

Effects of perceptual balance on aesthetic appreciation

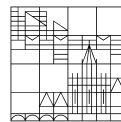
**Dissertation submitted for the degree of
Doctor of Natural Sciences
(Dr.rer.nat.)**

by

Martin Georg Fillinger

at the

Universität
Konstanz



Faculty of Sciences
Department of Psychology

Konstanz, 2019

Date of the oral examination: 14.02.2020

First referee: Prof. Dr. Ronald Hübner

Second referee: Prof. Dr. Christian G. Allesch

Acknowledgments

I would like to express my deep gratitude to my supervisor Ronald Hübner for his guidance, support and patience throughout the different phases of this work. Thank you for the possibility to realize this research project, for the training of my scientific thinking and working, and for constantly providing me with your knowledgeable and helpful advice.

I would also like to thank the rest of my dissertation committee. My special thanks go to Christian Allesch, whose extensive work on psychological aesthetics was always an important and inspiring scientific source for me. I would also like to thank Ulf-Dietrich Reips for his research on internet-based methodologies, which was also a solid foundation for my own online experiments.

I am especially grateful to my colleagues at the department of psychology at the University of Konstanz. I particularly would like to thank Tim Kuhlmann, Sebastian Dummel and Michel Druey, who supported me with their helpful comments, suggestions and corrections on this dissertation.

Finally, I would like to thank my parents for their unconditional support as well as my closest friends, especially Peter Ottmann, who was always there when times were tough. Moreover, I would like to offer my special thanks to Bernadette Prieschl, who encouraged me on my way to the decision to write this dissertation.

Abstract

In this dissertation, the effect of perceptual balance on aesthetic appreciation was investigated. Perceptual balance is a complex visual feature that depends on several factors and mechanisms, whose details remain largely unknown. Consequently, in four research papers, the concept and psychological mechanisms of perceptual balance and its effect on the aesthetic appreciation were investigated. In this context, different types of perceptual balance were examined. Moreover, the relations and interactions between balance and other visual features were studied. Thus, this dissertation aimed to contribute to a more universal theory and agreement about a general concept of perceptual balance.

For this purpose, in the first research paper, objective measures based on the notion of mechanical balance were used to examine how well they account for balance, symmetry and preference judgments of simple pictures. Based on this mechanical metaphor, it is assumed that each element in a picture has a certain perceptual weight that depends on its low-level features such as color, size and form. At least for simple stimulus material, the concept of mechanical balance and its positive association with aesthetic appreciation and subjective balance ratings could be confirmed.

In order to further investigate the predictive power of the objective measures as well as the relation of balance and aesthetic appreciation, in the second research paper Japanese calligraphies were used as stimuli. The results showed that perceptual balance was uncorrelated with the measures and with liking. However, discounting the effects of prototypicality on liking revealed a negative effect of stability on liking (for atypical calligraphies). These findings demonstrate that visual features can compete and interact in a complex way, which is why it can be difficult to detect the effect of perceptual balance on liking.

The third research paper aimed to examine the extent to which the mechanical concept of balance holds for more complex pictures. Therefore, a set of non-representational abstract paintings were used, which could be divided into single-element, multiple-element and dynamic-pattern pictures. Whereas mechanical balance was applied to assess single-element pictures, the balance of multiple-element and dynamic-pattern pictures was rated more in the sense of gravitational stability. Only for the multiple-element stimuli, there was a positive relation between balance, stability and

liking. Consequently, there are different types of balance, and their relation with liking depends on the picture type.

Finally, the fourth research paper addressed the relation between perceived stability and aesthetic appreciation because the previous research papers suggested two types of instability: first, a gravitational instability that is disliked; and second, an instability that implies movement and is liked. Systematical investigations of these two different types of visual instability confirmed that movement is responsible for the positive effect of instability on aesthetic appreciation because movement is positively correlated with emotionality. Consequently, instability reduces the aesthetical appreciation of a picture unless it implies movement.

Taken together, the findings of this dissertation support the notion that perceptual balance is understood as a mechanical balance at least for simple pictures. Objective measures that reflect this mechanical metaphor were significantly correlated with subjective balance ratings and aesthetic appreciation. However, depending on the picture type, the relation and meaning of balance can vary. As an alternative construct of perceptual balance, perceived stability was introduced. Moreover, the findings uncovered that perceptual balance could interact and compete with other visual features in a complex way. Consequently, this dissertation provides a deeper insight into the mechanism of balance perception, which is relevant for psychological aesthetics as well as art theory.

Zusammenfassung

In dieser Dissertation wurde der Einfluss der perzeptuellen Balance auf die ästhetische Wahrnehmung untersucht. Die perzeptuelle Balance ist eine komplexe visuelle Eigenschaft, welche von mehreren Faktoren und Mechanismen abhängt, deren Details noch weitgehend unbekannt sind. In vier Forschungsarbeiten wurde daher das Konzept und die psychologischen Mechanismen der perzeptuellen Balance und ihre Auswirkungen auf die ästhetische Wahrnehmung erforscht. In diesem Zusammenhang wurden verschiedene Arten der perzeptuellen Balance untersucht. Darüber hinaus wurden die Beziehungen und Wechselwirkungen zwischen Balance und anderen visuellen Eigenschaften analysiert. Auf diese Weise wurde mit dieser Dissertation, ein Beitrag zu einer universelleren Theorie und Einigung über ein allgemeines Konzept der perzeptuellen Balance angestrebt.

Zu diesem Zweck wurden in der ersten Forschungsarbeit objektive Maße, welche auf dem Prinzip der mechanischen Balance beruhen, verwendet, um zu untersuchen, wie gut sie die Beurteilung von Balance, Symmetrie und Präferenz von einfachen Bildern vorhersagen können. Basierend auf dieser mechanischen Metapher wird davon ausgegangen, dass jedes Element in einem Bild ein bestimmtes perzeptuelles Gewicht besitzt, das von Eigenschaften wie Farbe, Größe und Form abhängt. Zumindest für einfaches Stimulusmaterial konnte das Konzept der mechanischen Balance und dessen positiver Zusammenhang mit ästhetischer Wertschätzung und subjektiver Balancebewertungen bestätigt werden.

Um die Vorhersagekraft der objektiven Maße sowie die Beziehung von Balance und ästhetischer Wertschätzung weiter zu untersuchen, wurden in der zweiten Forschungsarbeit japanische Kalligrafie als Reize verwendet. Die Ergebnisse zeigten, dass die perzeptuelle Balance nicht mit den objektiven Maßen und den ästhetischen Urteilen korreliert war. Nachdem die Auswirkungen der Prototypikalität auf die ästhetischen Urteile kontrolliert wurden, zeigte sich jedoch ein negativer Einfluss der Stabilität auf die ästhetischen Urteile (bei atypischer Kalligrafie). Diese Ergebnisse zeigen, dass visuelle Eigenschaften auf komplexe Weise konkurrieren und interagieren können, weshalb es schwierig sein kann, den Effekt der perzeptuellen Balance auf das ästhetische Urteil zu erkennen.

Die dritte Forschungsarbeit zielte darauf ab, zu untersuchen, inwieweit das mechanische Balancekonzept für komplexere Bilder gilt. Daher wurde eine Reihe von nicht repräsentationalen abstrakten Gemälden verwendet, die sich in Bilder mit Einzelelementen, mehreren Bildkomponenten und dynamischen Mustern unterteilen ließen. Während bei der Beurteilung von Bildern mit Einzelelementen die mechanische Balance angewendet wurde, wurde die Balance von Bildern mit mehreren Bildkomponenten und dynamischen Mustern eher im Sinne der Gravitationsstabilität bewertet. Nur bei den Reizen mit mehreren Bildkomponenten gab es einen positiven Zusammenhang zwischen Balance, Stabilität und ästhetischer Wertschätzung. Folglich gibt es verschiedene Arten von Balance, und ihre Beziehung zum ästhetischen Urteil hängt vom Bildtyp ab.

Schließlich beschäftigte sich die vierte Forschungsarbeit mit dem Zusammenhang zwischen wahrgenommener Stabilität und ästhetischer Wertschätzung, da die vorangegangenen Forschungsarbeiten zwei Arten von Instabilität nahelegten. Erstens, eine Gravitationsinstabilität, die nicht gemocht wird. Zweitens, eine Instabilität, die Bewegung impliziert und gemocht wird. Systematische Untersuchungen dieser beiden unterschiedlichen Arten visueller Instabilität bestätigten, dass Bewegung für den positiven Effekt der Instabilität auf die ästhetische Wahrnehmung verantwortlich ist, da Bewegung positiv mit Emotionalität korreliert ist. Folglich reduziert Instabilität die ästhetische Wertschätzung eines Bildes, es sei denn, sie impliziert Bewegung.

In ihrer Gesamtheit unterstützen die Ergebnisse dieser Dissertation die Annahme, dass perzeptuelle Balance zumindest für einfache Bilder als mechanische Balance verstanden wird, da die objektiven Maße, welche dieses mechanische Prinzip widerspiegeln, signifikant mit subjektiven Balancebewertungen und ästhetischer Wertschätzung korrelierten. Je nach Bildtyp können jedoch die Beziehung und Bedeutung der Balance variieren. Als alternatives Konzept der perzeptuellen Balance wurde die wahrgenommene Stabilität eingeführt. Darüber hinaus zeigten die Ergebnisse, dass die perzeptuelle Balance auf komplexe Weise mit anderen visuellen Eigenschaften interagieren und konkurrieren kann. Folglich bietet diese Dissertation einen tieferen Einblick in den Mechanismus der Balancewahrnehmung, der sowohl für die psychologische Ästhetik als auch für die Kunsttheorie relevant ist.

Contents

Acknowledgments	I
Abstract.....	II
Zusammenfassung	IV
Contents.....	VI
List of Tables.....	X
List of Figures.....	XI
1 Introduction	1
1.1 Aesthetic appreciation.....	3
1.2 Origins of psychological aesthetics.....	4
1.3 The empirical turn in aesthetics	5
1.3.1 Fechner’s psychophysical theory	6
1.3.2 Gestalt psychology and aesthetics.....	8
1.4 Perceptual balance and aesthetic appreciation	10
1.5 Experimental work on perceptual balance	12
1.5.1 Testing Arnheim’s assumption of perceptual balance	13
1.5.2 Objective measures of balance.....	14
1.6 Alternative concepts of balance	15
1.7 Relations and interactions with other visual features.....	16
1.8 The aim of the present work	17
2 Summary of Research Papers.....	20
2.1 Research Paper I.....	20
2.2 Research Paper II	27
2.3 Research Paper III.....	32
2.4 Research Paper IV.....	38
3 General Discussion.....	44
3.1 Concepts of perceptual balance and its effect on the aesthetic appreciation ..	45
3.2 Examination of stability as an alternative concept of perceptual balance	47
3.3 Relations between perceptual balance and other visual features	48
3.4 Limitations and recommendations for further research	50
3.5 Conclusion	52
4 Research Paper I.....	54
4.1 Abstract	55
4.2 Introduction.....	55
4.2.1 Assessment of Preference for Balance (APB)	57
4.2.2 Deviation of the Center of “Mass” (DCM).....	58

4.2.3	Mirror Symmetry (MS).....	58
4.2.4	Homogeneity (HG).....	58
4.3	Experiment 1.....	58
4.3.1	Method.....	59
4.3.1.1	Apparatus and Stimuli.....	59
4.3.1.2	Procedure.....	59
4.3.2	Results.....	59
4.3.2.1	Balance Ratings.....	59
4.3.2.2	Symmetry Ratings.....	61
4.3.3	Discussion.....	61
4.4	Experiment 2.....	62
4.4.1	Method.....	62
4.4.2	Results.....	62
4.4.3	Discussion.....	63
4.5	Experiment 3.....	64
4.5.1	Method.....	64
4.5.2	Results.....	64
4.5.2.1	Balance Ratings.....	65
4.5.2.2	Preference Ratings.....	65
4.5.3	Discussion.....	66
4.6	General Discussion.....	66
4.7	Author Contribution.....	68
4.8	Acknowledgments.....	68
4.9	References.....	68
4.10	Conflict of Interest Statement.....	69
5	Research Paper II.....	70
5.1	Abstract.....	71
5.2	Introduction.....	72
5.3	Experiment 1.....	73
5.3.1	Method.....	74
5.3.1.1	Participants.....	74
5.3.1.2	Stimuli.....	74
5.3.1.3	Procedure.....	75
5.3.2	Results.....	75
5.3.3	Discussion.....	77
5.4	Experiment 2.....	78
5.4.1	Method.....	79
5.4.1.1	Participants and Stimuli.....	79
5.4.1.2	Procedure.....	79
5.4.2	Results.....	80
5.4.3	Discussion.....	83
5.5	General Discussion.....	85

5.6	Authors' Note.....	87
5.7	Acknowledgments.....	87
5.8	Declaration of Conflicting Interests.....	87
5.9	Funding.....	87
5.10	ORCID ID.....	87
5.11	References.....	87
5.12	Author Biographies.....	89
6	Research Paper III.....	90
6.1	Abstract.....	91
6.2	Introduction.....	91
6.3	Experiment 1.....	93
6.3.1	Method.....	94
6.3.1.1	Participants.....	94
6.3.1.2	Stimuli.....	94
6.3.1.3	Objective Measures.....	94
6.3.1.4	Procedure.....	94
6.3.2	Results.....	95
6.3.2.1	Categorization Check.....	96
6.3.3	Discussion.....	98
6.4	Experiment 2.....	99
6.4.1	Method.....	99
6.4.2	Results.....	99
6.4.3	Discussion.....	100
6.5	General Discussion.....	101
6.6	Authors' Note.....	103
6.7	Declaration of Conflicting Interests.....	103
6.8	Funding.....	103
6.9	ORCID ID.....	103
6.10	References.....	104
6.11	Appendix A.....	105
6.11.1	Method.....	105
6.11.2	Stimuli and Procedure.....	105
6.11.3	Results and Discussion.....	105
6.12	Appendix B.....	106
6.12.1	Experiment 1.....	106
6.12.1.1	Liking Ratings.....	106
6.12.1.2	Balance Ratings.....	106
6.12.2	Experiment 2.....	107
6.12.2.1	Stability Ratings.....	107
6.12.2.2	Balance Ratings.....	107
7	Research Paper IV.....	108
7.1	Abstract.....	109

7.2	Introduction.....	110
7.3	Experiment 1.....	112
7.3.1	Method.....	113
7.3.1.1	Participants.....	113
7.3.1.2	Stimuli.....	113
7.3.1.3	Procedure.....	113
7.3.2	Results.....	114
7.3.3	Discussion.....	116
7.4	Experiment 2.....	117
7.4.1	Method.....	117
7.4.1.1	Preliminary study.....	117
7.4.1.2	Roundness.....	118
7.4.1.3	Participants.....	118
7.4.1.4	Procedure.....	118
7.4.2	Results.....	119
7.4.3	Discussion.....	121
7.5	General Discussion.....	121
7.6	Authors' Note.....	124
7.7	Acknowledgments.....	124
7.8	Author Contributions.....	124
7.9	References.....	124
	Contributions to the research papers.....	128
	Research Paper I.....	128
	Research Paper II.....	128
	Research Paper III.....	129
	Research Paper IV.....	129
	References.....	130

List of Tables

Table 1. Mean ratings (balance and liking of Experiment 1 as well as prototypicality and stability of Experiment 2) for the Japanese calligraphies.	29
Table 2. Results of Experiment 1.	34
Table 3. The thumbnails of the four basic stimuli (first row) and corresponding ratings.	41
Table 4. Example stimuli used in Experiment 2 with diametrically opposite movement and stability ratings.	42

List of Figures

- Figure 1. Example stimuli used in the experiments. Left panel: Pictures from the APB (Wilson and Chatterjee, 2005). The first and second number below each picture indicates the APB and DCM score, respectively. Note: the lower the value the higher the balance. In the top left figure additionally the different axes are shown for the demonstration of how the APB score is computed (see text for details). Right panel: Examples of the new stimuli. The intersections of the long lines in each picture indicate the respective center of mass (see text for details). The short lines imply the corresponding geometric center. 21
- Figure 2. Relation between preference ratings in Experiment 2 for the two picture types (circles, and hexagons) and the APB scores and DCM scores. 24
- Figure 3. Relation between preference ratings in Experiment 3 for the two picture types (circles, and hexagons) and the APB scores and DCM scores. 26
- Figure 4. The interaction between prototypicality and stability for the prediction of liking. 31
- Figure 5. The relation between balance ratings and liking ratings in Experiment 1 for the different picture categories obtained in the supplementary categorization study. Each data point represents the picture according to its number (see Table 2). The colors and shapes of the symbols indicate the corresponding category. The lines are the corresponding regression lines. 36
- Figure 6. The relations between stability ratings and liking ratings from Experiment 1 for the categories obtained in the supplementary categorization study. Each number represents one of the 42 stimuli. The colors indicate the stimulus categories. The lines represent the regression lines. 37
- Figure 7. Left: Relation between stability and liking ratings in Experiment 2. Right: Relation between stability and liking after the effect of implied movement has been removed. 43

1 Introduction

This dissertation explores the effects of balance on aesthetic appreciation. Balance is an essential state that plays an important role in many domains of our daily life. For instance, without a sense of balance, man would not be able to move upright. Interestingly, in contrast to other senses like seeing and hearing, which are represented more dominantly, most of the time we only become aware of the sense of balance when it is lost, which can result in dizziness and disorientation. Our biological equilibrium organ, the vestibular apparatus of the inner ear, is responsible for the sense of balance. It allows an adequate spatial orientation considering gravitation (Schönhammer, 2013, p. 77). Additionally, human perception is also affected by this physical reality of gravitation. The phenomenon that particular orientations are preferred within the visual space (e.g., vertical gradients of horizontal disparity are perceived more easily than horizontal ones; Mitchison & McKee, 1990) is called visual anisotropy (I. P. Howard & Templeton, 1966). However, the state of balance not only plays a significant role in physics and perception. Indeed, balance is also mentioned and discussed in many other domains and scientific disciplines, as will be briefly outlined below.

In Greek mythology, *Kairos* is the god and the personification of the right or opportune moment (Schaffner, 2006), namely the moment when the past and the future stand in balanced relation to each other. Kairos is shown as a winged youth who is balancing a scale on a razor. This bronze sculpture that was created from the Greek sculptor Lysippos should symbolize the perfection achieved through balance. Due to this urgent need for perfection, man is striving for the continuous maintenance of balance in life.

Moreover, in *Aristotelian Philosophy*, balance plays a fundamental role in the concern of reaching happiness. In Aristotle's Nicomachean ethics, the finding of the middle between excess and lack is the basic requirement for a virtuous and fulfilled life (Wolf, 2019). Consequently, this virtuous balanced state in life is worth striving for.

Furthermore, from an affective or psychodynamic perspective, balance is a significant concern. Josef Breuer and Sigmund Freud (1895) established the term *mental balance* in their studies on hysteria. They postulated the existence of primary and secondary consciousness, whereas the secondary is the inhomogeneous and symptom-producing consciousness that must be harmonized with the primary everyday con-

sciousness. Moreover, in later years, Freud highlighted that the pleasure principle should be in balance with the so-called death drive because otherwise this leads to mental illness (Freud, 1920).

These three examples demonstrate that the notion of balance is a widespread concept in human experience that is strongly influenced philosophical, art-historical and psychological concepts. However, in the context of aesthetics, the question emerges whether balance can also be understood as something that is dependent on the biological sense of balance: in other words, whether the balance of pictorial compositions relates to our bodily sense of balance. Allesch (2009) argued that an image from an unbalanced building would not cause dizziness but instead an aesthetical missense. Consequently, from a phenomenological perspective, it is necessary to distinguish between *perceived* and *bodily felt balance*. Following this differentiation of balance, in this dissertation I aim to focus on the phenomenon of perceived balance in pictorial compositions and its effects on aesthetic appreciation. Given that this balance is a visually perceived one, in the following it is called *perceptual balance*.

The results of the studies of this dissertation will reveal that the relation between perceptual balance and aesthetic appreciation always depends on the pictorial elements and their arrangement. Moreover, regarding the theoretical concept of balance, it must be said that there are several types of perceptual balance. In addition, because visual features (e.g., balance) can interact and influence each other, the importance of extracting of all features of a composition will be highlighted. Only then can misinterpretations of the relation between perceptual balance and aesthetic appreciation be avoided by taking the relations with other visual features into account.

For a better understanding of the results and their interpretation of the empirical part of this dissertation, in the following I will first introduce the research field of aesthetic appreciation and experience. For this purpose, after briefly discussing the object of investigation, its methods as well as models, I will describe the origins of psychological aesthetics and its development as a natural science. In this context, I will discuss two of the most important experimental movements of the field: Fechner's psychophysical approach and Gestalt psychology. Second, I will introduce the relationship between perceptual balance and aesthetic appreciation. In this context, the experimental history of perceptual balance is shown, which also includes the testing of Arnheim's theory of visual balance. Moreover, the importance of developing objective measures of balance,

alternative theoretical concepts of balance, and finally the relation and interaction of balance with other visual features will be explained.

1.1 Aesthetic appreciation

Aesthetics influences many of our daily decisions. For example, a more attractive face will stand out from the crowd, well-designed consumer products will animate us to buy them, and the beauty of artwork will determine whether it is worth investing time to look at it. All of these human phenomena are fundamentally psychological processes. The psychological discipline that investigates *aesthetic appreciation* on all of its different levels is *psychological aesthetic*, as “the study of our interactions with artworks; our reactions to paintings, literature, poetry, music, movies, and performances; our experiences of beauty and ugliness, our preferences and dislikes; and our everyday perceptions of things in our world—of natural and built environments, design objects, consumer products, and of course, people” (Smith & Tinio, 2014, p. 3).

As reflected by this definition, contemporary empirical aesthetics takes both the science of aesthetics and the science of art into account (Vartanian, 2014). Even though these two fields are obviously overlapping, in the past these two fields have often been investigated separately (Bullot & Reber, 2013). For this purpose, researchers have used a variety of methodical approaches to investigate human aesthetic appreciation. Besides experimental and physiological methods, qualitative analyses and imaging techniques are applied to examine aesthetic appreciation. Because aesthetic experience is ubiquitous in our lives, it also leads to a wide range of psychological responses that are investigated. These responses involve typical aesthetic feelings such as pleasure and liking, but also knowledge emotions (e.g., interest or surprise), hostile emotions (like anger or disgust), and self-conscious emotions (such as pride or shame) (Silvia, 2009). Moreover, aesthetic experiences from different domains have been compared between art experts and novices (e.g., Winston & Cupchik, 1992), showing that novices prefer popular art and experts high art.

Furthermore, psychological aesthetics aims to develop a general model of human aesthetic appreciation. Leder, Belke, Oeberst & Augustin (2004) developed a very prominent *model of aesthetic experience and aesthetic judgments*, which includes multiple stages of information processing. At the beginning of information processing, the model provides a perceptual analysis and memory integration, which are followed by an

explicit classification, cognitive mastering and evaluation of the aesthetic phenomena. Another *framework for the psychology of aesthetics* was proposed by Jacobsen (2006, 2010). This model includes seven different vantage points (diachronia, ipsichronia, mind, body, content, person, and situation) that can be understood as an organizational scheme to embed findings in a more coherent theory of aesthetic experience. Finally, the *processing fluency model* (Reber, Schwarz, & Winkielman, 2004) serves as a framework for explaining aesthetic experiences. It has been shown that stimuli that can be processed more fluently are preferred (Winkielman, Halberstadt, Fazendeiro, & Catty, 2006).

However, investigating human aesthetic appreciation has a long tradition, which should be briefly outlined before I continue with introducing experimental psychological aesthetics, in which the studies of this dissertation are embedded.

1.2 Origins of psychological aesthetics

Aesthetics has its origins in philosophy, and it can be traced back to the Ancient Greeks, where Plato is often mentioned as the founder of aesthetics. In Plato's theory of ideas, he assigned beauty a special place. He claimed that *beauty, goodness, and justice* are the highest of all ideas (Kurz, 2015, p. 19). However, he did not explicitly investigate the nature of art and beauty, nor did he establish a theory of aesthetics. Instead, it was the Pythagorean school that first showed that beauty has a function of order, which can be traced back to mathematical proportions (Allesch, 1987, p. 8).

However, especially for psychological aesthetics, the German-speaking cultural area of the 18th and 19th century can be described as its cradle. In this context, two philosophical personalities must be emphasized. First, the German philosopher Alexander Gottlieb Baumgarten, who was a professor at the university in Frankfurt a. d. Oder in Germany. In 1750/58, he published his book *Aesthetica* (Baumgarten, 2007) in two volumes. In his work, aesthetics was understood as a science of sensual experiences (*scientia cognitionis sensitivae*). His understanding of aesthetics referred to the Aristotelian differentiation between *aisthesis*, which is experience mediated by the senses, and *noesis*, which represents mental experience and abstract reflections that are independent of the senses. Because aesthetics deals with the sensual experiences, he was able to establish it as an autonomous philosophical discipline parallel to logics that is mainly about abstract reflections beyond the senses.

The second philosophical person who significantly influenced the science of aesthetics is Immanuel Kant (1724–1804). In his *Critique of Judgment* (Kant & Pluhar, 1987), he dealt with the subjective character of aesthetic judgments, which was already reflected by Baumgarten. In his opinion, beauty is explicitly identified with the beholder's sensual experiences rather than objective properties of the perceived object. For him, aesthetics is a critique of taste, which is a judgment beyond the principles of reason and hence merely empirical. In addition, he also claimed that an aesthetic judgment also has to be characterized by its disinterested nature. A judgment can be disinterested in the sense that no desires or other interests are associated with the percept. Instead, one is pleased by an object or painting simply for what it is. Kant rejected the idea of a metaphysical concept of beauty and instead focused on the role of the perceiver.

Consequently, if the aesthetic judgment is anchored in the subject rather than the object, it cannot be generalized by a universal rule or norm. With his subjectivism, which refuses a metaphysical law for judgments of beauty, Kant was largely an important pioneer in psychological aesthetics.

1.3 The empirical turn in aesthetics

Since the beginning, the main goal in the field of aesthetics was to discover a finite set of universal laws with which to describe and predict aesthetic experience. However, the turn from a philosophical to an empirical approach aimed to work against criticism of the solely theoretical approach of philosophical aesthetics. Only with an empirical aesthetics would it be possible to capture individual differences in aesthetical perception and appreciation. The analysis of the history of empirical aesthetics shows that four single persons significantly influenced the research field with their theories and methods (Vartanian, 2014). The first person to mention is Gustav Theodor Fechner (1801–1887), the founder of psychophysics, who developed the first psychological experiments for investigating aesthetic experiences. The second is Rudolf Arnheim (1904–2007), who worked on aesthetics from the perspective of Gestalt psychology. Next, with his behavioral sight on aesthetics, Daniel E. Berlyne (1924–1976) mostly contributed to the field by publishing his book *Aesthetics and Psychobiology* (1971). Finally, Colin Martindale (1943–2008) contributed to empirical aesthetics with his prototype theory of preference and a neural network theory of aesthetic preferences.

Although all of these mentioned personalities made essential contributions to empirical aesthetics, in this context I will only discuss Fechner's psychophysical approach and the approach of Gestalt psychology, especially Arnheim's Gestalt theory. Because the following studies are experimental psychological work using stimulus material with more or less complex pictorial compositions, these two theories are the most important ones regarding this dissertation. This distinction also reflects the traditional contrasting perspectives on psychological aesthetics: an experimental and informational (Fechner's psychophysical theory) as well as an interpretative (Gestalt psychology) approach (Cupchik, 2014).

1.3.1 Fechner's psychophysical theory

In 1876, Gustav Theodor Fechner published the seminal work *Vorschule der Aesthetik* (1876), which marks the beginning of the research field of psychological aesthetics. With this work, he opposed the speculative aesthetics, which was at that time still mostly represented in Germany. Until that point in time, mainly the deductive approach of philosophical aesthetics, which derives the aesthetic effect and appreciation from the most general ideas and concepts to the particular, has been advocated. However, Fechner, who was a psychophysicist, had the claim to examine single aesthetic phenomena inductively "from below" instead of "from above," which was the procedure of the deductive aesthetics (Allesch, 2018b, p. 2). Consequently, with his inductive proceeding, he established an experimental aesthetics, which is until today characterized by an empirical approach.

In the context of investigating mental phenomena, Fechner's central assumption was that the physical properties of a stimulus could directly cause sensations. Moreover, he distinguished between *outer* and *inner psychophysics* (Allesch, 2018b, p. 26). With the outer psychophysics, he embraced processes that refer to the relationship between variations in physical properties and the mental. In contrast, inner psychophysics deals with the relationship between mental phenomena and the neural activities that underlie them. It should be emphasized that with this distinction of psychophysics, Fechner was truly ahead of his time. Even though there was no possibility to observe neuronal activities, he took the interaction between neural and perceptual processes explicitly into account—a perspective which until today is one of the main goals of neuropsychology.

Noteworthy are the three fundamental methods, which he already discussed in the book *Zur experimentalen Aesthetik* (1871) and which he eventually used for conducting his experiments documented in the *Vorschule der Aesthetik* (1876). First, with his *method of choice*, he asked participants to indicate their liking for a certain stimulus set. For this purpose, they could be asked to choose between two stimuli (paired comparison), order the stimuli according to their preference (ordering method), or indicate their preference for every single stimulus on a scale (rating method for a single stimulus). Second, the *method of production* involves asking participants to create a picture or an object that shows features or properties which are confirming their preference. Third, the *method of use* involved an examination of a broad range of objects that are used very often. It is suggested that their features reflect aesthetic preference due to their frequency. Until now, these methods are widely used by researchers in the field of empirical aesthetics. Moreover, all studies presented in Research Paper I-IV are based on Fechner's method of choice.

With the intention to establish an empirical aesthetics, Fechner wanted to underpin the perception of beauty with general properties. Because he was aware that he was only able to show the regularities of aesthetic effects, which appear in the experimental practice, he did not talk about laws. Instead, to avoid the notion of causality or lawfulness, he stated a *theory of principles of aesthetic appeal* in the *Vorschule der Aesthetik* (1876).

One of Fechner's key principles is the *principle of the aesthetic threshold*, which can be understood as a continuation of psychophysics. In this context, he distinguished between an outer and an inner threshold which must be exceeded to provoke an actual aesthetical reaction of pleasure or displeasure. Either stimulus intensity must be sufficiently strong to make an impression on the nervous system (outer threshold), or the individual susceptibility must be exceeded (inner threshold).

If single things that arouse pleasure coincide, the resulting pleasure is much higher than for the individual parts themselves. This phenomenon is what Fechner called the *principle of aesthetic help or enhancement*. Consequently, the beauty of some more or less beautiful elements together is higher than the sum of beautifulness of every single element.

With his third *principle of the unified connection of the many* Fechner states that a pleasing aesthetic stimulus must show an appropriate balance between complexity and

order. With other words, visual elements must be structured in a way that unity in variety can be perceived.

With the *principle of consistency, unanimity or truth*, Fechner claims that a pleasing aesthetic stimulus must dissolve an existing or conceivable contradiction between competing ideas or perceptual aspects of a stimulus. Based on this reduction of ambiguity, a stimulus gains in pleasure. In the context of Fechner's fourth principle, the truth can be understood as a kind of coherence between the ideas or perceptual aspects of the stimulus.

The *principle of clarity* is a result of Fechner's third and fourth principle. It is stated that the aesthetic impression depends on its equality, coherence and contradictoriness. As a result, it could happen that aesthetic pleasure is grounded in the felt clarity rather than in the nature of visual elements.

Finally, the *aesthetical principle of association* can be described best using an example. In most cases, an orange is considered more beautiful than a wooden ball of the same color. Fechner explains this fact by taking the taste and individual experiences of the perceiver into account. Following Fechner's sixth principle, it can be concluded that aesthetic experience can only be understood when the interaction of direct and associative factors is considered.

Consequently, it was Fechner who established a classical experimental and informational foundation of psychological aesthetics which was a milestone for the field of empirical aesthetics and which was later completed by the psychobiological approach of Berlyne (Cupchik, 2014). However, as a contrasting research tradition, Gestalt psychology also significantly contributed to psychological aesthetics.

1.3.2 Gestalt psychology and aesthetics

In short, it can be argued that in contrast to Fechner's psychophysical approach, which involves the bottom-up processing of information, the approach of Gestalt psychology could be understood as an interpretatively top-down approach (Vartanian, 2014). The main focus of the studies of Gestalt psychology has always been on the wholeness of conscious experience. Until today, Gestalt psychology significantly influences the psychology of art and aesthetics (Jacobsen, 2006).

The term *Gestalt* was established by Christian von Ehrenfels (1859–1932). With his article *Über Gestaltqualitäten* (1890), von Ehrenfels gave the go-ahead for a theory

beyond elementarism (Ash, 1995, p. 88). Instead of presupposing the possibility of an addition of simple perceptions, as it was supposed by the atomistic psychology of the 19th century, von Ehrenfels emphasizes that the perceived whole of something is different from its single properties and parts in isolation. This may not seem surprising since von Ehrenfels was a student of Franz Brentano (1838–1917), who stressed the importance of the description of the direct mental experiences and interestingly also had warned Fechner of elementalism (Brentano, 1959). He argued that Fechner’s empirical aesthetics “from below” does not necessarily imply to start investigating an object by its single visual features (Allesch, 2018a). Instead, a more complex piece of art can be subjected to the examination “from below,” which is in accordance with the notion of Gestalt psychology.

Another student of Brentano was Carl Stumpf, who was both a professor of philosophy and, additionally, interested in experimental psychology (Ash, 1995, p. 28). He conducted experiments in psychoacoustic, where he worked with fundamental musical stimuli as intervals and chords (Stumpf, 1883/1890). Based on his studies, he developed a phenomenological psychology of sound, albeit which was more of a contribution to music psychology than to psychological aesthetics (Allesch, 2000/2001). Stumpf’s students Max Wertheimer, Wolfgang Köhler and Kurt Koffka finally founded the school of Gestalt psychology that experimentally pursued the notion of Gestalt. They dealt with the description and analysis of phenomena of perceptual organization that has been formulated in a series of *Gestalt laws* (Wagemans et al., 2012). Popular laws of Gestalt are, for instance, the law of the goodness of Gestalt, the law of similarity or proximity (Wertheimer, 1923).

However, Gestalt psychologists also “attempted to introduce an aesthetic dimension of inherent order, meaning, and simplicity into the evaluation of scientific theories, and into the fabric of experience and nature itself” (Ash, 1995, p. 1). It was Rudolf Arnheim (1904–2007) who tried to describe the aesthetic value of a good Gestalt. Arnheim came out of the Berlin School of Gestalt psychology and was mainly influenced by the writings of Kurt Koffka (Allesch, 1987, p. 468). With his main work *Art and Visual Perception* (1954), he tried to use the notion of Gestalt to establish a theory that makes it possible to interpret a piece of art as its own microcosm. In this context, he argued that “vision is not a mechanical recording of elements but the grasping of significant structural patterns” (Arnheim, 1954, Introduction, pp. viii). Moreover, he explained that in the psychology of arts and aesthetics, the understanding of elemental visual processes

would not be sufficient. In contrast to Fechner's physical approach, Arnheim described the act of seeing as an active exploration beyond a micro-analysis of elementary visual features. Instead, to understand aesthetic experience, for him, it was more important to describe a perceptual system that attempts to catch and scan an object as a whole (Vartanian, 2014).

To describe this optical process, he also made use of the Gestalt principles, rules of grouping and the effect of *principal phenomena*. Arnheim claimed that artists have knowledge about these principles and rules, which they can use to influence not only the formal visual organization of an image but also its aesthetics. Therefore, it can be argued that he introduced a theory of art and aesthetics that can satisfy both artists as well as empiricists.

One of these principal phenomena that Arnheim described against the background of his theory of aesthetics is *striving for balance*. The importance of balance is demonstrated by the fact that even the first chapter in his book *Art and Visual Perception* (1954), which can be seen as the foundation of the Gestalt psychological theory of aesthetics, is dedicated to this visual feature.

1.4 Perceptual balance and aesthetic appreciation

How well a picture is liked usually depends on various factors such as form, color and symmetry, as well as on more conceptual factors like semantic content and prototypicality (for an overview see Palmer, Schloss, & Sammartino, 2013). Moreover, the perceptual mechanisms involved in visual aesthetics and preference judgments have long been a matter of debate. However, since the seminal work of Gustav Theodor Fechner (1876), one approach of corresponding experimental studies has been to find *aesthetic primitives*, that are relatively simple perceptual features that determine a stimulus' attraction (Latto, 1995; Munar, Gómez-Puerto, López-Navarro, & Nadal, 2014). A prominent candidate in this respect is *perceptual balance*, because also in art theory and related fields, it is widely assumed that the aesthetic appreciation of a picture depends, among others, on how well it is balanced (Arnheim, 1982; Bouleau, 1980; Kandinsky, 1926/1979; Poore, 1903; Ross, 1907).

The artist and writer Henry Poore (1859–1940, Poore, 1903), for example, assumed that a balanced picture must show an equal amount of elements around the center of the picture. With his steelyard metaphor, he stated that “in the balance of a picture it

will be found that a very important object placed but a short distance from the centre may be balanced by a very small object on the other side of the centre and further removed from it” (p. 14).

Moreover, Denman Ross (1853–1935, Ross, 1907), who was a painter and art theorist, considered balance as part of harmony. He writes that a balanced arrangement of elements “is a Harmony of Positions due to the coincidence of two centers, the center of the attractions and the center of the framing” (p. 24). Moreover, he reported that balance is a state of equilibrium that leads to stillness and, consequently, to the abolition of change and movement. However, when something is inducing movement, this momentum of balance would be lost.

Based on the ideas of Henry Poore and Denman Ross, Rudolf Arnheim hypothesized that each rectangular frame has a hidden structure or field of invisible forces which can be understood analogously to a magnetic field in physics (1954). A visual element placed in a picture will be influenced by the psychological forces of the hidden structure. Consequently, an element is assumed to be stably settled when it is placed on one of the main or diagonal axes of the field. However, the most balanced position for a visual element is the center of the field where the four main axes are crossed. Placing an element at the center of a picture will lead to a compensation of the field’s forces and, consequently, to a perfectly balanced pictorial composition. However, the situation is more complex if more than one element is in the field. In this case, elements are not only influenced by the forces of the hidden structure but also by the interaction of the elements. Moreover, grouping, color or shape can further influence the element’s *visual weight* and, hence, the forces between the elements. As a result, Arnheim defined a pictorial composition as “an arrangement of visual elements creating a self-contained, balanced whole, which is structured in such a way that the configuration of forces reflects the meaning of the artistic statement” (Arnheim, 1982, pp. 215-216).

Following the idea that pictorial elements differ in their visual weight, an alternative characterization of perceptual balance is the equilibration between these weights. According to Arnheim (1954), elements lying close to the center pull less compositional weight than elements lying off the center. Consequently, the larger the distance between the perceptual and geometrical center, the less balanced that an element will appear. Moreover, a heavy weight located on one side of the picture’s fulcrum (the pivot point of a scale) can be balanced by a lighter weight positioned further away on the other side. Usually, a picture is considered as balanced if all its elements are arranged in such a

way that their perceptual forces are in equilibrium about a fulcrum and if the fulcrum coincides with the center of the picture (Locher, 2006).

Thus, for a pictorial composition, Arnheim defined balance as “the dynamic state in which the forces constituting a visual configuration compensating for one another” (Arnheim, 1982, p. 237). Moreover, he emphasizes the importance of the visual center of an image because it neutralizes the directed tensions and forces by producing an effect of standstill and immobility.

Taken together, it could be shown that there is a broad consensus among aestheticians that balance has a significant effect on the appreciation of a picture (e.g., Arnheim, 1954; Poore, 1903; Ross, 1907). Consequently, balance can be considered as one of the universal aesthetic principles (Vartanian, Martindale, Podsiadlo, Overbay, & Borkum, 2005). However, the mechanisms of balance perception remain largely unknown. Whereas most researchers agree on a global level what perceptual balance is, they disagree on the details of how balance is determined. At the one hand, balance can be understood more abstractly as a principle of visual order (Van Geert & Wagemans, 2019). At the other side, perceptual balance can also be interpreted more physically, which is reflected by Arnheim’s theory of visual balance. By using *mechanical balance* as a metaphor, he assumed that each element in a picture has a certain perceptual weight that depends on its low-level features such as color, size and form.

In the following section, I will outline the history of the experimental investigation of perceptual balance which also embraces the testing of Arnheim’s assumptions, especially the concept of visual weights in the context of the notion of mechanical balance.

1.5 Experimental work on perceptual balance

If we go back in history, then, as far as it is known, perceptual balance and its connection with aesthetic appreciation have first been systematically examined by Pierce (1894, 1896) who at that time was a graduate student in Hugo Münsterberg’s lab at Harvard University. In his studies, Pierce actually wanted to investigate perceptual balance by asking participants to adjust the horizontal position of a movable object (e.g., a line) on one side of a display such that the side was aesthetically equal to the opposite side, where another object (line of different size, etc.) was fixed at a certain position. The results of Pierce’s experiment revealed that at least under some condi-

tions, adjustments were made in accord with mechanical balance. This was conceptually replicated in an experiment by Ethel D. Puffer (1903), where participants were asked to shift an object on their own until the arrangement was felt to be pleasing (Puffer, 1905, p. 109).

Consequently, by adopting this mechanical metaphor of perceptual balance, the relative perceptual weight of pictorial elements can easily be assessed, at least within a simple context. One merely has to ask persons to adjust the horizontal position of a target element on a seesaw so that it is in equilibrium to a fixed element opposite to a fulcrum. This was one of the first methods applied in experimental research on perceptual balance as described above.

Moreover, in the meantime, further research in the field of perceptual balance has demonstrated that balance could be perceived rapidly (Locher, Krupinski, Mello-Thoms, & Nodine, 2007; Locher & Nagy, 1996; Locher & Stappers, 2002; McManus, Edmondson, & Rodger, 1985). Furthermore, it has been shown that humans develop a sense of balance around the age of nine (Golomb, 1987).

However, the problem remains to calculate the visual weights Arnheim mentioned in the context of his theory of perceptual balance. Therefore, in the following section, I will outline the empirical testing of Arnheim's theory of perceptual balance, which especially includes the consideration of the notion of perceptual weight. It will come to light that investigating the concept of visual weight has laid the foundation for computing objective measures of perceptual balance.

1.5.1 Testing Arnheim's assumption of perceptual balance

Arnheim's (1954) main assumptions have already been tested. McManus, Edmondson, and Rodger (1985), for instance, presented reproductions of artwork as well as plain stimuli and had their participants place a fulcrum beneath each picture so that it looked balanced (horizontally). For the reproductions of artwork, they found that the adjusted position of the fulcrum varied considerably, suggesting that artwork is not generally well balanced. Moreover, when participants had to locate the perceptual center for unchanged pictures and for pictures where a portion was removed, the locations were rather similar. From these results, McManus et al. (1985) concluded that the balance of a picture depends more "upon a global integration of the picture as a whole than of any individual element of it" (p. 314).

Even for their plain stimuli, McManus et al. (1985) found no simple relation. Whereas element position was crucial for positioning the fulcrum, size and color of the elements were less important. Furthermore, although the distance of an element from the frame's geometric center and its size led to a larger shift of the fulcrum, these two factors were not correctly integrated for the judgment of balance.

In a later study, Locher, Gray, and Nodin (1996) used reproductions of twentieth-century art paintings and a manipulated less-balanced version of each. Art experts and non-experts had to rate the balance of each picture and determine the (two-dimensional) center of perceptual mass, which depends on the perceptual weight of the pictorial elements. If this center coincides with the geometric center of the frame, then according to Arnheim (1954) the picture is balanced. As a result, both groups moved the center for the disrupted version, but only the experts judged this version as less balanced. Locher et al. (1996) concluded that the center of perceptual mass and the overall judgment of balance are not as close as thought.

Because these results barely supported Arnheim's theory, Cupchik (2007) speculated that the terms of the theory were only meant metaphorically. McManus, Stöver, and Kim (2011), however, believed that Arnheim wanted his theory to be taken literally, namely in a physical sense. To test their conjecture, they even went a step further and, instead of asking participants to indicate the fulcrum of pictures, they calculated the center of perceptual mass by assuming that the mass of each pixel in a gray-level picture corresponds to the inverse of the pixel's gray level. Consequently, regarding perceptual weight, a black pixel would be heavier than a white one. They then examined whether the center was closer to an axis for art photographs than for control images, which was indeed the case.

Other tests, however, failed. For instance, in one experiment where McManus et al. (2011) presented simple pictures with only two discs but of a different gray level, the performance was incompatible with a physical interpretation of mechanical balance. Given these results, McManus et al. (2011) also concluded that the terms in Arnheim's theory could not be taken literally.

