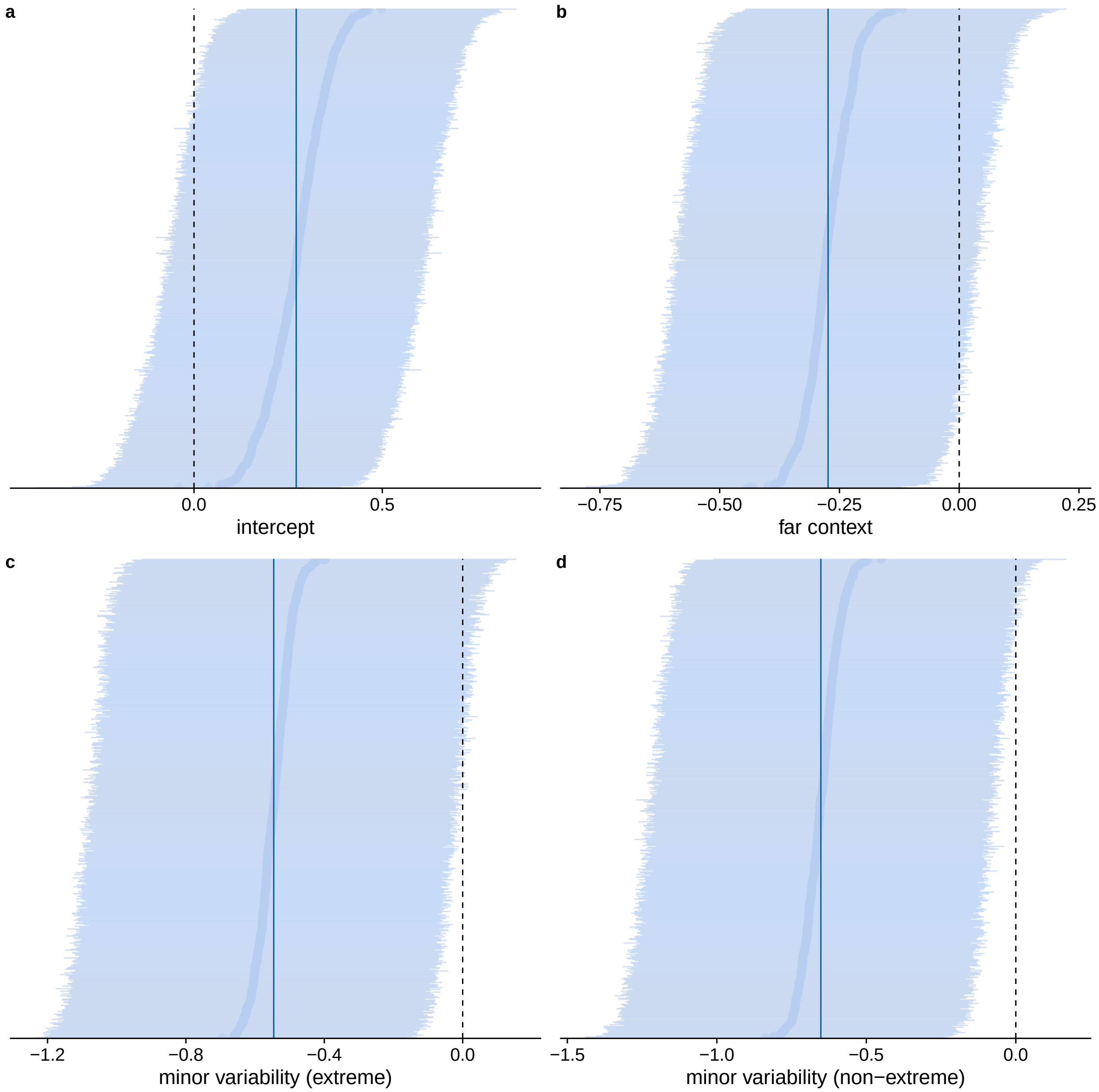


Figure D.3: Slopes for the intercept, the effect of context, the effect of a minor extreme feature value compared to a major extreme feature value, and the effect of a minor non-extreme feature value compared to a major extreme feature value on the proportion of times a feature was sharpened, with sharpening defined as a drawn value equal to the target value or more extreme, per participant. Mean and 95% highest density continuous intervals are shown. The blue line indicates the average mean value across participants. The dashed black line indicates a value of zero.

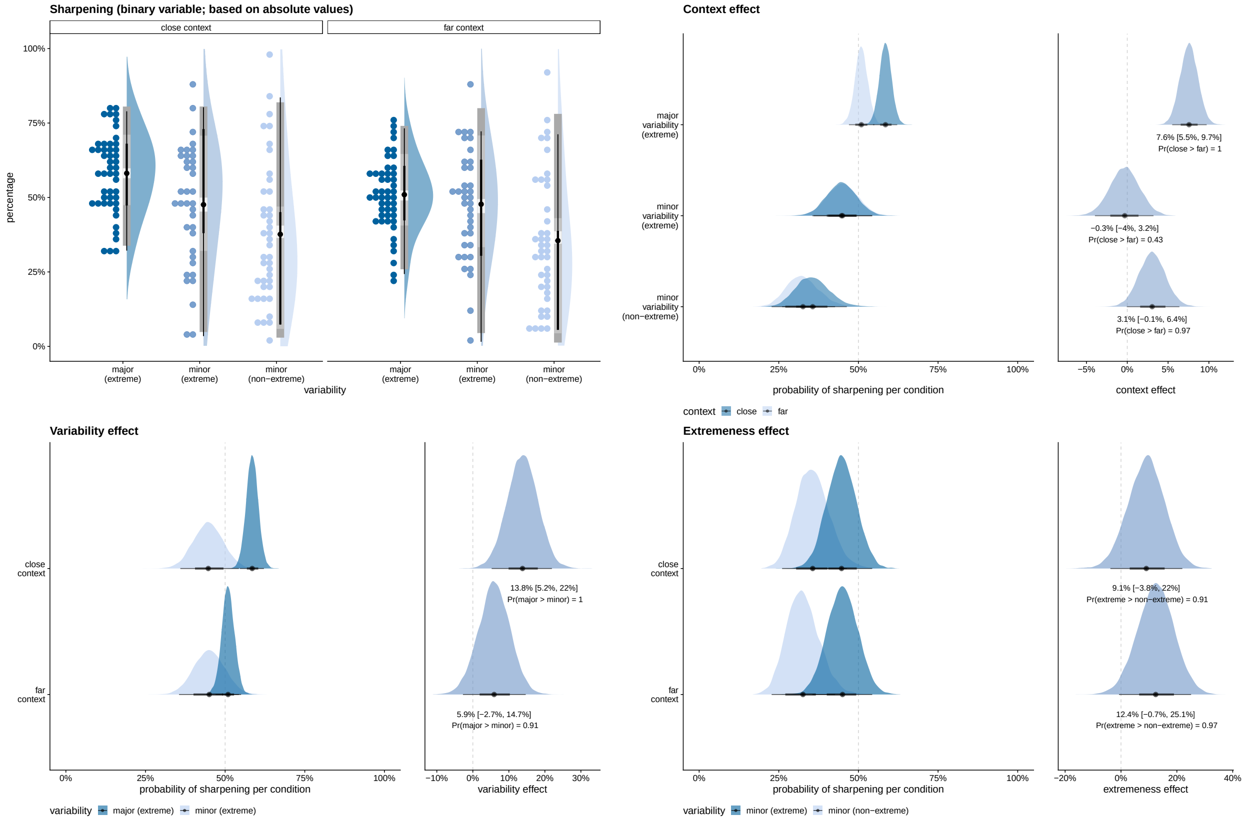


Figure D.4: Results of the qualitative model of sharpening, defined as a drawn value equal to the target or more extreme, based on the absolute feature values.

