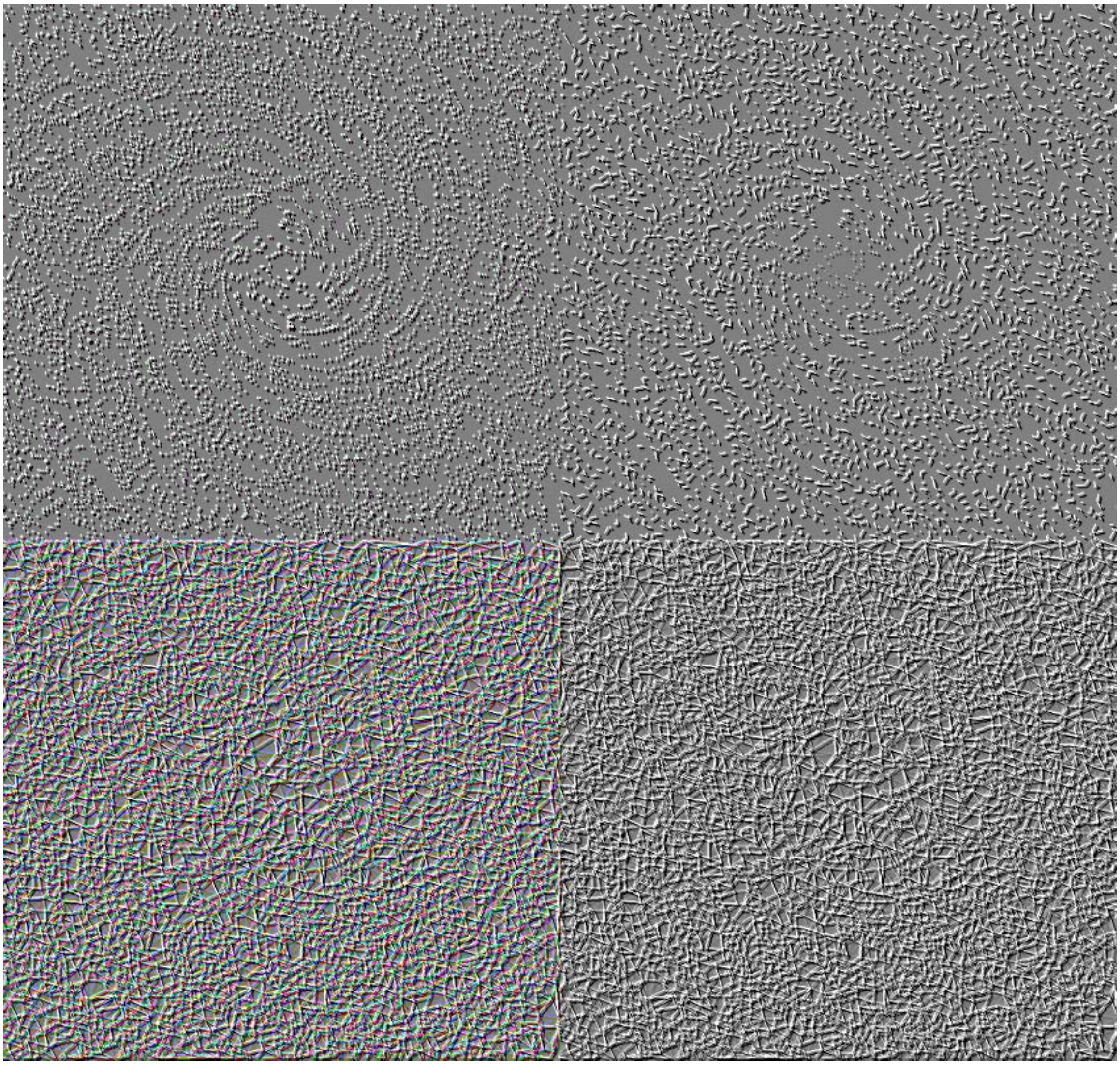
A Theoretical and Computational Investigation into Aspects of Human Visual Perception: Proximity and Transformations in Pattern Detection and Discrimination



Adrian K. Preiss

A Theoretical and Computational Investigation into Aspects of Human Visual Perception: Proximity and Transformations in Pattern Detection and Discrimination

Adrian K. Preiss BSc. (Hons.)

Department of Psychology

University of Adelaide

June 2006

Preface

The research effort that forms the basis of this thesis has its roots in other disciplines. Distance measuring and pattern classification methodology borrowed from geology, botany, geography, and the like, applied to questions raised by cognitive psychology, prove useful, and appear to enlarge the basis upon which visual pattern detection in psychology currently rests. Moreover some methods yield results that beg to be used in novel ways related to research questions in cognitive psychology. One aim of the thesis, therefore, is to introduce, explain, and develop for psychology, methods mostly encountered elsewhere; and yet another aim is to develop some new methods.

In keeping with a modern approach to vision science, vision is considered as a kind of computation from an interdisciplinary viewpoint. This does not necessarily present well because it is essentially theoretical. There is always the difficulty of presenting quantitative material clearly without overwhelming the reader, who may not have a computational background. In an endeavour to ease the load I have tried to make many compromises.

The thesis is conveniently divided into two broad categories. The subject of the first develops some useful statistical approaches to questions in visual pattern detection and the subject of the second develops a generative transformational approach to visual perception. The significant element common to the two categories is the nearest neighbour metric. However, this is only one of a hierarchy of distance measures usefully applied, not to mention other kinds of measure. Although 'nearest neighbour' can have limitations, it appears widespread as a component.