1.5.2 Objective measures of balance

The considered studies suggest that perceptual balance is a complex feature of pictures that depends on several factors, whose details remain mostly unknown. How-

ever, the studies also demonstrate that computing objective measures for predicting subjective balance and preference is a promising approach for investigating these matters. As we have seen, McManus et al.'s (2011) physical interpretation of perceptual mass was not successful in this respect. Nevertheless, this could be the starting point for developing more sophisticated measures that also apply to complex pictures. Consequently, one aim of this dissertation will be to optimize this approach of balance computation.

However, there are other measures. In recent years, even objective measures of perceptual balance have been developed based on the mechanical metaphor. One such measure is the *Assessment of Preference for Balance* (APB), developed by Wilson and Chatterjee (2005). These researchers assumed that the perceptual weight of each pixel in a picture is inversely related to its gray level, i.e., dark pixels are heavier than bright ones. Additionally, they divide a picture into four symmetrical areas around the horizontal, the vertical and the two diagonal axes, respectively. The differences between the summed weights in opposite areas are then computed, and the mean of the eight differences is taken as the picture's balance score.

Wilson and Chatterjee (2005) have shown that APB scores cannot only predict balance ratings but are also related to the aesthetic appreciation for simple pictures. However, in a study by Silvia and Barona (2009), who used a subset of Wilson and Chatterjee's (2005) stimuli, no substantial correlation between APB scores and liking was observed. That the applicability of the APB measure might indeed be restricted to simple pictures is also suggested by results of Gershoni and Hochstein (2011), who found that for Japanese calligraphies, the APB even completely failed to predict perceptual balance ratings. Consequently, another goal of this dissertation is to evaluate the predictive power of the APB score in further detail.

1.6 Alternative concepts of balance

In art theory, it is suggested that there are different types of balance. For instance, the perceived stability of a visual object is understood as an alternative concept of balance. In this context, stability is considered as a visual and aesthetic habit that plays an important role in composing pictures (R. Howard, 1914; Liu, Dong, Zhang, & Jiang, 2017; Ross, 1907). It is assumed that this habit is formed in individuals through their lifetime by living with gravitation, which forces them to arrange things in a stable way. In this respect, others also speak of *gravitational stability* (van der Helm, 2015).

Moreover, on a more abstract level, in artwork stability can also be used as a dominant law of structure that expresses visible meaning and, therefore, it also has symbolic character (Arnheim, 1954, p. 376). Stability reveals a meaningful skeleton of forces because it can offer either a stable framework that leads to order or instability that suggests chaos. From a Gestalt psychological perspective, the tendency toward stability is equal to the tendency toward the goodness of shape (Verstegen, 2005, p. 16). The more stable an object, the more simple, more *prägnant* (Wertheimer, 1923) and, consequently, easier to recognize the object is.

Interestingly, in his seminal experimental work on perceptual balance and its connection with aesthetic appreciation, Pierce (1894, 1896) rotated the display by 90° so that the elements appeared as stacked; his findings showed that balance adjustments chanced. For vertical layouts, stability was important. For instance, pictures were preferred when they had more weight in their lower half rather than in their upper one. Thus, it seems that persons prefer pictures whose elements are arranged in a gravitationally stable way. This finding is in line with more recent research by Friedenber (2012), who proposed a *perceptual instability hypothesis*, stating that objects perceived as more fragile are less attractive.

Consequently, this dissertation aims to investigate the concept of visual stability in further detail. For instance, it would be interesting to know to which picture types the concept of visual stability applies. Until now, it is unclear which kind of pictorial arrangement is necessary to evoke the perception of instability. Moreover, the question emerges whether stability perception comes into play when different figural elements in a picture are connected rather than disconnected. Additionally, another important concern that warrants further investigation is the relationship between stability, balance and aesthetic appreciation.

1.7 Relations and interactions with other visual features

An important aspect of the research field of psychological aesthetics is the question of how certain visual features are related and entangled with each other. Most empirical research assumes a reductive psychophysical approach where all stimulus dimensions (e.g., balance) are understood as orthogonal variables which affect aesthetic appreciation independently. However, in reality, this is not the case because these single

stimulus dimensions are always entangled, twisted and can compete and interact in a complex way. This phenomenon is also known as *Gestalt nightmare* (Makin, 2017).

Consequently, a pure bottom-up approach, like the psychophysical one, cannot be sufficient to investigate the relationship between visual features and aesthetic appreciation. An approach inspired by Gestalt psychology, however, which takes the whole stimulus into account, might help to overcome the problem of interacting and competing visual features. In case such competing dimensions are known, their effects can be discounted, which allows analyzing the relation between the unexplained variance and the variable of interest. Thus, following this approach, perceptual balance is not investigated in an isolated way. Instead, we obtain a more in-depth insight into a visual feature that only can be understood as a part of the whole.

As we will see, in this dissertation, perceptual balance is investigated in the context of several features. Beyond visual stability, visual complexity (Gartus & Leder, 2017; Jacobsen, 2004), and prototypicality (Hekkert, Snelders, & Van Wieringen, 2003; Whitfield & Slatter, 1979) are also taken into account. Moreover, it will be demonstrated that stability as an alternative concept of balance can imply movement (Osaka, Matsuyoshi, Ikeda, & Osaka, 2010). This perceived movement and dynamics can evoke emotions that ultimately positively influence aesthetic appreciation.

1.8 The aim of the present work

The preceding introduction has shown that perceptual balance is a complex feature that depends on several factors and mechanisms, whose details remain largely unknown. Despite the notion that in art theory perceptual balance is considered as an essential feature that affects aesthetic appreciation (e.g., Arnheim, 1954; Kandinsky, 1926/1979; Poore, 1903), there is no universal theory or agreement about a general concept of perceptual balance. Consequently, the aim of this dissertation is to investigate the fundamental factors and mechanisms that determine perceptual balance and its connection to aesthetic appreciation. Moreover, the interaction between these factors are examined and how they are related to other (visual) features.

First, the concept of perceptual balance should be investigated in further detail. For this purpose, in the field of psychological aesthetics, it has been proven to develop objective measures that are based on certain theoretical concepts. If these measures show reliable correlations with the subjective rating of interest, the question of what the

nature of the specific visual factor is can be better answered. This approach is used in Research Paper I where an already established objective measure of balance (the APB score) is compared with three other objective measures that have been proposed for measuring balance. One of these other measures is based on the notion of Arnheim's (1954) mechanical balance, which makes it possible to test his theoretical concept of perceptual balance again. Consequently, Research Paper I will shed light on the relationship between objective measures of perceptual balance and subjective balance, symmetry and aesthetic preference.

Next, the second research paper examines the relations between perceptual balance, prototypicality and aesthetic appreciation for Japanese calligraphy. Upon first glance, the results showed no significant relationship between perceptual balance and aesthetic appreciation. However, after discounting the effects of prototypicality on liking, there was a negative correlation between liking and an alternative concept of balance. Consequently, Research Paper II makes two important contributions to the research field of perceptual balance. First, by disentangling the visual features, we can point to the importance of considering the so-called Gestalt nightmare (Makin, 2017). Without taking the problem of interacting and competing visual features into account, a misinterpretation of the relationship between balance and aesthetic appreciation could be the result. Second, the second research paper discusses stability as an alternative concept of balance in the context of psychological aesthetic for the first time.

Research Paper III continues to pursue the idea that beyond the mechanical metaphor, perceptual balance can be interpreted alternatively. A study with single-element, multiple-element and dynamic-pattern pictures showed that the mechanistic interpretation of balance only held for single-element pictures whereas the balance of multiple-element and dynamic-pattern pictures were rated more in the sense of gravitational stability. Moreover, there was only a positive relation between stability and liking for the multiple-element stimuli. These findings show that there are different types of balance and that their relationship with liking depends on the picture type.

Finally, in Research Paper IV, we deeper investigated the relation between stability as an alternative concept of balance and its relation to aesthetic appreciation. Moreover, we addressed the conflicting findings of Research Paper II and III, which show that the relationship between (in)stability and liking can be both positive and negative. It is suggested that at least two types of instability can be differentiated. One type is associated with gravitation and is disliked; the other type is associated with move-

ment that evokes emotion and, finally, liking. By systematically investigating these two different types of stability, we created a more holistic understanding of perceptual balance, its relationship with aesthetic appreciation, and additional dimensions as pictorial dynamics and emotion.

In the next section, these four research papers are summarized and described in further detail.

2 Summary of Research Papers

2.1 Research Paper I

As outlined in the introduction, there is a wide consensus among aestheticians that balance has a major effect on the appreciation of a picture (Arnheim, 1954; Poore, 1903; Ross, 1907). Nevertheless, the mechanisms of balance perception remain largely unknown. Most researchers agree on what perceptual balance is in general. However, they disagree on the details of how balance is determined. In order to investigate the mechanisms of perceptual balance in further detail, in Research Paper I we considered objective measures of balance and examined how they are related to subjective balance, symmetry and aesthetic preference.

One of these measures is the so-called *APB*, which strongly correlated with balance and preference ratings (Wilson & Chatterjee, 2005). Consequently, one aim of the present study was to replicate Wilson and Chatterjee's (2005) results by applying complete sets of their original images. Furthermore, because the *APB* measure is the average of eight components, it was possible to use multiple regression analyses to examine the extent to which the components are related to perceptual balance and aesthetic preference. Such analyses have not been conducted before.

The second aim of our study was to compare the *APB* measure to three other objective measures that have also been proposed for measuring balance: a measure of balance that is based on the physical interpretation of perceptual mass, a measure of mirror symmetry, and a measure of heterogeneity. Finally, we wanted to examine the extent to which the results can be generalized. For this purpose, we applied the measures to new sets of stimuli.

To replicate Wilson and Chatterjee's (2005) results and compare the *APB* measure with alternative measures, we conducted three experiments. In the first one, we collected balance and symmetry ratings for the original pictures from the *APB* (see Figure 1) and examined how well the different measures can account for the judgments. In the second experiment, different participants rated the same pictures with respect to aesthetic preference (liking). Finally, in a third experiment, participants had to rate the balance as well as the liking of new stimuli (see Figure 1). Before the experiments and results are reported in further detail, the applied measures are briefly introduced.

APB: The test for the APB comprises images containing seven black elements of varying size that are scattered on a white background. There are 65 images with circles, hexagons and squares. All elements within an image have the same shape (see Figure 1 for example stimuli). The APB measures perceptual balance by testing how well the visual weight is distributed by taking the average of eight symmetry measures over the four axes of a picture (horizontal, vertical and the two diagonals) into account. These eight measures of symmetry derived by first dividing the picture into two equal parts along the principal axis, and second dividing the two equal parts in an inner (close to the center of the picture) and an outer area (close to the outside area of the picture). The different axes and areas are depicted in Figure 1. Every single symmetry measure reflects high symmetry when the number of pixels in the corresponding parts are equal (e.g., inner vs. outer area). The mean of the eight partial symmetry measures defines the APB score. Note that a low score means high balance, whereas a high score reflects poor balance (for a more detailed description of the calculation of the APB, see Research Paper I).

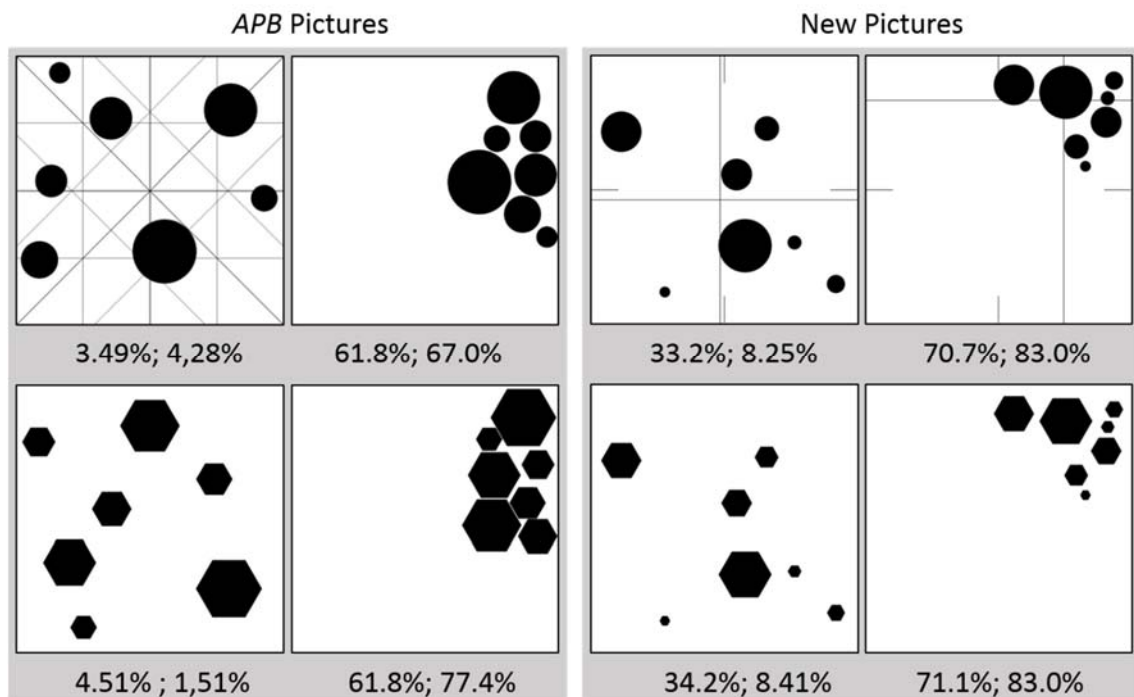


Figure 1. Example stimuli used in the experiments. Left panel: Pictures from the APB (Wilson and Chatterjee, 2005). The first and second number below each picture indicates the APB and DCM score, respectively. Note: the lower the value the higher the balance. In the top left figure additionally the different axes are shown for the demonstration of how the APB score is computed (see text for details). Right panel: Examples of the new stimuli. The intersections of the long lines in each picture indicate the respective center of mass (see text for details). The short lines imply the corresponding geometric center.

Deviation of the Center of Mass (DCM): As a more physical related measure of balance, the deviation of the center of mass from the picture's geometrical center is calculated. The DCM is defined by the Euclidean distance of the two-dimensional center of visual mass to the geometrical center of the image. For the black-and-white pictures used in this study, it was assumed that the mass of a black pixel is one, whereas that of a white pixel is zero. Consequently, in this context, the term mass is used synonymously for perceptual weight of the black pixels in an image.

Mirror Symmetry (MS): We also considered mirror symmetry as an objective measure. By definition, mirror symmetry is an exact reflection of pixels over a certain axis. We computed the mirror symmetry for the vertical, horizontal and two diagonal axes. The resulting MS score is the mean of mirror-symmetry measures around these four axes.

Homogeneity (HG): Because it holds for many pictures that the less scattered the elements are, the less balanced the image is, homogeneity was considered as an alternative measure for balance. For this purpose, the image was divided into ten bins along the horizontal and vertical axis. Maximum homogeneity was reached when all bins contained the same number of black pixels.

In Experiment 1, balance (from “not balanced” to “balanced”) and symmetry ratings (from “not symmetrical” to “symmetrical”) of two sets of pictures, which showed circles and hexagons, taken from the APB (Wilson & Chatterjee, 2005), were gathered in a lab experiment. The stimuli were presented on a monitor and participants rated each picture on a continuous scale (0 to 100 slider bar). Balance and symmetry were assessed in counterbalanced blocks of 130 trials (65 with circles and 65 with hexagons). The 130 pictures were randomized within each block and the experiment lasted approximately 50 minutes.

The results of Experiment 1 showed a strong correlation between balance and symmetry ratings ($r = .929, p < 0.001$), which indicates a very high overlap of both ratings. Moreover, the APB, DCM and HG scores correlated higher with the symmetry ratings than with the balance ratings. Concerning the APB scores, we replicated the results of Wilson and Chatterjee (2005). Although the score strongly correlated with balance ($r = -.784, p > 0.001$), regression analysis with the eight components of the APB (symmetry measures over the four axes of a picture) revealed that the horizontal and vertical ones had the largest effect. In other words, a differential weighting of the individual components increased the explained variance compared with the original score.

A closer look at the pictures revealed inconsistencies for the APB score. Pictures whose elements are only located in the central area received a relative poor APB score, although they were rated as highly balanced. Responsible for this inconsistency are the inner-outer components of the APB measure, which largely reflect homogeneity.

Moreover, the DCM score (balance: $r = -.822, p < 0.001$; symmetry: $r = -.926, p < 0.001$) correlated numerically even higher with the ratings than the APB (balance: $r = -.784, p < 0.001$; symmetry: $r = -.909, p < 0.001$). Hence, it was actually not necessary to invent a new score as the APB for measuring balance. Furthermore, the HG score also correlated substantially with the balance ($r = .707, p < 0.001$) and symmetry ratings ($r = .833, p < 0.001$) followed by the MS score, which had the weakest relations to the ratings (balance: $r = .418, p < 0.001$; symmetry: $r = .314, p > 0.001$). This finding suggests that perception does not take mirror symmetry into account, at least not for the current image type.

Taken together, the results of Experiment 1 show that objective measures can be constructed that reflect balance and symmetry ratings. However, due to the numerically lower correlations of the APB score and some inconsistencies (the APB did not reflect the balance of some images), the DCM score is strongly favored as a measure for perceptual balance.

The second experiment of this study pursued the goal to replicate the results of Wilson and Chatterjee (2005), who found a reliable correlation between their APB score and liking ratings. This was an important concern given that Silvia and Barona (2009) only found a rather low correlation between the APB score and preference ratings. However, in their study, they only used a selection of nine pictures with circles and nine images with hexagons because their main goal was to test the hypothesis that pictures with curved elements are preferred to those with angular elements (Bar & Neta, 2006). In Experiment 2, we used the entire picture set from the APB as in the previous experiment to ask participants for their liking ratings (from “I do not like it” to “I like it”). The apparatus and the procedure was similar to that in the first experiment.

Analyzing data from Experiment 2 showed that pictures with circles were liked more than those with hexagons. This finding supports the hypothesis from Silvia and Barona (2009) that curved objects are preferred to angular ones. Using the ratings from Experiment 1, we showed that the liking rating correlated higher with the pictures' rated symmetry ($r = .900, p < 0.001$) than with their rated balance ($r = .816, p < 0.001$). These results suggest that aesthetic preference is more affected by symmetry than the balance

perception. However, it remains unclear how participants operationalized these two concepts. In addition, because the liking rating correlated substantially with the APB scores ($r = -.867, p < 0.001$), we replicated Wilson and Chatterjee's (2005) results (see Figure 2). This seems to contradict Silvia and Barona's non-significant results. A reason for this discrepancy could be their small selection of eighteen pictures of the APB and their data analyses on an individual rather than a picture level.

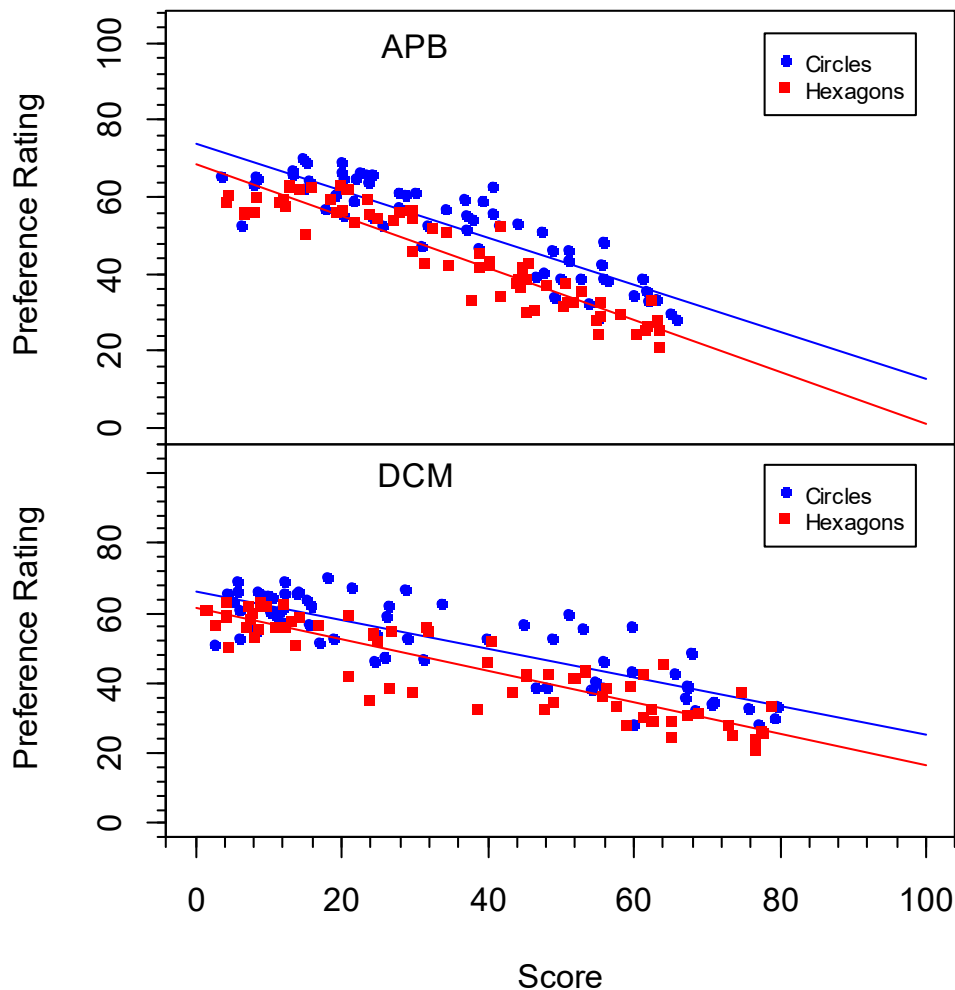


Figure 2. Relation between preference ratings in Experiment 2 for the two picture types (circles, and hexagons) and the APB scores and DCM scores.

Furthermore, the DCM score also correlated high with the preference ratings ($r = -.844, p < .001$), although significantly less than the APB scores (see Figure 2). The smallest correlation occurred between the MS scores and liking ratings ($r = .199, p > 0.05$). Multiple linear regression of the preference ratings on the different components of the APB score revealed that again the notion of homogeneity plays an im-

portant role. This interpretation is in accordance with the correlation between HG scores and liking, which was significant.

Until now, we only used the original stimuli from the APB, for which reason we did not know the extent to which results depend on the specific stimulus type. Wilson and Chatterjee's (2005) stimuli were manually constructed with a drawing program in such a way that the set covered a large range of the APB score. However, this process might have produced systematic relations between stimulus features which are favorable for the correlations between the APB score and the ratings. In contrast, the DCM score was not as evenly distributed as the APB score and clustered somewhat at smaller values. Hence, in Experiment 3, we applied a new picture set that was randomly drawn and whose DCM scores were more evenly distributed. The pictures also contained seven circles or hexagons of different size. Different from the original stimulus set, we used the same positions for constructing pictures with circles and for those with hexagons, which led to a more reliable comparison between the ratings for these stimulus types. In our third experiment, we required preference as well as balance ratings from the same participants. The experimental procedure was similar to that in the previous experiments.

Our results of Experiment 3 showed that on average the new pictures were less balanced, mirror-symmetric and homogeneous compared with the APB stimuli. The reduced calculated balance was also reflected by the somewhat smaller balance ratings, compared with Experiment 1. Moreover, balance ratings correlated highest with DCM scores ($r = -.916$, $p < 0.001$), followed by the APB ($r = -.836$, $p < 0.001$) and HG scores ($r = .364$, $p < 0.001$). Moreover, these correlations were significantly higher for pictures with circles than for those with hexagons.

Furthermore, the range and mean of the preference ratings for the new pictures are similar to those in Experiment 2. Same as in Experiment 2, pictures with circles were again preferred to those with hexagons. With respect to the correlations of the preference ratings with the different measures, we found no reliable difference between the APB ($r = -.737$, $p < 0.001$), DCM ($r = -.742$, $p < 0.001$), and HG ($r = .697$, $p < 0.001$) scores. The relations between preference ratings and objective measures (APB and DCM) are shown graphically in Figure 3.

Consequently, the results of Experiment 3 were similar to those obtained with the APB stimuli in the previous two experiments. However, the advantage of the DCM scores as measure of balance, compared with the APB scores, came out more clearly.

Furthermore, the advantage of the APB scores over the DCM scores for predicting liking vanished. This demonstrates that the extent to which the different measures account for balance and preference ratings does depend on the specific selection of pictures, even if the pictures are rather similar.

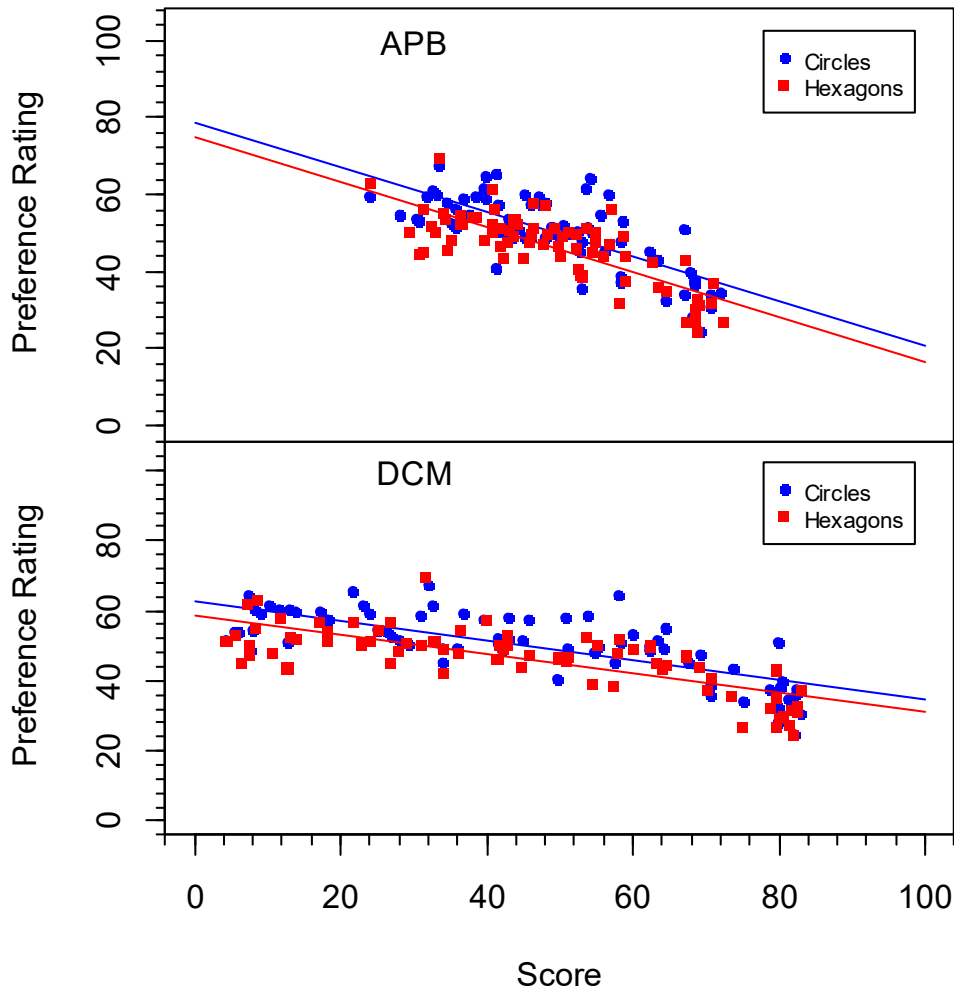


Figure 3. Relation between preference ratings in Experiment 3 for the two picture types (circles, and hexagons) and the APB scores and DCM scores.

Taken together, in Research Paper I, we aimed to investigate the mechanism of balance perception by comparing objective measures of balance and examining how well they account for balance, symmetry and liking judgments. For this purpose, we not only used the pictures from the APB but also a new stimulus set to examine the extent to which the results can be generalized. Our experiments and analyses demonstrate that the APB score is not a pure measure of balance. Therefore, if one is interested in predicting perceptual balance, the DCM measure is the better choice, mainly because it is less affected by homogeneity. If the goal is to predict preference ratings for pictures,

then the APB score is appropriate. Our results indicate that preference not only depends on balance but also on homogeneity, which is taken into account by the APB measure. However, APB scores can be substituted by HG scores, which produced comparable results. Moreover, the measure for mirror symmetry (MS) showed the weakest correlations with balance and liking ratings. This finding suggests that the visual system is not very sensitive to mirror symmetry, at least if the visual elements are scattered as in the present study.

2.2 Research Paper II

As Research Paper I showed, computational measures have been proposed to represent the balance and aesthetic appreciation of a picture (Hübner & Fillinger, 2016). However, there is also negative evidence. Gershoni and Hochstein (2011), for instance, applied the APB to Japanese calligraphies and found no substantial correlation with balance ratings. These results suggest that the proposed objective balance measures are not valid for all picture types. Instead, it seems that there are concepts of balance not reflected by the objective measures. If this were indeed the case, then it would be important to know the different concepts of balance and the respective relevant stimulus features.

The aim of Research Paper II was to investigate these issues. In a first step, we wanted to replicate and extend the results of Gershoni and Hochstein (2011). These researchers presented their stimuli very briefly (200 ms), and limited processing duration by masking. Moreover, the stimuli occurred at a random position (spatial uncertainty). Although it is known that balance can be perceived rapidly (Locher et al., 2007; Locher & Nagy, 1996; Locher & Stappers, 2002; McManus et al., 1985), it cannot be excluded that Gershoni and Hochstein's procedure was suboptimal for obtaining balance ratings related to the measures. Therefore, we used the same Japanese calligraphies as in the original study but presented the stimuli for a longer time and without spatial uncertainty.

Moreover, in addition to the APB, we also computed the DCM and homogeneity (HG) scores as measures of balance (Hübner & Fillinger, 2016). Because visual complexity might be related to balance as well (Gartus & Leder, 2017; Jacobsen, 2004), we additionally used the file size (FS) of the jpeg-compressed image (Donderi, 2006; Forsythe, Nadal, Sheehy, Cela-Conde, & Sawey, 2011; Machado et al., 2015) and the

degree of mirror symmetry (MS) in a picture (see Research Paper I) as corresponding measures.

We wanted to examine the relationship between balance and liking for the Japanese calligraphies. As mentioned, in art theory (Arnheim, 1982; Bouleau, 1980; Kandinsky, 1926/1979) pictorial balance is considered as a crucial variable for aesthetic appreciation. This relation has been confirmed by Wilson and Chatterjee (2005) and Hübner and Fillinger (2016). Surprisingly, Gershoni and Hochstein merely asked participants to assess the balance of the calligraphies, but not how much they liked them. Perhaps they supposed that perceptual balance and liking are generally closely related. To test whether this is indeed the case, we additionally asked our participants to indicate the extent to which they liked the Japanese calligraphies.

The first experiment was conducted to investigate whether the absent relation between the APB scores and balance ratings in the study of Gershoni and Hochstein (2011) was due to the specific picture type or the relatively short and spatially uncertain presentation of stimuli. Consequently, in Experiment 1, stimuli were presented for a longer duration and without spatial uncertainty. In contrast to the original study, we also collected liking ratings to examine whether the positive relationship between perceptual balance and aesthetic appreciation also holds for Japanese calligraphy.

















In the online experiment, the sixteen stimuli (see Table 1) were randomly presented in two counterbalanced blocks. In one block participants had to rate how much they liked the stimuli (from “I do not like it” to “I like it”) and in the other how well the stimuli were balanced (from “not balanced” to “balanced”). They entered their ratings by clicking on one of six scale positions with the mouse.

The results of Experiment 1 revealed a strong correlation between our and Gershoni and Hochstein’s (2011) balance ratings ($r = .943$, $p < .001$), which shows that balance assessments are largely unaffected by stimulus duration and spatial uncertainty. Moreover, as in the original study, we also found no significant correlation between balance and the APB score ($r = .071$, $p > .793$), which also holds for the other balance measures (DCM and HG, Hübner & Fillinger, 2016). Thus, the theoretical concepts of balance, as reflected by the balance measures, do not correspond to the balance perceived for Japanese calligraphy. This suggests that balance is not a unique concept. Instead, it seems that for different picture types, observers apply different concepts of balance, which are based on different features of feature combinations. Only complexity

















was negatively correlated with the balance ratings ($r = -.573, p = .020$), which means that the less complex the picture, the more it was perceived as balanced.

Table 1. Mean ratings (balance and liking of Experiment 1 as well as prototypicality and stability of Experiment 2) for the Japanese calligraphies.

















Mean balance ratings in ascending order

															
8	16	13	14	5	6	4	15	1	7	11	12	9	10	3	2
42	47	47	48	48	48	51	59	59	61	63	65	71	78	79	80

















Mean liking ratings in ascending order

															
5	2	8	9	6	14	7	1	12	13	16	11	3	4	10	15
38	39	43	51	52	53	54	56	56	58	60	61	63	64	65	66

Mean prototypicality ratings in ascending order

															
5	13	7	14	8	6	2	12	1	9	16	3	11	4	15	10
27	27	37	40	46	53	57	59	62	66	67	82	87	90	91	91

Mean stability ratings in ascending order

															
13	6	1	12	16	14	7	4	11	8	5	9	15	3	10	2
33	39	45	51	51	54	54	57	61	63	63	65	68	74	82	85

Note. Bold numbers represent the picture IDs.

Interestingly, balance ratings were also not related to liking ratings ($r = .168, p = .533$). As a reason, we hypothesized that Japanese calligraphy could also vary in other visual features which affect their aesthetic appreciation. Consequently, even if balance affected liking, its effect might have remained undetected, because it was masked by other dimensions that affected liking to a much stronger extent.

Because in Experiment 1 the objective measures of perceptual balance did not correlate with balance ratings, we further investigated the concept of perceptual balance for Japanese calligraphy and its relation to aesthetic appreciation in a second experiment. To gain an idea about which feature could have determined balance perception, we inspected the calligraphies and their order with respect to the balance ratings (see Table 1). It seemed to us that less-balanced stimuli showed some instability and flexibility, whereas more balanced stimuli looked more stable and rigid. In art theory, stability

is concerned as a visual and aesthetic habit that plays an essential role in composing pictures (Liu et al., 2017). To examine whether the visual balance of Japanese calligraphy was understood as stability, in Experiment 2, we asked the participants to assess the stability of these stimuli.

Additionally, we aimed to find a variable that strongly determines the liking ratings of the Japanese calligraphy. Such a variable would allow us to control its effect and test whether the unexplained variance can be accounted for by balance. Inspecting the stimuli and their order with respect to liking (see Table 1), let us hypothesize that prototypicality (Rosch, 1975) could be related to liking. It has been shown before that prototypicality can have a positive effect on preference for design (Hekkert et al., 2003; Whitfield & Slatter, 1979). Consequently, in Experiment 2, we additionally asked participants for their prototypicality ratings.

For the second experiment, we used the same stimuli as in Experiment 1. Furthermore, the procedure of the online experiment was almost the same. Again, we used a counterbalanced block design, where in one block participants had to rate the prototypicality (from “completely untypical” to “very typical”) and in the other block the stability of the pictorial arrangement (from “unstable” to “stable”).

The results of the second experiment revealed a reliable correlation between stability ratings of Experiment 2 and balance ratings of Experiment 1 ($r = .699, p < .01$), which indicates that these two concepts are closely related. Moreover, stability did not correlate with liking ($r = -.088, p > .05$), similar to perceptual balance. However, this only held for their direct linear relation. Moreover, prototypicality ratings were strongly correlated with liking ratings from Experiment 1 ($r = .700, p < .01$), which supports the idea that people prefer typical members of a category to less typical ones (Hekkert et al., 2003; Whitfield & Slatter, 1979).

Interestingly, with hierarchical multiple regression analysis, we were able to discount the effect of prototypicality on the liking ratings and examine the extent to which other variables can explain the remaining variance. When we conducted this for stability, it emerged that the variable had a significant effect on liking. Moreover, there was also a two-way interaction between prototypicality and stability, indicating that stability had a negative effect on liking, albeit only for less typical calligraphies (see Figure 4). For more typical exemplars, there was a trend in the opposite direction. However, for the balance ratings from Experiment 1, an analogous hierarchical regression analysis revealed no significant effects on liking, even when the effects of prototypicality was

controlled. This suggests that the perception of stability and balance are similar but not identical, which indicates that they are based, at least partly, on different visual features.

For the different objective measures, hierarchical regression revealed no reliable relations with liking, except for the DCM, which in combination with its interaction with prototypicality significantly predicted liking. Further analyses revealed that for less typical calligraphies liking decreased with an increasing DCM, namely with an increasing distance of the center of perceptual mass from the geometrical center of the picture. This relation was much weaker and unreliable for prototypical calligraphies.

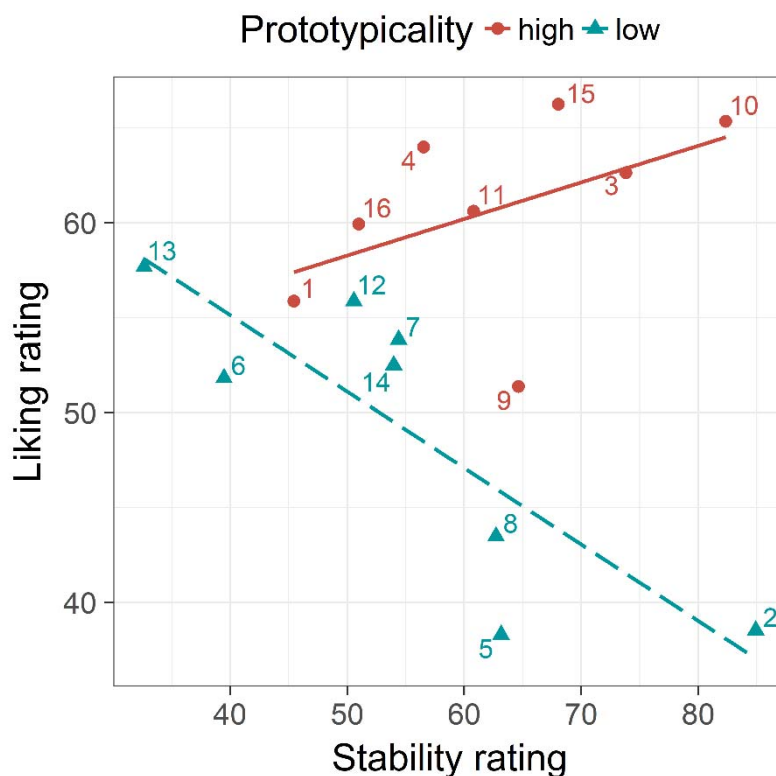


Figure 4. The interaction between prototypicality and stability for the prediction of liking.

Taken together, the results of Research Paper II show that the meaning of balance can differ between different stimulus types. For the Japanese calligraphies used here, it seems that balance had a different meaning than the usual interpretation in art theory (Arnheim, 1982; Bouleau, 1980; Kandinsky, 1926/1979). It was neither related to the objective measures nor to the liking ratings. However, the related concept of stability was predictive for liking, even if only for the more atypical members of the calligraphies. Interestingly, this relationship between stability and liking for atypical calligraphies was negative. This is an important concern because it is not in accordance with previous research, which showed that stability positively affects liking (Friedenberg,

2012; Pierce, 1896). However, for more prototypical members, stability had no effect on aesthetic appreciation. The fact that for more atypical calligraphies the DCM was related to liking suggests that this measure reflects a relatively general formal concept of balance, although it is obviously not always taken into account for the assessment of balance.

Research Paper II did not answer the question of how perceptual balance and stability are related. Moreover, due to the negative relation between stability and liking for atypical calligraphies, it prompts a closer investigation of the concept of stability and its relation to aesthetic appreciation. For instance, it would be interesting to know to which picture types the concept of visual stability applies or which kind of pictorial arrangement is necessary to evoke the perception of instability. Therefore, the further investigation of perceptual balance and perceived stability will be in the focus of Research Paper III. The issue of the relation between stability and aesthetic appreciation is addressed in Research Paper IV.

2.3 Research Paper III

In Research Paper III, the concept of perceptual balance and its generalizability for more complex images was investigated more closely. As already outlined in the introduction, it is commonly accepted that the aesthetic appreciation of an image depends on how well it is balanced (Arnheim, 1982; Bouleau, 1980; Kandinsky, 1926/1979; Ross, 1907). In this context, perceptual balance has often been defined analogously to mechanical balance (Arnheim, 1954; Pierce, 1894; Puffer, 1903). It is assumed that each element in a picture has a certain visual weight depending on its features like size, shape, and color (Arnheim, 1954).

Based on the mechanical metaphor, objective measures of perceptual balance have been developed; e.g., the DCM (Research Paper I; McManus et al., 2011) or the APB (Wilson & Chatterjee, 2005), which more or less rely on the mechanical balance metaphor. In Research Paper I, we have shown that for the pictures used by Wilson and Chatterjee (2005), the DCM predicts balance and liking ratings similarly well as the APB. However, it should be noted that these strong predictions are mainly due to the specific pictures applied in Research Paper I and in Wilson and Chatterjee (2005). They included only homogeneous elements with a simple shape and with an identical gray

level. Such pictures have the advantage that balance can be varied strongly without affecting other characteristics.

The obtained results say little about how far they can be generalized for more complex images. Some results (McManus et al., 2011; Thömmes & Hübner, 2018) suggest that the correlation between perceptual balance and aesthetic appreciation is much less for photographs. Thömmes and Hübner (2018), for instance, analyzed about 700 architectural photographs posted on Instagram by different photographers. For photographs representing a three-dimensional scene, they found that the scores (DCM and APB) significantly correlated with the number of Instagram likes. However, the explained variance was only about 10%. Although this percentage is small, it must be taken into account that balance is only one of many factors usually determining aesthetic appreciation. Therefore, for more complex artwork, one cannot expect the same large correlations between balance and liking as for specifically constructed simple pictures.

There are also counterexamples. For instance, Gershoni and Hochstein (2011) found that the APB completely failed to predict perceptual balance ratings for Japanese calligraphies. In Research Paper II, we replicated this result and observed similar negative results also for the DCM. However, they further showed that perceptual balance ratings were completely unrelated with liking ratings for these pictures.











































Two conclusions can be drawn from these results. First, the absent relation between balance ratings and the balance measures indicates that persons sometimes apply concepts of perceptual balance that are not reflected by the APB and DCM measures. In Research Paper II, for instance, we provided some evidence that under certain conditions balance is interpreted more in the sense of stability. Second, because the effect of perceptual balance can be relatively small or even absent. In case these factors are known, it can be helpful to discount their effects. Research Paper II, for instance, observed that prototypicality strongly determined the aesthetic appreciation of Japanese calligraphies. After taking this factor into account, the DCM showed again a significant relation with liking, but only for less prototypical calligraphies.

The aim of the present study was to further investigate the relations between objective measures, perceptual balance and aesthetic appreciation. For this objective, we used pictures from the Visual Aesthetic Sensitivity Test (VAST; Götz, 1985), developed by Götz, Borisy, Lynn, and Eysenck (1979). They comprise various configurations of different element types and, therefore, are more complex than those applied in Research Paper I and Wilson and Chatterjee (2005), and are more heterogeneous than the Japa-

nese calligraphies used in Research Paper II. Thumbnails of the VAST pictures can be seen in Table 2.

Because stimulus complexity is an important factor for aesthetic appreciation (Leder et al., 2004; Palmer et al., 2013; Tinio & Leder, 2009), we also applied a corresponding objective measure. Therefore, we used the ratio between the file size of the jpeg-compressed image and that of the uncompressed image (Palumbo, Ogden, Makin, & Bertamini, 2014). Additionally, we computed the APB and DCM scores (see Research Paper I) for each stimulus as objective measures of balance.

Table 2. Results of Experiment 1.