D

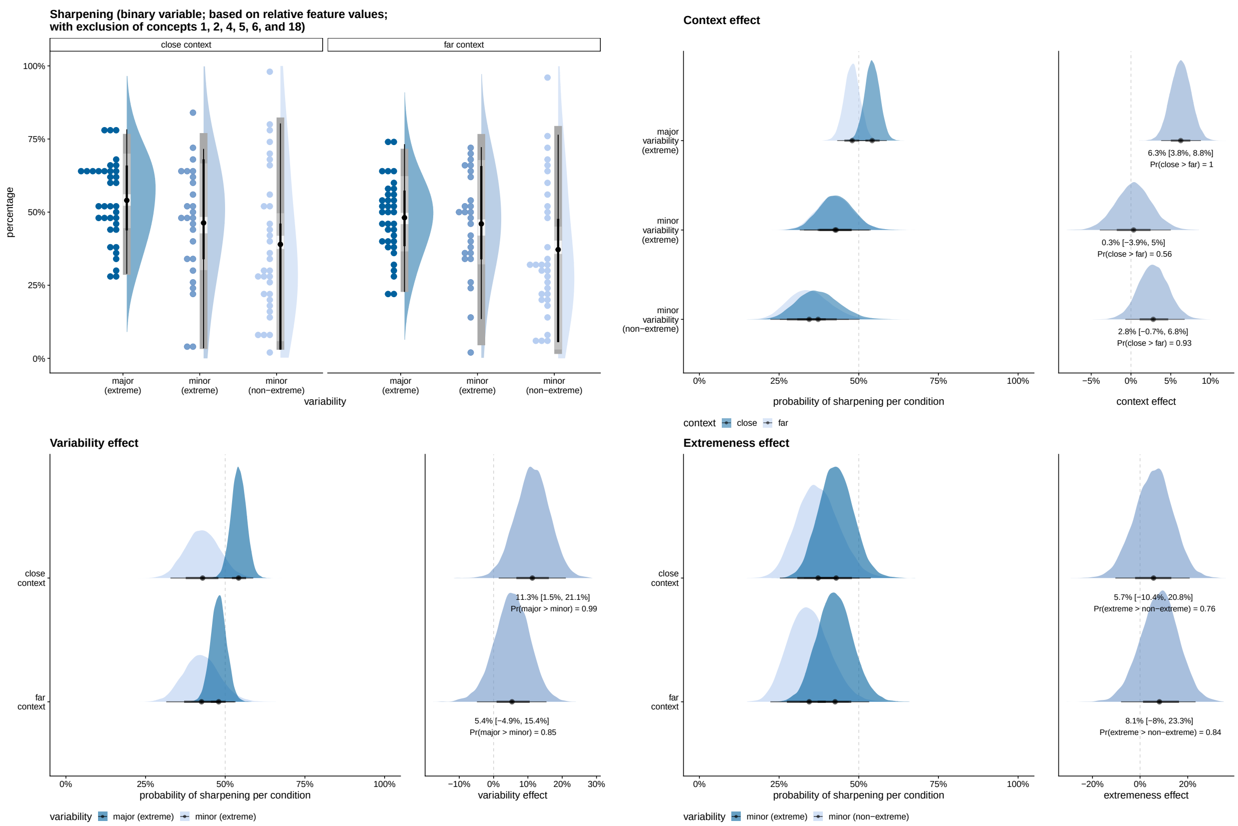


Figure D.5: Results of the qualitative model of sharpening, defined as a drawn value equal to the target or more extreme, based on the relative feature values, with exclusion of concepts 1, 2, 4, 5, 6, and 18.

The effect of *variability*, however, was very large given this alternative interpretation of leveling and sharpening (see Figure D.8). In all posterior samples, the probability for an extreme value on the major feature dimension to be sharpened was larger than for an extreme value on a minor feature dimension. Given that the model is a good approximation of the data, the data provide evidence for a clear effect of variability in both the close and far context conditions.

Also the effect of being one of the *extrema* on the feature dimension was very large given this alternative interpretation of leveling and sharpening (see Figure D.9). In all posterior samples, the probability for an extreme value on the minor feature dimension to be sharpened was larger than for a non-extreme value on a minor feature dimension. Given that the model is a good approximation of the data, the data provide evidence for a clear effect of extremeness in both the close and far context conditions.

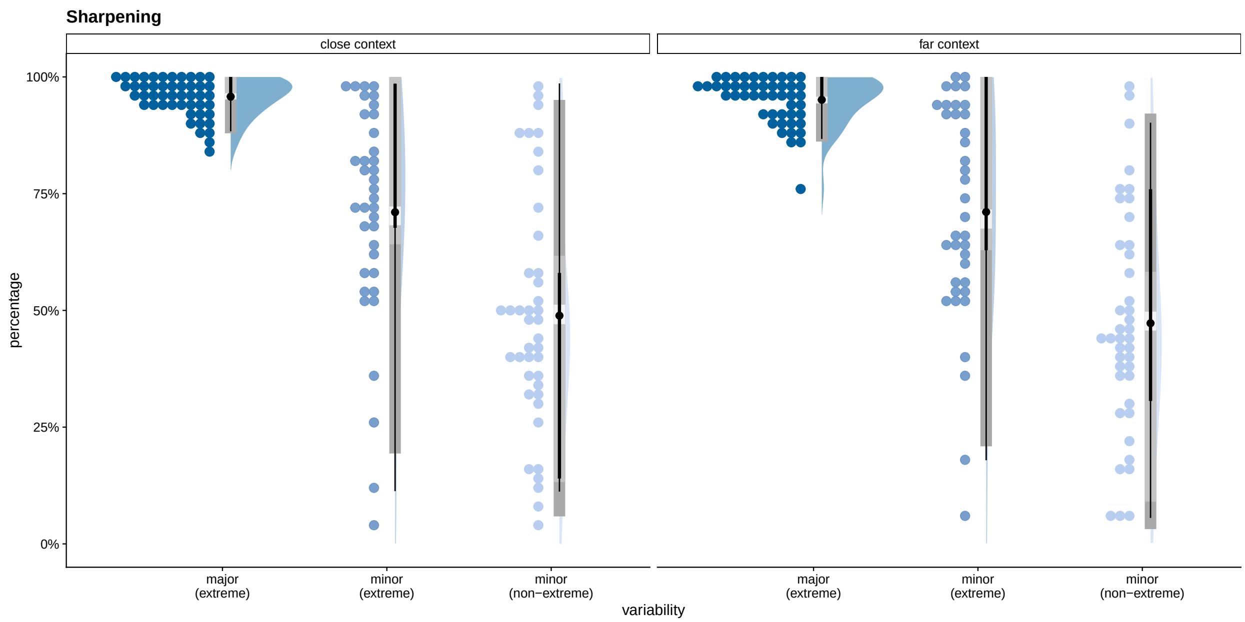


Figure D.6: Distribution of percentage of times a feature dimension was sharpened, per context and variability condition, with feature dimensions as individual data points. The black point and intervals indicate mean and 66% and 95% highest density continuous intervals for the data distribution. The grey intervals indicate the 66% and 95% highest density continuous intervals for the posterior predictive distribution, based on the qualitative model of sharpening as binary variable using the relative feature values. In addition, the white interval indicates the 95% highest density continuous interval for the mean posterior prediction.