														
#	33	24	16	4	39	26	11	30	42	5	20	36	41	12
C	0.54	0.47	0.46	0.46	0.46	0.42	0.41	0.40	0.40	0.37	0.35	0.35	0.35	0.33
L	<u>63</u>	<u>57</u>	<u>58</u>	<u>63</u>	<u>61</u>	<u>60</u>	<u>62</u>	<u>54</u>	<u>59</u>	<u>55</u>	<u>54</u>	<u>60</u>	<u>54</u>	<u>54</u>
B	58	52	56	46	73	46	68	44	55	37	37	48	43	34
														
#	10	19	7	22	8	37	15	3	23	29	18	9	25	34
C	0.31	0.30	0.25	0.24	0.22	0.22	0.20	0.19	0.19	0.19	0.18	0.17	0.17	0.17
L	<u>59</u>	<u>57</u>	<u>31</u>	<u>59</u>	<u>62</u>	<u>38</u>	<u>40</u>	<u>57</u>	<u>32</u>	<u>40</u>	<u>32</u>	<u>50</u>	<u>42</u>	<u>46</u>
B	38	57	39	75	68	60	40	72	37	56	33	48	56	58
														
#	38	14	2	1	35	40	17	28	27	31	6	21	32	13
C	0.16	0.16	0.16	0.15	0.14	0.13	0.12	0.11	0.11	0.11	0.10	0.10	0.10	0.08
L	<u>45</u>	<u>36</u>	<u>52</u>	<u>48</u>	<u>38</u>	<u>37</u>	<u>29</u>	<u>27</u>	<u>30</u>	<u>27</u>	<u>22</u>	<u>27</u>	<u>25</u>	<u>27</u>
B	53	45	61	54	46	36	53	33	64	45	45	59	71	76

Note. The thumbnails of the VAST pictures, shown in the continuing first row, are ordered by complexity. The corresponding scores are given in row 3 (C). The bold numbers in row 2 (#) represent the corresponding picture numbers. Mean ratings for liking (underlined) and balance are shown in the fourth (L) and fifth (B) row, respectively (for the works of Karl Otto Götz: © VG Bild-Kunst, Bonn 2017).

Experiment 1, which was an online experiment, comprised two blocks in each of which the 42 VAST pictures (see Table 2) were presented in random order. In the first block, participants had to rate how much they liked the stimuli (from “I do not like it” to “I like it”) on a visual analogue scale (Reips & Funke, 2008), and in the second block

how well the stimuli were balanced (from “not balanced” to “balanced”). Altogether, the experiment comprised 84 trials (2×42 trials) and lasted about 10 minutes.

Upon first glance, the results of Experiment 1 revealed that, overall, balance ratings did not correlate with liking ratings ($r = -.154, p > .05$). However, liking was strongly correlated with complexity ($r = .809, p < .001$); the more complex a picture was, the more it was liked, which is in accordance with former findings (e.g., Jacobsen, 2004; Tinio & Leder, 2009). Moreover, after we ordered the pictures by complexity, it became obvious that complexity increased with the number of elements in a picture (see Table 2). This visual inspection suggested that the ordered VAST pictures can easily be divided into three categories: single-element (pictures contain only one element), multiple-element (images composed of multiple elements), and dynamic-pattern pictures (comprising countless elements, and also representing some dynamics or implied motion). In a supplementary categorization study, where participants were asked to sort the 42 VAST pictures into three categories by the pictures’ features, we confirmed our preliminary categorization.

After dividing the images into categories, we found that perceptual balance and liking significantly correlated for multiple-element ($r = .844, p < .001$) and dynamic-pattern pictures ($r = .592, p < .05$), but not for single-element images (see Figure 5). This finding indicates that the effect of balance depends on the picture type, and more specifically, that balance only seems to be an essential factor for liking for pictures with a sufficient number of elements or complexity. However, across all picture types, the objective measures of balance (APB and DCM) were unrelated to perceptual balance and liking. Consequently, the concept of balance reflected by these measures is different from that applied by our participants. Concerning the relationship between complexity and liking within categories, it was reliable only for dynamic-pattern ($r = .486, p < .05$) and single-object pictures ($r = .830, p < .01$).

Our first experiment showed that the relationship between balance and liking depends on the picture type. However, because the objective measures of balance did not correlate with balance ratings, participants must have used another concept of balance. Inspecting the multiple-element pictures revealed that in some pictures rated as well balanced, the elements are piled up in a more stable way than in those rated as less balanced. Because this gradation of stability also was apparent for the dynamic-pattern pictures, we hypothesized that stability could have been applied by our participants as an alternative concept of balance.

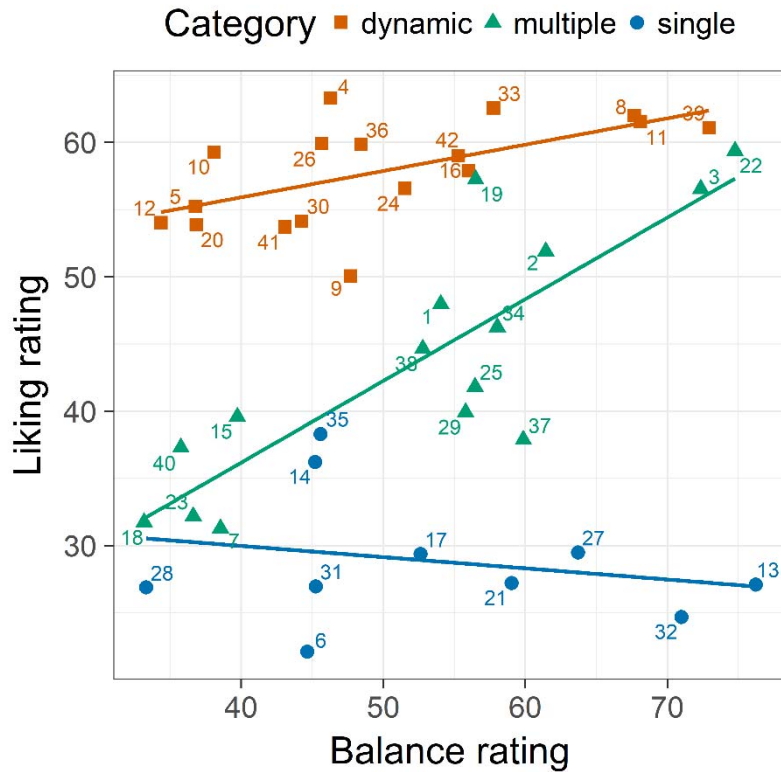


Figure 5. The relation between balance ratings and liking ratings in Experiment 1 for the different picture categories obtained in the supplementary categorization study. Each data point represents the picture according to its number (see Table 2). The colors and shapes of the symbols indicate the corresponding category. The lines are the corresponding regression lines.

Based on the findings of the first experiment, in Experiment 2, we asked the participants to rate stability directly. Because also Research Paper II has shown that perceptual stability is a variant of balance and, moreover, that more fragile and unstable objects are perceived as less attractive (Friedenberg, 2012), we expected a strong correlation between stability, balance and liking ratings. Additionally, because we expected a carryover effect from liking to balance ratings in the not counterbalanced first experiment, in the second experiment, we again collected balance ratings for the VAST pictures, but this time from an independent sample of participants. For this purpose, we conducted an experiment which was similar to the first one, except that half of the participants rated the pictures with respect to balance (from “not balanced” to “balanced”), whereas the other half rated how stable the composition of the pictures was (from “unstable” to “stable”).

The results of the second experiment showed that balance and stability ratings significantly correlated ($r = .601$, $p < .001$); however, their relationship was far from perfect. Balance and stability ratings shared only about 36% of their variance, which indicates that these two concepts are related but not identical. This conclusion is also

confirmed by the negative overall correlation between stability and liking ($r = -.535$, $p < .01$). This was because, different from balance, the range of stability differed systematically between the different stimulus categories (see Figure 6). The most complex pictures, namely the dynamic patterns, which were liked most, were rated as least stable, whereas the least complex single-element pictures, which were liked least, were judged as the most stable. Only the multiple-element pictures cover the whole range of stability. For these stimuli, the correlation between stability and liking was positive ($r = .664$, $p < .01$), as expected, namely the more stable a configuration, the more it was liked. Because a similar relation held for the balance ratings, it can be assumed that for this picture type balance was interpreted in the sense of stability and vice versa. Moreover, for single-element pictures ($r = -.464$, $p > .05$) as well as for the dynamic-pattern pictures ($r = .406$, $p > .05$) liking was independent of stability.

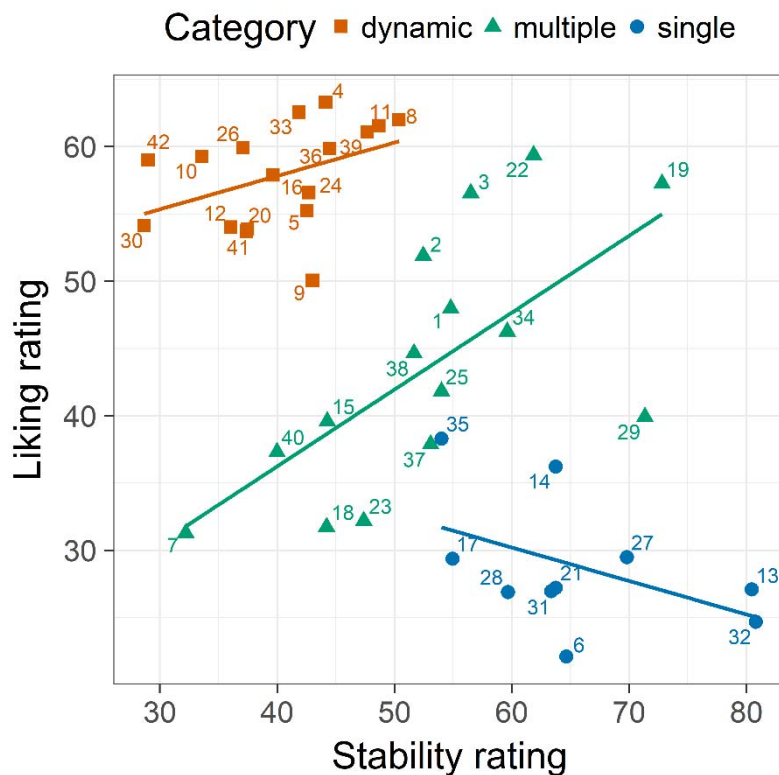


Figure 6. The relations between stability ratings and liking ratings from Experiment 1 for the categories obtained in the supplementary categorization study. Each number represents one of the 42 stimuli. The colors indicate the stimulus categories. The lines represent the regression lines.

However, given that across all stimuli, the stability ratings negatively correlated with liking, whereas the ratings positively correlated within the multiple-element pictures, it seems that the concept of stability was not used consistently. It was assumed, that the negative overall correlation is due to a confound with other variables. The result

that the dynamic patterns were liked most is probably not due to their low stability, but to the fact that in this specific case instability goes along with implied motion, which is usually liked. Interestingly, for Japanese calligraphies in Research Paper II, we observed a similar result. At least atypical calligraphies were liked the more, the less stable they were. As shown by Dubal et al. (2014), dynamic brushstrokes in calligraphies convey emotions. Thus, it seems that instability has no negative effects on liking, if it results from a dynamic that implies motion associated with positive emotion.

Moreover, the results of Experiment 2 revealed a strong correlation between the balance ratings of the first and second experiment ($r = -.928, p < .001$), which indicated that there was no carryover effect from liking to balance ratings in Experiment 1. Nevertheless, the independent balance ratings for the single-element pictures now negatively correlated with the DCM scores ($r = -.754, p < .05$), which means that the pictures were perceived as more balanced the less their center of perceptual mass deviated from their geometrical center.

Taken together, it seems that three types of perceptual balance/stability occurred in our study. For assessing the balance of the single-element pictures, the participants applied mechanical balance, namely the deviation of the center of mass from the geometrical center, as reflected by the DCM. However, this balance was unrelated to liking. For rating the balance of the multiple-element pictures, the gravitational stability of the configuration was assessed, which was positively associated with liking. Finally, the low stability ratings of the dynamic patterns were presumably not due to their perceived low gravitational stability, but to their dynamics and implied motion. We assumed, that the corresponding high emotional expressivity led these pictures to be liked most.

Consequently, Research Paper III shows that perceptual balance and aesthetic appreciation are related in a complex way. How balance is interpreted and assessed depends largely on the content of the picture. Moreover, other factors such as complexity can be dominant or at least modulate the relation between balance/stability and liking. The more factors are known and, therefore, can be discounted, the better the pure effect of balance can be isolated.

2.4 Research Paper IV

The previous research paper revealed that the concept of perceived stability, which is related to perceptual balance (Ross, 1907), can affect aesthetic appreciation

negatively (see untypical calligraphies in Research Paper II) as well as positively (see multiple-element pictures in Research Paper III). Consequently, in Research Paper IV we aimed to investigate the relation between perceived stability and aesthetic appreciation more closely.

As already shown in Research Paper I, perceptual balance is positively related to the aesthetic appreciation of an image. A widely used approach in this respect is to use mechanical balance as a metaphor. It is assumed that each element in a picture has a certain perceptual weight that depends on its low-level features such as color, size and form, as well as on its semantic content (e.g., Arnheim, 1982). This mechanical approach is also used to compute objective measures of balance (see Research Paper I for the APB and DCM score). These measures have successfully been applied to predict balance ratings and liking (e.g., Research Paper I, Thömmes & Hübner, 2018; Wilson & Chatterjee, 2005). However, there are also negative results. Gershoni and Hochstein (2011), for instance, used Japanese calligraphies as stimuli and found that the APB failed to predict balance ratings. In Research Paper II we replicated this result. Moreover, the results showed that balance ratings were also unrelated to liking ratings for these pictures. This suggests that there are different types of balance. Indeed, further data collection and analyses in Research Paper II revealed that the liking of calligraphies was affected by perceived stability, which is considered as different from but closely related to balance (Ross, 1907).

The difference between balance and stability has first been examined systematically by Pierce (1894, 1896). He observed that balance is mainly applied for the horizontal arrangement of elements, whereas for vertical arrangements stability plays a greater role. For instance, pictures were preferred when they had more weight in their lower part rather than in their upper half. This is in line with more recent research by Friedenbergl (2012), who observed that triangular shapes perceived as unstable (e.g., because they stood on one of the edges instead of one of the baselines), were rated as less attractive.

The research question remains why the positive association between stability and liking reverses for Japanese calligraphies. Or, more generally, why does instability increase liking in some pictures and decrease it in others? One possible explanation is that there are (at least) two different types of stability. One type is gravitational stability (van der Helm, 2015), which is preferred because it follows laws of gravitation thus

preventing damage and injuries. Due to corresponding associations, this preference is also generalized to the content of pictures.

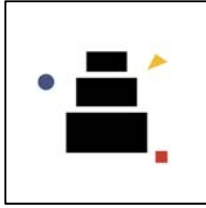

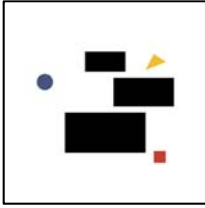

We hypothesized that the other type is somehow related to perceived movement and dynamics. In Chinese art theory, for instance, it is assumed that a brushstroke expresses the painter's emotion, which also holds for calligraphies (Dubal et al., 2014). Thus, expressive brushstrokes represent dynamics and imply movement, while at the same time they may appear unstable. Nonetheless, due to the implied movement (Osaka et al., 2010), which usually evokes positive emotions, the aesthetic appeal of the corresponding pictures is high.

This conjecture is also supported by Research Paper III, where we used pictures from the VAST as stimuli. The VAST pictures can be categorized into single-element, multiple-element and dynamic pictures. We found that overall, there was a negative correlation between stability and liking. In particular, the dynamic pictures were rated as highly unstable but were nevertheless liked most. The multiple-element pictures were liked less and were rated, on average, as more stable. Interestingly, within this category, there was a positive relation between stability and liking.

Taken together, the results presented thus far suggest that at least two types of instability can be differentiated: one type is associated with gravitation, whereas the other type is associated with movement. To systematically investigate these two types of instability, in our first experiment we show that gravitational instability in multiple-element pictures has a negative effect on aesthetic appreciation. In our second experiment, we demonstrate that instability is liked for dynamic pictures due to its implied movement.

In Experiment 1, we constructed a basic set of four pictures, each showing three rectangles and three colored decorative elements (see Table 3 for example stimuli). Although these basic stimuli already differed in stability, this property was further varied by rotating the stimuli. For the stimuli we also computed the objective balance measures (APB and DCM). In the online experiment, on a visual analogue scale (Reips & Funke, 2008) participants had to rate balance and stability for each of these stimuli, and they were also asked to rate how much they liked the pictures. We expected that balance and stability should be positively correlated.

Table 3. The thumbnails of the four basic stimuli (first row) and corresponding ratings.

				
#	1	9	17	25
S	93	55	57	50
L	55	44	40	46
B	83	43	43	58

Note. The full stimulus set contained 32 images because it additionally included the rotated versions (30, 90, 150, 180, 210, 270, and 330 degrees) of the basic stimuli. Row two (#) shows the identification number of the pictures. The corresponding mean ratings for stability (S), liking (L) and balance (B) are listed in the rows below.

The results of Experiment 1 revealed a positive correlation between stability and liking ratings ($r = .763$, $p < .001$). The same held for balance and liking ($r = .866$, $p < .001$). Thus, our results support the notion that pictures showing balanced and gravitationally stable compositions are preferred (Friedenberg, 2012; Pierce, 1896). Accordingly, there was also a strong relationship between balance and stability ($r = .912$, $p < .001$), which replicates the results from Research Paper II and III. The fact that this latter correlation was also quite strong indicates that the concepts of balance and stability are rather similar. However, it is also clear from our results that they nevertheless differ. Whereas the DCM scores significantly correlated with balance ($r = -.365$, $p < .05$), the correlation with stability was considerably lower and not significant ($r = -.145$, $p > .05$). Moreover, balance correlated much lower with the DCM scores than with stability. Therefore, we assumed that the balance ratings reflect both the usual mechanical balance as well as gravitational stability.



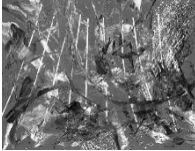
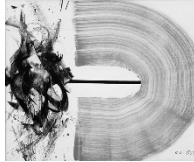
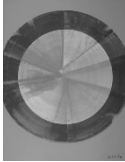
Interestingly, the DCM scores also significantly correlated with liking ($r = -.386$, $p < .05$). That means that a picture was liked more the closer the center of perceptual mass was located to the picture's geometric center. The APB scores did not correlate with any rating.

Consequently, the results of Experiment 1 support our hypothesis that, in multi-element pictures, gravitational stability is related to perceived balance, but nevertheless is also different from it. Moreover, it is positively related to liking.

In Experiment 1, instability was not liked. These results are in line with art theory (Liu et al., 2017), as well as early (Pierce, 1896) and more recent (Friedenberg, 2012, Research Paper III) experiments. However, the result is different from that observed for

Japanese calligraphies, for which instability was preferred (Research Paper II). Thus, in Experiment 2, we not only intended to replicate this latter result with a different type of stimuli but also to examine further the reasons for the diverging outcomes.

Table 4. Example stimuli used in Experiment 2 with diametrically opposite movement and stability ratings.

					
#	85	9	91	96	16
M	81	73	58	43	24
S	30	33	35	59	87
L	55	54	42	36	38
E	55	46	40	33	30

Note. The bold numbers in row two (#) represent the corresponding picture numbers. The following rows show movement (M), stability (S), liking (L), and emotion (E) ratings, respectively.

In order to achieve this aim, we used artwork by Karl Otto Götz (1914–2017) as stimulus material (see Table 4 for example paintings). In a preliminary study, we selected a representative set of 44 pictures. We expected that stability was negatively related to liking for these stimuli, and that this relation depended on movement and emotionality.

Moreover, in order to avoid pictures with a high movement rating being liked more simply because they are displaying rounder objects, in another preliminary study we asked participants for their roundness ratings of the selected pictures. The results of this preliminary study showed that roundness had no effect on liking ratings, thus excluding the possibility that the relation between movement and liking can be traced back to roundness. Finally, in the main study, the 44 paintings of Götz were rated for perceived stability and liking, as well as movement and emotionality.

As expected, the results of Experiment 2 showed that pictures that were rated as unstable were liked more than those rated as stable ($r = -.311, p < .05$). This is in line with observations in our previous studies (Research Paper II & III). Here, however, the relation was more direct and systematic. Given that this result is in contrast with the results of Experiment 1, we had the participants also rate emotionality and implied movement, since these variables might account for the reversal of the usual pattern. Consequently, as hypothesized, implied movement was responsible for the negative relation between stability and liking. If the variance accounted for by movement was

taken out from the liking ratings, then stability positively correlated again with liking (see Figure 7). Moreover, the positive correlation between movement and emotionality ($r = .826, p < .001$) further supports the notion that instability related to movement is liked because it induces emotions.

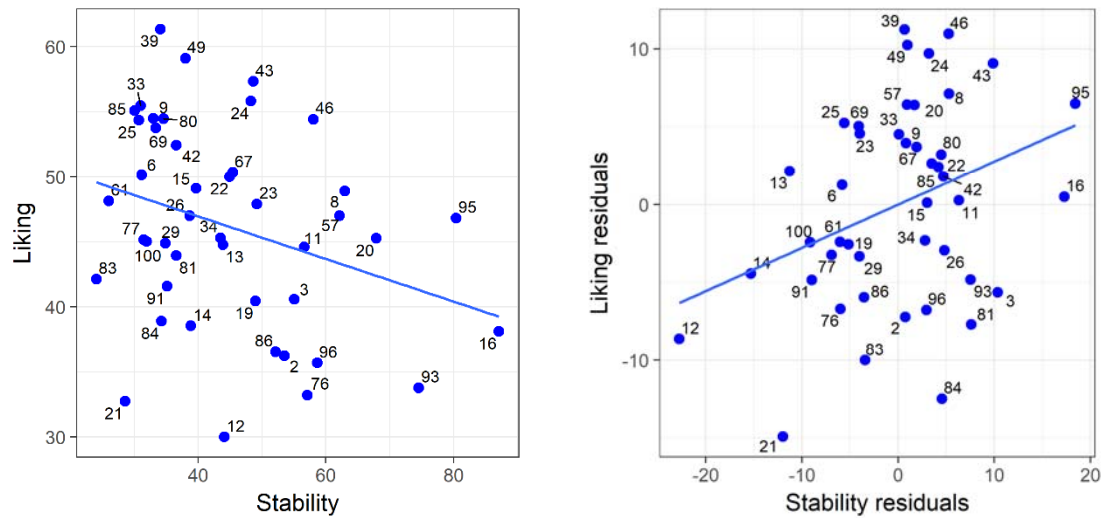


Figure 7. Left: Relation between stability and liking ratings in Experiment 2. Right: Relation between stability and liking after the effect of implied movement has been removed.

In sum, we found that perceived (in)stability as a concept closely related to perceptual balance can be attributed to gravitational instability as well as movement. Our findings suggest that in multi-element pictures, gravitational stability is liked because this preference has been formed through living with gravitation that forces individuals to arrange things in a stable way. In contrast, perceived instability of dynamic images is preferred if it is related to implied movement. It is assumed that instability related to movement is liked because it is correlated with emotionality. Because our study shed light on the relation between perceived stability, balance and liking, it has relevance for both empirical aesthetics as well as the arts.

3 General Discussion

In this dissertation, the effect of perceptual balance on the aesthetic appreciation was investigated. Perceptual balance is a complex visual feature that depends on several factors and mechanisms, whose details remain largely unknown. In four research papers, this visual feature and its relation to aesthetic appreciation was examined extensively. For this purpose, objective measures that are based on the notion of mechanical balance were used. Based on this mechanical metaphor it is assumed that each element in a picture has a certain perceptual weight that depends on its low-level features such as color, size and form. The objective measures were applied to verify their predictive power on subjective balance and liking ratings. At least for simple stimulus material, the concept of mechanical balance and its positive association with aesthetic appreciation could be confirmed. To extend the knowledge about the relation between perceptual balance and aesthetic appreciation, more complex pictures such as Japanese calligraphy or non-representational abstract paintings were used. The findings suggest that the meaning of balance could vary depending on the picture type. Beyond the mechanical balance, perceived stability has been revealed as an alternative concept of perceptual balance. Interestingly, the studies uncovered two types of instability. First, an instability that follows laws of gravitation and is disliked. Second, an instability that implies movement and is liked because this movement is correlated with emotionality.

The dissertation addressed the following main objectives:

- investigating the concept and psychological mechanisms of perceptual balance and its effect on the aesthetic appreciation;
- examining stability as an alternative concept of perceptual balance; and
- exploring the relations and interactions between balance and other visual features.

By addressing these three main objectives, this dissertation aims to contribute to a more universal theory and agreement about perceptual balance, as well as how this visual feature affects aesthetic appreciation.

3.1 Concepts of perceptual balance and its effect on the aesthetic appreciation

Despite wide agreement that pictorial balance is crucial for the aesthetic appreciation of pictures (e.g., Arnheim, 1954; Poore, 1903; Ross, 1907), the psychological mechanisms of balance perception and their relation to aesthetics remain not well understood. Consequently, the main objective of this dissertation was to investigate the concept of perceptual balance more closely.

Perceptual balance can be interpreted in a physical sense, which is reflected by Arnheim's theory of visual balance (Arnheim, 1954). By using mechanical balance as a metaphor, he assumed that each element in a picture has a certain perceptual weight that depends on its low-level features such as color, size and form. The idea of Arnheim was that each picture has a center of perceptual mass and that the more that this center deviates from the picture's geometric center, the less balanced the picture. In order to test this theoretical concept, we used objective measures based on the mechanical metaphor, and our research aimed to verify their predictability for subjective balance and aesthetic preference. One of the considered measures of balance, the DCM (McManus et al., 2011), is based on Arnheim's mechanical metaphor. Another measure that was less inspired by a physical metaphor is the APB score (Wilson & Chatterjee, 2005), which is the average of eight symmetry measures over the four axes of a picture.

For simple stimulus material, our research revealed strong correlations between the DCM and APB scores and the subjective balance as well as preference ratings. However, the APB score also reflects the concept of homogeneity. Consequently, the APB score is not a pure measure of balance. The DCM measure is recommended for predicting perceptual balance because it is less affected by homogeneity. If the goal is to predict preference ratings for pictures, then the APB score is appropriate because it includes homogeneity, which also strongly correlates with aesthetic appreciation.

Up to this point, the notion of mechanical balance as a metaphor of perceptual balance holds at least for simple pictures. Whereas in some studies, the measures could successfully predict the aesthetic appreciation of more complex photographs (McManus et al., 2011; Thömmes & Hübner, 2018), there are also negative results. Gershoni and Hochstein (2011), for instance, examined Japanese calligraphies and found no correlation between APB scores and balance ratings. This is an important result because it seems to demonstrate possible limits of the balance measures. We replicated this finding because our balance ratings of Japanese calligraphies did also not correlate with the

objective measures of balance (APB and DCM). Moreover, we showed that perceptual balance ratings were completely unrelated with liking ratings for these pictures. However, we could demonstrate that if the prototypicality of the calligraphies is taken into account, then at least the DCM is related to liking. This association could only be found for less prototypical calligraphies.

To further investigate the relations between perceptual balance and aesthetic appreciation of more complex images, we used pictures from the VAST (Götz, 1985). Our results suggest that at least two different types of perceptual balance were used by participants, depending on the stimulus material. First, for assessing the balance of pictures that show single elements, the participants applied the concept of mechanical balance, namely the deviation of the center of mass from the geometrical center, as reflected by the DCM. However, this balance was unrelated to liking. Second, for rating the balance of pictures that show multiple elements, the concept of gravitational stability, namely when its elements are ordered in a way that follows the laws of gravitation (van der Helm, 2015), was assessed, which was positively associated with liking.

Taken together, these findings show that perceptual balance and aesthetic appreciation are related in a complex way. How balance is interpreted and assessed depends largely on the content of the picture. Moreover, the meaning of balance can differ between different stimulus types. Because we used objective measures that were computed based on the theoretical assumption of Arnheim (1954), which includes the notion of perceptual weight, we can confirm the concept of mechanical balance at least for simple pictures. However, the studies with more complex stimulus material do not support the predictive power of the objective measures. This finding suggests that perceptual balance cannot be reduced to the notion of mechanical balance. Rather, this research is evidence of at least two different concepts of perceptual balance: First, a concept of balance that follows the notion of mechanics, which is also reflected by objective measures as the DCM and APB. Second, the concept of gravitational stability.

Consequently, this dissertation supports the broad consensus among aestheticians that balance has a significant effect on the appreciation of a picture (e.g., Arnheim, 1954; Poore, 1903; Ross, 1907). Moreover, by expanding the concept of perceptual balance by stability, the dissertation makes a relevant contribution to the mechanisms of balance perception.

3.2 Examination of stability as an alternative concept of perceptual balance

In a very early stage of investigating the concept of perceptual balance, it became clear that perceived stability is a closely related visual concept. For Japanese calligraphy, the balance ratings did not correlate with any of the balance measures. This indicates that the balance perceived when looking at calligraphies is unrelated to the concept of balance reflected by the applied objective measures of balance. Unexpectedly, the balance ratings, as well as the objective measures, did also not correlate with the liking ratings. Further investigations showed that perceived stability (Liu et al., 2017) strongly correlated with balance ratings. This suggested that instead of assessing how well the picture elements are balanced, the participants judged how stable or flexible the calligraphies look like, at least to a strong extent. Given this specific concept of balance, it is unsurprising that the objective measures did not correlate with the ratings.

Moreover, when non-representational abstract VAST pictures (Götz, 1985) were used as stimulus material, balance and stability ratings also significantly correlated, although their relation was far from perfect. Both variables share only about 36% of their variance, which indicates that the two concepts are related but not identical. A reason for that could be that the concept of stability was not used consistently, as suggested by different correlation between stability and liking in the picture categories. More evidence of the difference between these two concepts was also found in research with different stimulus material.

For Japanese calligraphies, discounting the effect of prototypicality revealed that stability was significantly related to liking. Interestingly, an analogous analysis of the balance ratings revealed no reliable relation with liking. This indicates that despite their substantial correlation, stability and balance ratings also differ in some relevant aspect.

For multiple-element pictures, a strong correlation between balance and stability was found. However, given that balance correlated with the DCM score that reflects the mechanical balance approach (e.g., Arnheim, 1982) but stability did not, we inferred that despite the substantial correlation between stability and balance these two concepts differ in some relevant aspects. Moreover, balance correlated much lower with the DCM scores than with stability. Based on this finding we assume that the balance ratings include both the usual mechanical balance as well as gravitational stability.

Interestingly, regarding the relation between stability and aesthetic appreciation, our findings were not consistent. We did find a positive correlation between gravitation-

al stability and liking for multiple-element pictures, which is in accordance with previous research (e.g., Friedenberg, 2012; Pierce, 1896). However, we also found a negative correlation between stability and liking for untypical Japanese calligraphies and the VAST pictures. Further research showed that the negative correlation is due to a confound with other variables. A study with artwork by Karl Otto Götz as stimuli showed that instability goes along with implied motion, which is usually liked. Thus, this dissertation showed that instability has no negative effects on liking if it results from a pictorial dynamic that implies motion associated with emotionality.

Taken together, this dissertation revealed perceived stability as a visual concept that is closely related to perceptual balance. This is an important concern because until now the concept of perceived stability is not extensively investigated in the context of psychological aesthetic. However, the findings of this dissertation provide evidence for the theoretical assumptions that consider stability as an essential feature for composing pictures (e.g., Arnheim, 1954; R. Howard, 1914; Liu et al., 2017). Moreover, with respect to the relation between stability and aesthetic appreciation, the findings of this dissertation suggest a differentiation between two types of instability: first, an instability that is associated with gravitation and therefore is disliked, because this preference has been formed through living with gravitation that forces individuals to arrange things in a stable way; and second, an instability that is related to movement that is liked because it is correlated with emotionality. Consequently, the differentiation of two types of instability support art theory because it emphasizes the symbolic character (Arnheim, 1954, p. 376) that a visual feature can have in an artwork. The effect on aesthetic appreciation can also differ depending on the interpretation and the meaning of instability.

3.3 Relations between perceptual balance and other visual features

Finally, in this dissertation, the relation and interaction between perceptual balance, stability and other visual features as well as its effect on aesthetic appreciation were investigated. This is an essential concern given that without taking interacting features into account a misinterpretation of the relationship between balance and aesthetic appreciation could be the result. This issue was observed when Japanese calligraphies were used as stimuli. In this study, participants assessed stability as an alternative construct of perceptual balance. Upon first glance, we did not find a reliable correlation

between stability, balance and liking. However, these conclusions were premature. With respect to the absent correlation between balance and liking, we hypothesized that a variable other than balance might strongly determine the aesthetic appreciation of Japanese calligraphies, which makes it difficult to detect a presumably relatively small effect of balance. As a likely candidate, we considered prototypicality, which has shown to be related to aesthetic preference (Hekkert et al., 2003; Whitfield & Slatter, 1979). As expected, prototypicality was strongly correlated with liking. Hence, we controlled for the effect of prototypicality on liking. When we did this with respect to stability, we found a reliable relation between stability and liking. Moreover, there was also a significant interaction between stability and prototypicality. However, stability was related to liking only for stimuli rated as more atypical Japanese calligraphies but not for more typical ones.

Thus, the present approach demonstrates that it can be difficult to detect effects of a certain variable on liking ratings if the considered pictures not only vary on the corresponding dimension but also on other dimensions affecting aesthetic appreciation. If this is the case, the dimensions can compete and interact in a complex way, a phenomenon called *Gestalt nightmare* (Makin, 2017). However, when such competing dimensions are known, their effects can be discounted, which allows analyzing the relation between the unexplained variance and the variable of interest.

Consequently, the research with Japanese calligraphy provided a positive example where the Gestalt nightmare could be avoided. Moreover, it emphasizes the importance of taking an approach inspired by Gestalt psychology into account when investigating the effect of visual features on aesthetic appreciation. Because certain visual features are related and entangled in a certain way, a reductive psychophysical approach where all stimulus dimensions are understood as orthogonal variables which affect aesthetic appreciation independently cannot be sufficient. This is becoming more important for more complex pictures whose aesthetic appreciation is determined by several features. The simple stimulus material from our studies was controlled for balance. However, studies have shown that, for instance, in more complex photographs of real-world scenes balance only accounts for 8–10% of the variance in aesthetic appreciation (Thömmes & Hübner, 2018). Moreover, this dissertation showed that beyond prototypicality, visual complexity (Gartus & Leder, 2017; Jacobsen, 2004), imply movement (Osaka et al., 2010) and emotionality (Lang, Greenwald, Bradley, & Hamm, 1993; Russell, 1980; Watson & Tellegen, 1985) are correlated with balance and stability. Con-

sequently, understanding visual features in a not-isolated sense, embedded in the wholeness of perception should be considered to overcome the problem of Gestalt nightmare.

3.4 Limitations and recommendations for further research

This dissertation revealed two main concepts of perceptual balance, which are closely related; mechanical balance and perceived stability. As could be shown, the meaning of balance often depends on the picture type. The concept of mechanical balance was successfully predicted by the DCM and APB measures as long as simple stimuli were used. On the contrary, for perceived stability, which was established as an alternative concept of perceptual balance, no adequate objective measure was available. Consequently, the computation of further objective measures is necessary, e.g., for perceived stability, to investigate the similarity as well as the diversity of these balance concepts more closely. Another important concern, especially in the context of more complex pictures, is the fact that several image contents influence balance. On a pictorial level, visual weights are not only modified by grayscale but also by color, contrast or position. However, on a more psychological level, it can also be influenced by semantics, meaning or emotionality. Consequently, future research will have to develop objective measures that can handle the ambiguity of the concept of balance as well as embraces the different levels of image perception.

This dissertation addressed the importance of interaction and competition among visual features. From a Gestalt psychological perspective, the reductionistic attempt to extract certain visual features is not satisfying. As already highlighted by Brentano (1959), it does not contradict an empirical aesthetics “from below” (Fechner, 1876) when instead of a single visual feature a more complex stimulus is investigated. Interestingly, in a recent paper by Makin (2017), this issue is described as follows: “Aesthetic experience is fundamentally about hot emotional reactions to wholes, but empirical aesthetics is stuck measuring cold evaluation of parts” (p. 208). Hence, this argument also emphasizes the importance of wholeness when aesthetic experiences are examined. Moreover, it stresses the importance of emotional reactions, for example, intense fascination or wondering (Fingerhut & Prinz, 2018), that is part of aesthetic experience but cannot be evoked in the context of a reductionistic approach investigated in the lab. Consequently, for a scientifically satisfactory result, future research will have to respect

subjective aspects as emotionality, including in the context of the relation between perceptual balance and aesthetic appreciation. For instance, a mixed-methods approach that also includes qualitative methods (Wertz, 2018) could help to investigate concepts like perceptual balance in a more holistic way.

Visual concepts such as balance are always culturally influenced. Consequently, for an integral theory of perceptual balance, this visual concept also has to be investigated from a cultural psychological as well as from a cross-cultural perspective. Typical stimulus-response paradigms as such as Fechner's psychophysics or Daniel Berlyne's experimental aesthetics have cut off the cultural aspect of aesthetic experience almost entirely (Allesch, 2018a). Culture psychological approaches can help to enrich theoretical concepts such as perceptual balance by understanding them as psychological phenomena that evolved in meaningful cultural environments (Valsiner, 2014, p. 6). Especially phenomenological approaches are well suited to capture and analyze subjective experiences and cultural meanings (for a concise historical introduction to phenomenological approaches, see Benetka & Joerchel, 2016).

Cross-cultural comparison can also help to expand the theoretical understanding of visual features and its impact on aesthetic appreciation. The advantage of this approach is that similarities in the findings would support the universality of a concept. Bode, Helmy, and Bertamini (2017), for instance, compared the preference for simple abstract symmetry between British and Egyptian non-experts. Their findings suggested universality in the preference for reflectional and rotational symmetry and emphasized a greater preference for simplicity in Egyptian participants. Furthermore, balance is a widespread concept in human experience that has been influenced culturally as well as art-historically. Hence, a cross-cultural comparison would also be a promising approach to further investigate the concept of perceptual balance.

Finally, the question of *why* perceptual balance is preferred was not part of this dissertation. Although the effect of perceptual balance on aesthetic appreciation has been extensively investigated and discussed in this dissertation, no theory or adequate explanation for this phenomenon has been provided. One possible explanation for the preference of balance compositions could be the human ability to empathic processes (Singer & Lamm, 2009). Empathy entails that the affective experiences of others can be shared and, consequently, understood. However, this ability is not limited to social interactions. Similarly to picking up the affects of others, humans are also able to feel into visual forms and images (Freedberg & Gallese, 2007; Lanzoni, 2009). Consequently,

when a viewer perceives visual art that shows emotional components, e.g., a scene with a person, a sense of bodily resonance can arise (Gerger, Pelowski, & Leder, 2018). Moreover, embodied simulations can also be triggered by abstract art where it is implied by the visible traces of the artist's creative gesture (Freedberg & Gallese, 2007). In other words, the observer can feel the painter's empathetic engagement of non-figurative abstract art. Thus, the roots of aesthetical feelings can probably be found in empathic processes (Gallese, 2019). Interestingly, this notion is not new. In the 19th century, Robert Vischer and Theodor Lipps established the *concept of empathy* in the context of psychological aesthetics (for a historical introduction to the concept of empathy, see Allesch, 2017). Although the concepts of bodily and perceived balance in the introduction were explicitly distinguished, the approach of embodiment could be a promising theoretical framework for explaining the effect of perceptual balance on aesthetic appreciation. Consequently, it could be argued that perceptual balance is preferred because it leads to a bodily simulation of balance that is liked. Due to the complexity of this approach, it will be challenging for future research to verify this hypothesis.

3.5 Conclusion

The findings of this dissertation support the notion that perceptual balance is understood as mechanical balance. Objective measures that reflect this mechanical metaphor were significantly correlated with subjective balance ratings and aesthetic appreciation. Moreover, we introduced perceived stability as an alternative concept of perceptual balance. By extending the concept of perceptual balance with stability, this dissertation provides a deeper insight into the mechanisms of balance perception. Furthermore, it was demonstrated that instability can be associated with gravitation, which leads to lower liking. It can also be related to movement that leads to higher liking because it is associated with emotionality. Regarding the relations and interactions between visual features, the findings recommend investigating perceptual balance as a visual feature integrated into wholeness of perception. Moreover, the empirical findings of this dissertation are not only relevant in the context of psychological aesthetics but also for art theory.

However, the findings of this dissertation are also limited. Because objective measures of balance can only handle simple pictures additional measures for more complex images are necessary. Furthermore, understanding balance as a cultural concept

would fruitfully expand the understanding of this visual feature. And finally, the question why balance is preferred must be answered. However, this dissertation not only provided many results for a better understanding of perceptual balance, it additionally offers interesting and novel ideas, which can be used as a reference point for future research.

4 Research Paper I

Comparison of Objective Measures for Predicting Perceptual Balance and Visual Aesthetic Preference

Authors:

Ronald Hübner

Martin G. Fillinger

Authors' affiliation:

Department of Psychology, Universität Konstanz, Germany

Citation:

Hübner, R., & Fillinger, M. G. (2016). Comparison of objective measures for predicting perceptual balance and visual aesthetic preference. *Frontiers in Psychology*, 7. doi:10.3389/fpsyg.2016.00335



Comparison of Objective Measures for Predicting Perceptual Balance and Visual Aesthetic Preference

Ronald Hübner* and Martin G. Fillinger

Department of Psychology, Universität Konstanz, Konstanz, Germany

OPEN ACCESS

Edited by:

Daniel Casasanto,
The University of Chicago, USA

Reviewed by:

Frouke Hermens,
University of Lincoln, UK
Gregory West,
University of Montreal, Canada
Anjan Chatterjee,
University of Pennsylvania, USA

*Correspondence:

Ronald Hübner
ronald.huebner@uni-konstanz.de

Specialty section:

This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

Received: 09 September 2015

Accepted: 23 February 2016

Published: 11 March 2016

Citation:

Hübner R and Fillinger MG (2016)
Comparison of Objective Measures for
Predicting Perceptual Balance and
Visual Aesthetic Preference.
Front. Psychol. 7:335.
doi: 10.3389/fpsyg.2016.00335

The aesthetic appreciation of a picture largely depends on the perceptual balance of its elements. The underlying mental mechanisms of this relation, however, are still poorly understood. For investigating these mechanisms, objective measures of balance have been constructed, such as the Assessment of Preference for Balance (*APB*) score of Wilson and Chatterjee (2005). In the present study we examined the *APB* measure and compared it to an alternative measure (*DCM*; Deviation of the Center of “Mass”) that represents the center of perceptual “mass” in a picture and its deviation from the geometric center. Additionally, we applied measures of homogeneity and of mirror symmetry. In a first experiment participants had to rate the balance and symmetry of simple pictures, whereas in a second experiment different participants rated their preference (liking) for these pictures. In a third experiment participants rated the balance as well as the preference of new pictures. Altogether, the results show that *DCM* scores accounted better for balance ratings than *APB* scores, whereas the opposite held with respect to preference. Detailed analyses revealed that these results were due to the fact that aesthetic preference does not only depend on balance but also on homogeneity, and that the *APB* measure takes this feature into account.

Keywords: visual aesthetics, perceptual balance, measures of balance, symmetry, homogeneity

INTRODUCTION

The perceptual mechanisms involved in visual aesthetics and preference judgments have long been a matter of debate (for an overview see Palmer et al., 2013). Since the seminal work of Gustav Theodor Fechner (Fechner, 1871, 1876) one approach of corresponding experimental studies has been to find aesthetic primitives, i.e., relatively simple perceptual features that determine the attraction of a stimulus (Latto, 1995; Munar et al., 2014). A prominent candidate in this respect is *perceptual balance*, i.e., how well the elements in a picture are arranged. There is wide consensus among aestheticians that balance has a great effect on the appreciation of a picture (Poore, 1903; Arnheim, 1954). Nevertheless, the mechanisms of balance perception are still largely unknown. Whereas most researchers agree on a global level what perceptual balance is, they disagree on the details of how balance is determined. In the present article we consider currently applied measures and examine how they are related to subjective balance, symmetry, and aesthetic preference.

Similar to early ideas of the artist and writer Henry Poore (1859-1940, Poore, 1903), and, as revealed by McManus et al. (2011), based on the work of Denman Ross (1853-1935, Ross, 1907), the Gestalt psychologist Rudolf Arnheim (1904-2007) hypothesized in his book *Art and Visual Perception* (Arnheim, 1954) that each rectangular frame has a hidden structure or field of invisible

forces (analog to a magnetic field in physics). The center of the frame has the strongest attraction, followed by the corners, the two main axes, and the diagonals. If an element is placed in the frame, then it is pulled by all the forces of the hidden structure, which produces an inner tension or *psychological force* of that element in relation to the square. For instance, if a single element is placed at the center, then all forces compensate each other and the picture is perfectly balanced. In contrast, if the element is placed off-center, then there is a pull toward the center, which results in imbalance. The situation is obviously more complex if several elements are placed in a frame. In this case each element has a relative perceptual weight resulting not only from the hidden forces of the frame, but also from forces originating from the other elements. A picture is perceived as balanced if these weights compensate each other. Furthermore, as proponent of Gestalt psychology, Arnheim (1954) also assumed that perceptual grouping (by similarity of form, color, etc.) modulates the forces between the elements.