Just as in the models using the other interpretations of leveling and sharpening, the estimated variability between feature dimensions was larger for the differences between the variability and extremeness conditions than for the context effect (see Figure D.10), and the estimated variability between feature dimensions was higher than the estimated variability between participants, which is potentially also a consequence of the limited number of data points per participant in comparison to the number of data points per feature dimension.

Figure D.11 shows the results for the equivalent qualitative model of sharpening based on the absolute rather than the relative feature values.

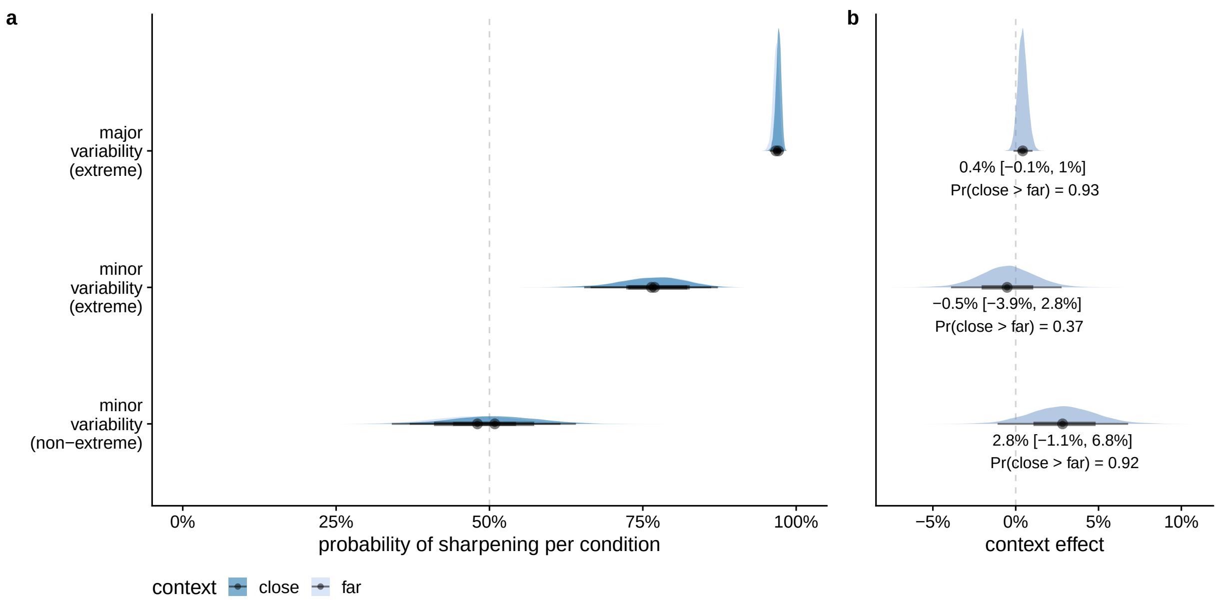


Figure D.7: (a) Posterior distribution for the probability of drawing a feature equal to the target value or more extreme, separately for each context and variability condition. (b) Estimated context effect in each variability condition. Mean, 66%, and 95% highest density continuous intervals are shown. The text also indicates the mean estimated difference with its 95% highest density continuous interval, as well as the posterior probability of the percentage of sharpening being higher in the close than in the far context, given a particular variability condition.

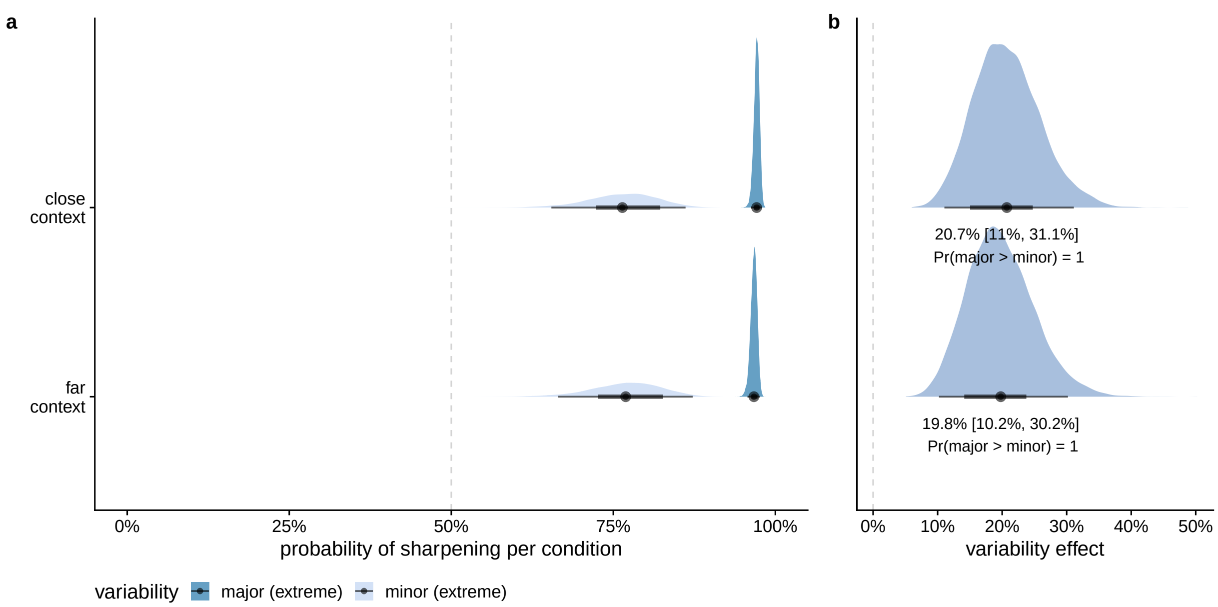


Figure D.8: (a) Posterior distribution for the probability of drawing a feature equal to the target value or more extreme, separately for the major (extreme) and minor (extreme) variability and context conditions. (b) Estimated variability effect in each context condition. Mean, 66%, and 95% highest density continuous intervals are shown. The text also indicates the mean estimated difference with its 95% highest density continuous interval, as well as the posterior probability of the percentage of sharpening being higher in case the feature exhibits major compared to minor variability, given a particular context condition.

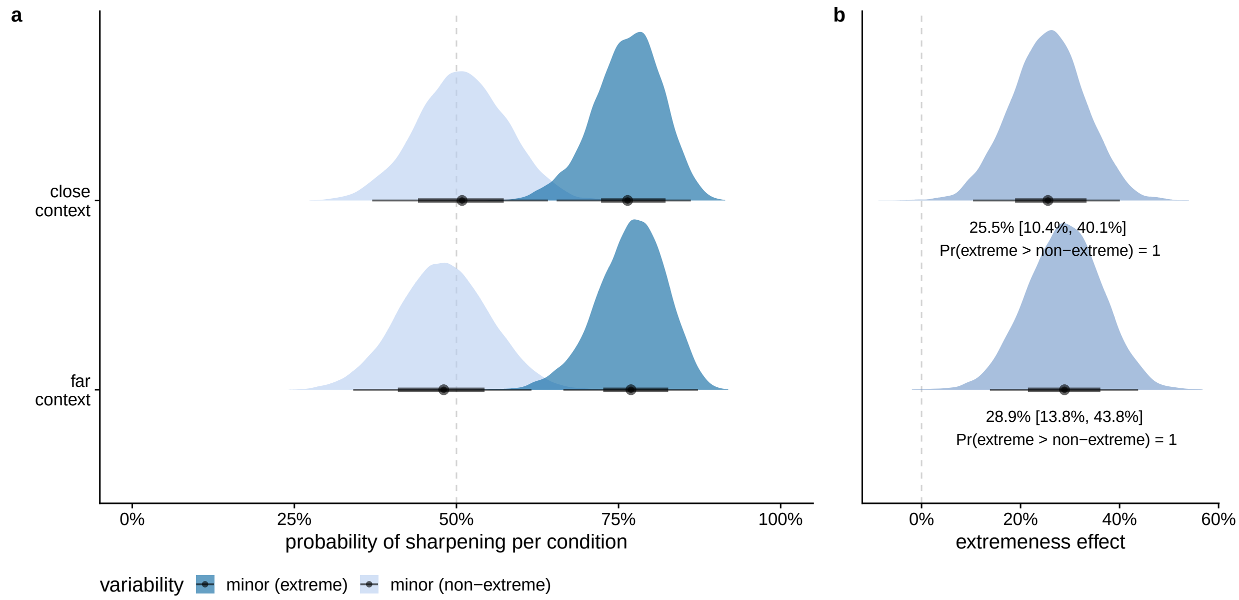


Figure D.9: (a) Posterior distribution for the probability of drawing a feature equal to the target value or more extreme, separately for the minor (extreme) and minor (non-extreme) variability and context conditions. (b) Estimated variability effect in each context condition. Mean, 66%, and 95% highest density continuous intervals are shown. The text also indicates the mean estimated difference with its 95% highest density continuous interval, as well as the posterior probability of the percentage of sharpening being higher in case the feature value is one of the extrema on the dimension, given a particular context condition.

D.2.3 Additional quantitative results

When inspecting the standard deviations for the feature dimension random effects (see Figure D.12), the estimated variability between feature dimensions was slightly larger for the differences between the variability conditions than for the context effect, but the difference in variability seemed less outspoken than in the qualitative analysis (see also Figure D.13). As in the other models, the estimated variability between feature dimensions was higher than the estimated variability between participants (see also Figure D.14).

Figure D.15 shows the results for the quantitative model of sharpening based on the relative feature values, but with exclusion of concepts 1, 2, 4, 5, 6, and 18.

D

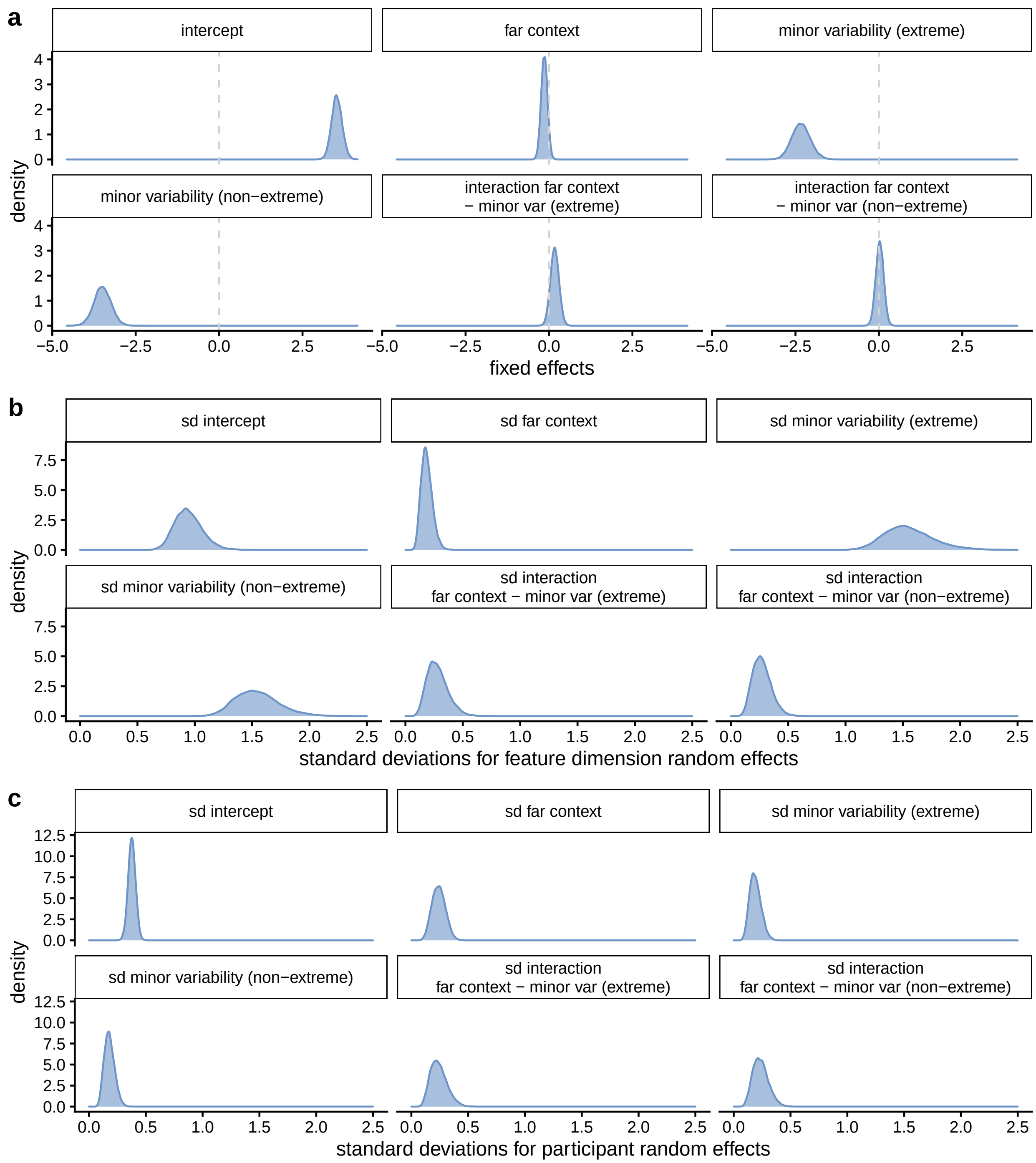


Figure D.10: Posterior distributions of fixed effects and standard deviations of random effects for concept and dimension combined, for the qualitative model, using the alternative interpretation of sharpening as a drawn value closer to the target value than to the mean value on the dimension (binarized; based on relative feature values).