An alternative characterization of perceptual balance is to consider the subjective equilibrium of a picture. According to Arnheim (1954), every visual pattern has a center of perceptual “mass,” which depends on the perceptual weight of the elements. If this center coincides with the geometric center of the frame, then the picture is balanced. It is assumed that the perceptual weight of an element increases proportionally to the element’s distance from the center of “mass” (lever principle in physics). However, the weight also depends on factors such as element size (larger elements are perceptually heavier than smaller ones), color (e.g., red is perceptually heavier than blue), and regularity (regular shapes are perceptually heavier than irregular ones). Arnheim conceded that most of these factors have to be verified, which is still valid today.

Some of Arnheim’s (1954) main assumptions have already been tested. McManus et al. (1985), for instance, presented reproductions of art work as well as plain stimuli, and had their participants to place a fulcrum beneath each picture so that it looked balanced (horizontally). For the reproductions of art work they found that the adjusted position of the fulcrum varied considerably, suggesting that art work is not generally well balanced. Moreover, when participants had to locate the perceptual center for unchanged pictures and for pictures where a portion was removed, the locations were rather similar. From these results McManus et al. (1985) concluded that the balance of a picture depends more “... upon a global integration of the picture as a whole, than of any individual element of it” (p. 314f).

Even for their plain stimuli McManus et al. (1985) found no simple relation. Whereas element position was crucial for positioning the fulcrum, size and color of the elements were less important. Furthermore, although the distance of an element from the frame’s geometric center and its size led to a larger shift of the fulcrum, these two factors were not correctly integrated for the judgment of balance.

In a later study, Locher et al. (1996) used reproductions of twentieth-century art paintings and a manipulated less-balanced version of each. Art experts and non-experts had to rate the balance of each picture and to determine the (two-dimensional) center of perceptual “mass.” As a result, both groups moved the

center for the disrupted version, but only the experts judged this version as less balanced. Locher et al. (1996) concluded that the center of perceptual “mass” and the overall judgment of balance are not as close as thought.

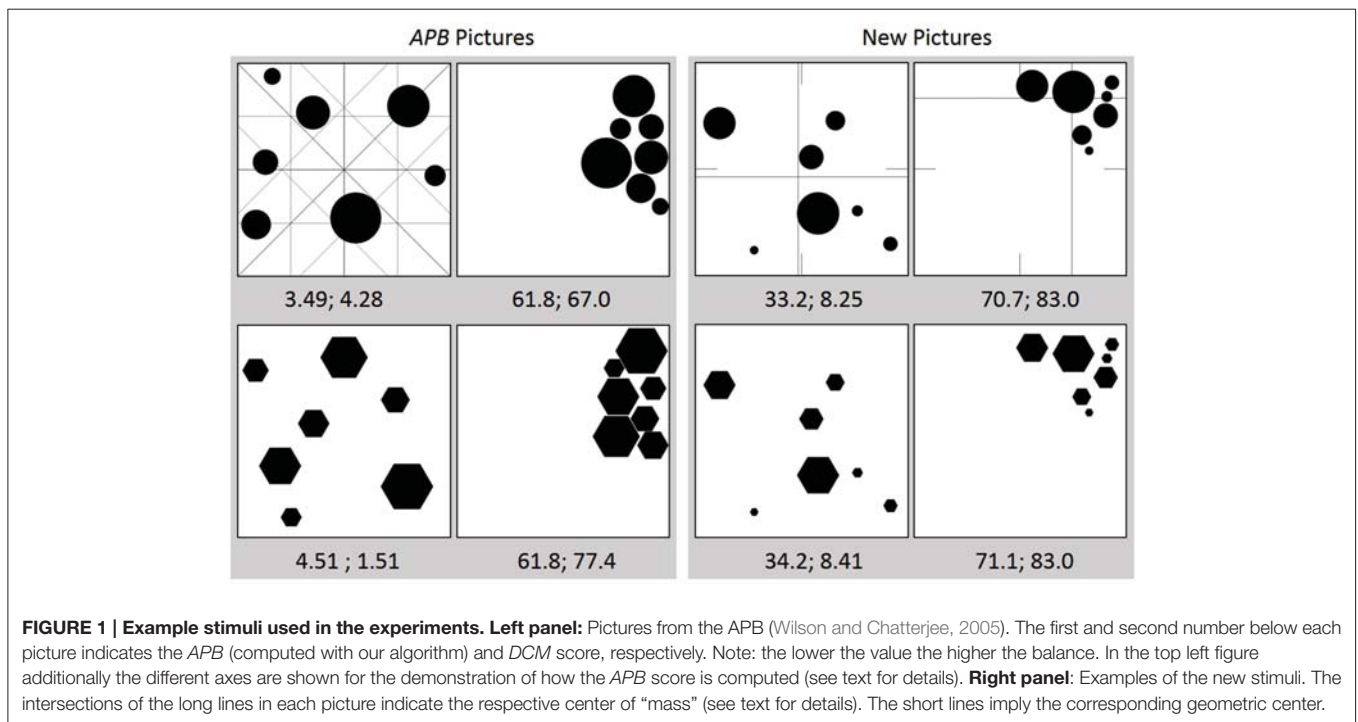
Because these results do hardly support Arnheim’s theory, Cupchik (2007) speculated that the terms of the theory were only meant metaphorically. McManus et al. (2011), however, believed that Arnheim wanted his theory to be taken literally, i.e., in a physical sense. To test their conjecture, they even went a step further and, instead of asking participants to indicate the fulcrum of pictures, calculated the center of “mass” by assuming that the “mass” of each pixel in a (gray-level) picture corresponds to the inverse of the pixel’s gray level. They then examined whether the center was closer to an axis for art photographs than for control images, which was indeed the case.

Other tests, however, failed. For instance, in one experiment where McManus et al. (2011) presented simple pictures with only two discs but of a different gray level, performance was incompatible with a physical interpretation of balance. In view of these results, McManus et al. (2011) also came to the conclusion that the terms in Arnheim’s theory cannot be taken literally.

The considered studies suggest that perceptual balance is a complex feature of pictures that depends on several factors, whose details are still largely unknown. However, the studies also demonstrate that computing objective measures for predicting subjective balance and preference is a promising approach for investigating these matters. As we have seen, McManus et al.’s (2011) physical interpretation of perceptual “mass” was not successful in this respect. However, there are other measures. Wilson and Chatterjee (2005), for instance, developed a test for the *Assessment of Preference for Balance* (APB). In connection with this test they introduced a measure, which we will call “APB” that highly correlated with perceptual balance and preference (liking), at least for simple pictures such as shown in **Figure 1**.

That the applicability of the APB measure might indeed be restricted to simple pictures is suggested by results of Gershoni and Hochstein (2011), who found only small correlations between this measure and ratings for Japanese calligraphy. Nevertheless, even if the measure predicts only preference between simple stimuli, it could be the starting point for the development of more sophisticated measures that also apply to complex pictures. Unfortunately, it is not even sure that the APB measure is valid for simple images. For instance, in a study by Silvia and Barona (2009), who used a subset of Wilson and Chatterjee’s (2005) stimuli, no substantial correlation between APB scores and liking was observed.

One aim of the present study was to replicate Wilson and Chatterjee’s (2005) results by applying complete sets of their original images. Furthermore, because the APB measure is the average of eight components, it was possible to use multiple regression analyses to examine the extent to which the components are related to perceptual balance and aesthetic preference. Such analyses have not been done before. A second aim of our study was to compare the APB measure to three other objective measures that have also been proposed for measuring balance: a measure of balance that is based on the physical interpretation of perceptual “mass,” a measure of mirror



symmetry, and a measure of heterogeneity. Finally, we wanted to examine to what extent the results can be generalized. Therefore, we also applied new sets of stimuli.

For replicating Wilson and Chatterjee’s (2005) results and for comparing the *APB* measure with alternative measures, we conducted two experiments. In the first one we collected balance and symmetry ratings for pictures from the APB and examined how well the different measures can account for the judgments. In the second experiment different participants rated the same pictures with respect to aesthetic preference (liking). The ratings were then correlated with the judgments from Experiment 1, and with the different measures. As we will show, our results were similar to those of Wilson and Chatterjee’s (2005). However, some of the alternative measures were also highly correlated with balance or preference ratings. A third experiment, where participants had to rate the balance as well as the liking of new stimuli, revealed that the specific selection of stimuli has some effects on the results. Before we report our results in detail, however, we introduce the applied measures.

Assessment of Preference for Balance (APB)

Wilson and Chatterjee’s (2005) test for the APB consists of images containing seven black elements of varying sizes that are scattered on a white quadratic background (750×750 pixels). There are 65 images with circles, hexagons, or squares, respectively. All elements within each image have the same shape (for examples see **Figure 1**). To also have an objective measure of balance for each picture, they created a specific score, defined by the mean of eight partial measures that are more or less related to symmetry. Relying on symmetry seems to be

reasonable, because this feature is strongly related to balance and preference. Mirror symmetry, for instance, is the simplest form of balance. Accordingly, symmetric pattern can not only be processed and remembered more easily than asymmetric ones (Garner and Clement, 1963), they are also judged as more “beautiful” (Jacobsen and Höfel, 2002). On the other hand, balance can be understood as a more complex form of symmetry (Locher and Nodine, 1989).

For obtaining the *APB* score, two symmetry measures are computed around the vertical and the horizontal axes, and around the two diagonal axes, respectively. Assume that a picture is divided along the horizontal dimension into four vertical, equally sized rectangles (see upper left picture in **Figure 1**), denoted by A_1 , A_2 , A_3 , and A_4 , from left to right, respectively. If f denotes a function that counts the number of black pixels in a given area, then the number N of all such pixels in a picture is $f(A_1) + f(A_2) + f(A_3) + f(A_4)$. The first partial symmetry measure for the horizontal dimension (around the vertical axis) is defined by $h = (|[f(A_1) + f(A_2)] - [f(A_3) + f(A_4)]|/N) \cdot 100$, i.e., the absolute difference between the number of black pixels in the left half and that in the right half of the picture in percent. The second measure for this dimension reflects the so-called horizontal *inner-outer relation* and is defined by $h_{io} = (|[f(A_1) + f(A_4)] - [f(A_2) + f(A_3)]|/N) \cdot 100$. Analogous partial measures are computed for each of the remaining three axes (the corresponding divisions of the picture area are shown in the upper left picture in **Figure 1**). The corresponding measures for the vertical dimension are denoted by v and v_{io} , those for the main diagonal (top left to bottom right) by md and md_{io} , and those for the anti-diagonal by ad and ad_{io} . Finally, the mean of the eight partial measures defines the *APB* score. Note that a

low score (percentage) means high balance, whereas a high score reflects poor balance.

Deviation of the Center of “Mass” (DCM)

Because the *APB* score is only loosely related to physics, we also applied a measure that is more strongly related to a physical interpretation of balance in the sense of Arnheim (1954). For this objective we computed a measure that represents the *deviation of the center of “mass” (DCM)* from the picture’s geometrical center. Assume two elements with visual “masses” m_1 and m_2 respectively, arranged on a beam. A point located between these objects at distance of d_1 and d_2 , respectively, is the center of “mass” (balance point, fulcrum) if $m_1d_1 = m_2d_2$. A practical way to calculate the center is to calculate the distances r_1 and r_2 of the “masses” from an arbitrary reference point (see McManus et al., 2011). The balance center is then located at distance $r = (m_1r_1 + m_2r_2)/(m_1 + m_2)$.

For the black-and-white pictures used in this study, we assumed that the “mass” of a black pixel is one, whereas that of a white pixel is zero. If we chose position $x = 0$ as reference point, then the center of “mass” b_x on the horizontal dimension is located at position:

$$b_x = \frac{\sum_{i=1}^w m_i r_i}{\sum_{i=1}^w m_i},$$

where w is the picture width, and m_i the number of black pixels in column i . The center for the vertical dimension is calculated analogously. In **Figure 1**, the line intersections in the two upper right pictures indicate the respective locations of the center of “mass.” The geometric centers are implied by the short lines.

In the present study we used the normalized location $b'_x = b_x/w$, which can vary from zero to one. For these coordinates the geometrical center is at 0.5, and the horizontal distance to the center of “mass” is $d_x = 0.5 - b'_x$. An analog distance d_y was calculated for the vertical dimension. The *DCM* measure of balance is then defined by the Euclidean distance of the two-dimensional center of visual “mass” to the geometrical center of the image. Specifically, we used the relative deviation in percent:

$$DCM = \left(\frac{\sqrt{d_x^2 + d_y^2}}{0.5} \right) 100.$$

Mirror Symmetry (MS)

As shown, the *APB* score is the mean of different measures most of which are based on the symmetry around some axis of the picture. Symmetry, however, is reflected only coarsely by these measures. Therefore, we also considered a measure of mirror symmetry (*MS*) that is defined by the mean of mirror-symmetry measures around different axes. The partial score for a given axis was computed by a formula suggested by Bauerly and Liu (2006). Assume that the vertical axis is the axis of reflection and that m and w denote the height and width of the image in pixels, respectively. The required number of comparisons n for each row is $w/2$, if w is even and $(w-1)/2$, if w odd. Assume further a binary variable X_{ij} that is 1 if there is a match between pixels and 0, otherwise. Finally, there is a factor

that reduces the weight of the match the farther away from the axis of reflection it is. The symmetry s for the vertical axis is then:

$$s = \frac{2}{3mn} \sum_{i=1}^m \sum_{j=1}^n X_{ij} \left(1 + \frac{j-1}{n-1} \right).$$

Analogous measures were computed for the horizontal axis and for each of the two diagonals. At the end, the four measures were multiplied by 100 and averaged. The resulting *MS* score is the mean symmetry in percent. The higher the value the more symmetric the picture.

Homogeneity (HG)

If we consider the pictures of the *APB* (for examples see the left panel in **Figure 1**), then it is obvious that balance is confounded to some extent with *homogeneity*. For many pictures it holds that, the less scattered the elements in the picture, the less balanced the picture. To investigate this relation in detail, we also wanted to include a measure of homogeneity. A measure that reflects this feature and that has widely been applied, among others for evaluating the design of user interfaces (e.g., Ngo et al., 2002), is information entropy (Shannon, 1948). Assume that we divide the picture area into M equally sized regions (bins). The entropy E is then defined by:

$$E = - \sum_{i=1}^M p_i \ln p_i,$$

Where p_i is the probability of black pixels in bin i , which is usually estimated by the corresponding relative frequency. For a given number of bins, the maximum entropy is reached if all bins contain the same number of black pixels. The value of this maximum is $\ln(M)$. Thus, a proper score of picture homogeneity can be obtained by calculating the relative entropy:

$$E_r = \frac{E}{\ln M}.$$

For the present study we computed separate values E_{rx} and E_{ry} for the horizontal and vertical dimension, respectively. For each dimension we divided the picture into 10 bins along the corresponding axis. The score *HG*, which reflects homogeneity in percentage, is then:

$$HG = \left(\frac{E_{rx} + E_{ry}}{2} \right) 100.$$

EXPERIMENT 1

In our first experiment we collected balance and symmetry ratings for two sets of pictures (constructed from circles or from hexagons) taken from the *APB* (Wilson and Chatterjee, 2005) and examined to what extent these ratings correlate with the objective measures of balance, symmetry, and homogeneity.

Method

Participants were 18 students from the University of Konstanz. They were recruited via an online system (ORSEE, Greiner, 2015) for participating in the experiment. The data of two participants were excluded from data analysis, because one of them produced many extreme values (0 and 100), and the other misunderstood the rating scales. The remaining 16 participants (3 males) had an average age of 23 years ($SD = 1.77$). All had normal or corrected-to-normal vision and were paid 8 € for their participation. The experiment was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. In agreement with the ethics and safety guidelines at the Universität Konstanz, we obtained a verbal informed consent statement from all individuals prior to their participation in the study. Potential participants were informed of their right to abstain from participation in the study or to withdraw consent to participate at any time without reprisal.

Apparatus and Stimuli

The stimuli were presented on a 19" LCD-monitor with a resolution of 1280×1024 pixels. A personal computer (PC) served for controlling stimulus presentation and response registration. As stimuli served all 65 pictures with circles and all 65 pictures with hexagons from Wilson and Chatterjee's (2005) APB. The pictures consisted of seven elements, which had the same shape, but varied in size. The APB score for each stimulus was calculated by our own algorithm, which produced values quite close ($r > 0.999$) to those provided by Wilson and Chatterjee (2005). Across pictures, the APB scores ranged from 3.49 to 65.9 ($M = 35.3$, $SD = 17.4$). DCM scores ranged from 1.51 to 79.5 ($M = 35.4$, $SD = 25.1$), homogeneity (entropy) from 64.1 to 94.8 ($M = 82.9$, $SD = 7.71$), and mirror symmetry from 1.13 to 10.9 ($M = 3.91$, $SD = 1.65$). The stimuli were presented at the center of the monitor on a black background. Each picture had an extension of 750×750 pixels, which approximately corresponded to a visual angle of 21° horizontally and vertically.

Procedure

After the participants had read the instruction and considered 6 example stimuli (3 with circles and 3 with hexagons), which were not used for the main task, they rated each picture with respect to balance and symmetry. Instead of a 1-to-5 rating scale, as in Wilson and Chatterjee (2005), we applied a continuous scale (1-to-100 slider bar) to reduce information loss (cf. Treiblmaier and Filzmoser, 2009). The scale went from "not balanced" to "balanced" for the balance rating, and from "not symmetrical" to "symmetrical" for the symmetry rating. The participants saw a horizontal slider located below the stimulus and had to move a computer mouse to adjust the position of the slider that corresponds to their subjective estimation of balance or of symmetry, respectively. The corresponding numeric value (not visible for the participants) of the chosen position was then entered by clicking the left mouse button. There was no time limit. Immediately after the value was entered, the next stimulus was displayed.

Balance and symmetry were assessed in alternating blocks of 130 trials. Half of the participants started with rating balance,

the other half with rating symmetry. There were two blocks for balance and symmetry rating, respectively. The 130 pictures (65 with circles and 65 with hexagons) were randomized within each block. The experiment lasted approximately 50 min.

Results

Balance Ratings

The mean balance ratings ranged from 16.2 to 79.3 ($M = 43.2$, $SD = 15.3$). They were subjected to a one-way within-participant ANOVA with factor *stimulus type* (circles, or hexagons). There was no significant difference (circles: 45.5, hexagons: 41.0) between the stimulus types, $F_{(1, 15)} = 2.88$, $p = 0.113$, $\eta_p^2 = 0.159$.

APB

The mean APB scores for pictures with circles and with hexagons were 34.9 and 35.7, respectively. In a first step we computed for each participant the correlation between the balance ratings and the scores across the 65 pictures with circles and across the 65 pictures with hexagons. For the pictures with circles the correlations ranged from -0.075 to -0.860 , and for those with hexagons from -0.132 to -0.855 . There was only one participant with non-significant ($p > 0.05$) correlations for both stimulus types. Three participants had a non-significant correlation for one of the stimulus types. The mean correlations are listed in **Table 1**.

Next, we computed the mean balance ratings across participants for each picture and correlated the obtained values across all 130 pictures with the different scores. The correlations and corresponding R^2 -values are shown in **Table 2**.

As mentioned, the APB scores are the mean of eight different measures. This implies that each component has the same weight. To examine whether this is appropriate, we also computed a multiple linear regression for each of the two stimulus types and for both types together. The results are shown in **Table 3**. If we consider the regression across both stimulus types, then we see that R^2 increased for the APB score (**Table 2**) from 0.615 to 0.751, which demonstrates that different weights for the components can improve the predictive power of the score. The obtained individual coefficients indicate that the horizontal component (symmetry over the vertical axis) has by far the largest weight. In contrast, the inner-outer components hardly explained variance.

DCM

The mean DCM scores for circles and hexagons were 33.3 and 37.3, respectively. Correlations of the DCM scores with the balance ratings for individual participants ranged from -0.008 to -0.786 for circles and from -0.153 to -0.861 for hexagons. The mean correlations are listed in **Table 1**. As can be seen, the correlations for the DCM scores were somewhat higher than those for the APB scores. However, a comparison across both stimulus types revealed no significant difference (-0.467 vs. -0.486), $F_{(1, 15)} = 2.11$, $p = 0.167$, $\eta_p^2 = 0.123$. If we consider the mean balance ratings (see **Table 2**), then their correlation with the DCM scores was also numerically larger than that with the APB scores.

TABLE 1 | Means (across participants) of the individual correlations (across pictures) between the balance and symmetry ratings and the different scores in Experiment 1.

	APB	DCM	MS	HG
Balance circles	-0.450 (0.222)	-0.479 (0.243)	0.269 (0.152)	0.378 (0.208)
Balance hexagons	-0.484 (0.207)	-0.493 (0.212)	0.252 (0.087)	0.444 (0.201)
Symmetry circles	-0.670 (0.231)	-0.677 (0.239)	0.234 (0.141)	0.592 (0.197)
Symmetry hexagons	-0.712 (0.200)	-0.723 (0.120)	0.247 (0.079)	0.661 (0.158)

The values in parenthesis are the standard deviations.

APB, Assessment of preference for balance; DCM, Deviation of the center of "mass"; MS, Mirror symmetry; HG, Homogeneity.

TABLE 2 | Correlations between the mean ratings and the different scores across both stimulus types in Experiment 1.

	Balance	Symmetry	APB	DCM	MS	HG
Balance	–	0.929***	-0.784***	-0.822***	0.418***	0.707***
Symmetry	0.864	–	-0.909***	-0.926***	0.314***	0.833***
APB	0.615	0.826***	–	0.866***	-0.177***	-0.852***
DCM	0.675	0.857***	0.751***	–	-0.389***	-0.798***
MS	0.175	0.099***	0.031***	0.151***	–	0.057***
HG	0.500	0.694***	0.725***	0.637***	0.003	–

The values below the diagonal are the corresponding R^2 -values.

*** $p < 0.001$. APB, Assessment of preference for balance; DCM, Deviation of the center of "mass"; MS, Mirror symmetry; HG, Homogeneity.

TABLE 3 | Regressions of balance ratings on the components of the APB scores in Experiment 1.

	Circles ($R^2 = 0.747$) $F_{(8, 56)} = 20.6$, $p < 0.001$ Intercept = 62.1		Hexagons ($R^2 = 0.784$) $F_{(8, 56)} = 25.3$, $p < 0.001$ Intercept = 63.5		Both ($R^2 = 0.751$) $F_{(8, 121)} = 45.5$, $p < 0.001$ Intercept = 63.1	
	β	$P(> t)$	β	$P(> t)$	β	$P(> t)$
h	-0.218	0.000***	-0.259	0.000***	-0.229	0.000***
v	-0.039	0.323	-0.1037	0.040*	-0.076	0.014*
ad	-0.066	0.195	-0.077	0.137	-0.082	0.019*
md	-0.096	0.026*	-0.064	0.204	-0.087	0.007**
h_{io}	0.015	0.769	0.039	0.443	0.032	0.348
v_{io}	-0.015	0.737	-0.098	0.051	-0.057	0.083
ad_{io}	0.041	0.384	-0.008	0.874	0.011	0.736
md_{io}	-0.070	0.165	0.003	0.949	-0.030	0.364

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Because the DCM score represents the Euclidian distance to the image center, it is interesting to examine whether a linear combination of the horizontal and the vertical distance would have been a better measure. Therefore, we computed a multiple linear regression with these two components. As a result, there was a strong contribution of the horizontal deviation (see Table 4). However, R^2 was smaller than the corresponding value for the DCM score (0.605 vs. 0.675). Thus, the Euclidian distance of the center of "mass" from the image center is a better measure than the linear combination of the horizontal and the vertical deviation.

TABLE 4 | Regressions of balance ratings on the components of the DCM score in Experiment 1.

	Circles ($R^2 = 0.629$) $F_{(2, 62)} = 55.0$, $p < 0.001$ Intercept = 63.9		Hexagons ($R^2 = 0.597$) $F_{(2, 62)} = 45.8$, $p < 0.001$ Intercept = 53.7		Both ($R^2 = 0.605$) $F_{(2, 127)} = 97.1$, $p < 0.001$ Intercept = 57.0	
	β	$P(> t)$	β	$P(> t)$	β	$P(> t)$
Horizontal	-52.1	0.000***	-60.3	0.000***	-56.8	0.000***
Vertical	-6.72	0.397	3.67	0.655	-0.969	0.867

*** $p < 0.001$.

MS

The mean scores of mirror symmetry for circles and hexagons were 3.21 and 3.36, respectively. Correlations between the symmetry ratings and the MS scores for the individual participants varied between 0.058 and 0.333 for circles and between -0.043 and 0.317 for hexagons. The mean values, which are shown in Table 1, were relatively small, as was the correlation between the mean balance ratings and the MS scores (see Table 2). A regression of the balance ratings on the four components of the MS score improved R^2 only slightly from 0.175 to 0.196 (see Table 5). Although the diagonals significantly accounted for the balance ratings, the horizontal dimension (vertical axis of reflection) had the strongest effect.

HG

The mean HG scores for circles and hexagons were 83.8 and 82.0, respectively. Correlations of the HG scores with the balance

TABLE 5 | Regressions of balance ratings on the components of the MS measure in Experiment 1.

Axis of reflection	Circles ($R^2 = 0.281$)		Hexagons ($R^2 = 0.189$)		Both ($R^2 = 0.196$)	
	$F(4, 60) = 5.86$, $p < 0.001$ Intercept = 32.6		$F(4, 60) = 3.49$, $p < 0.05$ Intercept = 31.8		$F(4, 125) = 7.61$, $p < 0.001$ Intercept = 29.0	
	β	$P(> t)$	β	$P(> t)$	β	$P(> t)$
Horizontal	0.078	0.882	1.017	0.134	0.570	0.179
Vertical	2.023	0.001*	1.595	0.007**	1.59	0.000***
Maj. diag.	0.472	0.322	1.122	0.073	0.851	0.028*
Min. diag.	0.972	0.049*	0.364	0.544	0.720	0.062

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

rating ranged for individual participants from 0.037 to 0.801 for circles and from -0.008 to 0.770 for hexagons. Mean correlations are shown in **Table 1**. The correlations for the *HG* measure are smaller than those for the *APB* and *DCM* scores. Such a pattern also occurred for the correlations with mean balance ratings (see **Table 2**). A statistical test revealed that the mean correlation for the *HG* scores was significantly smaller than that for the *APB* scores (-0.411 vs. -0.467), $F(1, 15) = 25.6$, $p > 0.001$, $\eta_p^2 = 0.631$.

To examine how the two dimensions of the *HG* measure are related to the balance ratings, we computed a multiple linear regression with the corresponding two components. It revealed that homogeneity along the vertical dimension was as important as that along the horizontal dimension. Accordingly, the regression did not increase R^2 .

Symmetry Ratings

The symmetry ratings differed significantly between the two stimulus types, $F(1, 15) = 14.0$, $p < 0.01$, $\eta_p^2 = 0.484$, indicating that the pictures with circles were perceived as more symmetric than those with hexagons (47.7 vs. 42.5). The mean correlations between the symmetry ratings and the different measures are shown in **Table 1**. Obviously, the symmetry ratings were rather weakly correlated with the *MS* scores. Six participants had at least one non-significant correlation ($p > 0.05$). Interestingly, the *APB* scores correlated higher with the symmetry ratings than with the balance ratings (-0.691 vs. -0.467), $F(1, 15) = 12.9$, $p < 0.01$, $\eta_p^2 = 0.462$, which was also the case for the *DCM* scores (-0.670 vs. -0.486), $F(1, 15) = 12.8$, $p < 0.01$, $\eta_p^2 = 0.461$, and for the *HG* scores (0.627 vs. 0.411), $F(1, 15) = 13.9$, $p < 0.01$, $\eta_p^2 = 0.481$.

The mean correlations were also somewhat larger for the *DCM* scores than for the *APB* scores (-0.691 vs. -0.700), which, however, was not significant, $F(1, 15) = 0.610$, $p = 0.447$, $\eta_p^2 = 0.039$. A similar pattern of correlations occurred for the mean scores. **Table 2** shows that the symmetry ratings correlated highly with the balance ratings (shared variance was 86%). In view of this correspondence we did not further analyze the symmetry ratings and their relation with the different measures.

Discussion

Our results show that the mean ratings for balance and symmetry were highly correlated, which suggests that it was difficult for the

participants to operationalize the two concepts differently. That the two ratings were nevertheless not identical is indicated by the fact that the symmetry ratings were significantly higher for pictures with circles than for those with hexagons, which was not the case for the balance ratings. Moreover, the *APB*, *DCM*, and *HG* scores correlated higher with the symmetry ratings than with the balance ratings.

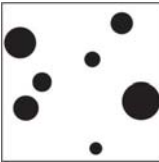


With respect to the *APB* scores, we replicated the result of Wilson and Chatterjee (2005). The mean scores correlated highly with the mean ratings of balance. However, it also became clear that the individual correlations were much smaller and varied considerably across participants. Furthermore, a regression analysis of the mean balance ratings on the components of the *APB* scores revealed that the components accounted differently for the balance ratings. The horizontal dimension had the largest effect, followed by the vertical one. Whereas the diagonal components also contributed to a small but significant extent, the inner-outer components had a negligible effect. In all, a differential weighting of the individual components increased the percentage of explained variance, compared to the original score with equal weights (averaging).

The *DCM* scores correlated surprisingly high with the ratings. The correlations were numerically even higher than those for the *APB* scores, which shows that it was actually not necessary to invent a new score for measuring balance. The *HG* scores also correlated substantially with subjective balance and symmetry, although not as high as the *APB* and the *DCM* scores. Interestingly, in contrast to the other measures, the vertical dimension was similarly importance for this correlation than the horizontal one. The *MS* scores had the weakest relation to the ratings, suggesting that perception does not take mirror symmetry into account, at least not for the current type of pictures.

Taken together, the results show that objective measures can be constructed that reflect perceptual balance (and symmetry), at least for the relatively simple pictures used here. A straightforward method is simply to compute how much the center of “mass” deviates from the geometric center of the picture. The larger the deviation the less balanced the picture. Another method would be to compute *APB* scores. However, although this measure also correlated highly with the ratings, a closer look at the pictures reveals an inconsistency. **Table 6** includes three pictures from the *APB* whose *APB* scores increase from left to right. Obviously, Pictures #45 is less balanced than Picture #27. However, it is hard to believe that picture #46 shall be less balanced than Picture #45. That this is inconsistent to one’s impression is also confirmed by our balance ratings. Picture #46 received a rating that was even higher than that of Picture #27. In contrast to the *APB* score, the *DCM* measure reflects this order. This strongly favors of the *DCM* score as measure for perceptual balance.

An analysis of the components of the *APB* measure revealed that the reason for this inconsistency are the inner-outer components, which represent the difference in black pixels between the inner and the outer areas. Consequently, if elements are present only in the center, as in Picture #46, then these components have a high value, indicating unbalance, which

TABLE 6 | Example pictures from the APB and corresponding ratings and objective scores.

Nr.			
	#27	#45	#46
Balance (Experiment 1)	54	36	61
Symmetry (Experiment 1)	76	32	52
Liking (Experiment 2)	66	39	51
<i>APB</i>	24	46	47
<i>DCM</i>	12	67	2.7
<i>MS</i>	2.2	3.7	11
<i>HG</i>	91	79	72

however, does not reflect subjective balance. If we consider the different measures in **Table 6**, then it is obvious that the inner-outer components correspond to homogeneity. Indeed, homogeneity is highest for Picture #27. Thus, the *APB* measure is not very well suited for representing balance.

EXPERIMENT 2

In our second experiment we wanted to collect preference ratings for the pictures shown in the first experiment. To avoid any influence from a second task, the participants had merely to indicate how much they liked each picture. The main goal was to examine to what extent the different ratings from Experiment 1 and the introduced measures can account for preference judgments.

In this context we also wanted to replicate the results of Wilson and Chatterjee (2005), who found a high correlation between their *APB* score and liking. In a subsequent study, Silvia and Barona (2009) could not replicate this result. However, their main goal was to test the hypothesis that pictures with curved elements are preferred to those with angular elements (Bar and Neta, 2006). Therefore, they applied only a selection of 9 pictures with circles and one of 9 pictures with hexagons from the APB to construct three different levels of balance. Whereas pictures with circles were indeed preferred to those with hexagons, the correlation between *APB* score and liking was rather low.

Method

Twenty-one students (5 male) from the University of Konstanz with an average age of 24 years ($SD = 3.11$) participated in the experiment. They were recruited in the same way as in Experiment 1, and no one had participated in Experiment 1. All had normal or corrected-to-normal vision and were paid with 4 € for their participation. The experiment was performed under the same ethical standards as the previous experiment.

Stimuli and apparatus were the same as in Experiment 1. Also the procedure was similar. A slider was again used for the assessment of picture preference (from “I do not like it” to “I like it”). The task consisted of two blocks of 130 trials (all 130 stimuli).

In each block the pictures were presented in a randomized order. The experiment lasted approximately 30 min.

Results

The mean preference ratings ranged from 21.2 to 70.2 ($M = 48.7$, $SD = 13.0$). They were subjected to a within-participant one-way ANOVA with factor *stimulus type* (circles, or hexagons). The analysis revealed a significant difference, $F_{(1, 20)} = 6.25$, $p < 0.05$, $\eta_p^2 = 0.238$, indicating that pictures with circles were liked more than those with hexagons (52.8 vs. 44.6).

We computed for each participant the correlation between the preference ratings and the objective measures (*APB*, *DCM*, *MS*, and *HG*). The mean correlations are shown in **Table 7**. As can be seen, they were substantial, except for the *MS* measure. However, the variability across participants was large (see **Figure 2**). The correlations between *APB* scores and liking ranged from -0.863 to 0.339 for the pictures with circles, and from -0.851 to 0.387 for the pictures with hexagons. Altogether, there were 5 participants with at least one non-significant ($p > 0.05$), correlation. Three participants produced at least one significant positive correlation. The correlations between *DCM* scores and liking ranged from -0.829 to 0.350 for the pictures with circles, and from -0.842 to 0.152 for the pictures with hexagons. There were 6 participants with at least one non-significant ($p > 0.05$) correlation, and two participants produced at least one significant positive correlation. The correlations were somewhat higher for the *APB* score than for the *DCM* score. A comparison across both stimulus types revealed a significant difference (-0.441 vs. -0.420), $F_{(1, 20)} = 4.59$, $p < 0.05$, $\eta_p^2 = 0.187$.

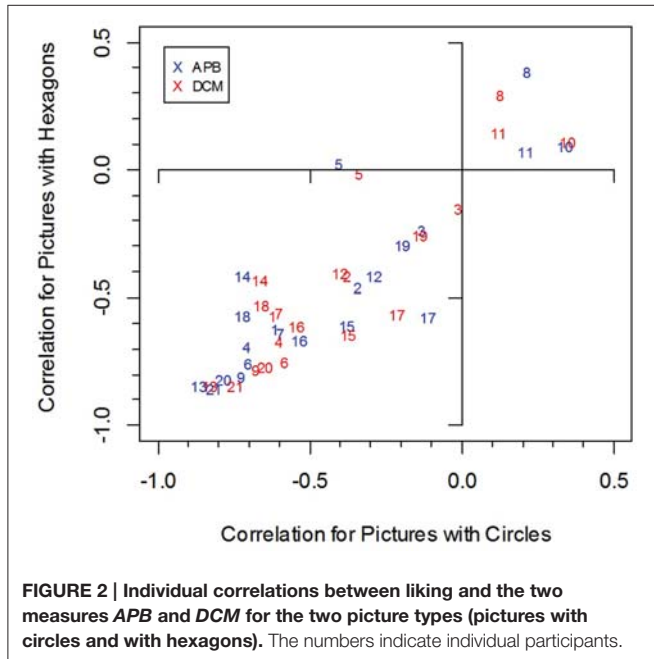
The correlations between *HG* scores and liking ranged from -0.353 to 0.829 for the pictures with circles, and from -0.435 to 0.785 for the pictures with hexagons. Altogether, there were 5 participants with at least one non-significant ($p > 0.05$) correlation, and three participants produced at least one significant negative correlation. The correlations with liking were somewhat lower for the *HG* score than for the *APB* score, but higher relative to the *DCM* score. A comparison between the correlations for the *APB* and the *HG* scores across both stimulus

TABLE 7 | Means (across participants) of the individual correlations (across pictures) between the preference ratings and the different scores in Experiment 2.

	APB	DCM	MS	HG
Liking Circles	-0.423 (0.363)	-0.401 (0.327)	0.084 (0.089)	0.408 (0.359)
Liking Hexagons	-0.459 (0.351)	-0.438 (0.340)	0.127 (0.110)	0.441 (0.320)

The values in parenthesis are the standard deviations.

APB, Assessment of preference for balance; DCM, Deviation of the center of mass; MS, Mirror symmetry; HG, Homogeneity.



types revealed no significant difference (-0.441 vs. -0.425), $F_{(1, 20)} = 1.41$, $p = 0.249$, $\eta_p^2 = 0.066$.

The correlations of the mean preference ratings with the different measures and previous ratings are shown in **Table 8**. Obviously, the correlations with the ratings from Experiment 1 were rather high. Interestingly, liking correlated higher with symmetry than with balance. The correlations with the objective measures were also high, except for *MS*. The relation of the mean preference ratings with the *APB* scores and that with the *DCM* scores are also illustrated in **Figure 3**.

For the *APB* measure we also computed a multiple linear regression for analyzing to what extent its components were related to the preference ratings. The result is shown in **Table 9**. As can be seen, the explained variance across both stimulus types increased only slightly from 75.2 to 78.5%, compared to the original scores. Interestingly, the inner-outer components had small but reliable effects.

Discussion

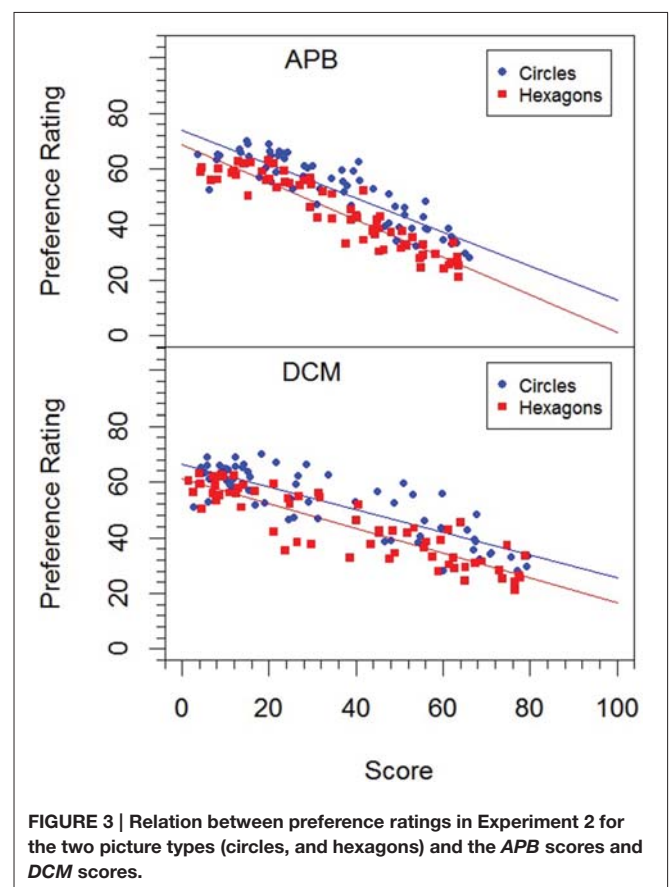
In this experiment the participants had to judge how much they liked the pictures from the *APB* also applied in Experiment 1. First of all, pictures with circles were liked more than those with hexagons, which supports the hypothesis of Silvia and Barona

TABLE 8 | Correlations between the mean preference rating in Experiment 2 and the different scores and ratings from Experiment 1 across both stimulus types.

	Bal.	Sym.	APB	DCM	MS	HG
Liking All	0.816	0.900	-0.867	-0.844	0.199	0.848
(R^2)	(0.666)	(0.810)	(0.752)	(0.713)	(0.039)	(0.719)
Circles	0.793	0.898	-0.882	-0.836	0.171 ^{n.s.}	0.843
(R^2)	(0.629)	(0.807)	(0.778)	(0.699)	(0.029)	(0.710)
Hexagons	0.843	0.930	-0.928	-0.892	0.256	0.875
(R^2)	(0.711)	(0.865)	(0.862)	(0.796)	(0.066)	(0.766)

The values in parenthesis are the corresponding R^2 -values.

^{n.s.} not significant; Balance, Bal. rating (Experiment 1); Sym., Symmetry Rating (Experiment 1); APB, Assessment of preference for balance; DCM, Deviation of the center of mass; MS, Mirror symmetry; HG, Homogeneity.



(2009) that curved objects are preferred to angular ones. If we consider the relation between liking and the judgments from Experiment 1, then liking correlated higher with the pictures' rated symmetry than with their rated balance. This suggests that aesthetic preference is more affected by symmetry perception than by balance perception. However, it remains somewhat unclear how the participants operationalized these two concepts.

With respect to the *APB* measure, we replicated Wilson and Chatterjee's (2005) result. The scores correlated substantially with the mean preference ratings (see **Figure 3**). This seems to contradict Silvia and Barona's non-significant results. However,

TABLE 9 | Regressions of preference ratings from Experiment 2 on the components of the APB score.

	Circles ($R^2 = 0.809$) $F_{(8, 56)} = 29.6$, $p < 0.001$ Intercept = 726.1		Hexagons ($R^2 = 0.885$) $F_{(8, 56)} = 53.6$, $p < 0.001$ Intercept = 67.0		Both ($R^2 = 0.785$) $F_{(8, 121)} = 55.2$, $p < 0.001$ Intercept = 69.8	
	β	$P(> t)$	β	$P(> t)$	β	$P(> t)$
<i>h</i>	-0.086	0.036*	-0.101	0.000***	-0.082	0.003**
<i>v</i>	-0.038	0.204	-0.101	0.001***	-0.074	0.003**
<i>ad</i>	-0.109	0.005**	-0.094	0.002**	-0.113	0.000***
<i>md</i>	-0.115	0.000***	-0.097	0.001**	-0.121	0.000***
<i>h_{io}</i>	-0.063	0.096	-0.062	0.038*	-0.050	0.067
<i>v_{io}</i>	-0.026	0.440	-0.034	0.230	-0.024	0.358
<i>ad_{io}</i>	-0.048	0.165	-0.071	0.017*	-0.066	0.013*
<i>md_{io}</i>	-0.078	0.041*	-0.037	0.183	-0.049	0.066

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

these researchers used only a small selection of 18 pictures from the APB and analyzed individual correlations rather than the correlation between the average ratings and APB scores. As we have seen, individual correlations can be much lower and vary considerably across participants (see **Figure 2**).

The DCM scores also correlated highly with the preference ratings, although significantly less than the APB scores. The smallest correlation occurred between the MS scores and the preference ratings.

A multiple linear regression of the preference ratings on the different components of the APB score revealed a reliable effect of inner-outer components (see **Table 9**). Because these components are related to homogeneity, there was also a corresponding high correlation between liking and the measure HG, which was not significantly different from that between liking and APB scores.

EXPERIMENT 3

In our third experiment we wanted to see how general the obtained results are, i.e., to what extent they depended on the specific stimulus set. Wilson and Chatterjee's (2005) constructed their stimuli manually with a drawing program in such a way that their sets covered a large range of APB scores. This construction process might have produced systematic relations between stimulus features which are favorable for the correlation of APB scores with ratings of balance and liking. For instance, if we consider **Figure 3**, then we see that the APB scores of the pictures are indeed rather evenly distributed whereas the DCM scores cluster somewhat at smaller values. That is, the center of "mass" of many APB stimuli is close to the geometric center. To additionally apply stimuli whose DCM scores are more evenly distributed, we constructed new pictures which also contained seven circles or seven hexagons of different size (examples are shown in the right panel of **Figure 1**). However, the positions of the elements in each picture were randomly drawn from a two-dimensional Gaussian distributions with specific mean and variance. For the new pictures the APB and DCM scores were

somewhat smaller, compared to the APB stimuli. Homogeneity and mirror symmetry were also reduced.

Different from Wilson and Chatterjee's (2005), we used the same positions for constructing pictures with circles and for those with hexagons, which allowed us a more reliable comparison between the ratings for these stimulus types. Finally, we required preference as well as balance ratings from the same participants.

Method

Twenty-seven students from the University of Konstanz were recruited via an online recruitment system (ORSEE, Greiner, 2015) for participating in the experiment. The data of four participants were excluded from data analysis, because their balance ratings were opposite to those of the other participants, which indicates that they did not understand the task correctly. The remaining 23 participants (4 males) had an average age of 22 years ($SD = 3.18$). All had normal or corrected-to-normal vision and were paid 5 € for their participation. The experiment was performed under the same ethical standards as the previous experiments.