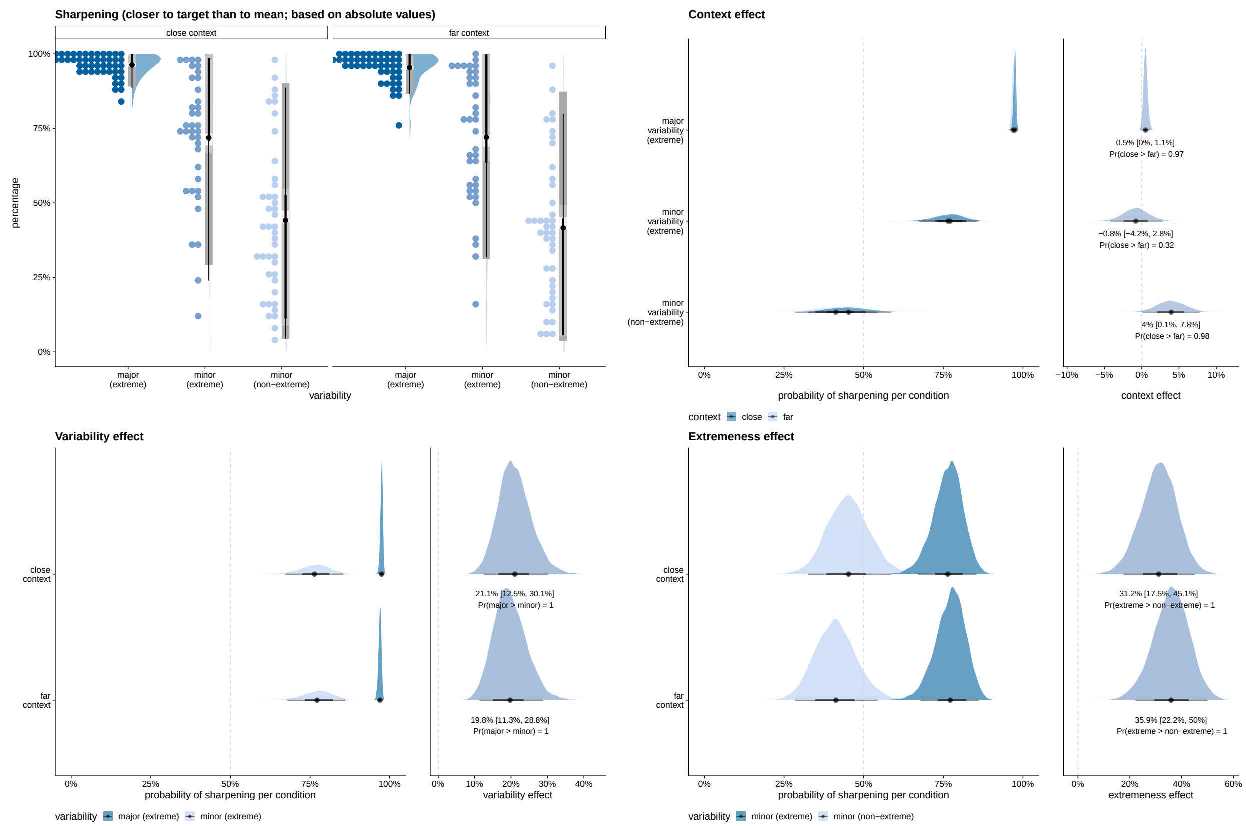


Figure D.11: Results of the qualitative model of sharpening, defined as a drawn value closer to the target than to the mean, based on the absolute feature values.

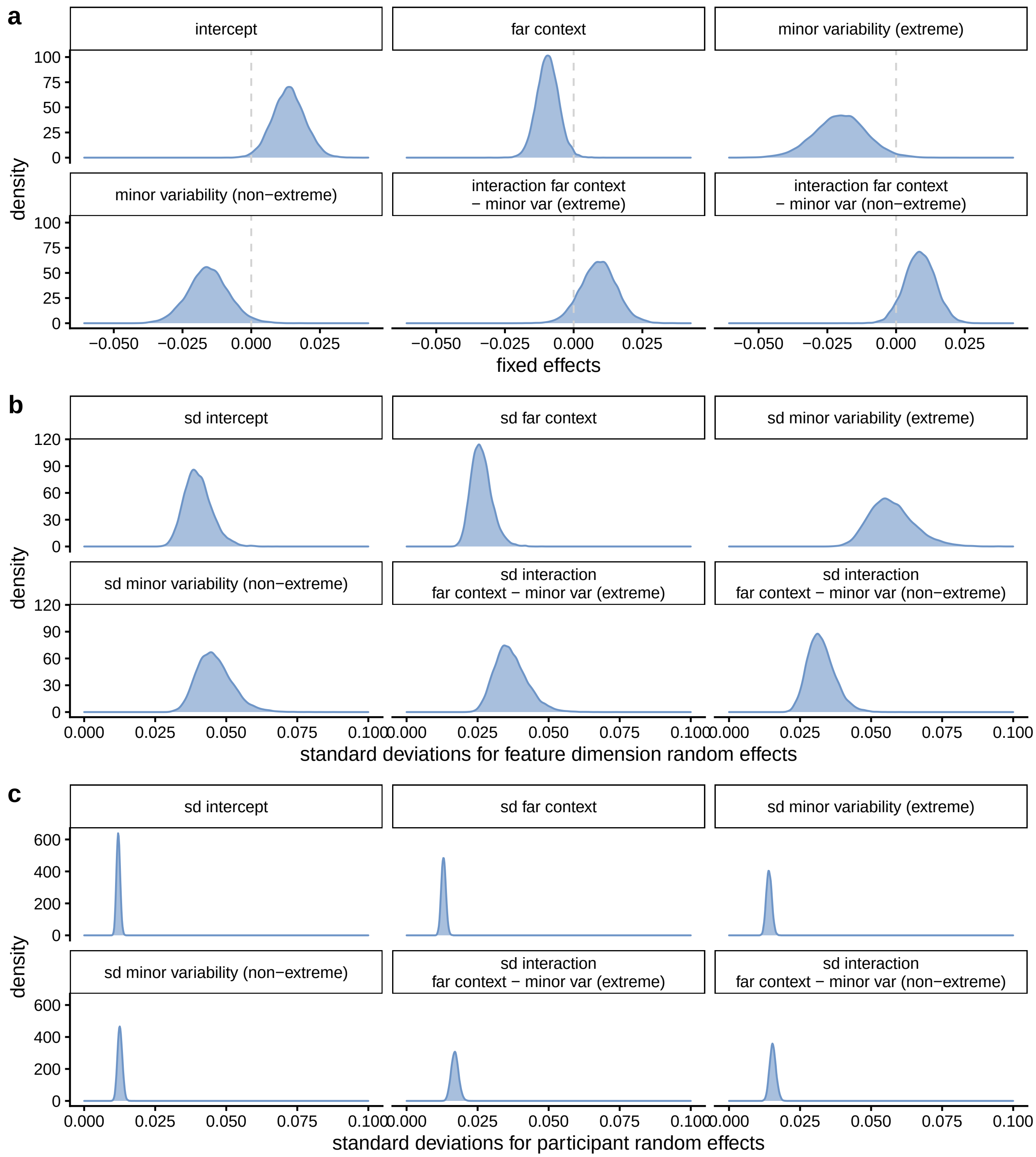


Figure D.12: Posterior distributions of fixed effects and standard deviations of random effects for concept and dimension combined, for the quantitative model (based on relative feature values).

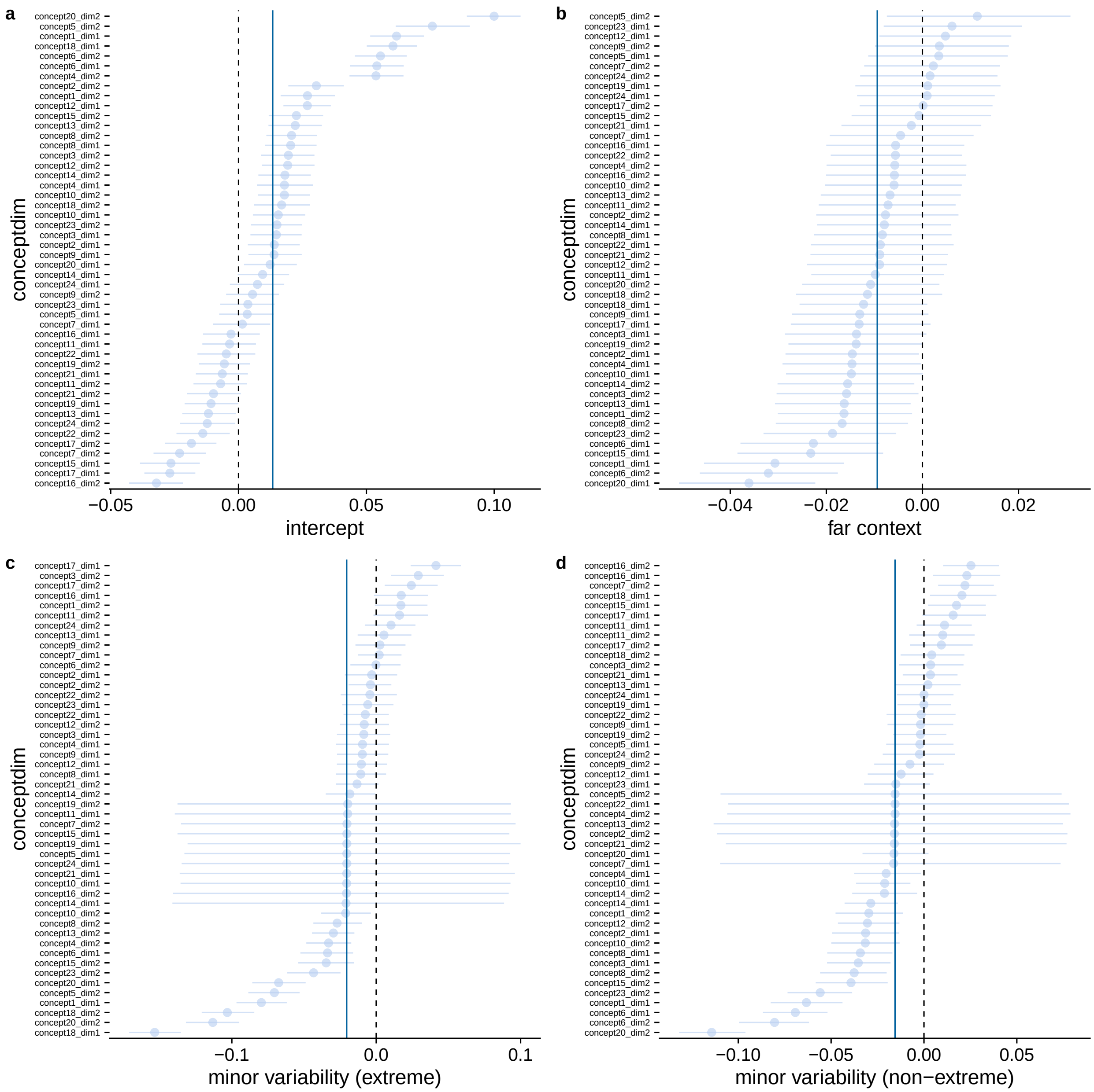


Figure D.13: Slopes for the intercept, the effect of context, the effect of a minor extreme feature value compared to a major extreme feature value, and the effect of a minor non-extreme feature value compared to a major extreme feature value on drawing a feature equal to the target value or more extreme, per feature dimension. Mean and 95% highest density continuous intervals are shown. The blue line indicates the average mean value across feature dimensions. The dashed black line indicates a value of zero. *Note.* The very wide intervals for some feature dimensions are the consequence of those feature dimensions not being present in the relevant condition, e.g., no minor variability (extreme) condition for that feature dimension.

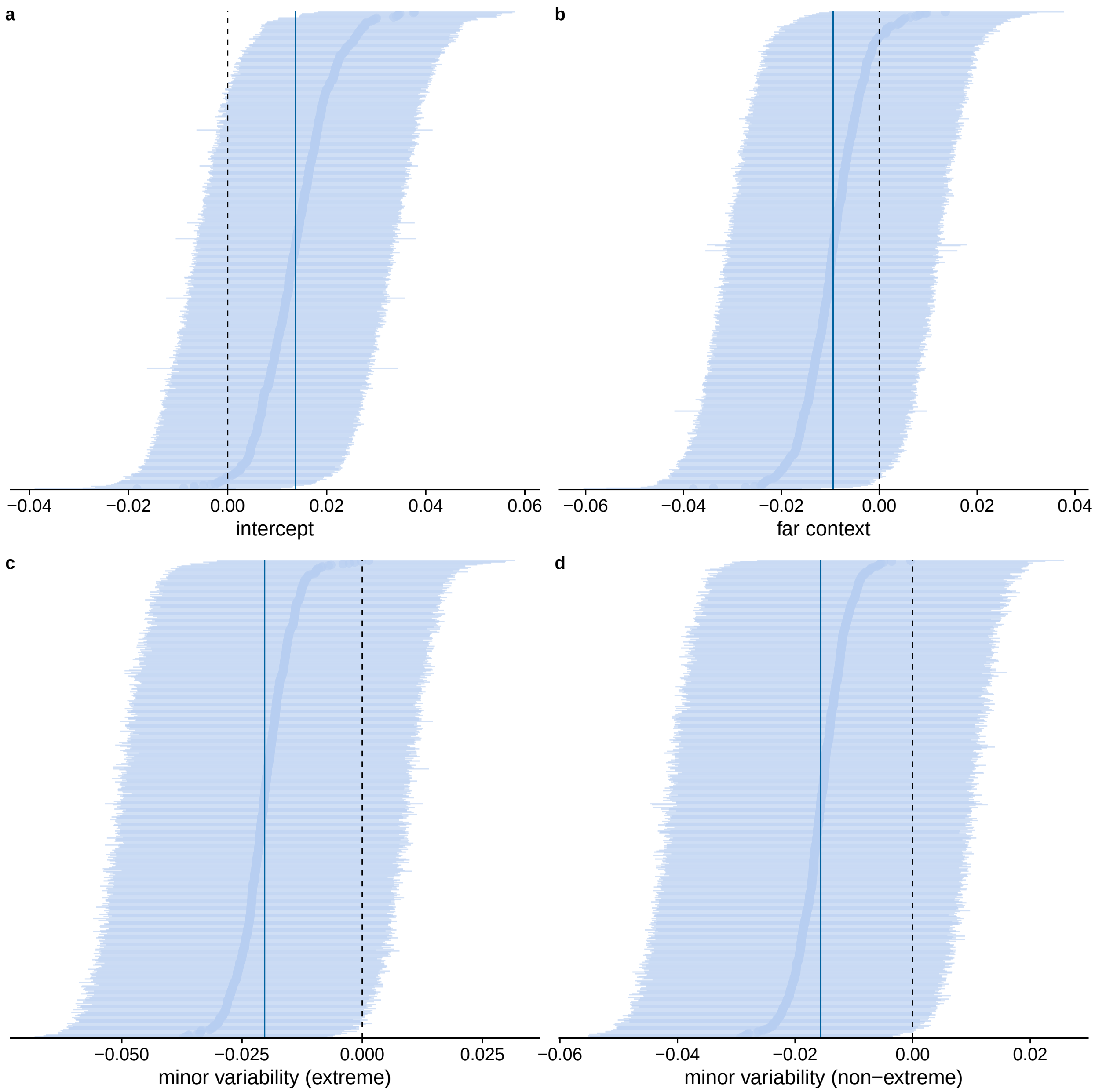


Figure D.14: Slopes for the intercept, the effect of context, the effect of a minor extreme feature value compared to a major extreme feature value, and the effect of a minor non-extreme feature value compared to a major extreme feature value on drawing a feature equal to the target value or more extreme, per participant. Mean and 95% highest density continuous intervals are shown. The blue line indicates the average mean value across participants. The dashed black line indicates a value of zero.