As stimuli served new sets of pictures. Each picture had an extension of 500×500 pixels, which approximately corresponded to a visual angle of 14° horizontally and vertically. As elements served either seven circles or seven hexagons of different size. Examples are shown in the right panel of **Figure 1**. As can be seen, the elements were somewhat smaller than those in the APB stimuli. The location of the elements was established by a random process. The positions for each picture were drawn from a two-dimensional Gaussian distribution under the restriction that the elements do not overlap. The mean of the distribution could be one of the nine combinations resulting from three horizontal (left, center, and right) and three vertical (top, center, and bottom) positions. Most pictures (53) were constructed from a distribution with a "center-center" mean, i.e., a mean that corresponded to the geometric center. To construct less balanced pictures, also the other means, e.g., "top-left" were used. Additionally, the variances were reduced. A set of 72 pictures for each element type (circles, hexagons) was put together such that DCM scores were evenly distributed. For the new stimulus sets the APB scores ranged from 23.9 to 72.4 ($M = 49.0$, $SD = 12.3$), and the DCM scores from 4.30 to 83.0 ($M = 44.7$, $SD = 25.3$). Homogeneity (entropy) ranged from 58.4 to 85.5 ($M = 70.3$, $SD = 5.48$), and mirror symmetry from 0.29 to 4.02 ($M = 1.22$, $SD = 0.616$).

The experimental procedure was similar to that in our previous experiments, except that participants rated each picture with respect to balance and liking. There were two blocks of 144 trials (all 144 stimuli) each. In each block the pictures were presented in a randomized order. In the first block the participants had to judge how much they liked each picture, and in the second block they had to rate the pictures' balance. The experiment lasted approximately 30 min.

Results

We first computed the mean balance ratings and the mean preference ratings across participants for each picture and

TABLE 10 | Correlations between the mean ratings in Experiment 3 and the different scores across both stimulus types.

	Liking	Balance	APB	DCM	MS	HG
Liking	–	0.865***	–0.737***	–0.742***	0.296***	0.697***
Balance	0.748	–	–0.836***	–0.916***	0.364***	0.661***
APB	0.544	0.699***	–	0.793***	–0.139**	–0.689***
DCM	0.552	0.838***	0.628***	–	–0.454***	–0.558***
MS	0.088	0.132***	0.019***	0.206	–	0.194**
HG	0.486	0.437***	0.474***	0.311	0.003	–

The values below the diagonal are the corresponding R^2 -values.

** $p < 0.01$; *** $p < 0.001$. APB, Assessment of preference for balance; DCM, Deviation of the center of "mass"; MS, Mirror symmetry; HG, Homogeneity.

correlated the obtained values across all 144 pictures with the different scores. The resulting correlations and corresponding R^2 -values are shown in **Table 10**. Moreover, the correlations between the different measures are also shown in this table. As can be seen, compared to the previous experiments, the correlations between the mean APB scores and the other measures were generally smaller. The same holds for the DCM measure, except for the correlation with MS. The relations between the ratings and the APB scores and DCM scores were further analyzed.

Balance Ratings

The mean balance ratings, which ranged from 8.09 to 69.3 ($M = 42.2$, $SD = 16.4$), were subjected to a within-participant ANOVA with factor *stimulus type* (circles, or hexagons). The analysis revealed no significant difference, $F_{(1, 23)} = 0.02$, $p = 0.56$, $\eta_p^2 = 0.015$. Mean balance for pictures with circles and for those with hexagons were almost identical (circles: 42.4, hexagons: 42.0). Next, we computed for each participant the correlation between the balance ratings and the relevant scores across the 72 pictures with circles and across the 72 pictures with hexagons.

APB

For the pictures with circles, the correlations between the APB scores and the balance ratings ranged from -0.818 to -0.187 , and for those with hexagons from -0.658 to -0.146 . There was only one participant with non-significant ($p > 0.05$) correlations for both stimulus types. Three participants had a non-significant correlation for one of the stimulus types. The mean correlations for the circles and hexagons were -0.601 ($SD = 0.169$) and -0.495 ($SD = 0.147$), respectively.

DCM

Correlations between the DCM scores and the balance ratings ranged from -0.853 to -0.258 for pictures with circles, and from -0.775 to -0.146 for pictures with hexagons. There was no participant with non-significant ($p > 0.05$) correlations for both stimulus types. Two participants had a non-significant correlation for one of the stimulus types. The mean correlations for the circles and hexagons were -0.669 ($SD = 0.172$) and -0.581 ($SD = 0.176$), respectively.

HG

Correlations between the HG scores and the ratings ranged from 0.128 to 0.756 for pictures with circles, and from 0.169 to 0.596 for pictures with hexagons. One participants had non-significant ($p > 0.05$) correlations for both stimulus types, and two participants had a non-significant correlation for one of the stimulus types. The mean correlations for the circles and hexagons were 0.514 ($SD = 0.161$) and 0.398 ($SD = 0.134$), respectively.

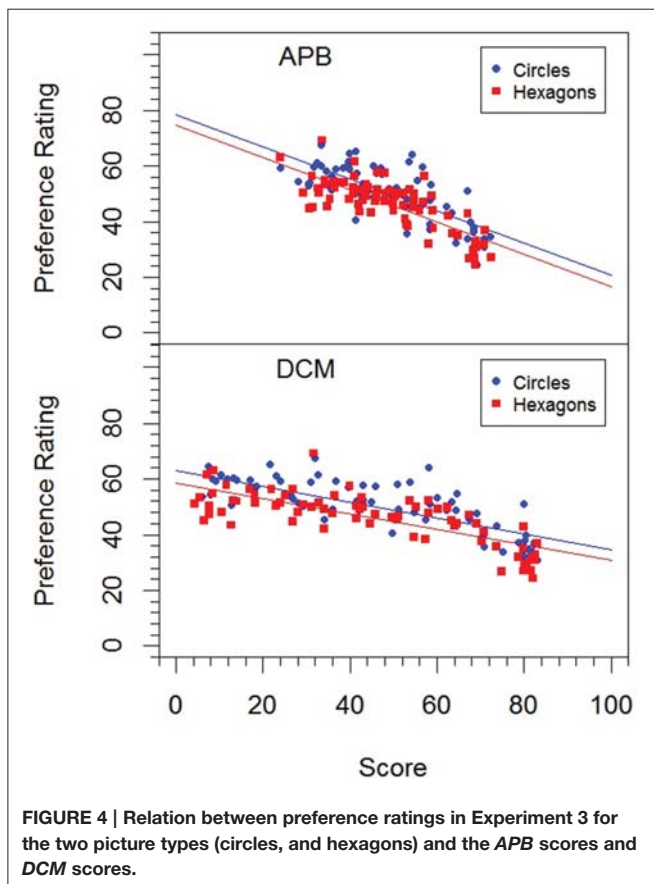
The correlation values (for HG the values were multiplied by -1) were entered to an ANOVA with the two within-participant factors *score* (APB, DCM, or HG), and *stimulus type* (circles, or hexagons). It revealed significant main effects. The correlations between the scores and ratings were significantly higher (-0.597 vs. -0.491) for pictures with circles than for those with hexagons, $F_{(1, 22)} = 65.8$, $p < 0.001$, $\eta_p^2 = 0.749$. Moreover, the correlations differed reliably between the scores $F_{(2, 44)} = 27.9$, $p < 0.001$, $\eta_p^2 = 0.559$. Further analyses showed that the correlations were significantly higher for the DCM scores than for the APB scores (-0.625 vs. -0.551), $F_{(1, 22)} = 13.2$, $p < 0.01$, $\eta_p^2 = 0.376$, and for APB scores than for the HG scores (-0.551 vs. -0.456), $F_{(1, 22)} = 42.5$, $p < 0.001$, $\eta_p^2 = 0.659$. These relations also corresponds to the correlations with the mean ratings (see **Table 10**).

Preference Ratings

The mean preference ratings ranged from 24.3 to 69.2 ($M = 48.2$, $SD = 9.61$). They were subjected to a within-participant one-way ANOVA with factor *stimulus type* (circles, or hexagons). The analysis revealed a significant difference, $F_{(1, 22)} = 7.75$, $p < 0.05$, $\eta_p^2 = 0.261$, indicating that pictures with circles were liked more than those with hexagons (50.3 vs. 46.2). This difference can also be seen in **Figure 4**, where the relations between the mean preference ratings and the APB scores and DCM scores are shown by scatterplots.

APB

The average of the correlations between the APB scores and the preference ratings was -0.276 . For the pictures with circles the correlations ranged from -0.715 to 0.419, and for those with hexagons from -0.658 to -0.534 . Four participants had non-significant ($p > 0.05$) correlations for both stimulus types. Three participants had a non-significant correlation for one of



the stimulus types. The mean correlations for the circles and hexagons were -0.310 ($SD = 0.356$) and -0.241 ($SD = 0.330$), respectively.

DCM

The average of the correlations between the APB scores and the preference ratings was -0.292 . The correlation ranged from -0.706 to 0.391 for pictures with circles, and for those with hexagons from -0.744 to 0.564 . Four participants had non-significant ($p > 0.05$) correlations for both stimulus types, and four participants had a non-significant correlation for one of the stimulus types. The mean correlations for the circles and hexagons were -0.315 ($SD = 0.318$) and -0.269 ($SD = 0.316$), respectively.

HG

The average of the correlations between the HG scores and the preference ratings was -0.265 . The correlation between the HG measures and the pictures with circles ranged from -0.484 to 0.642 , and for those with hexagons from -0.467 to 0.552 . Five participants had non-significant ($p > 0.05$) correlations for both stimulus types, and four participants had a non-significant correlation for one of the stimulus types. The mean correlations for the circles and hexagons were 0.310 ($SD = 0.345$) and 0.220 ($SD = 0.282$), respectively.

The correlation values (for HG the values were multiplied by -1) were entered to an ANOVA with the

two within-participant factors *score* (APB, DCM, or HG), and *stimulus type* (circles, or hexagons). It revealed a significant main effect of *stimulus type*. The correlations between the scores and the preference ratings were significantly higher (-0.312 vs. -0.243) for pictures with circles than for those with hexagons, $F_{(1, 22)} = 5.26$, $p < 0.05$, $\eta_p^2 = 0.193$. The factor *score* (DCM = -0.292 , APB = -0.276 , HG = -0.265) was not significant, $F_{(2, 44)} = 1.39$, $p = 0.260$, $\eta_p^2 = 0.059$.

Discussion

In this experiment we applied new sets of pictures with circles or hexagons, whose element positions were selected by a random process. On average, the pictures were less balanced, mirror symmetric, and homogenous, compared to the APB stimuli. The reduced objective balance is also reflected by the somewhat smaller balance ratings, compared to Experiment 1. If we consider the different measures, then their correlations with the balance ratings were significantly higher for the DCM scores than for the APB scores, and those for the APB scores were significantly higher than those for the HG scores. This relation also holds for the correlation with the mean balance ratings (see Table 10). These results demonstrate again that the DCM measure is well-suited for assessing the balance of simple pictures. However, our results also show that the correlations depended on the form of the picture elements. They were significantly higher for pictures with circles than for those with hexagons. We can take this result seriously, because in our stimuli circles and hexagons had identical positions in the corresponding pictures.

In contrast to the balance ratings, the range and mean of the preference ratings for the new pictures are similar to those in Experiment 2. Moreover, pictures with circles were again preferred to those with hexagons. This effect was even more pronounced than in Experiment 2, presumably because our picture types had identical element locations.

With respect to the correlations of the preference ratings with the different measures, we found no reliable difference between the APB, DCM, and BH scores. However, stimulus type again modulated the correlations. For pictures with circles they were reliably higher than for those with hexagons.

Taken together, our results with the new stimuli are similar to those obtained with the APB stimuli in the previous two experiments. However, the advantage of the DCM scores as measure of balance, compared to the APB scores, came out more clearly. Moreover, the advantage of the APB scores over the DCM scores for predicting liking vanished. This demonstrates that the extent to which the different measures account for balance and preference ratings, depends on the specific selection of pictures, even if the pictures are rather similar.

GENERAL DISCUSSION

There is a wide agreement that pictorial balance is crucial for the aesthetic appreciation of pictures (e.g., Poore, 1903; Arnheim, 1954; Locher et al., 1996). However, the mechanisms of balance perception and their relation to aesthetics are still not well understood. A promising approach to shed some light onto these

mechanisms has been to create objective measures of balance that correlate with aesthetic preference. Therefore, the aim of the present study was to compare currently applied measures by examining how well they account for balance, symmetry, and preference judgments of simple pictures.

One of the considered measures of balance, the *DCM*, is based on the theory of Arnheim (1954) and his precursors (e.g., Poore, 1903; Ross, 1907). The idea of these researchers was that each picture has a center of perceptual “mass,” and that the more this center deviates from the picture’s geometric center, the less balanced the picture is. Although it has been shown that such a measure does not generally correlate with aesthetic preference (e.g., McManus et al., 2011), we hypothesized that it might work for the simple pictures used in this study.

Another measure that was less inspired by a physical metaphor was the *APB* score (Wilson and Chatterjee, 2005). It is defined by the mean of eight partial measures that are more or less related to symmetry. For comparison, we also included a more strict measure of symmetry. Specifically, we computed a score *MS* that represents the degree of mirror symmetry around four axes. Finally, we applied a measure *HG* that reflects the homogeneity of the elements in a picture.

In our first experiment, participants judged the perceptual balance and symmetry of simple pictures taken from Wilson and Chatterjee’s (2005) *APB* test. The pictures consisted of seven circles or hexagons of different size. Our results show that, although there was some variance across participants, the *DCM* and *APB* scores correlated rather high with the balance and symmetry mean ratings. The high correlation of the *APB* scores with the balance ratings replicates the result of Wilson and Chatterjee’s (2005). For examining to what extent the individual components of this measure were related to the balance ratings, we entered the components into a multiple linear regression. The analyses revealed that the horizontal dimension had the largest weight, followed by the vertical dimension. The other components had little or no effect. This result suggests that a differential weighting of the components can improve the predictive power of the *APB* scores with respect to balance ratings. Such a modification of the original measure would also remedy one of its deficits. As has been shown, balance perception strongly depends on the orientation of a picture (Gershoni and Hochstein, 2011). The *APB* measure, however, is invariant with respect to image rotation.

Interestingly, the *DCM* scores correlated numerically higher with the balance and symmetry ratings than the *APB* scores. This demonstrates that the traditional idea of a center of perceptual “mass” is an adequate account of perceptual balance, at least for simple pictures. Inventing a new score would not have been necessary. Moreover, a detailed analysis revealed that the *APB* measure is inconsistent to some extent. Pictures whose elements are only located in the central area receive a relatively high score (poor balance), although they were rated as highly balanced. Responsible for this inconsistency are the inner-outer components of the *APB* measure, which reflect to a large degree homogeneity. In the present case the effects of the inconsistency remained relatively weak, because there were only few pictures of this type.

Obviously, only if the center of “mass” is near the geometric center, homogeneity can vary freely from minimum to maximum. The closer the center of “mass” moves to the border, i.e., the less balanced a picture, the more restricted homogeneity. That is, in the less balanced pictures the elements are less scattered. Consequently, balance and homogeneity are correlated across stimuli. This confound was responsible for the correlation between the homogeneity scores and the balance ratings. However, the correlation was smaller than those for the *DCM* and *APB* scores. The smallest correlations were found for the *MS* scores. This suggests that the visual system is not very sensitive to mirror symmetry, at least if the elements are scattered as in the present study.

To see how the ratings and measures are related to aesthetic preference ratings, we conducted a second experiment in which (different) participants had to rate how much they liked the *APB* pictures. First of all, we found that the pictures with circles were liked more than those with hexagons. This replicates results from Silvia and Barona (2009) and supports the hypothesis that pictures with curved elements are preferred to those with angular elements (Bar and Neta, 2006). Furthermore, correlating the preference ratings with the ratings from Experiment 1 revealed that liking correlated higher with symmetry than with balance. However, this result is not easy to interpret, because symmetry and balance ratings were highly correlated. Moreover, rating the balance of pictures was presumably more difficult to conceptualize than rating their symmetry.

The correlations between the preference ratings and the different measures were generally rather high except for the *MS* scores. However, the correlations with the *APB* scores were significantly higher than those with the *DCM* scores. Interestingly, the correlation with the *HG* scores did not differ significantly from those with the *APB* scores. This indicates that homogeneity, in addition to balance, also affected preference, which, in turn, explains why the *APB* scores correlated higher with liking than the *DCM* scores. For predicting linking, the inner-outer components of the *APB* measure, which largely reflect homogeneity, came favorably into play.

Thus, the first two experiments show that the *DCM* scores can account similarly well as the *APB* scores for the balance ratings, but might be preferred, because the latter measure can lead to inconsistencies. For predicting preference, the *APB* scores were superior to the *DCM* scores, mainly because they take homogeneity into account. That homogeneity is a crucial feature is supported by the fact that the *HG* scores also accounted well for the preferences and, therefore, could be used alternatively. As one reviewer pointed out, the fact that homogeneity played an important role for preference is compatible with Arnheim’s (1954) idea that also the corners of a frame exert some “force” on the elements.

Because these results in our first two experiments were obtained with a specific selection of stimuli that were presumably constructed to be optimal for the *APB* scores, we wanted to examine to what extent the results generalize to a different set of stimuli. For this objective we conducted a third experiment

with new pictures that were also constructed from circles and hexagons, but whose element locations were drawn randomly from a two-dimensional Gaussian distribution. The resulting pictures are less balanced than those from the APB, but the corresponding DCM scores are more evenly distributed. Moreover, the correlations between the different measures are reduced, except for mirror symmetry. The participants in Experiment 3 had to rate both balance and preference. Whereas the obtained balance ratings were somewhat smaller than those in Experiment 1, the preference ratings were similar to those in Experiment 2. Pictures with circles were again preferred to those with hexagons. Moreover, the correlation between the balance ratings and the DCM scores was significantly higher than that with the APB scores, whereas the correlations with the preference ratings did not significantly differ between DCM, APB, and HG scores. These results demonstrate that the performance of the different scores depend, at least to some extent, on the specific selection of pictures, even if they look rather similar.

Taken together, our experiments and analyses demonstrate that the APB score is not a pure measure of balance. Therefore, if one is interested in predicting perceptual balance, then the DCM measure is the better choice, mainly because it is less affected by homogeneity. If the goal is to predict preference ratings for pictures, then the APB score is appropriate. Our results indicate that preference not only depends on balance, but also on homogeneity, which is taken into account by the APB measure. However, APB scores can be substituted

by HG scores, which produced comparable results. Their advantage is that they are rather simple to compute and easy to comprehend.

Further studies will have to show to what extent the considered measures can also predict aesthetic preferences for more complex pictures, e.g., for those also including objects. A related question is whether the present ratings were the result of a global impression formed during the “first glance” (Locher, 2015), or also of a deeper processing. The registration of eye movements during the rating period would presumably be helpful in this respect. Finally, it should be noted that we considered a selection of proposed or possible objective measures for predicting perceptual balance and aesthetic preference. Therefore, it would be interesting to compare other scores as well. For instance, for the type of pictures applied in this study measures reflecting the goodness of dot patterns (e.g., Van Der Helm and Leeuwenberg, 1996) are promising candidates.

AUTHOR CONTRIBUTIONS

RH: conception, design, final approval, accountable for quality. MF: conception, data acquisition, drafting, revising, interpretation.

ACKNOWLEDGMENTS

We thank Lydia Wenger for her conceptual contribution to this article.

REFERENCES

- Arnheim, R. (1954). *Art and Visual Perception: A Psychology of the Creative Eye*. Berkeley, CA; Los Angeles, CA: University of California Press.
- Bar, M., and Neta, M. (2006). Humans prefer curved visual objects. *Psychol. Sci.* 17, 645–648. doi: 10.1111/j.1467-9280.2006.01759.x
- Bauerly, M., and Liu, Y. (2006). Computational modeling and experimental investigation of effects of compositional elements on interface and design aesthetics. *Int. J. Hum. Comput. Stud.* 64, 670–682. doi: 10.1016/j.ijhcs.2006.01.002
- Cupchik, G. C. (2007). A critical reflection on Arnheim's Gestalt theory of aesthetics. *Psychol. Aesthet. Creat. Arts* 1:16. doi: 10.1037/1931-3896.1.1.16
- Fechner, G. T. (1871). Zur experimentalen Ästhetik. *Abh. der Königlich Sächsischen Ges. der Wissenschaften* 9, 555–635.
- Fechner, G. T. (1876). *Vorschule der Ästhetik*. Leipzig: Breitkopf and Härtel.
- Garner, W. R., and Clement, D. E. (1963). Goodness of pattern and pattern uncertainty. *J. Verbal Learn. Verbal Behav.* 2, 446–452.
- Gershoni, S., and Hochstein, S. (2011). Measuring pictorial balance perception at first glance using Japanese calligraphy. *i-Perception* 2, 508. doi: 10.1068/i0472aap
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *J. Econ. Sci. Assoc.* 1, 114–125. doi: 10.1007/s40881-015-0004-4
- Jacobsen, T., and Höfel, L. (2002). Aesthetic judgments of novel graphic patterns: analyses of individual judgments. *Percept. Mot. Skills* 95, 755–766. doi: 10.2466/pms.2002.95.3.755
- Latto, R. (1995). “The brain of the beholder,” in *The Artful Eye*, eds L. R. Gregory, J. J. Harris, P. Heard, and D. Rose (Oxford: Oxford University Press), 66–94.
- Locher, P. (2015). “The aesthetic experience with visual art ‘at first glance,’” in *Investigations into the Phenomenology and the Ontology of the Work of Art*, Vol. 81, eds P. F. Bundgaard and F. Stjernfelt (New York, NY: Springer International Publishing), 75–88.
- Locher, P., Gray, S., and Nodine, C. (1996). The structural framework of pictorial balance. *Perception* 25, 1419–1436. doi: 10.1068/p251419
- Locher, P., and Nodine, C. (1989). The perceptual value of symmetry. *Comput. Math. Appl.* 17, 475–484. doi: 10.1016/0898-1221(89)90246-0
- McManus, I., Edmondson, D., and Rodger, J. (1985). Balance in pictures. *Br. J. Psychol.* 76, 311–324. doi: 10.1111/j.2044-8295.1985.tb01955.x
- McManus, I., Stöver, K., and Kim, D. (2011). Arnheim's Gestalt theory of visual balance: examining the compositional structure of art photographs and abstract images. *i-Perception* 2, 615. doi: 10.1068/i0445aap
- Munar, E., Gómez-Puerto, G., López-Navarro, E., and Nadal, M. (2014). “Visual preference for curvature as a potential aesthetic primitive,” in *Proceedings of the Twenty-Third Biennial Congress of the International Association of Empirical Aesthetics*, ed A. Kotzbelt (New York, NY), 316–319.
- Ngo, D. C. L., Teo, L. S., and Byrne, J. G. (2002). Evaluating interface esthetics. *Knowl. Inf. Syst.* 4, 46–79. doi: 10.1007/s10115-002-8193-6
- Palmer, S. E., Schloss, K. B., and Sammartino, J. (2013). Visual aesthetics and human preference. *Annu. Rev. Psychol.* 64, 77–107. doi: 10.1146/annurev-psych-120710-100504
- Poore, H. R. (1903). *Pictorial Composition and the Critical Judgement of Pictures*. New York, NY: Baker and Taylor.
- Ross, D. W. (1907). *A Theory of Pure Design: Harmony, Balance, Rhythm*. Boston, MA: Houghton, Mifflin.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423, 623–656. doi: 10.1002/j.1538-7305.1948.tb00917.x
- Silvia, P. J., and Barona, C. M. (2009). Do people prefer curved objects? Angularity, expertise, and aesthetic preference. *Empir. Stud. Arts* 27, 25–42. doi: 10.2190/EM.27.1.b

- Treiblmaier, H., and Filzmoser, P. (2009). *Benefits from Using Continuous Rating Scales in Online Survey Research*. Vienna: University of Technology, Research Report SM-2009-M-2004.
- Van Der Helm, P. A., and Leeuwenberg, E. L. (1996). Goodness of visual regularities: a nontransformational approach. *Psychol. Rev.* 103:429. doi: 10.1037/0033-295X.103.3.429
- Wilson, A., and Chatterjee, A. (2005). The assessment of preference for balance: introducing a new test. *Empir. Stud. Arts* 23, 165–180. doi: 10.2190/B1LR-MVF3-F36X-XR64

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Hübner and Fillinger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

5 Research Paper II

Relations Between Balance, Prototypicality, and Aesthetic Appreciation for Japanese Calligraphy

Authors:

Martin G. Fillinger

Ronald Hübner

Authors' affiliation:

Department of Psychology, Universität Konstanz, Germany

Citation:

Fillinger, M. G., & Hübner, R. (2018). Relations Between Balance, Prototypicality, and Aesthetic Appreciation for Japanese Calligraphy. *Empirical Studies of the Arts*. doi:10.1177/0276237418805656

Relations Between Balance, Prototypicality, and Aesthetic Appreciation for Japanese Calligraphy

Empirical Studies of the Arts
0(0) 1–19

© The Author(s) 2018

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0276237418805656

journals.sagepub.com/home/art



Martin G. Fillinger¹  and
Ronald Hübner¹

Abstract

Aesthetic appreciation of pictures partly depends on the perceptual balance of their elements. This relation has also been supported by objective measures predicting balance ratings as well as preference. Gershoni and Hochstein, however, applied these measures to Japanese calligraphies and failed to find such a relation, which questions the generality of these balance concepts. In our first experiment, we, therefore, tried to replicate these results with a slightly different method. In addition, we calculated further balance measures and collected liking ratings. As result, perceptual balance was again uncorrelated with the measures and with liking. In a second experiment, participants assessed the perceptual stability of the calligraphies, which was considered as alternative concept of balance, and their prototypicality. After discounting the effects of prototypicality on liking, there were correlations between liking and stability and between liking and one of the balance measures. However, the correlations were reliable only for atypical calligraphies.

Keywords

empirical aesthetics, preference, prototypicality, visual stability, perceptual balance

¹Universität Konstanz, Germany

Corresponding Author:

Martin G. Fillinger, Universität Konstanz, Fachbereich Psychologie, Fach D29. D-78457 Konstanz, Germany.

Email: martin.fillinger@uni-konstanz.de

Introduction

The aesthetic appreciation of pictures depends on perceptual factors such as form, color, and symmetry as well as on more conceptual factors like semantic content and prototypicality (for an overview, see Palmer, Schloss, & Sammartino, 2013). A perceptual factor that has received great attention in art theory and related fields is pictorial balance, that is, how well the picture's elements are arranged (Arnheim, 1982; Bouleau, 1980; Kandinsky, 1926/1979). A picture is commonly considered as balanced if the perceptual "weights" of the pictorial elements are in equilibrium. In recent years, computational measures have even been proposed for representing the balance of a picture. One is the *Assessment of Preference for Balance* (APB), suggested by Wilson and Chatterjee (2005). The APB score is defined as the average of eight symmetry measures over the four axes of a picture (horizontal, vertical, and the two diagonals). As an alternative measure of balance, Hübner and Fillinger (2016) proposed the *Deviation of the Center of "Mass"* (DCM) from the picture's geometrical center. The DCM is based on Arnheim's (1954) idea that the more the center of perceptual "mass" deviates from the geometric center of the picture, the less the picture is perceived as balanced.

Results obtained with the APB and DCM show that these measures do not only correlate with balance ratings but also predict liking ratings. However, the pictures applied in these studies were relatively simple and included only basic (e.g. circles, or squares) and unrelated geometrical elements, which raises the question to what extent the results can be generalized. In an attempt to answer this question, McManus, Stöver, and Kim (2011) could show that the center of mass in art photographs deviated less from the image center than that in random photographs. Yet, there is also negative evidence. Gershoni and Hochstein (2011), for instance, applied the APB to Japanese calligraphies and found no substantial correlation with balance ratings. These results suggest that the proposed formal balance measures are not valid for all picture types. Moreover, because the participants were obviously able to rate perceptual balance, it seems that there are concepts of balance not reflected by the formal measures. If this were indeed the case, then it would be important to know the different concepts of balance and the respective relevant stimulus features.

The aim of the present study was to investigate these issues. In a first step, we wanted to replicate and extend the results of Gershoni and Hochstein (2011). These researchers presented their stimuli very briefly (200 ms) and limited processing duration by masking. Moreover, the stimuli occurred at a random position (spatial uncertainty). Although it is known that balance can be perceived rapidly (Locher, Krupinski, Mello-Thoms, & Nodine, 2007; Locher & Nagy, 1996; Locher & Stappers, 2002; McManus, Edmondson, & Rodger, 1985), it cannot be excluded that Gershoni and Hochstein's procedure was suboptimal for obtaining balance ratings related to the measures. Therefore, we used the

same Japanese calligraphies as in the original study but presented the stimuli for a longer time and without spatial uncertainty.

Moreover, in addition to the APB, we also computed the DCM and homogeneity (HG) scores as measures of balance, where HG reflects the distribution of mass within a picture (Hübner & Fillinger, 2016). Because visual complexity might be related to balance as well (Gartus & Leder, 2017; Jacobsen, 2004), we additionally used the file size (FS) of the jpeg compressed image (Donderi, 2006; Forsythe, Nadal, Sheehy, Cela-Conde, & Sawey, 2011; Machado et al., 2015) and the degree of mirror symmetry (MS) in a picture (Hübner & Fillinger, 2016) as corresponding measures.

Furthermore, we wanted to examine the relation between balance and liking for the Japanese calligraphies. As mentioned in art theory (Arnheim, 1982; Bouleau, 1980; Kandinsky, 1926/1979), pictorial balance is considered as a crucial variable for aesthetic appreciation. This relation has been confirmed by Wilson and Chatterjee (2005) and by Hübner and Fillinger (2016). Surprisingly, Gershoni and Hochstein merely asked participants to assess the balance of the calligraphies, but not how much they liked them. Perhaps they supposed that perceptual balance and liking are generally closely related. To test whether this is indeed the case, we additionally asked our participants to indicate how much they liked the Japanese calligraphies.

As we will show, the balance ratings in our first experiment are similar to those of Gershoni and Hochstein (2011). Accordingly, there were again no substantial relations between the balance ratings and the different measures. This suggests that the concept of balance for Japanese calligraphies differs from that reflected by the measures. To investigate these issues further, we conducted a second experiment in which we examined perceived “stability” (Liu, Dong, Zhang, & Jiang, 2017) as an alternative concept of balance.

A further result of our first experiment was that there was no correlation between balance ratings and liking ratings, which indicates that other variables than balance determined the aesthetic differences between the Japanese calligraphies. To examine a possible variable in this respect, in our second experiment we also tested to what extent prototypicality (Rosch, 1975) determines how much Japanese calligraphies are liked. At least in product design, it has been shown that prototypicality can affect preference judgments (e.g., Hekkert, Snelders, & Van Wieringen, 2003; Whitfield & Slatter, 1979).

Experiment I

The main goal of our first experiment was to investigate whether the absent relation between balance ratings and APB scores in the study of Gershoni and Hochstein (2011) was due to the specific picture type (Japanese calligraphies) or to the relatively short and spatially uncertain stimulus presentation. Therefore, we presented the stimuli for a longer duration and without spatial uncertainty.

Moreover, we also collected liking ratings for the stimuli to examine whether the widely assumed positive relation between perceptual balance and aesthetic appreciation also holds for Japanese calligraphies. As computational measures for balance, we applied not only APB scores but also DCM and HG scores (Hübner & Fillinger, 2016). Furthermore, for additionally assessing effects of visual complexity, we considered the FS (file size) of our images (Donderi, 2006) and MS (mirror symmetry) scores (Gartus & Leder, 2017; Hübner & Fillinger, 2016) as corresponding measures.

















Method

Participants. One hundred forty-nine participants (32 men, mean age 23.6 years, $SD = 6.61$) were recruited via an online system (ORSEE; Greiner, 2015) for participation in an online experiment. The pool of persons includes students of all disciplines from the Universität Konstanz as well as persons not associated with our university. As incentive, participants had the chance to win a voucher for their participation. This study was carried out in accordance with the ethical guidelines of the Universität Konstanz and the Declaration of Helsinki. Participants were informed of their right to abstain from participation in the study or to withdraw consent to participate at any time without reprisal.

















Stimuli. As stimuli, we used the same 16 Japanese calligraphies (see Table 1) as Gershoni and Hochstein (2011). The calligraphies were positioned in a white square of 450×450 pixels and presented in the center of the monitor on a gray background. Stimulus presentation and response registration were controlled by SoSci Survey (Leiner, 2014). The device (computer, smartphone, etc.) and screen size at the user frontend were also registered.

Table 1. Mean Ratings (Balance and Liking) for the Japanese Calligraphies in Experiment 1.

Mean balance ratings in ascending order

															
8	16	13	14	5	6	4	15	1	7	11	12	9	10	3	2
42	47	47	48	48	48	51	59	59	61	63	65	71	78	79	80

Mean liking ratings in ascending order

															
5	2	8	9	6	14	7	1	12	13	16	11	3	4	10	15
38	39	43	51	52	53	54	56	56	58	60	61	63	64	65	66

Note. Bold numbers represent the picture IDs.

For all stimuli, we calculated objective measures of balance (APB, DCM, and HG scores) and complexity (FS and MS scores). The APB scores ranged from 10.5 to 38.9 ($M = 23.1$, $SD = 7.65$), the DCM scores from 2.51 to 19.6 ($M = 9.57$, $SD = 5.59$), the HG scores from 65 to 87 ($M = 79.8$, $SD = 6.37$), the FS scores from 12 to 36 ($M = 25.6$, $SD = 6.15$), and the MS scores from 1.34 to 19.1 ($M = 11.8$, $SD = 4.67$).

Procedure. The online experiment started with an instruction that informed the participants about the task. In addition, we used a seriousness check (Reips, 2002, 2009) to control for participants' involvement on the task. Subsequently, we used a counterbalanced design with two blocks in each of which the 16 stimuli were presented in random order. Participants had to rate how much they liked the stimuli (1 = *I do not like it* to 6 = *I like it*) in one block and how well the stimuli were balanced (1 = *not balanced* to 6 = *balanced*) in the other block. The ratings were entered by clicking on one of the six scale positions with the mouse. Shortly after the response, the next stimulus was displayed. There was no time limit. In total, the experiment lasted about 5 minutes.

Results

All 149 participants indicated in the seriousness check that they wanted to participate seriously; 40.3% conducted the online experiment on a computer (screen width ranged from 1,138 to 2,048 pixels, and screen height ranged from 640 and 1,152 pixels) and 59.7% on a smartphone (screen width between 320 and 412 pixels, and screen height between 568 and 846 pixels). Despite the use of different devices with unequal screen sizes, ratings of computer and smartphone users produced similar ratings: balance ratings: $r = .967$, $p < .001$, 95% CI [0.906, 0.989]; and liking ratings: $r = .918$, $p < .001$, 95% CI [0.776, 0.972] (correlations of mean values of every single stimuli across participants). Thus, our data did not depend on the use of a specific device.

For comparison with the original study, we also multiplied the ratings by 16, in order to adjust the values to the range of the balance scores (0 to 100). We then computed the mean balance and liking ratings across participants for each of the 16 stimuli. The correlation between balance and liking ratings was not reliable ($r = .374$, $p = .154$, 95% CI [-0.149, 0.734]). To test whether there were sequential effects between the rating types, we also analyzed the between-participants data separately for each block of ratings. For the first block, the correlation between balance (B_1) and liking (L_1) across pictures was not significant ($r = .168$, $p = .533$, 95% CI [-0.357, 0.613]). However, for the second block, after which the participants had already assessed balance or liking, respectively, the correlation (between B_2 and L_2) was significant ($r = .529$, $p = .035$, 95% CI [0.044, 0.812]). This increased correlation was due to influences from picture rating in Block 1 to that in Block 2. When we used partial correlation to control

for the variances of the ratings from Block 1, then the correlation between the rating in Block 2 changed to $r = -.170$, $p = .545$. This confirms the substantial interference between the blocks. Thus, to avoid potential biases and misinterpretations, we used only the between-participants data for further analyses.

For these data, the mean balance and liking ratings were 59.1 ($SD = 12.7$) and 54.9 ($SD = 8.76$), respectively. The mean ratings for each stimulus are shown in Table 1. Importantly, our balance ratings correlated highly ($r = .943$, $p < .001$, 95% CI [0.840, 0.980]) with those observed by Gershoni and Hochstein (2011).

The correlations between our mean ratings and the considered scores of balance and complexity are shown in Table 4. As can be seen, there was no reliable correlation ($r = .071$, $p = .793$, 95% CI [-0.440, 0.548]) between perceptual balance and the APB scores, which replicates the results from Gershoni and Hochstein (2011). Moreover, the absent correlation also held for the DCM and HG scores. Merely the complexity measure FS correlated significantly with the balance ratings ($r = -.573$, $p = .020$, 95% CI [-0.832, -0.108]). Finally, the relation between balance ratings and liking ratings, which is also shown graphically in Figure 1, was not significant ($r = .168$, $p = .533$, 95% CI [-0.357, 0.613]).

By averaging the ratings for each picture across participants, much information is lost. Therefore, to increase statistical power, we also analyzed the data by applying linear mixed-effects models (LMMs). These models use all data from each participant and allow to take interindividual differences and stimulus-specific effects into account by treating participants and stimuli as random factors.

The LMMs were computed with R (R Core Team, 2017) and the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015). We analyzed the extent to which the formal scores (fixed effects, without interaction term) account for the liking and balance ratings, respectively. Before examining the specific associations, we checked which model specification contributes significantly to the model's goodness of fit (Baayen, Davidson, & Bates, 2008). Therefore, for every association, we conducted a likelihood ratio test that compared the version of the model in which intercepts were allowed to vary across participants and stimuli (random-intercept model), with the version in which participants and stimuli were also allowed to have different slopes (random-slope model). For most of the associations, the random-slope model revealed no significant improvement in the model's goodness of fit (p -values $> .05$). Exceptions were the models with liking ratings as criterion and HG, MS, and FS as predictors. For these models, assuming varying slopes seemed justified. For the remaining models, we assumed only random intercepts for participants and stimuli. In addition, all predictors were used after a grand-mean-centering. Moreover, we applied the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017) to calculate the degrees of freedom using the Satterthwaite's approximations for the t test and corresponding p values. The results, except the variance components of participant and stimulus random slopes, because they were smaller than 1%, are

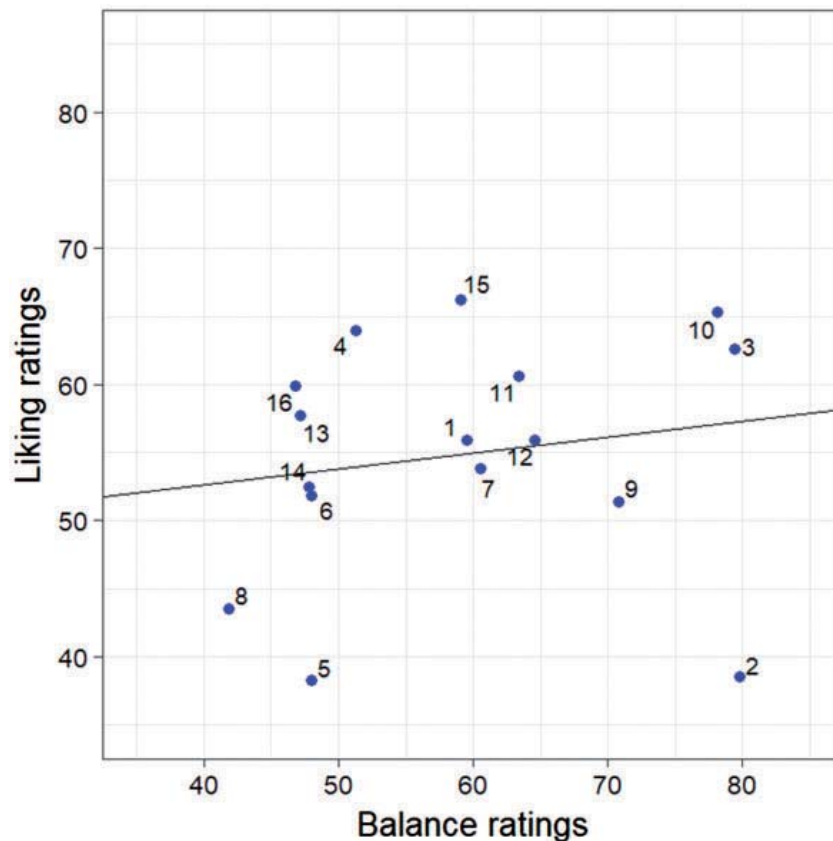


Figure 1. The relation between the mean balance ratings and mean liking ratings for the 16 calligraphies (see Table 1).

shown in Table 2. Our analyses revealed that no measure predicts balance or liking ratings, except FS, which significantly predicts the balance ratings ($p = .013$).

Discussion

Despite the methodological differences between our and Gershoni and Hochstein's (2011) study, the balance ratings were highly correlated. This shows that balance assessments are largely unaffected by stimulus duration and spatial uncertainty. Accordingly, in our experiment, there was also no significant relation between balance ratings and the APB scores, as in Gershoni and Hochstein, which also holds for the other balance measures (Hübner & Fillinger, 2016). Merely complexity (FS) was reliably correlated with the balance ratings. Because the correlation was negative, this means that the less complex the picture (smaller file size), the more it was perceived as balanced.

Thus, our data replicate and generalize Gershoni and Hochstein's (2011) result that the theoretical concepts of balance, as reflected by the considered computational measures, do not correspond to the balance perceived for the Japanese calligraphies. This suggests that perceptual balance is not a unique

Table 2. Results of the LMMs with the Objective Measures as Fixed Effect, and Participants and Stimuli as Random Effects.

Predictor	Criterion	Coefficient (SE)	df	t-value	p	σ_P^2 (%)	σ_S^2 (%)
ABP	Balance	0.118 (0.415)	16.0	0.285	.779	12.0	25.7
	Liking	0.025 (0.288)	16.0	0.088	.931	10.8	11.4
DCM	Balance	0.086 (0.570)	16.0	0.152	.881	12.0	25.8
	Liking	-0.436 (0.379)	16.0	-1.15	.266	10.9	10.6
HG	Balance	-0.592 (0.454)	16.0	-1.31	.210	12.3	23.9
	Liking	1.28 (0.921)	12.9	1.39	.189	11.3	12.1
MS	Balance	0.330 (0.677)	16.0	-0.488	.632	12.0	25.5
	Liking	0.442 (0.558)	11.9	0.792	.444	11.9	6.64
FS	Balance	-1.18 (0.424)	16.0	-2.78	.013	13.1	18.8
	Liking	0.075 (0.479)	7.30	0.157	.879	12.6	1.88

Note. ABP = Assessment of Preference for Balance; DCM = Deviation of the Center of "Mass"; HG = homogeneity; MS = mirror symmetry; FS = file size; SE = Standard error; df = degrees of freedom; σ^2 = variance components of participant (P) and stimulus (S) random intercepts.

concept. Rather, it seems that for different picture types, observers apply different concepts of balance, which are based on different features or feature combinations. One possible alternative concept will be considered in the next experiment.

An unexpected result was that the balance ratings were not related to the liking ratings. Several reasons are conceivable why this was the case. For instance, it is possible that only specific concepts of balance are related to aesthetic appreciation and that the concept used for assessing the calligraphies is not among them. However, there could also have been methodological reasons. In the studies of Wilson and Chatterjee (2005) and Hübner and Fillinger (2016), simple stimuli were used that were specifically constructed to vary mainly in the balance dimension. In contrast, the real calligraphies used here and in Gershoni and Hochstein (2011) vary also in other dimensions. Thus, even if balance affected liking, its effect might have remained undetected, because it was masked by other dimensions that affected liking to a much larger extent. If we would know these dimensions, then we could control for their contribution to liking and test, whether the remaining variance correlates significantly with balance. In the next experiment, we examined such a dimension.

Experiment 2

The goal of our second experiment was to further investigate the concept of perceptual balance for Japanese calligraphies and its relation to liking. As we

have seen in Experiment 1, the balance ratings were highly correlated with those observed by Gershoni and Hochstein (2011). This indicates that persons agree on what “balance” means for these stimuli. However, the fact that the ratings did not correlate with the formal measures of balance suggests that balance has a specific meaning for Japanese calligraphies that is not reflected by these measures. To get an idea which feature could have determined balance perception, we inspected the calligraphies in Table 1 and their order with respect to the balance ratings. It seemed to us that the stimuli assessed as less balanced show some kind of instability and flexibility, whereas calligraphies rated as more balanced look more stable and rigid. In art theory, stability is considered as a visual and aesthetic habit that plays an important role in composing pictures (Liu et al., 2017). This habit is formed in individuals through their lifetime by living with gravitation, which forces them to arrange things in a stable way. In this respect, others also speak of “gravitational stability” (van der Helm, 2015). To examine whether visual balance actually means perceived stability for the Japanese calligraphies, we asked our participants to assess the stability of these stimuli.

Another aim of the present experiment was to find a variable that strongly determined the liking ratings of the calligraphies in Experiment 1. Such a variable would allow us to control its effect and test whether the unexplained variance can be accounted for by balance. However, which property of the calligraphies strongly determines how much they are liked? After inspecting the pictures and their order with respect to liking (Table 1), we hypothesized that prototypicality (Rosch, 1975) could be related to liking. It has been shown before that prototypicality can have a positive effect on preference for design (Hekkert et al., 2003; Whitfield & Slatter, 1979). Furthermore, this property is related to perceptual fluency, which also affects liking (e.g., Graf & Landwehr, 2017). Therefore, we asked our participants to also assess the prototypicality of each stimulus.

Method

Participants and Stimuli. We again used ORSEE (Greiner, 2015) for the recruitment of 152 (39 men, mean age 24.3 years, $SD = 7.65$) participants. As incentive, participants had the chance to win a voucher for their participation. The experiment was performed under the same ethical standards as the previous experiment. We used the same stimuli as in Experiment 1. Moreover, stimuli were presented online with SoSci Survey (Leiner, 2014) under same conditions as in the previous experiment.

Procedure. The procedure was similar to that in Experiment 1. We used a counterbalanced block design. In one block, participants had to rate prototypicality of the stimuli, that is, we asked “Is this a typical Asian character?” (1 = *completely untypical* to 6 = *very typical*). In addition, in the other block,

we asked the participants how stable the arrangement of pictorial elements appears (1 = *instable* to 6 = *stable*). Participants entered their judgment by clicking on a 6-point scale. There was no time limit. In total, the experiment lasted about 5 minutes.

Results

Two of the 152 participants indicated in the seriousness check that they did not want to participate seriously. Consequently, the corresponding data were excluded from analysis. Furthermore, 49.3% of the participants used a computer (screen width ranged from 1,280 to 2,560 pixels, and screen height ranged from 720 and 1,440 pixels) for the task, 48% a smartphone (screen width between 320 and 732 pixels, and screen height between 412 and 1,107 pixels), and 2.0% a tablet (screen width between 601 and 768 pixels, and screen height between 962 and 1,024 pixels). One persons' device was not identifiable. Importantly, regardless of the used device and screen size, the performance showed a high consistency: prototypicality ratings: $r = .992$, $p < .001$, 95% CI [0.977, 0.997]; and stability ratings: $r = .980$, $p < .001$, 95% CI [0.942, 0.993] (correlations of mean values of every single stimuli across participants).

We first computed the mean prototypicality and stability ratings (multiplied by 16) across participants for each of the 16 stimuli. There was a reliable correlation between these two variables ($r = .688$, $p = .003$, 95% CI [0.292, 0.883]). However, when we analyzed the between-participants data for the first and second block separately, then the correlation was smaller for the first block ($r = .467$, $p = .068$, 95% CI [-0.037, 0.782]) than for the second one ($r = .818$, $p < .001$, 95% CI [0.543, 0.934]). This indicates that the ratings in the second block were again strongly influenced by the ratings in the first block. When we controlled for the ratings in Block 1, the correlation between prototypicality and stability decreased to $r = -.163$, $p = .562$. Therefore, we used only the ratings from the first block (between participants) for the further analyses.

















Means of prototypicality and stability ratings were 61.4 ($SD = 22.3$) and 59.0 ($SD = 14.2$), respectively. The means for each stimulus are shown in Table 3.

As can be seen in Table 4, prototypicality strongly correlated with the liking ratings from Experiment 1 (see also Figure 2). Moreover, there was a reliable correlation of prototypicality with the balance ratings from the previous experiment. As expected, the stability ratings also correlated highly with the balance ratings. However, stability did not correlate significantly with liking. Finally, there were no reliable correlations between the present ratings and the objective measures of balance and complexity.



Analogous to Experiment 1, LMMs were computed with typicality or stability ratings as dependent variable and the formal measures as predictors. The analyses revealed that no measure was able to predict prototypicality or stability ratings.

Table 3. Mean Prototypicality and Stability Ratings of the Japanese Calligraphies in Experiment 2.

Mean prototypicality ratings in ascending order

															
5	13	7	14	8	6	2	12	1	9	16	3	11	4	15	10
27	27	37	40	46	53	57	59	62	66	67	82	87	90	91	91

Mean stability ratings in ascending order

															
13	6	1	12	16	14	7	4	11	8	5	9	15	3	10	2
33	39	45	51	51	54	54	57	61	63	63	65	68	74	82	85

Note. Bold numbers represent the picture IDs.

Table 4. Correlations (Across Stimuli) Between the Mean Balance, Liking (Experiment 1), Prototypicality, and Stability Ratings (Experiment 2) and Objective Measures for Balance and Complexity.

	Liking	Prototypicality	Stability	APB	DCM	HG	MS	FS
Balance	.168	.498*	.699**	.071	.038	-.311	-.121	-.573*
Liking	–	.700**	-.088	.022	-.278	.268	.047	.095
Prototypicality	–	–	.467	.232	.072	.044	.122	-.201
Stability	–	–	–	.106	.213	-.127	.220	-.353

Note. APB = Assessment of Preference for Balance; DCM = Deviation of the Center of “Mass”; HG = homogeneity; MS = mirror symmetry; FS = file size.

* $p < .05$. ** $p < .01$.

The relationships between liking and the other ratings were further analyzed by hierarchical linear regressions with liking as dependent variable. To control for the effect of prototypicality, this variable was always entered first as predictor ($R^2 = .490$, $F(1, 14) = 13.5$, $p = .003$), followed by the variable of interest in conjunction with the two-way interaction between prototypicality and the variable of interest. In addition, we separately analyzed the contributions of the single terms of the conjunction (variable of interest and interaction). When we did the analysis with stability, the conjunction of the predictor and the two-way interaction accounts for 38% of the variance, $\Delta R^2 = .377$, $\Delta F = 17.1$, $p < .001$. Furthermore, the separate analysis of the conjunctions’ components showed that stability accounts for about 22% of the variance, $\Delta R^2 = .221$, $\Delta F = 9.93$, $p = .008$, and, moreover, the two-way interaction explains 16% of the variation in liking, $\Delta R^2 = .157$, $\Delta F = 14.2$, $p = .003$. This significant interaction indicates that the effect of stability on liking depends on the prototypicality of the stimuli.

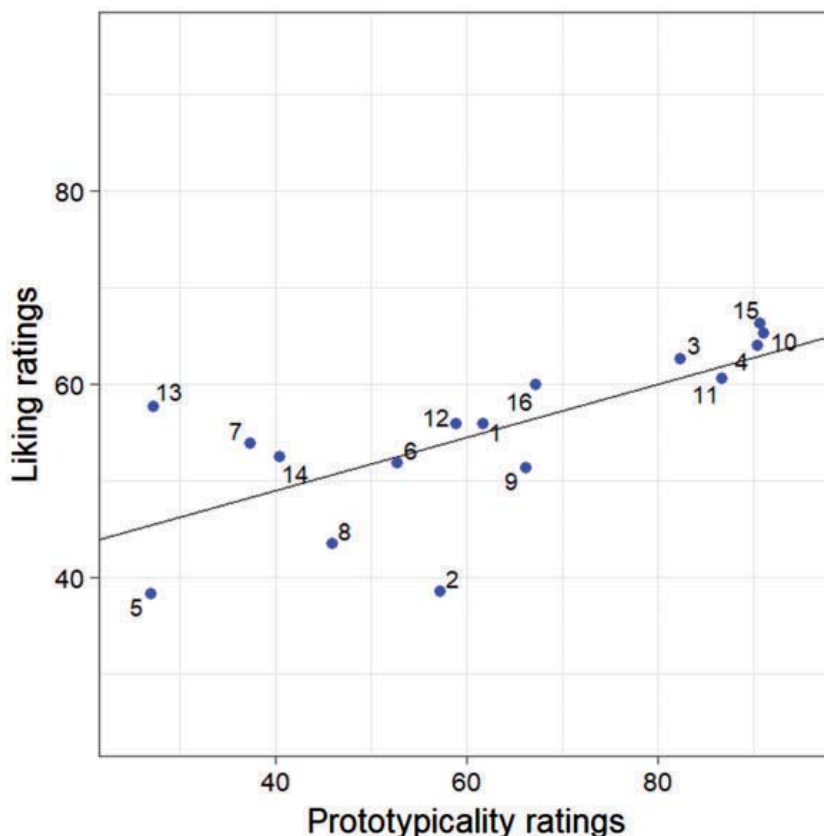


Figure 2. The relation between the mean prototypicality (Experiment 2) and mean liking ratings (Experiment 1) for 16 calligraphies (see Table 3).

To illustrate this interaction graphically, we used a median split to divide the calligraphies into two categories of more and less typical exemplars, respectively. We then computed the correlations between liking and stability separately for the two categories and plotted the data of each category together with their respective regression lines. As can be seen in Figure 3, if prototypicality was low, stability had a negative effect on liking ($r = -.829$, $p = .011$, 95% CI $[-0.968, -0.299]$). However, if it was high, stability had no significant effect on liking ($r = .462$, $p = .250$, 95% CI $[-0.360, 0.880]$).

In a further analogous regression analysis, we substituted stability by perceptual balance (Experiment 1). The overall effect of balance and its two-way interaction with prototypicality did not contribute significantly to liking, $\Delta R^2 = .072$, $\Delta F = 0.989$, $p = .400$, which also held for the individual components (balance: $\Delta R^2 = .043$, $\Delta F = 1.21$, $p = .291$; and Balance \times Prototypicality: $\Delta R^2 = .029$, $\Delta F = 0.787$, $p = .392$).

With the same schema of hierarchical regression, we tested whether the objective measures predict liking when controlled for prototypicality. For the measures APB, HG, MS, and FS, there were no effects. However, the DCM together with its interaction with prototypicality significantly predicts liking, $\Delta R^2 = .206$, $\Delta F = 4.05$, $p = .045$. To illustrate this interactive relation graphically, we

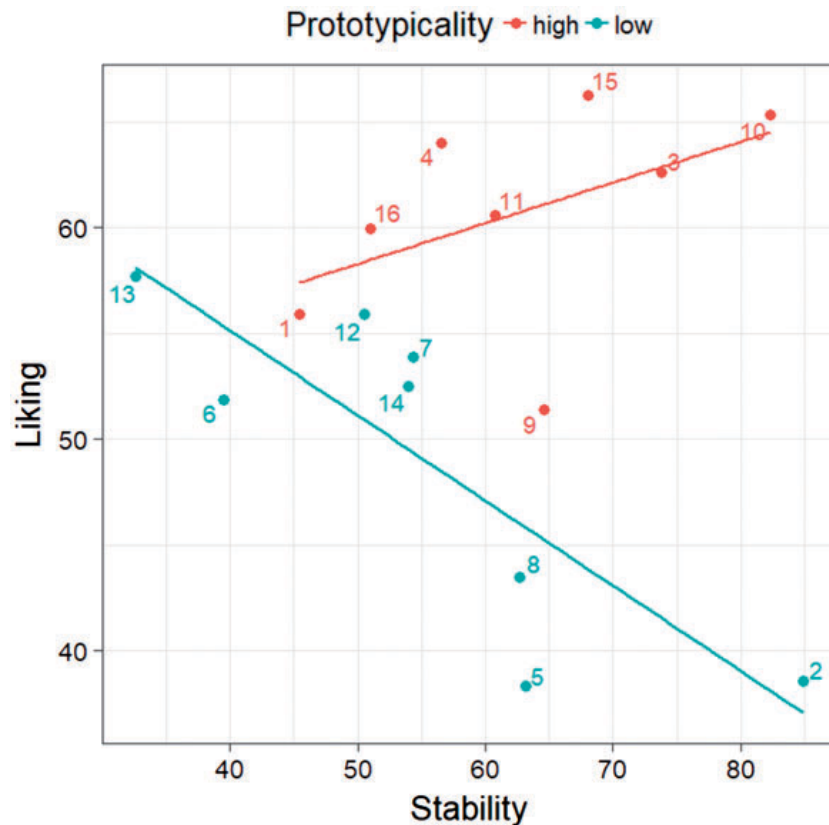


Figure 3. The interaction between prototypicality and stability for the prediction of liking (see Table 3).

computed the correlation between DCM and liking separately for typical and atypical calligraphies. As can be seen in Figure 4, for atypical calligraphies liking substantially increased with a decreasing DCM ($r = -.861$, $p = .006$, 95% CI $[-0.974, -0.397]$), whereas no significant relation was present for typical calligraphies ($r = -.550$, $p = .158$, 95% CI $[-0.904, 0.252]$).

Discussion

In this experiment, we collected ratings of stability (Liu et al., 2017; van der Helm, 2015) as well as of prototypicality (Rosch, 1975) for the Japanese calligraphies already used in our first experiment. Stability ratings were used to examine to what extent this dimension corresponds to the concept of perceptual balance for the calligraphies. Indeed, stability ratings correlated highly with the balance ratings from Experiment 1, which indicates that the two concepts are closely related. Accordingly, stability did not correlate with liking, similar to perceptual balance. However, this only held for their direct linear relation.

As expected, prototypicality ratings were highly correlated with the liking ratings from Experiment 1, which supports the idea that people prefer typical members of a category to less typical ones (Hekkert et al., 2003; Whitfield &

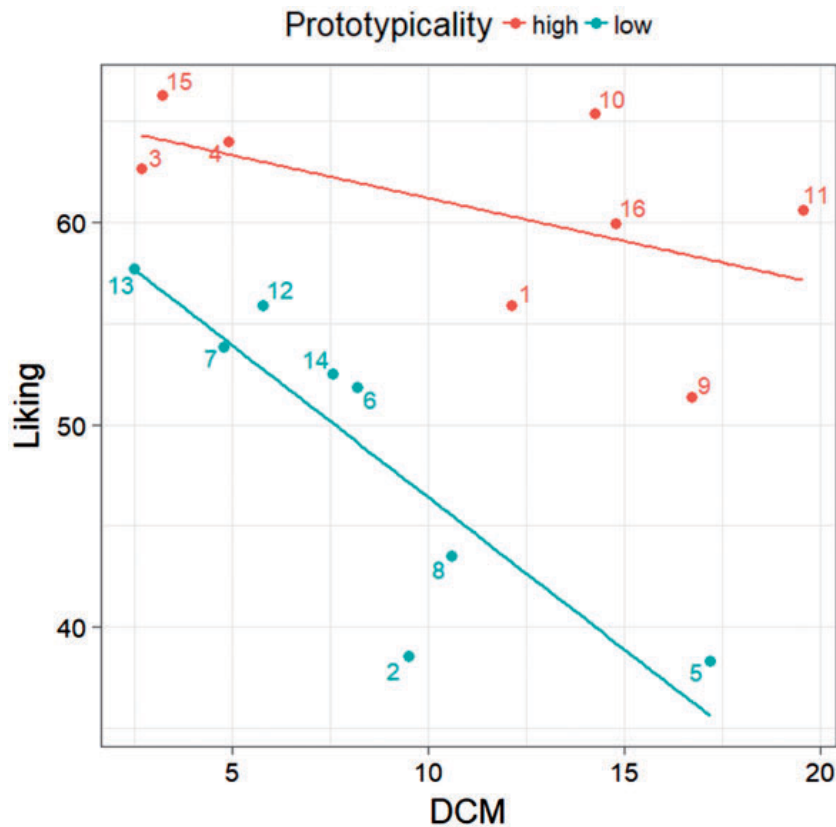


Figure 4. The interaction between prototypicality and DCM for the prediction of liking (see Table 3).

Slatter, 1979). Most importantly, though, this relation allowed us by means of hierarchical multiple regression to discount the effect of prototypicality on the liking ratings and to examine to what extent the remaining variance can be explained by the other variables. When we did this for stability, it turned out that the variable had a significant effect on liking. Moreover, there was also a reliable two-way interaction between prototypicality and stability indicating that stability had a negative effect on liking, but only for less typical calligraphies. For more typical exemplars, there was a trend in the opposite direction (see Figure 3).

For the balance ratings from our first experiment, an analogous hierarchical regression analysis revealed no significant effects on liking, even when the effect of prototypicality was controlled. This shows that the perception of stability and balance are similar but not identical, which indicates that they are based, at least partly, on different visual features.

For the different formal measures, hierarchical regression revealed no reliable relations with liking, except for the DCM, which in combination with its interaction with prototypicality significantly predicted liking. Further analyses revealed that for less typical calligraphies, liking decreased with an increasing DCM, that is, with an increasing distance of the center of mass from the

geometrical center of the picture. This relation was much weaker and unreliable for prototypical calligraphies (see Figure 4).

General Discussion

In this study, we investigated the proposed positive relation between the perceptual balance of a picture and its aesthetic appreciation (Arnheim, 1982; Bouleau, 1980; Kandinsky, 1926/1979). Up to now, this hypothesis has mainly been supported by a study with photographs (McManus et al., 2011), and a few studies with simple and specifically constructed stimuli (Hübner & Fillinger, 2016; Wilson & Chatterjee, 2005). In the latter studies, it has also been shown that computational measures of balance can be constructed that predict not only balance but also liking ratings. Gershoni and Hochstein (2011), however, applied one of such measures (APB) to Japanese calligraphies and failed to find a relation with balance ratings. This is an important result because it seems to demonstrate possible limits of the balance measures. Therefore, the first aim of this study was to replicate and extend Gershoni and Hochstein's result. Specifically, because these researchers presented their stimuli for a relatively short time and with spatial uncertainty, we investigated in our first experiment whether the results also occur with a longer stimulus presentation time and without spatial uncertainty. Furthermore, we wanted to apply different measures of balance as well as measures related to visual complexity. In addition, we collected liking ratings to examine the relation between balance and aesthetic appreciation for Japanese calligraphies. Because Gershoni and Hochstein (2011) used only 16 calligraphies, the statistical power of their tests was relatively small. Nonetheless, we used the same small set to ensure comparability between the original and our study.

As the results show, our balance ratings correlated highly with those from the original study, which demonstrates that balance perception is rather robust against procedural variations. Moreover, the balance ratings did not correlate with any of the balance measures. This indicates that the balance perceived when looking at calligraphies is unrelated to the concept of balance reflected by the applied formal measures. Finally and unexpectedly, our balance ratings, as well as the formal measures, did also not correlate with the liking ratings.

The results of our first experiment raised at least two questions. First, what does the concept of balance as assessed by our participants mean for calligraphies? Second, why was there no relation between balance and liking? For answering the first question, we inspected the order of pictures with respect to balance. Our impression was that perceived stability (Liu et al., 2017; van der Helm, 2015) might explain some of the variance. Therefore, in our second experiment, we asked our participants to assess the stability of the calligraphies.

As result, the stability ratings correlated highly with the balance ratings from Experiment 1. This suggested that, instead of assessing how well the picture

elements are balanced, the participants judged, at least to a large extent, how stable or flexible the calligraphies look like. Given this specific concept of balance, it is not surprising that the formal measures did not correlate with the ratings. Because stability did also not correlate with liking, this concept of balance seemed to be unrelated to aesthetic appreciation. However, these conclusions are premature. With respect to the absent correlation between balance and liking, we also hypothesized that a variable other than balance might strongly determine the aesthetic appreciation of Japanese calligraphies, which makes it difficult to detect a presumably relatively small effect of balance. As likely candidate, we considered prototypicality, which has shown to be related to aesthetic preference (Hekkert et al., 2003; Whitfield & Slatter, 1979). Therefore, in our second experiment, we additionally collected ratings for this property.

As expected, prototypicality was highly correlated with liking. This allowed us by means of hierarchical multiple regression to control the effects of prototypicality on liking. When we did this with respect to stability, we found a reliable relation between stability and liking. However, there was also a significant interaction between stability and prototypicality. Stability was related to liking only for stimuli rated as more atypical for Japanese calligraphies but not for more typical ones. Interestingly, an analogous analysis of the balance ratings from Experiment 1 revealed no reliable relation with liking. This indicates that despite their substantial correlation, stability and balance ratings also differ in some relevant aspect.

Discounting the effects of prototypicality also revealed an interesting relationship between the DCM and the liking ratings observed in Experiment 1. It turned out that the DCM scores affected only the liking ratings for more atypical members of the calligraphies.

Thus, the present approach demonstrates that it can be difficult to detect effects of a certain variable on liking ratings if the considered pictures not only vary on the corresponding dimension but also on other dimensions affecting aesthetic appreciation. In this case, the dimensions can compete and interact in a complex way, a phenomenon called “gestalt nightmare” by Makin (2017). However, in case such competing dimensions are known, their effects can be discounted, which allows analyzing the relation between the unexplained variance and the variable of interest. Thus, our results provide a positive example where the gestalt nightmare could be avoided.

Our conclusions are based on the between-participants data only to avoid influences of sequential effects between the two blocks of ratings. This reduced the statistical power of our tests. However, because the sequential effects were statistically significant, discarding the data of the respective second block was the only way to prevent biases and misinterpretations. Beyond that, the sequential interferences between the different ratings is an important observation that should be taken into account when designing future studies in this area.

Taken together, the results of the present study show that the meaning of balance can differ between different stimulus types. For the Japanese calligraphies used here, it seems that balance had a different meaning than the usual interpretation in art theory (Arnheim, 1982; Bouleau, 1980; Kandinsky, 1926/1979). It was neither related to the considered formal measures nor to the liking ratings. However, the related concept of stability was predictive for liking, even if only for the more atypical members of the calligraphies. For more prototypical members, stability had no effect on aesthetic appreciation. In future research, the concept of stability and its relation to other concepts of balance and to aesthetic appreciation should be investigated further. For instance, it would be interesting to know to which picture types the concept of visual stability applies. Until now, it is not clear which kind of pictorial arrangement is needed to evoke the perception of instability. Our results suggest that stability perception comes into play when different figural elements in a picture are connected rather than disconnected.

The fact that for more atypical calligraphies the DCM was related to liking suggests that this measure reflects a relatively general formal concept of balance, although it is obviously not always taken into account for the assessment of balance.

Authors' Note

The full data set for this article is accessible on <https://osf.io/urmbk/>.

Acknowledgments

The authors thank Sebastian Dummel for his methodical advice regarding the linear mixed-effects models.


Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Martin G. Fillinger  <http://orcid.org/0000-0001-9532-1699>

References

Arnheim, R. (1954). *Art and visual perception: A psychology of the creative eye*. Berkeley: University of California Press.

- Arnheim, R. (1982). *The power of the center: A study of composition in the visual arts*. Berkeley: University of California Press.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. doi:10.1016/j.jml.2007.12.005
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01
- Bouleau, C. (1980). *The painter's secret geometry: A study of composition in art*. New York, NY: Hacker Art Books.
- Donderi, D. C. (2006). An information theory analysis of visual complexity and dissimilarity. *Perception*, 35(6), 823–835. doi:10.1068/p5249
- Forsythe, A., Nadal, M., Sheehy, N., Cela-Conde, C. J., & Sawey, M. (2011). Predicting beauty: Fractal dimension and visual complexity in art. *British Journal of Psychology*, 102(1), 49–70. doi:10.1348/000712610X498958
- Gartus, A., & Leder, H. (2017). Predicting perceived visual complexity of abstract patterns using computational measures: The influence of mirror symmetry on complexity perception. *PLoS One*, 12(11), e0185276. doi:10.1371/journal.pone.0185276
- Gershoni, S., & Hochstein, S. (2011). Measuring pictorial balance perception at first glance using Japanese calligraphy. *i-Perception*, 2(6), 508–527. doi:10.1068/i0472aap
- Graf, L. K. M., & Landwehr, J. R. (2017). Aesthetic pleasure versus aesthetic interest: The two routes to aesthetic liking. *Frontiers in Psychology*, 8, 15. doi:10.3389/fpsyg.2017.00015
- Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1), 114–125. doi:10.1007/s40881-015-0004-4
- Hekkert, P., Snelders, D., & Van Wieringen, P. C. W. (2003). Most advanced, yet acceptable: Typicality and novelty as joint predictors of aesthetic preference in industrial design. *British Journal of Psychology*, 94(1), 111–124. doi:10.1348/000712603762842147
- Hübner, R., & Fillinger, M. G. (2016). Comparison of objective measures for predicting perceptual balance and visual aesthetic preference. *Frontiers in Psychology*, 7, 335. doi:10.3389/fpsyg.2016.00335
- Jacobsen, T. (2004). Individual and group modelling of aesthetic judgment strategies. *British Journal of Psychology*, 95(1), 41–56.
- Kandinsky, W. (1926/1979). *Point and line to plane*. New York, NY: Dover.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 26. doi:10.18637/jss.v082.i13
- Leiner, D. J. (2014). SoSci Survey (Version 2.5.00). Retrieved from <https://www.soscisurvey.de>
- Liu, J., Dong, W., Zhang, X., & Jiang, Z. (2017). Orientation judgment for abstract paintings. *Multimedia Tools and Applications*, 76(1), 1017–1036. doi:10.1007/s11042-015-3104-5
- Locher, P. J., Krupinski, E. A., Mello-Thoms, C., & Nodine, C. F. (2007). Visual interest in pictorial art during an aesthetic experience. *Spatial Vision*, 21(1/2), 55–77. doi:10.1163/156856808782713762

- Locher, P. J., & Nagy, Y. (1996). Vision spontaneously establishes the percept of pictorial balance. *Empirical Studies of the Arts*, 14(1), 17–31.
- Locher, P. J., & Stappers, P. J. (2002). Factors contributing to the implicit dynamic quality of static abstract designs. *Perception*, 31(9), 1093–1107. doi:10.1068/p3299
- Machado, P., Romero, J., Nadal, M., Santos, A., Correia, J., & Carballal, A. (2015). Computerized measures of visual complexity. *Acta Psychologica*, 160, 43–57. doi:10.1016/j.actpsy.2015.06.005
- Makin, A. D. J. (2017). The gap between aesthetic science and aesthetic experience. *Journal of Consciousness Studies*, 24(1-2), 184–213.
- McManus, I., Edmondson, D., & Rodger, J. (1985). Balance in pictures. *British Journal of Psychology*, 76(3), 311–324. doi:10.1111/j.2044-8295.1985.tb01955.x
- McManus, I., Stöver, K., & Kim, D. (2011). Arnheim's Gestalt theory of visual balance: Examining the compositional structure of art photographs and abstract images. *i-Perception*, 2(6), 615–647. doi:10.1068/i0445aap
- Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human preference. *Annual Review of Psychology*, 64, 77–107.
- R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Reips, U.-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology*, 49(4), 243–256. doi:10.1026//1618-3169.49.4.243
- Reips, U.-D. (2009). Internet experiments: Methods, guidelines, metadata. *Human Vision and Electronic Imaging XIV, Proceedings of SPIE*, 7240, 724008.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3), 192–233. doi:10.1037/0096-3445.104.3.192
- van der Helm, P. A. (2015). Symmetry perception. In J. Wagemans (Ed.), *Oxford handbook of perceptual organization* (pp. 108–128). Oxford, England: Oxford University Press.
- Whitfield, T. W. A., & Slatter, P. E. (1979). The effects of categorization and prototypicality on aesthetic choice in a furniture selection task. *British Journal of Psychology*, 70(1), 65–75. doi:10.1111/j.2044-8295.1979.tb02144.x
- Wilson, A., & Chatterjee, A. (2005). The assessment of preference for balance: Introducing a new test. *Empirical Studies of the Arts*, 23(2), 165–180.

Author Biographies

Martin G. Fillinger is a research associate at the Department of Psychology, Universität Konstanz. His work focuses on the connection between visual balance and aesthetic appreciation.

Ronald Hübner is a professor of Cognitive Psychology at the Department of Psychology, Universität Konstanz. His work focuses on empirical aesthetics and visual attention.

6 Research Paper III

Perceptual Balance, Stability, and Aesthetic Appreciation: Their Relations Depend on the Picture Type

Authors:

Ronald Hübner

Martin G. Fillinger

Authors' affiliation:

Department of Psychology, Universität Konstanz, Germany

Citation:

Hübner, R., & Fillinger, M. G. (2019). Perceptual Balance, Stability, and Aesthetic Appreciation: Their Relations Depend on the Picture Type. *i-Perception*, 10(3). doi:10.1177/2041669519856040

Perceptual Balance, Stability, and Aesthetic Appreciation: Their Relations Depend on the Picture Type

i-Perception

2019, Vol. 10(3), 1–17

© The Author(s) 2019

DOI: 10.1177/2041669519856040

journals.sagepub.com/home/ipe

**Ronald Hübner**  and **Martin G. Fillinger** 

Universität Konstanz, Germany

Abstract

It is widely assumed that the aesthetic appreciation of a picture depends, among others, on how well the picture's composition is perceptually balanced, where "perceptual balance" is often defined analogous to mechanics. To what extent this metaphor holds for different picture types, however, is still open. Therefore, in this study, we examined the relationship between balance, liking, and some objective measures with pictures from an aesthetic sensitivity test. These stimuli could be divided into single-element, multiple-element, and dynamic-pattern pictures. The results show that "balance" is interpreted differently, depending on the stimulus type. Whereas "mechanical" balance was applied to assess single-element pictures, the balance of multiple-element and dynamic-pattern pictures was rated more in the sense of gravitational stability. Only for the multiple-element stimuli, there was a positive relation between balance/stability and liking. Together, our findings show that there are different types of balance, and that their relation with liking depends on the picture type.

Keywords

aesthetics, preference, perceptual balance, complexity, perceptual stability

Date received 18 January 2019; accepted 17 May 2019

Introduction

In art theory and related fields, it is widely assumed that the aesthetic appreciation of a picture depends, among others, on how well it is balanced (Arnheim, 1983; Bouleau, 2014; Kandinsky, 1926/1979; Ross, 1907). To define perceptual balance, *mechanical balance* is

Corresponding author:

Ronald Hübner, Universität Konstanz, Fachbereich Psychologie, Fach D29, D-78457 Konstanz, Germany.

Email: ronald.huebner@uni-konstanz.de



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<http://www.creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

often used as metaphor. It is assumed that each element in a picture has a certain visual “weight” depending on its features like size, shape, and color (Arnheim, 1954). The weight exerts a corresponding perceptual force that increases with its distance from the picture’s “center of gravity.” Accordingly, a heavy weight located on one side of the fulcrum can be balanced by a lighter weight positioned further away on the other side. Usually, a picture is considered as balanced if all its elements are arranged in such a way that their perceptual forces are in equilibrium about a fulcrum and if the fulcrum coincides with the center of the picture (Locher, 2006).

By adopting the mechanical metaphor, the relative perceptual weight of pictorial elements can easily be assessed, at least within a simple context. One merely has to ask persons to adjust the horizontal position of a target element on a seesaw so that it is in equilibrium to a fixed element opposite to a fulcrum. This was one of the first methods applied in experimental research on perceptual balance (Pierce, 1894).

In recent years, even formal measures of perceptual balance have been developed based on the mechanical metaphor. One such measure is the *Assessment of Preference for Balance* (APB), developed by Wilson and Chatterjee (2005). These researchers assumed that the perceptual weight of each pixel in a picture is inversely related to its gray level, that is, dark pixels are heavier than bright ones. In addition, they divide a picture into four symmetrical areas around the horizontal, the vertical, and the two diagonal axes, respectively. The differences between the summed weights in opposite areas are then computed, and the mean of the eight differences is taken as the picture’s balance score. Wilson and Chatterjee (2005) have shown that APB scores not only can predict balance ratings but are also related to aesthetic appreciation.

A measure even more closely related to the mechanical metaphor is the *Deviation of the Center of “Mass”* (DCM) from the picture’s geometrical center (Hübner & Fillinger, 2016; McManus, Stöver, & Kim, 2011). Based on the pixels’ gray level, the center of perceptual mass (or the center of gravity) is computed analogously to mechanics. It is further assumed that the less this center deviates from the geometric center of the picture, the more the picture is perceived as balanced and liked. Hübner and Fillinger (2016) have shown that for the pictures used by Wilson and Chatterjee (2005), the DCM predicts balance and liking ratings similarly well as the APB. Averaged across the pictures, the DCM scores explained up to 68% of the variance for the balance ratings, and up to 86% of the variance for the liking ratings.

It should be noted, though, that these strong predictions are mainly due to the specific pictures applied in Hübner and Fillinger (2016) and in Wilson and Chatterjee (2005). They included only homogeneous elements with a simple shape and with an identical gray level. Such pictures have the advantage that balance can be varied strongly without affecting other characteristics. However, the obtained results say little about how far they can be generalized for more complex images. Some results (McManus et al., 2011; Thömmes & Hübner, 2018) suggest that the correlation between perceptual balance and aesthetic appreciation is much less for photographs. Thömmes and Hübner (2018), for instance, analyzed about 700 architectural photographs posted on Instagram by different photographers. For photographs representing a three-dimensional scene, they found that the scores (DCM and APB) correlated significantly with the number of Instagram Likes. However, the explained variance was only about 10%. Although this percentage is small, it must be taken into account that balance is only one of many factors usually determining aesthetic appreciation. Therefore, for more complex artwork, one cannot expect the same large correlations between balance and liking as for specifically constructed simple pictures.

Gershoni and Hochstein (2011) found that for Japanese calligraphies the APB even completely failed to predict perceptual balance ratings. Recently, Fillinger and Hübner (2018) replicated this result and observed similar negative results also for the DCM. However, they further showed that for these pictures perceptual balance ratings were completely unrelated with liking ratings.

Two conclusions can be drawn from these results. First, the absent relation between balance ratings and the balance measures indicates that persons sometimes apply concepts of perceptual balance that are not reflected by the APB and DCM measures. Fillinger and Hübner (2018), for instance, provided some evidence that under certain conditions balance is interpreted more in the sense of stability. Already Pierce (1894) observed that balance is mainly applied for the horizontal arrangement of elements, whereas for vertical arrangements stability plays a greater role. For instance, pictures were preferred when they had more weight in their lower part rather than in their upper half. Second, because the aesthetic appreciation of more complex pictures is usually determined by multiple factors, the effect of perceptual balance can be relatively small or even absent. In case these factors are known, it can be helpful to discount their effects. Fillinger and Hübner (2018), for instance, observed that prototypicality strongly determined the aesthetic appreciation of Japanese calligraphies. After taking this factor into account, the DCM showed again a significant relation with liking, but only for less prototypical calligraphies.

The aim of the present study was to further investigate the relations between the formal measures, perceptual balance, and aesthetic appreciation. For this objective, we used pictures from the *Visual Aesthetic Sensitivity Test* (VAST; Götz, 1981), developed by Götz, Borisy, Lynn, and Eysenck (1979). They consist of various configurations of different element types and, therefore, are more complex than those applied by Hübner and Fillinger (2016) and Wilson and Chatterjee (2005), and are more heterogeneous than the Japanese calligraphies used by Gershoni and Hochstein (2011). Because stimulus complexity is an important factor for aesthetic appreciation (Leder, Belke, Oeberst, & Augustin, 2004; Palmer, Schloss, & Sammartino, 2013; Tinio & Leder, 2009), we also applied a corresponding objective measure.

In two experiments, we collected three different ratings for each of the pictures: perceptual balance, liking, and stability. In our first experiment, we could not find a significant correlation between balance and liking ratings. However, liking was strongly related to complexity. After discounting this effect, a reliable relationship between balance and liking ratings emerged.

However, our objective measures of perceptual balance did still not show a significant correlation with the ratings. Given these results, we reasoned that for VAST images there might be a different concept of perceptual balance that is not related to our objective measures of balance. Therefore, in our second experiment, we considered an alternative concept of balance.

Experiment I

In this experiment, we wanted to further investigate the relationship between perceptual balance, aesthetic appreciation, and balance-related measures. As stimulus set we used pictures from the VAST (Götz, 1981). This test consists of 42 pairs of nonrepresentational gray-level pictures. The pictures in each pair are the same, except that one is better configured than the other. Usually, the pictures are used for measuring aesthetic sensitivity. The person under examination has to decide which picture of each pair is the “correct” one. For the present objective, though, we applied only the 42 correct pictures as stimuli, and our participants had to rate each picture with respect to liking and perceptual balance.

The VAST (Götz, 1981) pictures are more complex and more diverse than those applied in several former studies on perceptual balance (e.g., Fillinger & Hübner, 2018; Gershoni & Hochstein, 2011; Hübner & Fillinger, 2016; Wilson & Chatterjee, 2005), but are still less complex than most of real artworks. To take the variation in complexity into account, we also wanted to have objective scores for this property. An often used measure in this respect is the Kolmogorov complexity, which is defined as the length of the shortest program that can describe an item (Donderi, 2006), and can be approximated by the jpeg compression algorithm (Faloutsos & Megalooikonomou, 2007). Therefore, we used the corresponding file sizes of the images to construct complexity scores.

We expected that the balance scores reliably predict perceptual balance ratings as well as liking ratings. Furthermore, in addition to balance, complexity should also affect liking (e.g., Tinio & Leder, 2009).

Method











































Participants. Fifty-two persons (16 male, mean age 24.6 years, $SD = 7.95$) participated in the rating task and received a 3-€ voucher as incentive. They were recruited via a local online system (ORSEE; Greiner, 2015). This study was carried out in accordance with the ethical guidelines of the Universität Konstanz, which is based on the Declaration of Helsinki. Participants were informed of their right to abstain from participation in the study or to withdraw consent to participate at any time without reprisal.

Stimuli. As stimuli, we used the 42 correct pictures from the VAST (Götz, 1981). Corresponding images were created by digitizing (264×330 pixels) a paper version of the VAST, which was ordered from our library. The copyright for the works of Karl Otto Götz are held by VG Bild-Kunst, Germany. Thumbnails of the images can be seen in Table 1. The images were positioned at the center of the screen on a gray background. Stimulus presentation and response registration, which occurred exclusively online, were controlled by the SoSci Survey System (Leiner, 2016).

Objective Measures. We computed the APB and DCM scores for each picture as objective measures of balance; see Hübner and Fillinger (2016) for details of calculation. The DCM scores ranged from 1.82 to 26.9 ($M = 9.85$, $SD = 6.56$), and the APB scores from 5.53 to 47.0 ($M = 21.8$, $SD = 8.62$). As measure of complexity, we used the ratio between the file size of the jpeg-compressed image and that of the uncompressed image (Palumbo, Ogden, Makin, & Bertamini, 2014), where a value close to 1 indicates a very complex image and a value close to 0 a very simple one. In our stimulus set, complexity varied between 0.08 and 0.54 ($M = 0.25$, $SD = 0.13$).

Procedure. The experiment, which was carried out exclusively online, started with an instruction that informed the participants about the task (the instruction in German and its English translation are provided in Appendix B). In addition, we used a seriousness check (Reips, 2009) to control for participants' involvement in the task. Next, there were two blocks in each of which the 42 stimuli were presented in random order. In the first block, the participants had to rate how much they liked the stimuli (from *I do not like it* to *I like it*), and in the second block how well the stimuli were balanced (from *not balanced* to *balanced*). Internally, the scale ranged from 0 to 100 (values not visible for the participants). The ratings were entered by clicking with the mouse on a continuous slider. Immediately after each response,

Table 1. Results of Experiment 1.

														
#	33	24	16	4	39	26	11	30	42	5	20	36	41	12
C	0.54	0.47	0.46	0.46	0.46	0.42	0.41	0.40	0.40	0.37	0.35	0.35	0.35	0.33
L	<u>63</u>	<u>57</u>	<u>58</u>	<u>63</u>	<u>61</u>	<u>60</u>	<u>62</u>	<u>54</u>	<u>59</u>	<u>55</u>	<u>54</u>	<u>60</u>	<u>54</u>	<u>54</u>
B	58	52	56	46	73	46	68	44	55	37	37	48	43	34
														
#	10	19	7	22	8	37	15	3	23	29	18	9	25	34
C	0.31	0.30	0.25	0.24	0.22	0.22	0.20	0.19	0.19	0.19	0.18	0.17	0.17	0.17
L	<u>59</u>	<u>57</u>	<u>31</u>	<u>59</u>	<u>62</u>	<u>38</u>	<u>40</u>	<u>57</u>	<u>32</u>	<u>40</u>	<u>32</u>	<u>50</u>	<u>42</u>	<u>46</u>
B	38	57	39	75	68	60	40	72	37	56	33	48	56	58
														
#	38	14	2	1	35	40	17	28	27	31	6	21	32	13
C	0.16	0.16	0.16	0.15	0.14	0.13	0.12	0.11	0.11	0.11	0.10	0.10	0.10	0.08
L	<u>45</u>	<u>36</u>	<u>52</u>	<u>48</u>	<u>38</u>	<u>37</u>	<u>29</u>	<u>27</u>	<u>30</u>	<u>27</u>	<u>22</u>	<u>27</u>	<u>25</u>	<u>27</u>
B	53	45	61	54	46	36	53	33	64	45	45	59	71	76

The thumbnails of the VAST pictures, shown in the continuing first row, are ordered by complexity. The corresponding scores are given in row 3(C). The bold numbers in row 2(#) represent the corresponding picture numbers. Mean ratings for liking (underlined) and balance are shown in the third (L) and fourth (B) row, respectively. (For the works of Karl Otto Götz: © VG Bild-Kunst, Bonn 2017.)

the next stimulus was displayed. Altogether, the experiment comprised 84 trials (2 × 42 trials) and lasted about 10 minutes.

Results

All 52 participants passed the seriousness check. Mean balance and liking ratings were 51.8 ($SD = 27.4$) and 45.9 ($SD = 28.2$), respectively. The means across participants for each picture are shown in Table 1. Correlational analyses revealed that overall there was no significant correlation between the two ratings (see Table 2). Concerning the relations with the objective measures, there was merely a significant and strong correlation between liking and complexity, indicating that the more complex a picture was, the more it was liked.

To get an idea in what sense complexity varies across the VAST pictures, we ordered the pictures accordingly. As can be seen in Table 1, complexity seems to be largely determined by the number of elements in a picture. Moreover, the order suggests that there are three categories of pictures. The 10 pictures with the least complexity contain only one element, whereas the following 17 pictures of medium complexity are composed of multiple elements. The remaining 15 pictures of highest complexity consist of countless elements, and also represent some dynamics or implied motion.

The relation between complexity and liking is also shown in Figure 1, where the three suggested picture categories are indicated by specific shapes and colors of the corresponding

Table 2. Correlations^a Between Stability, Balance, and Liking Ratings (Means Across Stimuli) and Objective Measures Separated for Every Visual Category.

		Balance ₂	Liking	Stability	Complexity	APB	DCM
Balance ₁	Overall	0.928***	0.154	0.524***	-0.048	0.108	-0.265
Balance ₂	(df = 40)	–	0.086	0.601***	-0.136	0.059	-0.305
Liking		0.086	–	-0.535***	0.809***	-0.261	0.274
Stability		0.601***	-0.535***	–	-0.639***	0.262	-0.236
Balance ₁	Dynamic	0.921***	0.592*	0.582*	0.223	0.232	-0.200
Balance ₂	pattern	–	0.597*	0.489*	0.172	0.124	-0.214
Liking	(df = 15)	0.597*	–	0.406	0.486*	0.084	0.148
Stability		0.489*	0.406	–	-0.055	0.325	0.152
Balance ₁	Multiple	0.951***	0.844***	0.669**	0.186	-0.258	-0.328
Balance ₂	elements	–	0.805***	0.799***	0.268	-0.311	-0.319
Liking	(df = 13)	0.805***	–	0.664**	0.236	-0.179	-0.267
Stability		0.799***	0.664**	–	0.298	-0.248	-0.033
Balance ₁	Single	0.922***	-0.228	0.802**	-0.541	0.384	-0.625
Balance ₂	element	–	-0.066	0.731*	-0.411	0.377	-0.754*
Liking	(df = 8)	-0.066	–	-0.464	0.830**	0.151	-0.044
Stability		0.731*	-0.464	–	-0.593	0.466	-0.313

Note. *df* = degrees of freedom; APB = Assessment of Preference for Balance; DCM = Deviation of the Center of “Mass.”

^aFor many of the correlations in this table we had a specific hypothesis. Therefore, we did not correct for multiple testing. When interpreting other significant correlations, however, one should be aware of the problem of inflating false-positive rates with multiple testing.

* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

data points. In view of this result, an interesting further question was to what extent the relations between the ratings and measures differ between the categories. However, before we tried to answer this question, we first wanted to test whether the assumed categories are common sense, and if so, exactly which pictures are assigned to each category.

Categorization Check. In order to examine how people spontaneously categorize the VAST pictures, we conducted a supplementary visual categorization study. The task of the participants was to sort each picture into one of three nonlabeled categories. The detailed method and results are provided in Appendix A. Here, it is sufficient to know that the majority of participants confirmed our overall categorization. Merely four of the pictures (nos. 40, 14, 8, and 9) were categorized differently relative to the categorization shown in Figure 1 (see Figure 2). In view of this result, we decided to name the three empirically validated categories “single element,” “multiple elements,” and “dynamic pattern,” respectively.