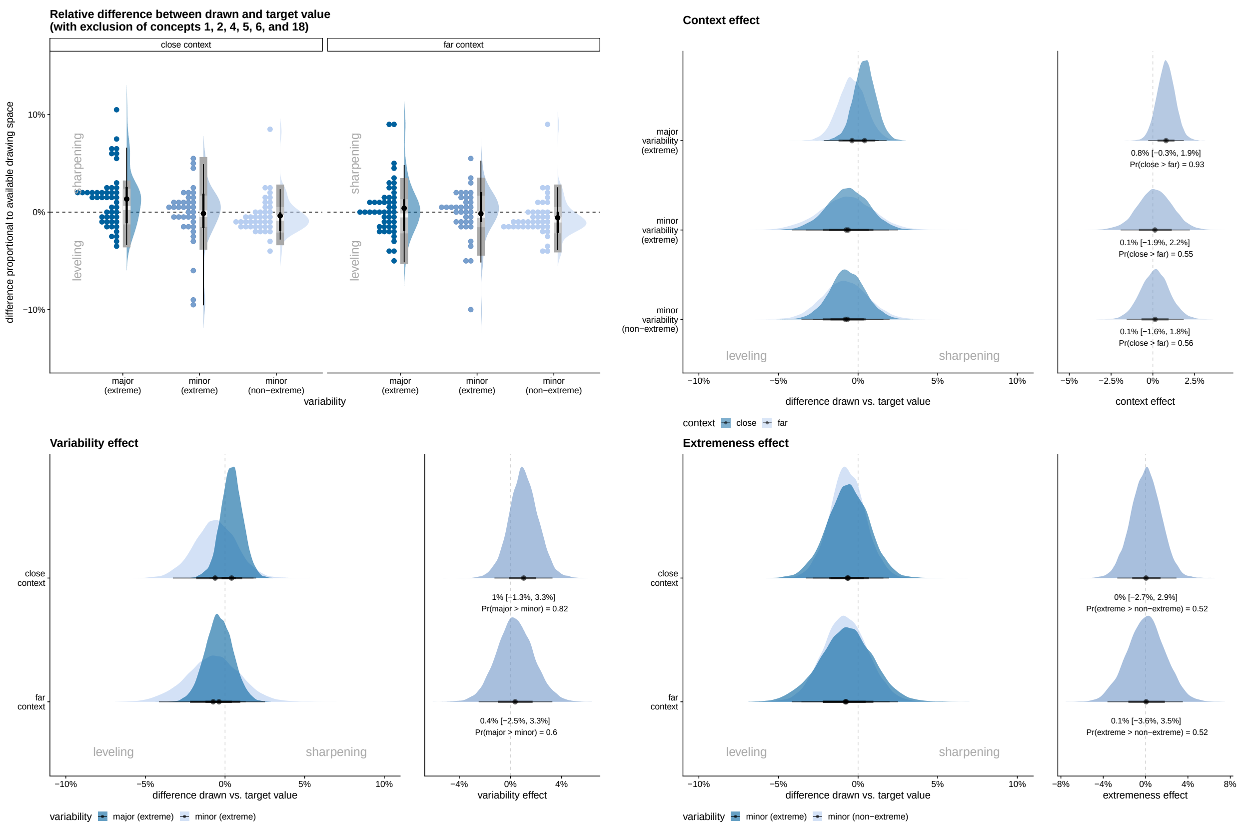


Figure D.15: Results of the quantitative model of sharpening, defined as a drawn value equal to the target or more extreme, based on the relative feature values, with exclusion of concepts 1, 2, 4, 5, 6, and 18.

D

D.3 Supplemental videos

In the HTML version of this Appendix, you can find screen recordings of some components that were part of this study (cf. Figure D.16).