The mean ratings for pictures in these categories are: single element (balance 53.7, $SD = 13.5$; liking 28.8, $SD = 4.95$), multiple elements (balance 52.4, $SD = 13.0$, liking 43.7, $SD = 9.36$), and dynamic pattern (balance 50.0, $SD = 11.6$, liking 57.9, $SD = 3.84$). One-way ANOVAs revealed that the balance ratings did not differ between the categories, $F(2, 39) = 0.291$; $p = 0.749$; $\eta^2 = 0.015$. However, liking differed significantly, $F(2, 39) = 62.9$; $p < 0.001$; $\eta^2 = 0.763$.

The correlations between the ratings and measures for each individual category are shown in Table 2. As can be seen, single-element pictures were not only liked least, but their aesthetic appreciation was also independent of perceptual balance. In contrast, for pictures with

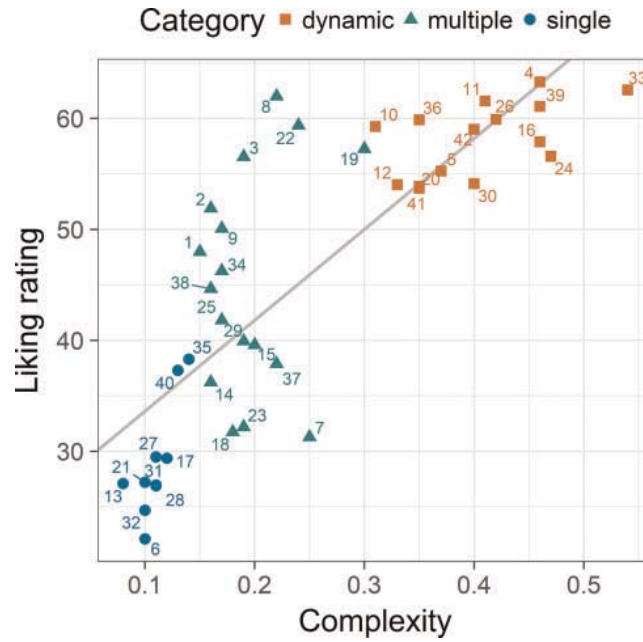


Figure 1. Relation between complexity (ratio between the file size of the jpeg-compressed image and that of the uncompressed image) and liking ratings in Experiment I and the corresponding linear regression line. Each data point represents the picture according to its number (see Table I). The colors and shapes of the symbols indicate the corresponding preliminary category.

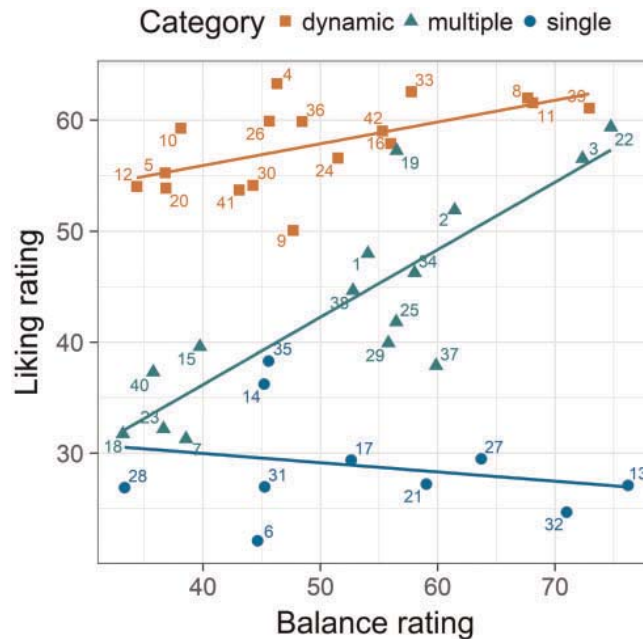


Figure 2. The relation between balance ratings and liking ratings in Experiment I for the different picture categories obtained in the supplementary categorization study. Each data point represents the picture according to its number (see Table I). The colors and shapes of the symbols indicate the corresponding category. The lines are the corresponding regression lines

multiple elements and for those with dynamic patterns, there was a significant correlation between liking and perceptual balance. The more the pictures were perceived as balanced, the more they were liked, where the relation was more pronounced for the multiple-element pictures.

Finally, it should be noted that, although there was a strong correlation between liking and complexity for the whole stimulus set, within categories such a correlation was reliable only for the dynamic-pattern and single-element pictures (see Table 2).

Discussion

In this experiment, we investigated the relationship between perceptual balance, aesthetic appreciation, and related objective measures. As stimuli, we presented pictures from the VAST (Götz, 1981; Götz et al., 1979). A first data analysis revealed that, overall, balance ratings did not correlate with liking ratings. However, liking strongly depended on the complexity of the pictures. Overall, the more complex a picture was, the more it was liked, which is in line with former results (e.g., Jacobsen, 2004; Tinio & Leder, 2009). Moreover, when we ordered the pictures by complexity, it became obvious that complexity increased with the number of elements in a picture. Already Berlyne (1971) observed such a relation (see also Nadal, Munar, Marty, & Cela-Conde, 2010). Interestingly, visual inspection suggested that the VAST pictures can easily be divided into three categories: single-element, multiple-element, and dynamic-pattern pictures. In a supplementary categorization study (Appendix A), we confirmed and refined our preliminary categorization.

The categorization of the pictures offered the possibility to analyze the relations between the variables separately for each category. As a result, we found that perceptual balance and liking correlated significantly for multiple-element and dynamic-pattern pictures, but not for single-element ones. This indicates that the effect of perceptual balance on liking depends on the picture type. More specifically, balance seems to be an important factor for liking only for pictures with a sufficient number of elements or complexity.

However, despite these positive relations, the formal measures of balance (APB and DCM) were unrelated to perceptual balance and liking. This held across all pictures as well as within each category and indicates that the concept of balance reflected by these measures is different from that applied by our participants.

Concerning complexity, despite the strong overall correlation with liking, within categories, it was reliable only for dynamic-pattern and single-object pictures. If we consider the order of these pictures with respect to complexity (Table 1), then it seems that complexity increases with the number of edges, which is in line with Berlyne (1971), who also observed that irregularity of shape affects complexity. Recently, Friedenberg and Bertamini (2015) showed that the preference increases with contour length and the number of concavities.

Taken together, the results of this experiment demonstrate that the relationship between balance and liking depends on the picture type. Furthermore, the fact that the balance ratings were unrelated to the formal measures of balance, even for those picture categories where balance and liking were correlated, shows that a different concept of perceptual balance was applied. But which property was used by our participants for assessing the balance of a picture? If one inspects the multiple-element pictures (Table 1), then it seems that at least in some pictures rated as well balanced, the elements are piled up in a more stable way than in those rated as less balanced. Even for the dynamic-pattern pictures, it appears that the most balanced ones look more stable than the less balanced ones.

Therefore, we hypothesized that stability could have been applied by our participants as concept of balance. To test whether this was the case, we conducted a further experiment, in which we collected stability ratings for the VAST pictures.

Experiment 2

In the previous experiment, we observed that perceptual balance correlated for some picture types with liking, but was generally unrelated to the formal measures of balance. This indicates that the participants applied some concept of balance that was different from that reflected by the measures. A possible alternative in this respect might be stability. As mentioned in the Introduction, already Pierce (1894) observed that perceptual stability is a variant of balance, especially for vertically arranged elements in a picture. More recently, Friedenbergs (2012) even proposed a *perceptual instability* hypothesis, stating that objects perceived as more fragile are less attractive.

Perhaps our participants rated the balance of the VAST pictures, at least the more complex ones, by assessing the compositions' stability. In order to test whether balance was indeed interpreted as stability in Experiment 1, we asked the participants in this experiment to rate stability directly. If our supposition is correct, then stability ratings should highly correlate with balance as well as with liking ratings, at least for multiple-element and dynamic-pattern pictures.

Because the participants in Experiment 1 rated balance after they had assessed the aesthetic appreciation of the pictures, there might have been some carryover effects from liking to balance ratings. Therefore, in the present experiment, we again collected balance ratings for the VAST pictures, but this time from an independent sample of participants.

Method

Altogether, 104 (20 male, mean age 24.0 years, $SD = 5.10$) persons participated in the online rating tasks. As incentive, each participant had the chance to win one of fifty 3-€ vouchers. The experiment was similar to the first one, except that half of the participants (9 male, mean age 24.2 years, $SD = 5.72$) rated the picture with respect to balance (from *not balanced* to *balanced*), whereas the other half (11 male, mean age 23.8 years, $SD = 4.45$) rated how stable the composition of the pictures was (from *unstable* to *stable*). The instruction is provided in Appendix B. Each task lasted about 5 minutes.

Results

All 104 participants passed the seriousness check. The mean balance rating was 53.0 ($SD = 11.1$), and the mean stability rating 50.8 ($SD = 13.2$). A one-way ANOVA revealed that there was again no significant difference between the balance ratings for the different picture categories (single element: $M = 55.5$, $SD = 11.4$; multiple elements: $M = 54.8$, $SD = 12.1$; dynamic pattern: $M = 50.0$, $SD = 9.78$), $F(2, 39) = 1.08$; $p = 0.351$; $\eta^2 = 0.052$. Moreover, the balance ratings were rather similar to those in Experiment 1 ($r(40) = 0.928$, $p < 0.001$). Accordingly, the present balance ratings correlated similarly strong as the previous balance ratings with the liking ratings from Experiment 1 (see Table 2).

This time, the DCM scores correlated significantly with the balance ratings, but only for single-element pictures. The corresponding correlation in Experiment 1 was already relatively high, but shortly failed to reach significance. Because the correlation between liking and the present balance ratings is reduced, compared to Experiment 1, the increased correlation with the DCM scores could indicate that the former balance ratings were indeed biased to some extent.

Other than the balance ratings, stability ratings differed significantly between the categories (single element: $M = 65.5$, $SD = 9.22$; multiple elements: $M = 53.1$, $SD = 10.9$; dynamic pattern: $M = 40.3$, $SD = 6.29$), $F(2, 39) = 26.4$; $p < 0.001$; $\eta^2 = 0.575$. Accordingly, the range

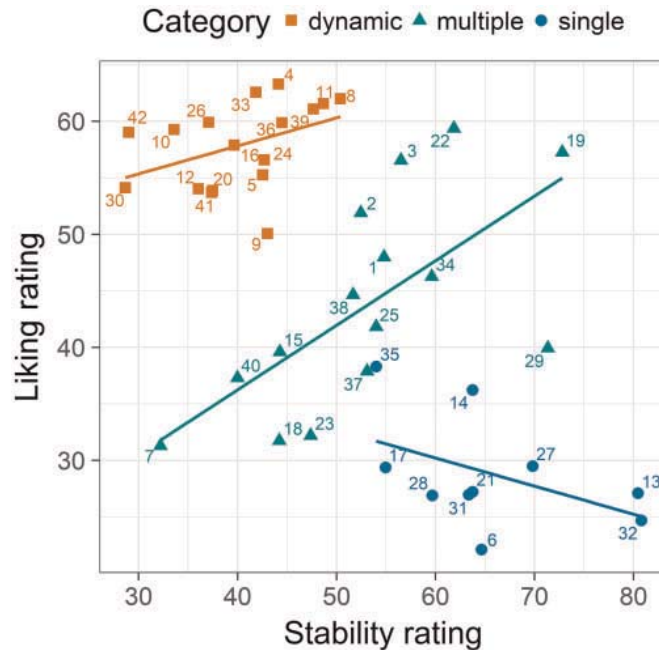


Figure 3. The relations between stability ratings and liking ratings from Experiment 1 for the categories obtained in the supplementary categorization study. Each number represents 1 of the 42 stimuli. The colors indicate the stimulus categories. The lines represent the regression lines.

of stability between the categories differed (Figure 3). Dynamic-pattern pictures were rated low and single-element stimuli high in stability, whereas the multiple-element pictures covered almost the entire range of stability. Together, the different ranges produced a negative overall correlation between stability and liking (Table 2). Given this result, it is not surprising that stability also correlated negatively with complexity. Thus, complex stimuli were rated low in stability. Nevertheless, they were much liked.

Analyses of the relations within the different stimulus categories show that for multiple-element pictures there was a significant positive relation between stability and liking (Table 2). The more stable the pictorial arrangement was rated, the more it was liked. However, multiple regression revealed that the stability ratings do not explain variance that is not already explained by balance, $\beta = .056$, $t(12) = 0.196$, $p = 0.848$. This indicates that at least for multiple-element pictures, balance and stability ratings were based on a rather similar concept.

Discussion

In this experiment, we collected stability and balance ratings for the VAST pictures. Our aim was to test whether the balance ratings in Experiment 1 were accomplished by assessing the pictures' stability. If that were the case, then stability ratings should not only be highly correlated with the balance ratings but also with the liking ratings for multiple-element and dynamic-pattern pictures. Such a result would explain why the formal balance measures (DCM and APB) did not correlate with the balance ratings in the previous experiment. They simply reflect a different concept of balance.

If we consider the results, then it is obvious that our hypothesis was only partially confirmed. Although balance and stability ratings correlate significantly, their relation was far from perfect. Both variables share only about 36% of their variance, which indicates that the

two concepts are related but not identical. This conclusion is also confirmed by the negative overall correlation between stability and liking. The dynamic patterns, which were liked most, were rated as relatively unstable. In contrast, single-element pictures, which were liked least, were rated as most stable. Only the multiple-element pictures cover the whole range of stability. Moreover, for these pictures, stability was positively related to liking, as was the case for balance. For single-element pictures as well as for the dynamic-pattern pictures liking was independent of stability.

Because of the within-participants design in Experiment 1, we speculated that there could have been carryover effects from liking to balance ratings. Therefore, we again collected balance ratings, but this time with a between-participants design. The high correlation between the two balance ratings indicates that carryover effects, if present at all, were rather small. Nevertheless, the independent balance ratings for the single-element pictures now correlated negatively with the DCM scores, which means that the pictures were perceived as more balanced the less their center of “mass” deviated from their geometrical center.

Taken together, the results of this experiment demonstrate that balance and stability ratings for the VAST pictures are related, but not identical. Across all pictures, there was a negative correlation between stability and liking, whereas balance was unrelated to liking. If one considers the relations within each picture type, then it seems that the participants performed the rating tasks differently, depending on the stimulus type. If pictures contained multiple elements, then balance was interpreted in the same way as stability. Accordingly, both variables were positively correlated with liking. However, for the other stimulus types, balance ratings and stability ratings were conceptualized and performed differently.

General Discussion

The objective of this study was to further investigate the relation between perceptual balance and aesthetic appreciation. It has widely been assumed that the aesthetic appreciation of a picture depends on how well it is balanced, where balance has often been defined analogously to mechanical balance (e.g., Arnheim, 1954; Pierce, 1894; Puffer, 1903). Recently, even objective balance measures have been proposed, like the DCM (Hübner & Fillinger, 2016; McManus et al., 2011), or the APB (Wilson & Chatterjee, 2005), which more or less also rely on the mechanical-balance metaphor. Although these measures have successfully been applied to predict balance and liking ratings, the considered pictures were mostly relatively simple. Therefore, an important question is to what extent these concepts and measures are also valid for more complex pictures. Up to now, the provided evidence in this area is equivocal.

Whereas in some studies the measures could successfully predict the aesthetic appreciation of various photographs (McManus et al., 2011; Thömmes & Hübner, 2018), there are also negative results. Gershoni and Hochstein (2011), for instance, examined Japanese calligraphies and found no correlation between APB scores and balance ratings. In a recent study with the same stimuli, we could show that if prototypicality is taken into account, then at least the DCM is related to liking, but only for less prototypical calligraphies (Fillinger & Hübner, 2018).

These results demonstrate that the relationships between aesthetic appreciation, perceptual balance, and liking are still largely unknown, especially for more complex pictures. Therefore, for the present experiments, we used pictures from the VAST (Götz, 1981; Götz et al., 1979). These stimuli are more complex than those applied in the studies that

introduced the formal balance measures, but are still less complex than photographs or most of real artworks.

In Experiment 1, the participants had first to rate how much they liked the VAST pictures and then to assess how well these pictures are balanced. As a result, across all pictures, there was no correlation between these two variables. This is similar to our result with Japanese calligraphies (Fillinger & Hübner, 2018). However, we found that the liking of the VAST pictures depended strongly on complexity, which is in line with other results (e.g., Berlyne, 1971; Tinio & Leder, 2009). When we inspected the pictures more closely, then it became clear that complexity was largely determined by the number of elements in the pictures (see also Jacobsen, 2004; Nadal et al., 2010). Even more, the VAST pictures could be divided into single-element, multiple-element, and dynamic-pattern pictures, the latter with countless elements. This categorization was confirmed and refined in a supplementary categorization study (Appendix A).

Therefore, in a next step we analyzed the relations between ratings and scores separately for each category and found that for single-element pictures, which were liked least, perceptual balance had no effect on liking. However, for dynamic-pattern pictures, which were liked most, liking increased significantly with perceptual balance. The numerically largest correlation, though, occurred for the multiple-element pictures, which makes sense, because balancing should play some role if a medium number of elements are present in a picture. These findings are in line with the principle of *unity-in-variety* (Fechner, 1876) and the related idea that complexity must go along with order for aesthetic appreciation (Van Geert & Wagemans, 2019). Accordingly, balance affects liking only when an image is sufficiently complex, which, in our case, is the case for dynamic-pattern and multiple-element pictures, but not for single-element pictures. Thus, the results from our first experiment demonstrate that perceptual balance has a positive effect on aesthetic appreciation, but only for certain types of pictures.

Despite these positive relations, there were no reliable correlations between the ratings and the applied formal measures of balance, which indicates that the participants used a concept of balance that was different from that reflected by the APB and DCM scores. A closer look at the multiple-element pictures suggested that visual (gravitational) stability could be a promising candidate as an alternative concept of balance. That stability affects liking, especially for vertically arranged picture elements, has already been proposed by Pierce (1894). Stability could even have been relevant for the dynamic-pattern and single-element pictures.

To test whether our participants indeed rated the balance of the pictures by assessing their stability, we conducted a second experiment where we asked our participants directly to rate the stability. As a result, stability correlated significantly with balance, but not very high. Moreover, across all stimuli, stability was negatively correlated with liking. This was due to the fact that, different from balance, the range of stability differed systematically between the different stimulus categories. The most complex pictures, that is, the dynamic patterns, which were liked most, were rated as least stable, whereas the least complex (single-element) pictures, which were liked least, were judged as the most stable. Within categories, stability was related to liking only for the multiple-element pictures. For these stimuli, the correlation was positive, as expected, that is, the more stable a configuration, the more it was liked. Because a similar relation held for the balance ratings, it can be assumed that for this picture type balance was interpreted in the sense of stability and vice versa. In any case, balance was not interpreted in the sense reflected by the formal balance measures.

In Experiment 2, we again collected balance ratings, but this time with a between-participants design. As a result, for the single-element pictures, there was a reliable

correlation between the balance ratings and the DCM scores, which indicates that pictures were rated as more balanced if the center of mass was closer to the geometric center of the picture.

Thus, it seems that the concept of stability was not used consistently. Across all stimuli, the corresponding ratings correlated negatively with liking, whereas they correlated positively within the multiple-element pictures. How can this discrepancy be explained? Because stability should actually have a positive effect on aesthetic appreciation (Friedenberg, 2012; Pierce, 1894), we think that the negative overall correlation is due to a confound with other variables. The result that the dynamic patterns were liked most is probably not due to their low stability, but to the fact that in this specific case instability goes along with implied motion, which is usually liked. Interestingly, for Japanese calligraphies, we observed a similar result (Fillinger & Hübner, 2018). At least atypical calligraphies were liked the more, the less stable they were. As shown by Dubal et al. (2014), dynamic brushstrokes in calligraphies convey emotions (Dubal et al., 2014). Thus, it seems that instability has no negative effects on liking, if it results from a dynamic that implies motion associated with positive emotion.

Moreover, the result that single-element pictures were liked least is presumably not due to their great stability, but rather to their restricted variety.

Thus, it seems that three types of perceptual balance/stability occurred in our study. For assessing the balance of the single-element pictures, the participants applied mechanical balance, that is, the deviation of the center of “mass” from the geometrical center, as reflected by the DCM. However, this balance was unrelated to liking. For rating the balance of the multiple-element pictures, the gravitational stability of the configuration was assessed, which was positively associated with liking. Finally, the low stability ratings of the dynamic patterns were presumably not due to their perceived low gravitational stability, but to their dynamics and implied motion. The corresponding high emotional expressivity led these pictures to be liked most.

Taken together, our study shows that perceptual balance and aesthetic appreciation are related in a complex way. How balance is interpreted and assessed depends largely on the content of the picture. Moreover, other factors such as complexity can be dominant or at least modulate the relation between balance/stability and liking. The more factors are known and, therefore, can be discounted, the better the pure effect of balance can be isolated.

Authors' Note

The data of both experiments are available at DOI OSF.IO/4BUQW.

Declaration of Conflicting Interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Ronald Hübner  <https://orcid.org/0000-0002-2788-6214>

Martin G. Fillinger  <https://orcid.org/0000-0001-9532-1699>

References

- Arnheim, R. (1954). *Art and visual perception: A psychology of the creative eye*. Berkeley and Los Angeles: University of California Press.
- Arnheim, R. (1983). *The power of the center: A study of composition in the visual arts*. Berkeley: University of California Press.
- Berlyne, D. E. (1971). *Aesthetics and psychology*. New York, NY: Appleton-Century-Crofts.
- Bouleau, C. (2014). *The painter's secret geometry: A study of composition in art*. New York, NY: Hacker Art Books.
- Donderi, D. C. (2006). An information theory analysis of visual complexity and dissimilarity. *Perception*, 35, 823–835. doi:10.1068/p5249
- Dubal, S., Lerebours, A.-E., Taffou, M., Pelletier, J., Escande, Y., & Knoblauch, K. (2014). A psychophysical exploration of the perception of emotion from abstract art. *Empirical Studies of the Arts*, 32, 27–41. doi:10.2190/EM.32.1.EOV.4
- Faloutsos, C., & Megalooikonomou, V. (2007). On data mining, compression, and Kolmogorov complexity. *Data Mining and Knowledge Discovery*, 15, 3–20. doi:10.1007/s10618-006-0057-3
- Fechner, G. T. (1876). *Vorschule der Ästhetik*. Leipzig, Germany: Breitkopf & Härtel.
- Fillinger, M. G., & Hübner, R. (2018). The relations between balance, prototypicality, and aesthetic appreciation for Japanese calligraphy. *Empirical Studies of the Art*. Advance online publication. doi:10.1177/0276237418805656
- Friedenberg, J. (2012). Aesthetic judgment of triangular shape: Compactness and not the golden ratio determines perceived attractiveness. *i-Perception*, 3, 163–175. doi:10.1068/i0484
- Friedenberg, J., & Bertamini, M. (2015). Aesthetic preference for polygon shape. *Empirical Studies of the Arts*, 33, 144–160. doi:10.1177/0276237415594708
- Gershoni, S., & Hochstein, S. (2011). Measuring pictorial balance perception at first glance using Japanese calligraphy. *i-Perception*, 2, 508–527. doi:10.1068/i0472
- Götz, K. O. (1981). *VAST: Visual Aesthetic Sensitivity Test*. Düsseldorf, Germany: Concept Verlag.
- Götz, K. O., Borisy, A. R., Lynn, R., & Eysenck, H. J. (1979). A new visual aesthetic sensitivity test: I. Construction and psychometric properties. *Perceptual and Motor Skills*, 49, 795–802.
- Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1, 114–125. doi:10.1007/s40881-015-0004-4
- Hübner, R., & Fillinger, M. G. (2016). Comparison of objective measures for predicting perceptual balance and visual aesthetic preference. *Frontiers in Psychology*, 7. doi:10.3389/fpsyg.2016.00335
- Jacobsen, T. (2004). Individual and group modelling of aesthetic judgment strategies. *British Journal of Psychology*, 95, 41–56. doi:10.1348/000712604322779451
- Kandinsky, W. (1926/1979). *Point and line to plane* (H. Dearstyne and H. Rebay, Trans.). New York, NY: Dover.
- Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, 95(4), 489–508. doi:10.1348/0007126042369811
- Leiner, D. J. (2016). SoSci Survey (Version 2.6.00). Retrieved from <https://www.sosicisurvey.de>
- Locher, P. (2006). Experimental scrutiny of the role of balance in the visual arts. In P. Locher, C. Martindale, & L. Dorfman (Eds.), *New Directions in Aesthetics, Creativity and the Arts* (pp. 19–32). Amityville, NY: Baywood Publishing Company.
- McManus, I., Stöver, K., & Kim, D. (2011). Arnheim's Gestalt theory of visual balance: Examining the compositional structure of art photographs and abstract images. *i-Perception*, 2, 615–647. doi:10.1068/i0445aap
- 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, 173–191. doi:10.2190/EM.28.2.d
- Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human preference. *Annual Review of Psychology*, 64, 77–107. doi:10.1146/annurev-psych-120710-100504
- Palumbo, L., Ogden, R., Makin, A. D., & Bertamini, M. (2014). Examining visual complexity and its influence on perceived duration. *Journal of Vision*, 14, 3.

- Pierce, E. (1894). Studies from the Harvard Psychological Laboratory (II): The aesthetics of simple forms. I. Symmetry. *The Psychological Review*, *1*, 483–495. doi:10.1037/h0069000
- Puffer, E. D. (1903). Studies in symmetry. *The Psychological Review: Monograph Suppl. 4 (Harvard Psychology Studies 1)*, 467–539.
- Reips, U.-D. (2009). Internet experiments: Methods, guidelines, metadata. *Proceedings of SPIE Human Vision and Electronic Imaging XIV*, 7240, 724008. doi:10.1117/12.823416
- Ross, D. W. (1907). *A theory of pure design: Harmony, balance, rhythm*. Boston, MA: Houghton, Mifflin.
- Thömmes, K., & Hübner, R. (2018). Instagram likes for architectural photos can be predicted by quantitative balance measures and curvature. *Frontiers in Psychology*, *9*. doi:10.3389/fpsyg.2018.01050
- Tinio, P. P. L., & Leder, H. (2009). Just how stable are stable aesthetic features? Symmetry, complexity, and the jaws of massive familiarization. *Acta Psychologica*, *130*, 241–250. doi:10.1016/j.actpsy.2009.01.001
- Van Geert, E., & Wagemans, J. (2019). Order, complexity, and aesthetic appreciation. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. doi:10.1037/aca0000224
- Wilson, A., & Chatterjee, A. (2005). The assessment of preference for balance: Introducing a new test. *Empirical Studies of the Arts*, *23*, 165–180. doi:10.2190/B1LR-MVF3-F36X-XR64

How to cite this article

Hübner, R., & Fillinger, M. G. (2019). Perceptual Balance, Stability, and Aesthetic Appreciation: Their Relations Depend on the Picture Type. *i-Perception*, *10*(3), 1–17. doi:10.1177/2041669519856040

Appendix A

The supplementary study reported here served to confirm and refine our preliminary categorization of the VAST pictures based on complexity and the number of elements.

Method

Twenty-one persons (12 male, mean age 25.7 years, $SD = 7.80$) from the Universität Konstanz participated and received goodies for their contribution. None of the persons had participated in Experiment 1. The task was performed under the same ethical standards as the main experiment.

Stimuli and Procedure

The digitized 42 pictures from the VAST (Götz, 1981) were printed on white cardboards (12×15 cm). The participants were first asked to spread all cards in front of them and then to sort the 42 cards into three categories by the pictures' features within approximately 1 minute. Subsequently, they were interviewed about the features they had used for categorization. In total, the session lasted about 5 minutes.

Results and Discussion

Fifteen of the 21 participants used the same stimulus features. Given the resulting categories, we assigned each picture to the category chosen most. A multipattern χ^2 -Test (3 categories \times 42 stimuli) revealed that the distribution of frequencies across categories was not

random, $\chi^2(82) = 906.90$, $p < 0.001$, $V = 0.188$. The assignments were slightly different from our preliminary ones. Picture #40 was sorted into the single-element category instead of the multiple-element category. For Picture #14 it was the other way round. Furthermore, Picture #8 and Picture #9 were assigned to the dynamic-pattern category instead of the multiple-element category.

In the single-element category, the APB score ranged from 18.7 to 38.8 ($M = 25.2$, $SD = 6.22$), the DCM score from 1.52 to 13.8 ($M = 5.65$, $SD = 3.98$), and complexity from 18 to 38 ($M = 26.5$, $SD = 5.56$). Moreover, the APB score ranged from 5.64 to 34.5 ($M = 22.6$, $SD = 8.21$), the DCM score from 3.20 to 26.4 ($M = 10.0$, $SD = 6.50$), and complexity from 32 to 71 ($M = 45.8$, $SD = 10.4$) for multiple-element pictures. Finally, in the dynamic-pattern category, the APB score ranged from 5.53 to 47.0 ($M = 19.3$, $SD = 9.76$), the DCM score from 5.47 to 68.3 ($M = 14.9$, $SD = 15.2$), and complexity from 41 to 127 ($M = 90.5$, $SD = 21.7$).

The pictures in the resulting single-element category were described by the participants as, for example, “clear shapes consisting of a big black area,” “single coherent elements,” or “chunky elements.” The multiple-element pictures were described as “several black elements,” “single shapes scattered all over,” or “multiple elements.” The dynamic-pattern category was described, for instance, as “abstract art,” “complex patterns made by brushstrokes,” or “movements.” The other six participants also used the dynamic-pattern category, but divided the other pictures into two categories that can be labeled as “curved elements” and “angular elements.”

Thus, by and large, this study confirmed our preliminary categorization. The majority of the participants used the number of elements and the picture’s dynamics for their categorization.

Appendix B

In this part, we provide the instructions (originally in German) and their English translations.

Experiment 1

Liking ratings. On the following pages, you will see simple pictures. Your task is to rate these images according to certain criteria. The survey consists of two parts. In the first part, you should rate how much you like the pictures. Make your decisions spontaneously! After you have decided, set the slider between “I do not like it” and “I like it” and click on “Next” to go to the next page. Since these are subjective assessments, there are no right or wrong answers.

[In German (original): Auf den folgenden Seiten werden Sie einfache Bilder sehen. Ihre Aufgabe besteht darin, diese Bilder nach gewissen Kriterien zu beurteilen. Die Befragung besteht aus zwei Teilen. Im ersten Teil sollen Sie beurteilen, wie sehr Ihnen die Bilder gefallen. Treffen Sie Ihre Entscheidungen ganz spontan! Nachdem Sie sich entschieden haben, stellen Sie den Regler zwischen “gefällt mir gar nicht” und “gefällt mir sehr gut” ein und klicken auf “Weiter” um zur nächsten Seite zu gelangen. Da es sich um subjektive Einschätzungen handelt, gibt es keine richtigen oder falschen Antworten.]

Balance ratings. In the second part of the survey, you will rate the images with regard to balance. Set the slider between “not balanced” and “balanced.”

[In German (original): Im zweiten Teil der Befragung sollen Sie die Bilder nun in Hinsicht auf Balance beurteilen. Stellen Sie hierzu den Regler zwischen “nicht balanciert” und “balanciert” ein.]

Experiment 2

Stability ratings. On the following pages, you will see simple pictures. Your task is to rate how stable the arrangement of the picture elements appears to you. Make your decisions spontaneously! After you have decided, set the slider between “unstable” and “stable” and click on “Next” to go to the next page. Since these are subjective assessments, there are no right or wrong answers.

[In German (original): Auf den folgenden Seiten werden Sie einfache Bilder sehen. Ihre Aufgabe besteht darin, zu beurteilen, wie stabil Ihnen die Anordnung der Bildelemente erscheint. Treffen Sie Ihre Entscheidungen ganz spontan! Nachdem Sie sich entschieden haben, stellen Sie den Regler zwischen “instabil” und “stabil” ein und klicken auf “Weiter” um zur nächsten Seite zu gelangen. Da es sich um subjektive Einschätzungen handelt, gibt es keine richtigen oder falschen Antworten.]

Balance ratings. On the following pages, you will see simple pictures. Your task is to rate them in terms of balance. Make your decisions spontaneously! After you have decided, set the slider between “not balanced” and “balanced” and click on “Next” to go to the next page. Since these are subjective assessments, there are no right or wrong answers.

[In German (original): Auf den folgenden Seiten werden Sie einfache Bilder sehen. Ihre Aufgabe besteht darin, diese in Hinsicht auf Balance zu beurteilen. Treffen Sie Ihre Entscheidungen ganz spontan! Nachdem Sie sich entschieden haben, stellen Sie den Regler zwischen “nicht balanciert” und “balanciert” ein und klicken auf “Weiter” um zur nächsten Seite zu gelangen. Da es sich um subjektive Einschätzungen handelt, gibt es keine richtigen oder falschen Antworten.]

7 Research Paper IV

On the relation between perceived stability and aesthetic appreciation

Authors:

Martin G. Fillinger

Ronald Hübner

Authors' affiliation:

Department of Psychology, Universität Konstanz, Germany

Citation:

Fillinger, M. G., & Hübner, R. (2019). *On the relation between perceived stability and aesthetic appreciation*. Unpublished manuscript, Department of Psychology, University of Konstanz, Germany.

7.1 Abstract

Perceived stability is an important feature of pictures with respect to their aesthetic appreciation. Pictures whose composition is perceived as gravitationally stable are usually liked more than those with unstable arrangements. However, there are exceptions. In a recent study, we found that unstable Japanese calligraphies were preferred to stable ones. From this result, we derived the hypothesis that instability is liked when it implies movement, which, presumably, evokes emotionality. Here, we systematically tested these two types of instability. In our first experiment, where we used multiple-element pictures of varying stability as stimuli, we showed that perceived gravitational instability has a negative effect on aesthetic appreciation. In a second experiment, we used paintings from the artist K.O. Götz, which largely varied in implied movement. The paintings were rated with respect to liking, emotionality, movement and stability. Our results show that movement is responsible for the positive effect of instability on aesthetic appreciation. After we discounted the effect of movement, instability and liking were again negatively correlated. We assume that instability related to movement is liked because movement is positively correlated with emotionality. Taken together, our findings indicate that perceived instability reduces the aesthetic appreciation of a picture unless it implies movement.

7.2 Introduction

How well a picture is liked usually depends on various factors. Some of them are related to the picture's perceptual features, others to its semantic content. An example of the first type is *perceptual balance*. The importance of balance has been proposed by art theorists (e.g., Arnheim, 1982; Bouleau, 1980; Kandinsky, 1926/1979; Ross, 1907) and intensively investigated by empirical researchers (e.g., Locher, 2006; Locher, Jan Stappers, & Overbeeke, 1998; McManus et al., 1985). Ross (1907), for example, considers balance as part of harmony. He writes that a balanced arrangement of elements "is a Harmony of Positions due to the coincidence of two centers, the center of the attractions and the center of the framing" (p. 24). Whereas the geometric center of a frame can easily be located, it is unclear how the center of attractions can be determined. A widely used approach in this respect is to use mechanical balance as a metaphor. It is assumed that each element in a picture has a certain perceptual weight that depends on its low-level features such as color, size and form, as well as on its semantic content (e.g., Arnheim, 1982). However, the problem remains to calculate the weights. In the simplest case, the weight of an element is determined by its area and gray level. The center of the attraction (perceptual weights) is then computed analogously to the center of mass in mechanics.

The mechanical approach is also used to compute objective measures of balance. For instance, for the DCM (Deviation of the Center of Mass), a measure proposed by Hübner and Fillinger (2016), it is assumed that the perceptual weight of each pixel is related to its gray level. The center of mass in a picture can then easily be computed analogously to mechanics (McManus et al., 2011). The DCM score, i.e., the degree of balance is finally defined, as suggested by Ross (1907), by the distance of the center of mass from the geometric center of the picture. A similar measure is the APB (Assessment of Preference for Balance), proposed by Wilson and Chatterjee (2005). It is also assumed that the weight of each pixel is related to its gray level. The APB score is then defined as the average of eight symmetry measures over the four axes of a picture (horizontal, vertical and the two diagonals).

At least for simple pictures that included only basic (e.g., circles, or squares) and unrelated geometrical elements, these measures have successfully been applied to predict balance ratings and liking. Moreover, Thömmes and Hübner (2018) analyzed about 700 architectural photographs and found that for those depicting a real scene (with a 3D

appearance), the scores (DCM and APB) significantly correlated with the number of Instagram likes.

However, there are also negative results. Gershoni and Hochstein (2011), for instance, used Japanese calligraphies as stimuli and found that the APB failed to predict balance ratings. Recently, Fillinger and Hübner (2018) replicated this result. Moreover, they showed that for these pictures, balance ratings were also unrelated to liking ratings. This suggests that there are different types of balance. Indeed, further data collection and analyses revealed that the liking of calligraphies was affected by perceived stability, which is considered as different from but closely related to balance (Ross, 1907).

If we go back in history, then, as far as we know, the difference between balance and stability has first been examined systematically by Pierce (1894, 1896), who at that time was a graduate student in Hugo Münsterberg's lab at Harvard University. In his studies, Pierce wanted to investigate perceptual balance by asking participants to adjust the horizontal position of a movable object (e.g., a line) on one side of a display such that this side was aesthetically equal to the opposite side, where another object (e.g., line of different size) was fixed at a certain position. At least under some conditions, adjustments were made in accord with mechanical balance. Interestingly, however, Pierce also rotated the displays by 90° and found that in this case, the adjustments changed. For vertical layouts, stability was more important than balance. Accordingly, pictures were preferred when they had more weight in their lower half rather than in their upper one. Thus, it seems that pictures are preferred whose elements are arranged in a gravitationally stable way. This is in line with more recent research by Friedenber (2012), who observed that triangular shapes perceived as unstable (e.g., because they stood on one of the edges instead of one of the baselines), were rated as less attractive.

Thus, going back, one may ask the question why, if stability usually affects liking positively (i.e., the higher the stability the more a picture is liked), this has been reversed in our previous study on the Japanese calligraphies, where the less stable ones were liked more? Alternatively, more generally, why does instability increase liking in some pictures and decrease it in others? One possible explanation is that there are (at least) two different types of stability. One type is gravitational stability (van der Helm, 2015), which is preferred because it follows laws of gravitation thus preventing damage and injuries. Due to corresponding associations, this preference is also generalized to the content of pictures.

However, what is the other type? We hypothesized that it is in some way related to perceived movement and dynamics. In Chinese art theory, for instance, it is assumed that a brushstroke expresses the painter's emotion, which also holds for calligraphies (Dubal et al., 2014). Thus, expressive brushstrokes represent dynamics and imply movement, while at the same time they may appear unstable. Nonetheless, due to the implied movement (Osaka et al., 2010), which usually evokes positive emotions, the aesthetic appeal of the corresponding pictures is high.

This conjecture is also supported by another of our recent studies (Hübner & Fillinger, 2019), where we used pictures from the Visual Aesthetic Sensitivity Test (VAST; Götz, 1985) as stimuli. The VAST pictures can be categorized into single-element, multiple-element and dynamic pictures. We found that overall, there was a negative correlation between stability and liking. In particular, the dynamic pictures were rated as highly unstable but were nevertheless liked most. The multiple-element pictures were liked less and were rated, on average, as more stable. Interestingly, within this category, there was a positive relation between stability and liking.

Taken together, the results presented thus far suggest that at least two types of instability can be differentiated: one type is associated with gravitation and is disliked, whereas the other type is associated with movement and liked.

The aim of the present study was to systematically investigate these two types of instability. In our first experiment, we show that gravitational instability in multiple-element pictures has a negative effect on aesthetic appreciation. In our second experiment, we demonstrate that instability is liked for dynamic pictures due to its implied movement.

7.3 Experiment 1

In our first experiment, we wanted to show that perceived gravitational instability has a negative effect on aesthetic appreciation. For this objective, we constructed a basic set of four pictures, each showing three rectangles and three colored decorative elements (see Figure 1). Although these basic stimuli already differed in stability, this property was further varied by rotating the stimuli. Participants had to rate balance and stability for each of these stimuli, and they were also asked to rate how much they liked the pictures.

We expected similar results as in Hübner and Fillinger (2019) for multiple-element pictures. Specifically, balance and stability should be positively correlated. Most importantly, however, participants should prefer balanced and stable pictures.

7.3.1 Method

7.3.1.1 Participants

Eighty-seven persons (26 males, mean age 22.4, $SD = 5.67$) were recruited via an online system (ORSEE; Greiner, 2015) for participation in the online experiment. All participants received a 3 € voucher as an incentive. The study was carried out in accordance with the ethical guidelines of the Universität Konstanz and the Declaration of Helsinki. Participants were informed of their right to abstain from participation in the study or withdraw consent to participate at any time without reprisal.

7.3.1.2 Stimuli

Stimuli were 32 pictures (500 x 500 pixels) composed of three black rectangles, whose formations were more or less stable. Additionally, the pictures contained three small decorative colored elements (for examples see Table 1). The stimulus set was based on four basic stimuli (#1, 9, 17, and 25 in Table 1) that were thought to differ in stability. Additional stimuli were created by rotating (30, 90, 150, 180, 210, 270, and 330 degrees) these stimuli. The stimuli were presented at the center of the display on a gray background. Stimulus presentation and response registration were controlled by SoSci Survey (Leiner, 2019).

For the stimuli we also computed the objective balance measures (APB and DCM). The APB ranged from 43.2 to 51.1 ($M = 47.3$, $SD = 2.18$), and the DCM from 3.77 to 12.8 ($M = 7.89$, $SD = 3.08$).

7.3.1.3 Procedure

The online experiment started with an instruction that informed the participants about the task. Subsequently, two blocks of trials were administered. In the first block, all participants rated how much they liked the stimuli (from “I do not like it” to “I like it”). The second block differed between the participants. About half of the participants (44) rated how well the stimuli were balanced (from “not balanced” to “balanced”), whereas the remaining participants (43) rated how stable the pictorial elements were arranged (from “unstable” to “stable”). The ratings were entered by clicking on a visual

analogue scale. Shortly after each response, the next stimulus was displayed. Altogether, the experiment comprised 64 trials (2×32 trials) and lasted about 7 minutes.

7.3.2 Results

Mean ratings for liking, balance and stability were 44.0 ($SD = 4.75$), 47.8 ($SD = 14.3$), and 46.9 ($SD = 18.3$), respectively. The mean ratings for each picture are presented in Table 1. Correlational analyses revealed significant correlations between all rated properties (see Table 2). As expected, and most important, there was a positive relation between stability and liking, which is also shown in Figure 1. Concerning the objective measures, there were merely significant correlations of balance and liking with the DCM scores.

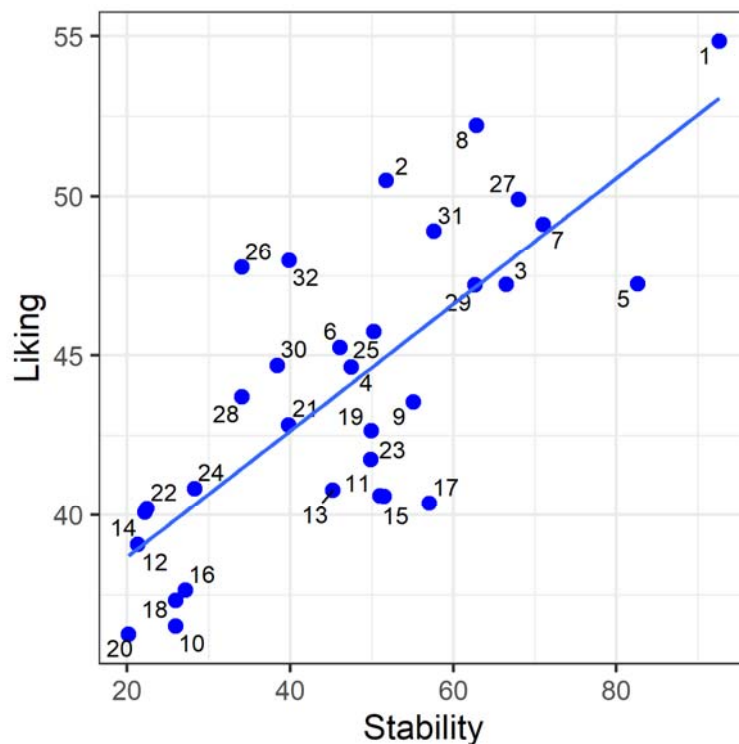


Figure 1. Relation between stability and liking ratings in Experiment 1. The line represents the corresponding linear regression line. Each data point represents the picture according to its number (see Table 1).

Table 1: Results of Experiment 1.