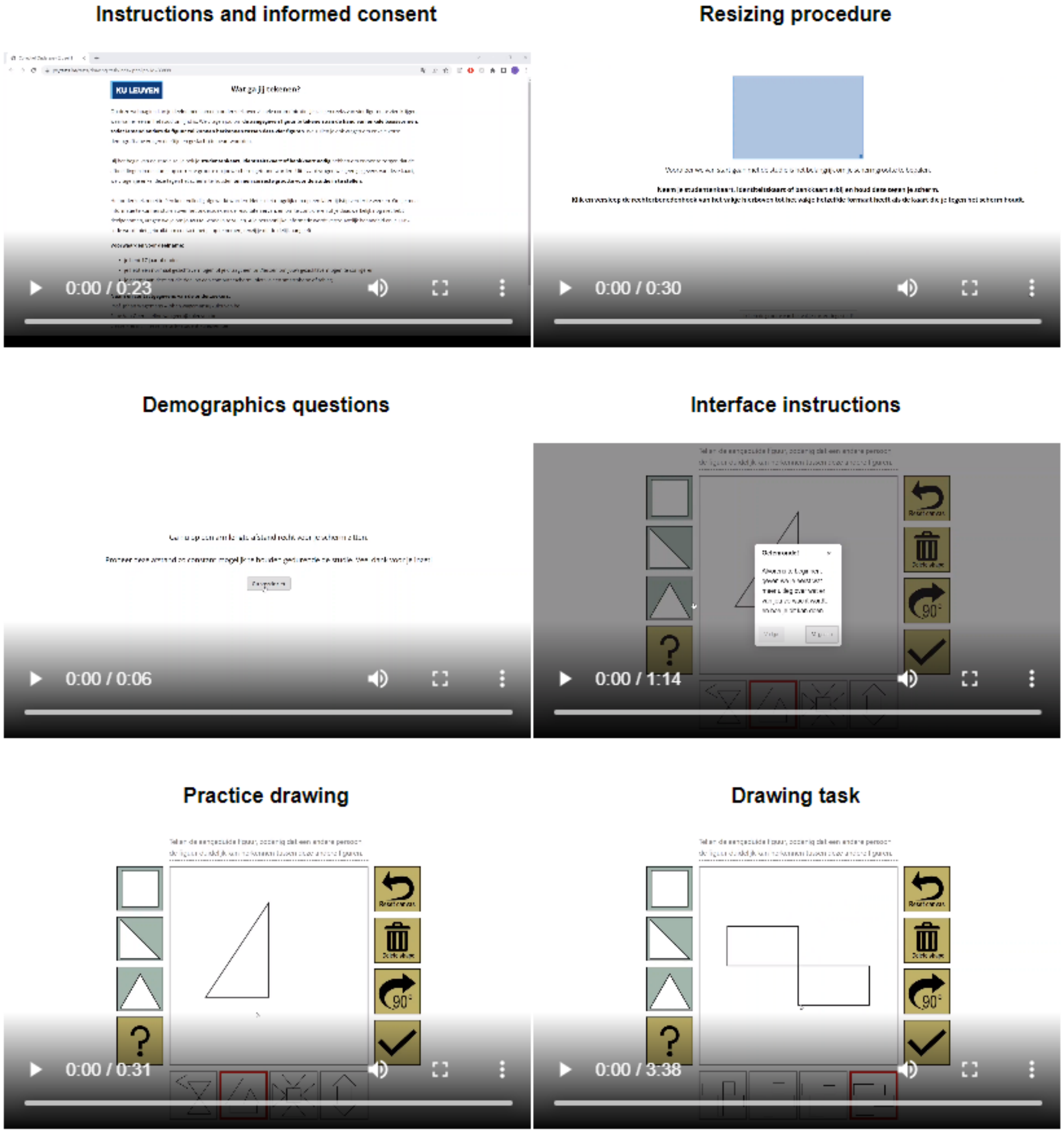


Figure D.16: Screen recordings of some components that were part of this study.

Appendix E
The Order & Complexity Toolbox for Aesthetics (OCTA)

E.1 The OCTA Shiny application

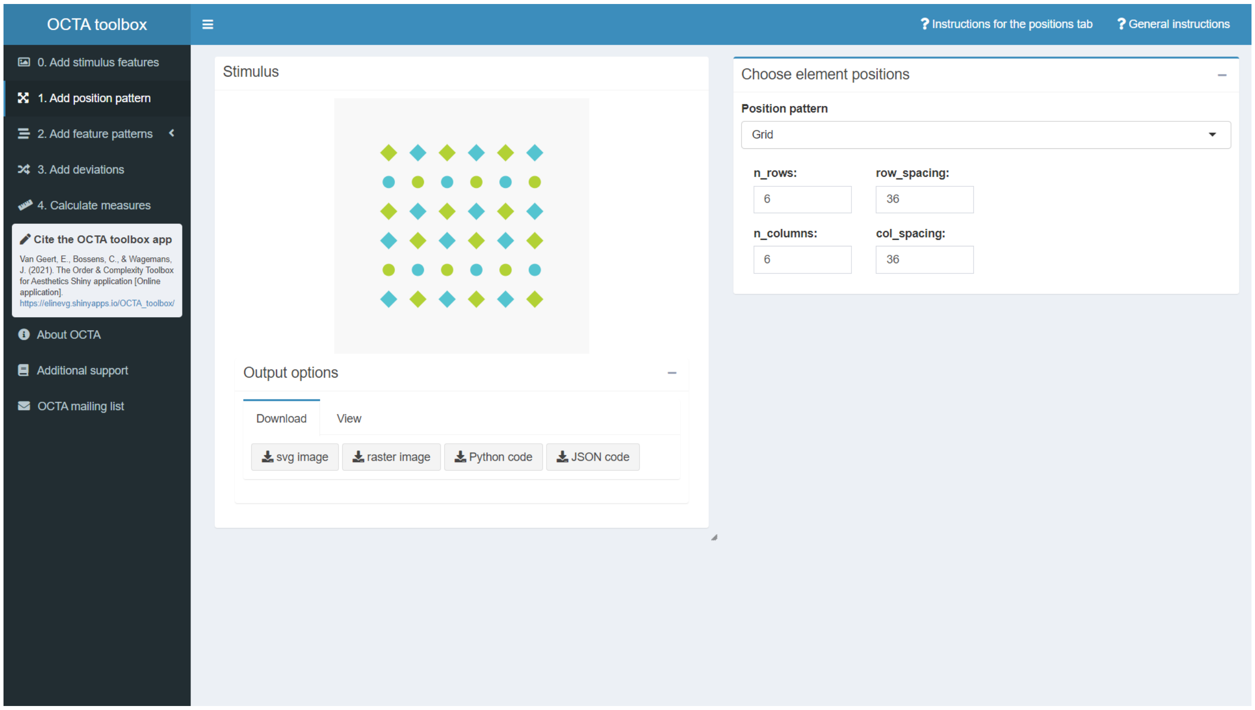


Figure E.1: The OCTA Shiny application. Click the link to try out the application.

E

E.2 Additional OCTA resources

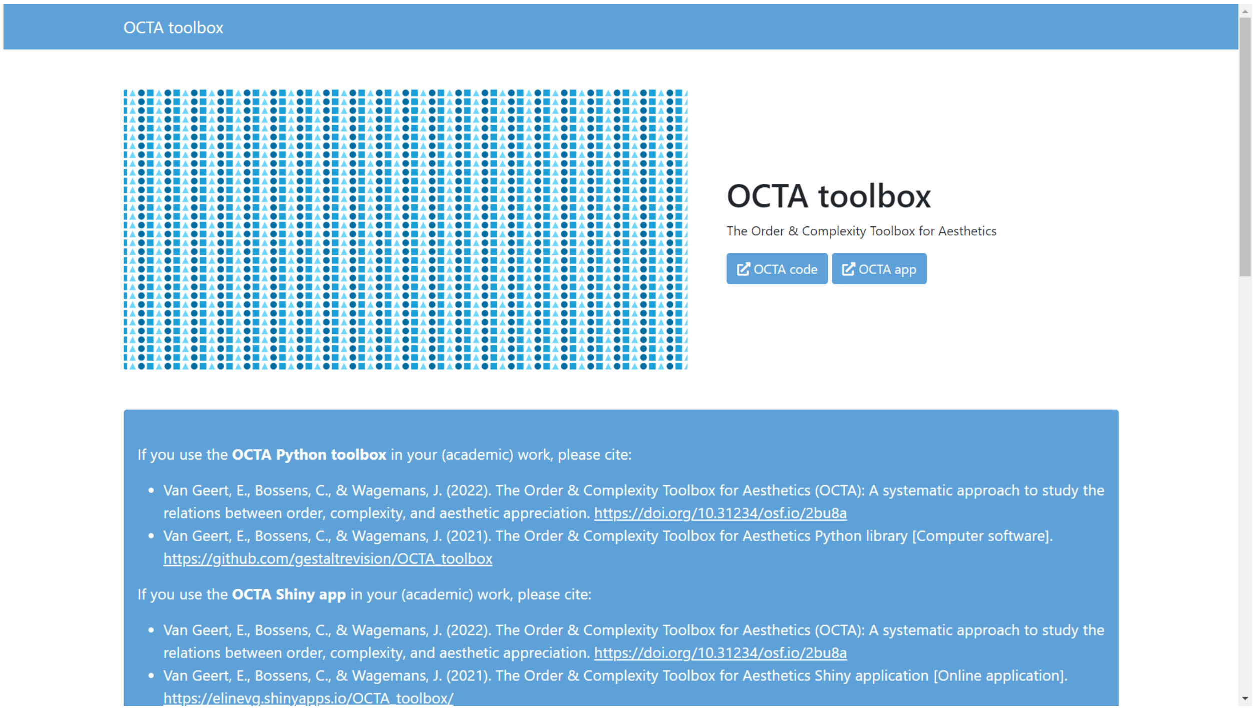


Figure E.2: The website collecting all OCTA resources available online. Click the link to visit the page.