#	1	5	7	27	3	8	29	31
S	93	83	71	68	66	63	63	58
L	55	47	49	50	47	52	47	49
B	83	78	68	61	64	59	58	55
#	17	9	2	15	11	25	19	23
S	57	55	52	52	51	50	50	50
L	40	44	50	41	41	46	43	42
B	43	43	58	45	50	58	43	40
#	4	6	13	32	21	30	28	26
S	47	46	45	40	40	38	34	34
L	45	45	41	48	43	45	44	48
B	54	55	40	46	37	50	47	48
#	24	16	18	10	22	14	12	20
S	28	27	26	26	22	22	21	20
L	41	38	37	37	40	40	39	36
B	32	34	27	31	28	34	32	27

Note. The thumbnails of the 32 stimuli, shown in the first row, are ordered by stability. Row two (#) shows the identification number of the pictures. The corresponding mean ratings for stability (S), liking (L) and balance (B) are listed in the rows below.

Table 2. Correlations between mean ratings for liking, balance and stability in Experiment 1. The correlations between the mean ratings and objective measures of balance (APB and DCM) are also shown.

	Balance	Stability	APB	DCM
Liking	.866***	.763***	-.311	-.386*
Balance	-	.912***	-.250	-.365*
Stability	-	-	-.122	-.145

Note. * $p < .05$, *** $p < .001$; APB = Assessment of Preference for Balance; DCM = Deviation of the Center of Mass.

7.3.3 Discussion

In this experiment, we used multi-element stimuli of varying stability to investigate the relationship of balance, stability and liking. As expected, stability positively correlated with liking ratings (see Figure 1). The same held for balance and liking. Thus, our results support the notion that pictures showing balanced and gravitationally stable compositions are preferred (Friedenberg, 2012; Pierce, 1896). Accordingly, there was also a strong relationship between balance and stability, which replicates results from our previous studies (Fillinger & Hübner, 2018; Hübner & Fillinger, 2019). The fact that this latter correlation was also quite high indicates that the concepts of balance and stability are rather similar. However, it is also clear from our results that they nevertheless differ: Whereas the DCM scores significantly correlated with balance, the correlation with stability was considerably lower and not significant. Moreover, balance correlated much lower with the DCM scores than with stability. This could mean that the balance ratings reflect the usual mechanical balance as well as gravitational stability.

Interestingly, the DCM scores also significantly correlated with liking. This is different from the Japanese calligraphies (Fillinger & Hübner, 2018) and means that a picture was liked more the closer the center of mass was located to the picture's geometric center. The fact that the DCM correlated with both balance and liking is different from our previous results (Fillinger & Hübner, 2018). The APB scores did not correlate with any rating.

Taken together, the results support our hypothesis that, in multiple-element pictures, gravitational stability is related to perceived balance, but nevertheless is also different from it. Moreover, it is positively related to liking.

7.4 Experiment 2

In the previous experiment with multi-element pictures, instability was not liked. These results are in line with art theory (Liu et al., 2017), as well as with early (Pierce, 1896) and more recent (Friedenberg, 2012; Hübner & Fillinger, 2019) experiments. However, the result is different from that observed for Japanese calligraphies. For these stimuli, instability was preferred (Fillinger & Hübner, 2018; Hübner & Fillinger, 2019). In the present experiment, we not only intended to replicate this latter result with a different type of stimuli, but also to examine further the reasons for the diverging outcomes. In order to achieve this aim, we used artwork by Karl Otto Götz (1914–2017), who was one of the most important members of the German Art Informel movement. Götz is known for his explosive and complex abstract forms consisting of a mixture of organic and geometric elements.

A selection of Götz's paintings was rated for perceived stability and liking, but also for movement and emotionality. We expected that for these stimuli, stability was negatively related to liking and that this relation was depending on movement and emotionality. If instability is related to implied movement, then it is presumably liked, because movement is associated with positive emotions (Dubal et al., 2014).

7.4.1 Method

Before describing the method of the main experiment, we report a preliminary study that was conducted to select an appropriate set of pictures.

7.4.1.1 Preliminary study

For the preliminary study, we selected 100 paintings by Karl Otto Götz, which he created between 1936 and 2010. Forty-two persons (9 males; mean age 22.9, $SD = 2.70$) were recruited for the study that was performed under the same ethical standards as the previous experiment. They received a 3 € voucher for their participation. The procedure was similar to the previous experiment. After a short instruction, gray-level pictures of the 100 paintings were presented in random order. Two visual analogue scales located below the picture were used to enter stability (from “unstable” to “stable”) and movement (from “static” to “dynamic”) ratings. There was no time limit. Overall, the study lasted about 10 minutes.

The mean stability and movement ratings were 45.3 ($SD = 7.10$) and 41.9 ($SD = 20.8$), respectively. The two ratings were strongly negatively correlated ($r = -.870, p < .001$), as expected. For the main study, we selected a representative set of 44 pictures (for example stimuli see Table 3; the 44 selected paintings are accessible on <https://osf.io/ebdah/>) that almost equally covered the range of perceived stability as well as perceived movement. For the selected set, the correlation between stability and movement ratings was $r = -.896, p < .001$.

7.4.1.2 Roundness

Because the selected stimuli (especially the more dynamic ones) are rounder than Japanese calligraphies (Fillinger & Hübner, 2018) or those from Experiment 1, and roundness usually affects liking strongly (e.g., Gómez-Puerto, Munar, & Nadal, 2016; Silvia & Barona, 2009), we wanted to control for roundness. This is important to prevent that pictures with a high movement rating are liked more simply because they are rounder, which we then may falsely interpret as caused by movement. Therefore, we conducted a small study in which fifteen participants (4 males; mean age 24.3, $SD = 4.8$, range 20-40) merely had to rate roundness versus angularity for the selected set of stimuli. Roundness did not correlate with stability, $r = -.109, p < .480$, nor with movement, $r = -.137, p < .375$.

7.4.1.3 Participants

Ninety-two persons (18 males; mean age 24.7, $SD = 6.30$) were recruited in the same way as in the previous experiments for participation in the online experiment. None of them had participated in the preliminary study. All other participation criteria were the same as in Experiment 1.

7.4.1.4 Procedure



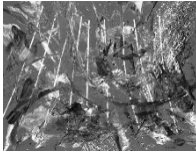
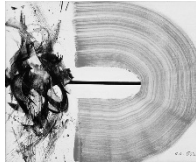
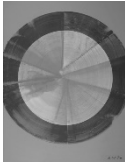
The participants were randomly assigned to four groups of 23 persons each. One group started with liking ratings in the first block, followed by movement ratings (from “static” to “dynamic”) in a second block. Another group started with liking ratings, followed by stability ratings. A third group started with emotionality ratings, followed by movement ratings and a final group started with emotionality ratings, followed by stability ratings. For the emotionality ratings, we asked participants to indicate how strongly a picture provokes emotions (from “not at all” to “very strongly”). In each block, the

gray-level pictures of the 44 selected paintings were presented in random order. There was no time limit. Overall, the two blocks lasted about 10 minutes.

7.4.2 Results

Mean liking, emotionality, stability and movement ratings were 46.2 ($SD = 7.70$), 44.2 ($SD = 9.20$), 44.9 ($SD = 14.7$), and 56.7 ($SD = 16.2$), respectively. For further analyses, we averaged the ratings across participants for each painting. Example stimuli and their mean ratings can be seen in Table 3. First, it should be noted that the stability as well as the movement ratings were strongly significantly correlated with the respective ratings from the preliminary study ($r = .756$, $p < .001$, for stability, and $r = .837$, $p < .001$, for movement).

Table 3. Example stimuli used in Experiment 2 with diametrically opposite movement and stability ratings.

					
#	85	9	91	96	16
M	81	73	58	43	24
S	30	33	35	59	87
L	55	54	42	36	38
E	55	46	40	33	30

Note. The bold numbers in row two (#) represent the corresponding picture numbers. The following rows show movement (M), stability (S), liking (L), and emotionality (E) ratings, respectively.

Further correlational analyses revealed significant correlations between all three variables (see Table 4): first, there was a positive relationship between liking and emotionality; second, stability ratings negatively correlated with liking, emotionality, and movement; and finally, movement was positively correlated with liking and with emotionality. Importantly, liking showed no significant correlation with the roundness ratings of the preliminary study ($r = 0.198$, $p = 0.198$).

The relation between liking, stability, and movement was further analyzed by multiple linear regressions (see Table 5), with liking as the dependent variable, and stability and movement as independent variables. As a result, stability, as well as movement, accounted significantly for the variance of liking. However, the regression coeffi-

cient for stability was now positive. This shows that if the effect of implied movement is controlled, stability still has an effect on liking, albeit a positive one (see Figure 2).

Table 4. Correlations between the different ratings in Experiment 2.

	Emotionality	Stability	Movement
Liking	.514***	-.311*	.549***
Emotionality	-	-.626***	.826***
Stability	-	-	-.843***

Note. * $p < .05$, *** $p < .001$.

Table 5. Results of the multiple regression analyses with liking ratings as dependent variables (DV).

DV	Predictor	B	SE	β	p	F	R^2	$R\Delta$
Liking	–	–	–	–	–	12.7***	.382	.285
	(Constant)	46.2	0.940	–	< .001	–	–	–
	Stability	0.277	0.120	0.527	.026	–	–	–
	Movement	0.476	0.109	0.994	< .001	–	–	–

Note. *** $p < .001$.

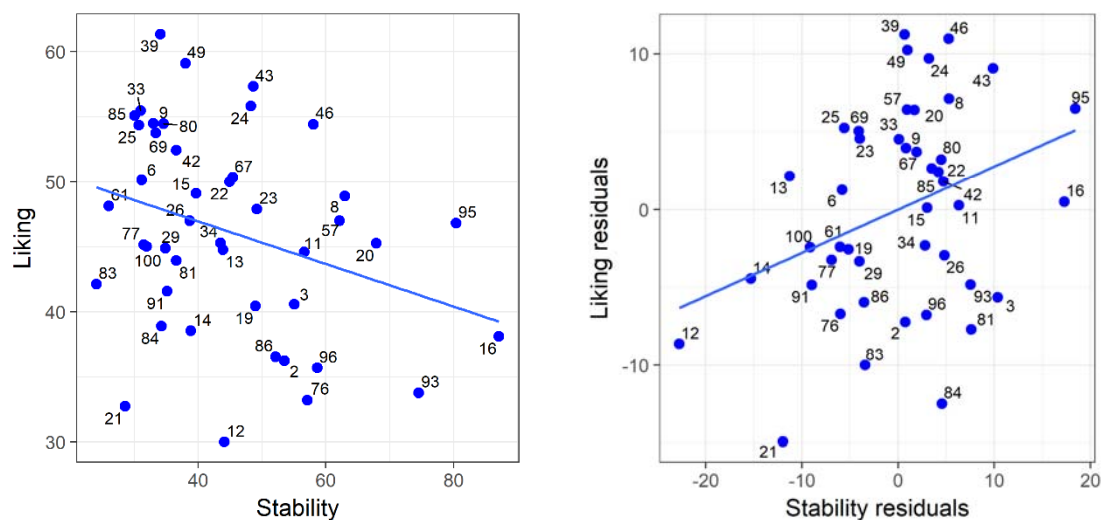


Figure 2. Left: Relation between stability and liking ratings in Experiment 2. Right: Relation between stability and liking after the effect of implied movement has been removed.

7.4.3 Discussion

In this experiment, we tested the hypothesis that perceived instability, if related to implied movement, increases the liking of an image. For this objective, we used artworks of Karl Otto Götz, which largely varied in implied movement. As expected, pictures that were rated as unstable were liked more than those rated as stable. This is in line with observations in our previous studies (Fillinger & Hübner, 2018; Hübner & Fillinger, 2019). Here, however, the relation was more direct and systematic.

Given that this result is in contrast with the results of Experiment 1, the participants also rated emotionality and implied movement, since these variables might account for the reversal of the usual pattern. In order to avoid pictures with a high movement rating being liked more simply because they are displaying rounder objects, in a preliminary study, we asked participants for their roundness ratings of the selected pictures. The results of this preliminary study showed that roundness had no effect on liking ratings, thus excluding the possibility that the relation between movement and liking can be traced back to roundness. Consequently, as hypothesized, implied movement was responsible for the negative relation between stability and liking. If the variance accounted for by movement was taken out from the liking ratings, then stability positively correlated again with liking.

Taken together, these results suggest that perceived instability can be separated into two parts: One of gravitational instability, which is disliked and another of dynamic instability, which is related to movement and liked. The positive correlation between movement and emotionality further supports the notion that instability related to movement is liked because it induces emotions. Although we did not have assessed the valence of emotionality (Lang et al., 1993; Russell, 1980; Watson & Tellegen, 1985), it is likely that the emotions in the present case were generally positive.

7.5 General Discussion

In the present study, we aimed to systematically produce two different types of visual instability and investigate how they affect aesthetic appreciation. Previous studies have suggested that instability could be associated with either gravitation stability or movement. For this purpose, in our first experiment we used multi-element pictures showing a pictorial composition containing three rectangles and three colored decora-

tive elements that vary in stability (see Table 1). The participants had to rate the balance and the stability of these compositions and how much they liked them. We observed a significant correlation between stability and liking, which supports the notion that gravitationally balanced and stable pictorial compositions are preferred (Friedenberg, 2012; Hübner & Fillinger, 2019; Pierce, 1896). This finding is in line with the assumption that gravitational stability (van der Helm, 2015) is preferred because, throughout our lives, we learned that it is important to arrange things in a way that withstands gravitational forces and thereby prevents damage and injuries. Based on this experience, art theory also considers stability as a visual and aesthetic habit that plays an important role in composing pictures (Liu et al., 2017).

Furthermore, our results indicate that the concepts of stability and balance are rather similar. However, given that balance correlated with the DCM score that reflects the mechanical balance approach (e.g., Arnheim, 1982) but stability did not, we inferred that despite the substantial correlation between stability and balance these two concepts differ in some relevant aspects. Moreover, balance correlated much lower with the DCM scores than with stability why we assumed that the balance ratings include both the usual mechanical balance as well as gravitational stability.

However, the large overlap between the concepts of stability and balance was also found using Japanese calligraphy as stimulus material (Fillinger & Hübner, 2018). Interestingly, in the same study, no correlation between the DCM and either of the two ratings was found. Consequently, the meaning of balance and stability can change from one picture to another one. Further studies will be needed to investigate the similarity as well as the diversity of these two concepts more closely.

The results of Experiment 1 were in line with art theory (Liu et al., 2017) and previous experiments (Friedenberg, 2012; Hübner & Fillinger, 2019). However, because the results are different from those observed for Japanese calligraphy (Fillinger & Hübner, 2018), we conducted a second experiment, which aimed at replicating and extending the finding of preference for instability. For this purpose, we used artwork by Karl Otto Götz, which had to be rated for perceived stability and liking, but also for movement and emotionality. For these pictures, we expected a negative relationship between stability and liking depending on movement.

As expected, the results of Experiment 2 revealed a negative correlation between stability and aesthetic appreciation, which is consistent with our previous finding for the calligraphies (Fillinger & Hübner, 2018). Regression analysis showed that implied

movement was responsible for this negative relationship, because after omitting variance accounted for by movement from the liking ratings, the relationship between stability and liking was again positive. Consequently, depending on stimulus material, both types of instability may be reflected within one and the same rating.

In order to avoid a misrepresentation of the relationship between movement and liking, in a preliminary study, we additionally asked participants to rate our selected images according to roundness. The results of this pre-study showed that roundness did not correlate with liking ratings. Consequently, we were able to exclude the possibility that pictures with a high movement rating are liked more simply because they are rounder. Interestingly, this is not in accord with previous research that showed that roundness is preferred (e.g., Gómez-Puerto et al., 2016; Silvia & Barona, 2009). This finding suggests that for this particular type of paintings roundness does not contribute to aesthetic appreciation.

Interestingly, our results also revealed that movement was positively correlated with emotionality, which supports the notion that instability related to movement is liked because movement is associated with emotionality. This impact of movement on aesthetic appreciation was already observed in previous research that used Chinese calligraphy as stimuli (Dubal et al., 2014). However, as already mentioned, we did not assess emotional valence, which is why we can only speculate that they have been positive.

One possible explanation for the evoked emotions by implied movement could be the human ability to empathize into observed actions (Singer & Lamm, 2009). Interestingly, it has been showed that these empathic processes can also be triggered by artwork (Freedberg & Gallese, 2007; Gerger et al., 2018). However, it is currently unclear whether the implied movement is the same for abstract and figurative art (e.g., an image of a person). Consequently, in future research, the concept of implied movement and its influence on emotions should be investigated further.

In sum, we found that perceived (in)stability as a concept closely related to perceptual balance can be attributed to gravitational instability as well as to movement. Our findings suggest that in multi-element pictures, gravitational stability is liked because this preference has been formed through living with gravitation that forces individuals to arrange things in a stable way. In contrast, perceived instability of dynamic images is preferred if it is related to implied movement. It is assumed that instability related to movement is liked because it is correlated with emotionality. Given that our study shed

light on the relation between perceived stability, balance and liking, it has relevance for both empirical aesthetics as well as the arts.

7.6 Authors' Note

The data of both experiments and the preliminary study are available at <https://osf.io/ebdah/>.

7.7 Acknowledgments

The artwork of Karl Otto Götz has been used with the kind permission of the K.O. Götz und Rissa-Stiftung, Germany.

7.8 Author Contributions

RH and MF conceived the study. MF constructed or selected the stimuli, collected the data, analyzed the data, and wrote a first draft of the paper. RH revised the paper and supervised the study at all stages.

7.9 References

- Arnheim, R. (1982). *The power of the center: a study of composition in the visual arts*. Berkeley: University of California Press.
- Bouleau, C. (1980). *The painter's secret geometry: a study of composition in art*. New York: Hacker Art Books.
- Dubal, S., Lerebours, A.-E., Taffou, M., Pelletier, J., Escande, Y., & Knoblauch, K. (2014). A Psychophysical Exploration of the Perception of Emotion from Abstract Art. *Empirical Studies of the Arts*, 32(1), 27-41.
doi:10.2190/EM.32.1.EOV.4
- Fillinger, M. G., & Hübner, R. (2018). Relations Between Balance, Prototypicality, and Aesthetic Appreciation for Japanese Calligraphy. *Empirical Studies of the Arts*.
doi:10.1177/0276237418805656

- Freedberg, D., & Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends in Cognitive Sciences*, *11*(5), 197-203.
doi:10.1016/j.tics.2007.02.003
- Friedenberg, J. (2012). Aesthetic Judgment of Triangular Shape: Compactness and Not the Golden Ratio Determines Perceived Attractiveness. *i-Perception*, *3*(3), 163-175. doi:10.1068/i0484
- Gerger, G., Pelowski, M., & Leder, H. (2018). Empathy, Einfühlung, and aesthetic experience: the effect of emotion contagion on appreciation of representational and abstract art using fEMG and SCR. *Cognitive Processing*, *19*(2), 147-165.
doi:10.1007/s10339-017-0800-2
- Gershoni, S., & Hochstein, S. (2011). Measuring pictorial balance perception at first glance using Japanese calligraphy. *i-Perception*, *2*(6), 508-527.
doi:10.1068/i0472aap
- Gómez-Puerto, G., Munar, E., & Nadal, M. (2016). Preference for Curvature: A Historical and Conceptual Framework. *Frontiers in Human Neuroscience*, *9*(712). doi:10.3389/fnhum.2015.00712
- Götz, K. O. (1985). *VAST: Visual Aesthetic Sensitivity Test* (4 ed.). Düsseldorf: Concept.
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114-125.
doi:10.1007/s40881-015-0004-4
- Hübner, R., & Fillinger, M. G. (2016). Comparison of objective measures for predicting perceptual balance and visual aesthetic preference. *Frontiers in Psychology*, *7*.
doi:10.3389/fpsyg.2016.00335
- Hübner, R., & Fillinger, M. G. (2019). Perceptual Balance, Stability, and Aesthetic Appreciation: Their Relations Depend on the Picture Type. *i-Perception*, *10*(3).
doi:10.1177/2041669519856040
- Kandinsky, W. (1926/1979). *Point and Line to Plane*. New York: Dover.
- Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*(3), 261-273. doi:10.1111/j.1469-8986.1993.tb03352.x
- Leiner, D. J. (2019). SoSci Survey (Version 3.1.06). Retrieved from <https://www.soscisurvey.de>

- Liu, J., Dong, W., Zhang, X., & Jiang, Z. (2017). Orientation judgment for abstract paintings. *Multimedia Tools and Applications*, 76(1), 1017-1036. doi:10.1007/s11042-015-3104-5
- Locher, P. J. (2006). Experimental Scrutiny of the Role of Balance in the Visual Arts. In P. Locher, C. Martindale, & L. Dorfman (Eds.), *New directions in aesthetics, creativity and the arts* (pp. 19-32). Amityville, New York: Baywood Publishing Company, Inc.
- Locher, P. J., Jan Stappers, P., & 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. doi:10.1016/S0001-6918(98)00008-0
- McManus, I., Edmondson, D., & Rodger, J. (1985). Balance in pictures. *British Journal of Psychology*, 76(3), 311-324.
- McManus, I., Stöver, K., & Kim, D. (2011). Arnheim's Gestalt Theory of Visual Balance: Examining the Compositional Structure of Art Photographs and Abstract Images. *i-Perception*, 2(6), 615-647. doi:10.1068/i0445aap
- Osaka, N., Matsuyoshi, D., Ikeda, T., & Osaka, M. (2010). Implied motion because of instability in Hokusai Manga activates the human motion-sensitive extrastriate visual cortex: an fMRI study of the impact of visual art. *NeuroReport*, 21(4), 264-267. doi:10.1097/WNR.0b013e328335b371
- Pierce, E. (1894). Studies from the Harvard Psychological Laboratory (II): Aesthetics of simple forms. I. Symmetry. *Psychological Review*, 1(5), 483-495. doi:10.1037/h0069000
- Pierce, E. (1896). Studies from the Harvard Psychological Laboratory (V): The aesthetics of simple forms. (II) The functions of the elements. *Psychological Review*, 3(3), 270-282. doi:10.1037/h0073983
- Ross, D. W. (1907). *A Theory of Pure Design: Harmony, Balance, Rhythm*. Boston, MA: Houghton, Mifflin.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161-1178. doi:10.1037/h0077714
- Silvia, P. J., & Barona, C. M. (2009). Do people prefer curved objects? Angularity, expertise, and aesthetic preference. *Empirical Studies of the Arts*, 27(1), 25-42.

- Singer, T., & Lamm, C. (2009). The Social Neuroscience of Empathy. *Annals of the New York Academy of Sciences*, 1156(1), 81-96. doi:10.1111/j.1749-6632.2009.04418.x
- Thömmes, K., & Hübner, R. (2018). Instagram Likes for Architectural Photos Can Be Predicted by Quantitative Balance Measures and Curvature. *Frontiers in Psychology*, 9(1050). doi:10.3389/fpsyg.2018.01050
- van der Helm, P. A. (2015). Symmetry perception. In J. Wagemans (Ed.), *Oxford Handbook of Perceptual Organization* (pp. 108–128). Oxford: Oxford University Press.
- Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98(2), 219-235. doi:10.1037/0033-2909.98.2.219
- Wilson, A., & Chatterjee, A. (2005). The Assessment of Preference for Balance: Introducing a New Test. *Empirical Studies of the Arts*, 23(2), 165-180. doi:10.2190/B1LR-MVF3-F36X-XR64

Contributions to the research papers

Research Paper I

- *Literature review:* Ronald Hübner & Martin Fillinger
- *Idea generation:* Ronald Hübner & Martin Fillinger
- *Formulating hypotheses:* Ronald Hübner & Martin Fillinger
- *Design of behavioral experiments:* Ronald Hübner
- *Data collection and analysis:* Martin Fillinger & Ronald Hübner
- *Interpretation of results:* Ronald Hübner & Martin Fillinger
- *Preparation of draft manuscript:* Ronald Hübner & Martin Fillinger
- *Preparation of final manuscript:* Ronald Hübner & Martin Fillinger

Research Paper II

- *Literature review:* Martin Fillinger
- *Idea generation:* Martin Fillinger & Ronald Hübner
- *Formulating hypotheses:* Martin Fillinger & Ronald Hübner
- *Design of behavioral experiments:* Martin Fillinger
- *Data collection and analysis:* Martin Fillinger
- *Interpretation of results:* Martin Fillinger & Ronald Hübner
- *Preparation of draft manuscript:* Martin Fillinger
- *Preparation of final manuscript:* Martin Fillinger & Ronald Hübner

Research Paper III

- *Literature review:* Martin Fillinger & Ronald Hübner
- *Idea generation:* Martin Fillinger & Ronald Hübner
- *Formulating hypotheses:* Martin Fillinger & Ronald Hübner
- *Design of behavioral experiments:* Martin Fillinger & Ronald Hübner
- *Data collection and analysis:* Martin Fillinger
- *Interpretation of results:* Martin Fillinger & Ronald Hübner
- *Preparation of draft manuscript:* Martin Fillinger
- *Preparation of final manuscript:* Ronald Hübner & Martin Fillinger

Research Paper IV

- *Literature review:* Martin Fillinger
- *Idea generation:* Martin Fillinger
- *Formulating hypotheses:* Martin Fillinger
- *Design of behavioral experiments:* Martin Fillinger
- *Data collection and analysis:* Martin Fillinger
- *Interpretation of results:* Martin Fillinger & Ronald Hübner
- *Preparation of draft manuscript:* Martin Fillinger
- *Preparation of final manuscript:* Martin Fillinger & Ronald Hübner

References

- Allesch, C. G. (1987). *Geschichte der psychologischen Ästhetik*. Göttingen: Hogrefe.
- Allesch, C. G. (2000/2001). Carl Stumpfs Ästhetik. *Brentano Studien*, 9, 199-214.
- Allesch, C. G. (2009). Gleichgewichtssinn, Balance und Ästhetik. In R. Schönhammer (Ed.), *Körper, Dinge und Bewegung: der Gleichgewichtssinn in materieller Kultur und Ästhetik* (pp. 243-250). Wien: Facultas.
- Allesch, C. G. (2017). Einfühlung – A Key Concept of Psychological Aesthetics. In V. Lux & S. Weigel (Eds.), *Empathy: Epistemic Problems and Cultural-Historical Perspectives of a Cross-Disciplinary Concept* (pp. 223-243). London: Palgrave Macmillan.
- Allesch, C. G. (2018a). Aesthetics and cultural psychology. In G. Jovanović, L. Allolio-Näcke, & C. Ratner (Eds.), *The Challenges of Cultural Psychology: Historical Legacies and Future Responsibilities* (pp. 318-336). London: Routledge.
- Allesch, C. G. (2018b). *Fechner: Vorschule der Ästhetik*. Berlin, Heidelberg: Springer Spektrum.
- Arnheim, R. (1954). *Art and visual perception: A psychology of the creative eye*. Berkeley: University of California Press.
- Arnheim, R. (1982). *The power of the center: a study of composition in the visual arts*. Berkeley: University of California Press.
- Ash, M. G. (1995). *Gestalt psychology in German culture, 1890 - 1967: holism and the quest for objectivity*. Cambridge: Cambridge University Press.
- Bar, M., & Neta, M. (2006). Humans prefer curved visual objects. *Psychological Science*, 17(8), 645-648.
- Baumgarten, A. G. (2007). *Ästhetik* (Vol. 1). Hamburg: Meiner.
- Benetka, G., & Joerchel, A. (2016). Psychology as a Phenomenological Science. In J. Valsiner, G. Marsico, N. Chaudhary, T. Sato, & V. Dazzani (Eds.), *Psychology as the Science of Human Being* (Vol. 13, pp. 17-32): Springer International Publishing.
- Berlyne, D. E. (1971). *Aesthetics and psychobiology*. New York: Appleton-Century-Crofts.

- Bode, C., Helmy, M., & Bertamini, M. (2017). A cross-cultural comparison for preference for symmetry: comparing British and Egyptians non-experts. *Psihologija*, *50*(3), 383-402. doi:10.2298/PSI1703383B
- Bouleau, C. (1980). *The painter's secret geometry: a study of composition in art*. New York: Hacker Art Books.
- Brentano, F. (1959). *Grundzüge der Ästhetik*. Bern: Francke.
- Breuer, J., & Freud, S. (1895). *Studien über Hysterie*. Leipzig; Wien: Deuticke.
- Bulot, N. J., & Reber, R. (2013). The artful mind meets art history: Toward a psycho-historical framework for the science of art appreciation. *Behavioral and brain sciences*, *36*(2), 123-137. doi:10.1017/S0140525X12000489
- Cupchik, G. C. (2007). A critical reflection on Arnheim's Gestalt theory of aesthetics. *Psychology of Aesthetics, Creativity, and the Arts*, *1*(1), 16-24. doi:10.1037/1931-3896.1.1.16
- Cupchik, G. C. (2014). Theoretical foundations for an empirical aesthetics. In J. K. Smith & P. P. L. Tinio (Eds.), *The Cambridge Handbook of the Psychology of Aesthetics and the Arts* (pp. 60-85). Cambridge: Cambridge University Press.
- Donderi, D. C. (2006). An Information Theory Analysis of Visual Complexity and Dissimilarity. *Perception*, *35*(6), 823-835. doi:10.1068/p5249
- Dubal, S., Lerebours, A.-E., Taffou, M., Pelletier, J., Escande, Y., & Knoblauch, K. (2014). A Psychophysical Exploration of the Perception of Emotion from Abstract Art. *Empirical Studies of the Arts*, *32*(1), 27-41. doi:10.2190/EM.32.1.EOV.4
- Ehrenfels, C. v. (1890). Über Gestaltqualitäten. *Vierteljahrsschrift für wissenschaftliche Philosophie*, *14*, 242-292.
- Fechner, G. T. (1871). *Zur experimentalen Aesthetik*. Leipzig: Hirzel.
- Fechner, G. T. (1876). *Vorschule der Ästhetik*. Leipzig: Breitkopf & Härtel.
- Fillinger, M. G., & Hübner, R. (2018). Relations Between Balance, Prototypicality, and Aesthetic Appreciation for Japanese Calligraphy. *Empirical Studies of the Arts*. doi:10.1177/0276237418805656
- Fingerhut, J., & Prinz, J. J. (2018). Wonder, appreciation, and the value of art. In J. F. Christensen & A. Gomila (Eds.), *Progress in Brain Research* (Vol. 237, pp. 107-128). Amsterdam: Elsevier.

- Forsythe, A., Nadal, M., Sheehy, N., Cela-Conde, C. J., & Sawey, M. (2011). Predicting beauty: Fractal dimension and visual complexity in art. *British Journal of Psychology*, *102*(1), 49-70. doi:10.1348/000712610X498958
- Freedberg, D., & Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends in Cognitive Sciences*, *11*(5), 197-203. doi:10.1016/j.tics.2007.02.003
- Freud, S. (1920). *Jenseits des Lustprinzips*. Leipzig: Internationaler Psychoanalytischer Verlag.
- Friedenberg, J. (2012). Aesthetic Judgment of Triangular Shape: Compactness and Not the Golden Ratio Determines Perceived Attractiveness. *i-Perception*, *3*(3), 163-175. doi:10.1068/i0484
- Gallese, V. (2019). Embodied Simulation. Its Bearing on Aesthetic Experience and the Dialogue Between Neuroscience and the Humanities. *GESTALT THEORY*, *41*(2), 113-127. doi:10.2478/gth-2019-0013
- Gartus, A., & Leder, H. (2017). Predicting perceived visual complexity of abstract patterns using computational measures: The influence of mirror symmetry on complexity perception. *Plos One*, *12*(11), e0185276. doi:10.1371/journal.pone.0185276
- Gerger, G., Pelowski, M., & Leder, H. (2018). Empathy, Einfühlung, and aesthetic experience: the effect of emotion contagion on appreciation of representational and abstract art using fEMG and SCR. *Cognitive Processing*, *19*(2), 147-165. doi:10.1007/s10339-017-0800-2
- Gershoni, S., & Hochstein, S. (2011). Measuring pictorial balance perception at first glance using Japanese calligraphy. *i-Perception*, *2*(6), 508-527. doi:10.1068/i0472aap
- Golomb, C. (1987). The Development of Compositional Strategies in Children's Drawings. *Visual Arts Research*, *13*(2), 42-52.
- Gómez-Puerto, G., Munar, E., & Nadal, M. (2016). Preference for Curvature: A Historical and Conceptual Framework. *Frontiers in Human Neuroscience*, *9*(712). doi:10.3389/fnhum.2015.00712
- Götz, K. O. (1985). *VAST: Visual Aesthetic Sensitivity Test* (4 ed.). Düsseldorf: Concept.

- Götz, K. O., Borisy, A. R., Lynn, R., & Eysenck, H. J. (1979). A new visual aesthetic sensitivity test: I. Construction and psychometric properties. *Perceptual and Motor Skills, 49*(3), 795-802.
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association, 1*(1), 114-125.
doi:10.1007/s40881-015-0004-4
- Hekkert, P., Snelders, D., & Van Wieringen, P. C. W. (2003). 'Most advanced, yet acceptable': Typicality and novelty as joint predictors of aesthetic preference in industrial design. *British Journal of Psychology, 94*(1), 111-124.
doi:10.1348/000712603762842147
- Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation*. London: Wiley.
- Howard, R. (1914). A note on pictorial balance. *British Journal of Psychology, 7*(1), 56-63.
- Hübner, R., & Fillinger, M. G. (2016). Comparison of objective measures for predicting perceptual balance and visual aesthetic preference. *Frontiers in Psychology, 7*.
doi:10.3389/fpsyg.2016.00335
- Hübner, R., & Fillinger, M. G. (2019). Perceptual Balance, Stability, and Aesthetic Appreciation: Their Relations Depend on the Picture Type. *i-Perception, 10*(3).
doi:10.1177/2041669519856040
- Jacobsen, T. (2004). Individual and group modelling of aesthetic judgment strategies. *British Journal of Psychology, 95*(1), 41-56.
- Jacobsen, T. (2006). Bridging the Arts and Sciences: A Framework for the Psychology of Aesthetics. *Leonardo, 39*(2), 155-162. doi:10.1162/leon.2006.39.2.155
- Jacobsen, T. (2010). Beauty and the brain: culture, history and individual differences in aesthetic appreciation. *Journal of Anatomy, 216*(2), 184-191.
doi:10.1111/j.1469-7580.2009.01164.x
- Kandinsky, W. (1926/1979). *Point and Line to Plane*. New York: Dover.
- Kant, I., & Pluhar, W. S. (1987). *Critique of judgment*. Indianapolis: Hackett Publishing.
- Kurz, G. (2015). *Das Wahre, Schöne, Gute*. Leiden: Wilhelm Fink Verlag.
- Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology, 30*(3), 261-273. doi:10.1111/j.1469-8986.1993.tb03352.x

- Lanzoni, S. (2009). Practicing psychology in the art gallery: Vernon Lee's aesthetics of empathy. *Journal of the History of the Behavioral Sciences*, *45*(4), 330-354. doi:10.1002/jhbs.20395
- Latto, R. (1995). The brain of the beholder. In L. R. Gregory, J. J. Harris, P. Heard, & D. Rose (Eds.), *The artful eye* (pp. 66–94): Oxford University Press.
- Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, *95*(4), 489-508. doi:10.1348/0007126042369811
- Leiner, D. J. (2019). SoSci Survey (Version 3.1.06). Retrieved from <https://www.soscisurvey.de>
- Liu, J., Dong, W., Zhang, X., & Jiang, Z. (2017). Orientation judgment for abstract paintings. *Multimedia Tools and Applications*, *76*(1), 1017-1036. doi:10.1007/s11042-015-3104-5
- Locher, P. J. (2006). Experimental Scrutiny of the Role of Balance in the Visual Arts. In P. Locher, C. Martindale, & L. Dorfman (Eds.), *New directions in aesthetics, creativity and the arts* (pp. 19-32). Amityville, New York: Baywood Publishing Company, Inc.
- Locher, P. J., Gray, S., & Nodine, C. (1996). The structural framework of pictorial balance. *Perception*, *25*(12), 1419-1436. doi:10.1068/p251419
- Locher, P. J., Jan Stappers, P., & 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. doi:10.1016/S0001-6918(98)00008-0
- Locher, P. J., Krupinski, E. A., Mello-Thoms, C., & Nodine, C. F. (2007). Visual interest in pictorial art during an aesthetic experience. *Spatial Vision*, *21*(1/2), 55-77. doi:10.1163/156856808782713762
- Locher, P. J., & Nagy, Y. (1996). Vision spontaneously establishes the percept of pictorial balance. *Empirical Studies of the Arts*, *14*(1), 17-31.
- Locher, P. J., & Stappers, P. J. (2002). Factors Contributing to the Implicit Dynamic Quality of Static Abstract Designs. *Perception*, *31*(9), 1093-1107. doi:10.1068/p3299
- Machado, P., Romero, J., Nadal, M., Santos, A., Correia, J., & Carballal, A. (2015). Computerized measures of visual complexity. *Acta Psychologica*, *160*, 43-57. doi:10.1016/j.actpsy.2015.06.005

- Makin, A. D. J. (2017). The Gap Between Aesthetic Science and Aesthetic Experience. *Journal of Consciousness Studies*, 24(1-2), 184-213.
- McManus, I., Edmondson, D., & Rodger, J. (1985). Balance in pictures. *British Journal of Psychology*, 76(3), 311-324.
- McManus, I., Stöver, K., & Kim, D. (2011). Arnheim's Gestalt Theory of Visual Balance: Examining the Compositional Structure of Art Photographs and Abstract Images. *i-Perception*, 2(6), 615-647. doi:10.1068/i0445aap
- Mitchison, G. J., & McKee, S. P. (1990). Mechanisms underlying the anisotropy of stereoscopic tilt perception. *Vision Research*, 30(11), 1781-1791. doi:10.1016/0042-6989(90)90159-I
- Munar, E., Gómez-Puerto, G., López-Navarro, E., & Nadal, M. (2014). Visual preference for curvature as a potential aesthetic primitive. In A. Kotzbelt (Ed.), *Twenty-third Biennial Congress of the International Association of Empirical Aesthetics* (pp. 316-319). New York.
- Osaka, N., Matsuyoshi, D., Ikeda, T., & Osaka, M. (2010). Implied motion because of instability in Hokusai Manga activates the human motion-sensitive extrastriate visual cortex: an fMRI study of the impact of visual art. *NeuroReport*, 21(4), 264-267. doi:10.1097/WNR.0b013e328335b371
- Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human preference. *Annual Review of Psychology*, 64, 77-107.
- Palumbo, L., Ogden, R., Makin, A. D. J., & Bertamini, M. (2014). Examining visual complexity and its influence on perceived duration. *Journal of Vision*, 14(14), 3-3. doi:10.1167/14.14.3
- Pierce, E. (1894). Studies from the Harvard Psychological Laboratory (II): Aesthetics of simple forms. I. Symmetry. *Psychological Review*, 1(5), 483-495. doi:10.1037/h0069000
- Pierce, E. (1896). Studies from the Harvard Psychological Laboratory (V): The aesthetics of simple forms. (II) The functions of the elements. *Psychological Review*, 3(3), 270-282. doi:10.1037/h0073983
- Poore, H. R. (1903). *Pictorial Composition and the Critical Judgement of Pictures*. New York: Baker and Taylor.
- Puffer, E. D. (1903). Studies in symmetry. *The Psychological Review: Monograph Supplements*, 4(1), 467-539.
- Puffer, E. D. (1905). *The psychology of beauty*. New York: Houghton, Mifflin and Co.

- Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, 8(4), 364-382. doi:10.1207/s15327957pspr0804_3
- Reips, U.-D., & Funke, F. (2008). Interval-level measurement with visual analogue scales in Internet-based research: VAS Generator. *Behavior Research Methods*, 40(3), 699-704. doi:10.3758/BRM.40.3.699
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3), 192-233. doi:10.1037/0096-3445.104.3.192
- Ross, D. W. (1907). *A Theory of Pure Design: Harmony, Balance, Rhythm*. Boston, MA: Houghton, Mifflin.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161-1178. doi:10.1037/h0077714
- Schaffner, B. (2006). Kairos. In H. Cancik, H. Schneider, & M. Landfester (Eds.), *Brill's New Pauly* (pp. 7:3-4). Leiden: Brill.
- Schönhammer, R. (2013). *Einführung in die Wahrnehmungspsychologie: Sinne, Körper, Bewegung* (2 ed.). Wien: Facultas.
- Silvia, P. J. (2009). Looking past pleasure: Anger, confusion, disgust, pride, surprise, and other unusual aesthetic emotions. *Psychology of Aesthetics, Creativity, and the Arts*, 3(1), 48-51. doi:10.1037/a0014632
- Silvia, P. J., & Barona, C. M. (2009). Do people prefer curved objects? Angularity, expertise, and aesthetic preference. *Empirical Studies of the Arts*, 27(1), 25-42.
- Singer, T., & Lamm, C. (2009). The Social Neuroscience of Empathy. *Annals of the New York Academy of Sciences*, 1156(1), 81-96. doi:10.1111/j.1749-6632.2009.04418.x
- Smith, J. K., & Tinio, P. P. L. (2014). Introduction by the editors. In J. K. Smith & P. P. L. Tinio (Eds.), *The Cambridge Handbook of the Psychology of Aesthetics and the Arts* (pp. 3-5). Cambridge: Cambridge University Press.
- Stumpf, C. (1883/1890). *Tonpsychologie* (Vol. 2). Leipzig: Hirzel.
- Thömmes, K., & Hübner, R. (2018). Instagram Likes for Architectural Photos Can Be Predicted by Quantitative Balance Measures and Curvature. *Frontiers in Psychology*, 9(1050). doi:10.3389/fpsyg.2018.01050

- Tinio, P. P. L., & Leder, H. (2009). Just how stable are stable aesthetic features? Symmetry, complexity, and the jaws of massive familiarization. *Acta Psychologica*, 130(3), 241-250. doi:10.1016/j.actpsy.2009.01.001
- Valsiner, J. (2014). *An Invitation to Cultural Psychology*. London: Sage.
- van der Helm, P. A. (2015). Symmetry perception. In J. Wagemans (Ed.), *Oxford Handbook of Perceptual Organization* (pp. 108–128). Oxford: Oxford University Press.
- Van Geert, E., & Wagemans, J. (2019). Order, complexity, and aesthetic appreciation. *Psychology of Aesthetics, Creativity, and the Arts*. doi:10.1037/aca0000224
- Vartanian, O. (2014). Empirical aesthetics: hindsight and foresight. In J. K. Smith & P. P. L. Tinio (Eds.), *The Cambridge Handbook of the Psychology of Aesthetics and the Arts* (pp. 6-34). Cambridge: Cambridge University Press.
- Vartanian, O., Martindale, C., Podsiadlo, J., Overbay, S., & Borkum, J. (2005). The link between composition and balance in masterworks vs. paintings of lower artistic quality. *British Journal of Psychology*, 96(4), 493-503. doi:10.1348/000712605X47927
- Verstegen, I. (2005). *Arnheim, Gestalt and Art: A Psychological Theory*. Vienna: Springer Vienna.
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, 138(6), 1172-1217. doi:10.1037/a0029333
- Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98(2), 219-235. doi:10.1037/0033-2909.98.2.219
- Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt II. *Psychologische Forschung*, 4(1), 301-350. doi:10.1007/bf00410640
- Wertz, F. J. (2018). Qualitative Methods as Fundamental Tools: Autonomy and Integration in Mixed Methods Research. In B. Schiff (Ed.), *Situating Qualitative Methods in Psychological Science* (pp. 43-60). New York, NY: Routledge.
- Whitfield, T. W. A., & Slatter, P. E. (1979). The effects of categorization and prototypicality on aesthetic choice in a furniture selection task. *British Journal of Psychology*, 70(1), 65-75. doi:10.1111/j.2044-8295.1979.tb02144.x

- Wilson, A., & Chatterjee, A. (2005). The Assessment of Preference for Balance: Introducing a New Test. *Empirical Studies of the Arts*, 23(2), 165-180. doi:10.2190/B1LR-MVF3-F36X-XR64
- Winkielman, P., Halberstadt, J., Fazendeiro, T., & Catty, S. (2006). Prototypes are attractive because they are easy on the mind. *Psychological Science*, 17(9), 799-806. doi:10.1111/j.1467-9280.2006.01785.x
- Winston, A. S., & Cupchik, G. C. (1992). The Evaluation of High Art and Popular Art By Naive and Experienced Viewers. *Visual Arts Research*, 18(1), 1-14.
- Wolf, U. (2019). Über den Sinn der Aristotelische Mesoteslehre (II). In O. Höffe (Ed.), *Aristoteles: Nikomachische Ethik* (4 ed., pp. 65-84). Berlin: De Gruyter.