It has been convenient to implement gleaned information and original ideas, along with statistical and mathematical procedures, in a series of computer programs (developed by me), from which many screen images are included. These were mostly produced using the Microsoft *Visual Basic 6* language, although a few were produced using Wolfram Research *Mathematica 4.2* and MathWorks *MATLAB 6.5.*

Use of single and plural possessive pronouns in the thesis depends upon whether a particular investigation has been conducted individually or in collaboration with D. Vickers. It was he who acquainted me with issues associated with the field of visual perception, and inspired me with his generative transformational approach. Some material in the thesis appears in Vickers (n.d.), which is a report to the *Defence Science & Technology Organisation,* and in which our collaboration is acknowledged.

In developing statistical approaches to visual pattern detection in psychology, a consolidation of relevant literature in spatial statistics is provided. Much of the resultant methodology has been widely used elsewhere, and other barely used, and some small amount of methodology I have introduced. Inasmuch as we have ascertained, these methods have not been applied to questions concerning visual patterns in psychology, and there appears to be no literature concerning their relationship to human visual perception. In the thesis, the relationship has been explored in a limited empirical manner. Beyond this, there are no explicit claims concerning the methodology and its relationship to human visual perception. However, *correspondences* between outcomes and the way we see patterns should not go unnoted. More so, the early chapters provide a technical and statistical background for concepts and preliminary methodology used in pattern detection as outlined in some later chapters.

The thesis developed as technology developed. Questions that could not be reasonably tested owing to earlier computer limitations became accessible at later stages. It is emphasized that computer methodology provides an opportunistic implementation of principles related to the visual system. There is no claim that the visual system operates after the fashion of computer software or hardware.

The thrust of the thesis is theoretical, owing much of its outcome to computational investigation. A common drawback with theoretical treatments is that they incline to proceed in some degree of isolation from experimental findings. Only here and there in the thesis are accounts explored by way of experiments with subjects. Nonetheless the theory responds to concerns raised in and by the literature, and many more experiments could be designed to test it, time and resources permitting.

Last, it would be a mistake to expect complete (or insular) discussions and justifications for various phenomena in any one treatment. Investigations are often ongoing, and punctuate the thesis in varying context as it becomes established. I have taken the approach of letting the story unfold, and many questions that arise from earlier episodes are developed and addressed. In short, earlier treatments have a role in laying the foundation for later treatments, and justification is not necessarily wholly realized upon first encounter.

Contents

Abstract

A variety of measures are enlisted in an explanation of some longstanding perceptual phenomena associated with an assortment of visual patterns. In following the proximity principle of Gestalt psychology, these are commonly based upon a statistical treatment applied to one or another of a hierarchy of distance measures. Following from this, some problems of visual perception are tackled in terms of an *active* perceiving mechanism, which generates transformations in the realization of object invariance in space and over time. This generative transformational approach is also employed in an account of perception of various patterns and visual illusions. Although a range of proximity measures is involved throughout, the nearest neighbour metric is staple. For perception of unstructured visual arrays, the contribution of distance mechanisms, particularly nearest neighbours, is shown to be important. For structured arrays, the contribution of distance mechanisms along with transformations is important. Information about relative positions of image elements permits the selection of transformations that reveal structure. With respect to such information, however, the proximity principle is taken to its limits.

Declaration

I declare this thesis contains no material that has been accepted for the award of any other degree or diploma in any University, and to the best of my knowledge and belief contains no material previously published or written by any other person except where due reference is made in the text of the thesis.

If accepted for the award of the degree for which it is submitted, I consent to the thesis being made available for photocopying and loan.

Adrian K. Preiss

June 2006

Acknowledgement

Thank you to my principal supervisor, Professor Douglas Vickers (sadly, now deceased), for his general direction and unstinting collaboration in developing the generative transformational approach to visual perception; which is the subject of much of Section 2 of my thesis. He was working on elements of the approach when I enrolled as a PhD student, and I can only hope that the matter has been advanced since then. Thank you to Associate Professor Michael Lee, who substituted as principal supervisor at the beginning of the review stage of the thesis. His many comments, suggestions, and insights have been invaluable. Thank you also to my second supervisor, Dr. Nick Burns, who ultimately became principle supervisor in the final stages of review. His valuable reviewing effort and personable responses to concerns about publication format are much appreciated. And thank you to Matthew Dry and Tamera Wiggins for their help in running experiments, and to Lama Chandrasena for his help in implementing an experiment. Lastly, thank you to Peter Hughes, who supplied raw data from other experiments that were performed within the discipline.

Introduction

Dodwell (1975, 1978, 1984) provides a convenient division of human visual processing into three levels: sensory, perceptual, and cognitive. Sensory, at level-1, is psycho-physiologically placed, and deals with basic structural elements: detection of change or difference in elementary properties; shape, brightness, or position, for example. Perceptual, at level-2, deals with organization of elements into structures; grouping via proximity, similarity, good continuation, or closure, for example. And cognitive, at level-3, deals with recognition and understanding of such structures. Essentially, cognition is judgment; an entity is distinguished from others and is characterized by some concept, for example. The subject of this thesis is chiefly concerned with modelling sensory (level-1) and perceptual (level-2) aspects of visual perception.

Grouping of elements into a pattern and separation of figure from background were investigated in the first half of the twentieth century by German psychologists, mainly concerned with perception of structure, or *Gestalt*. Although the Gestalt movement made important contributions to learning and the creative process, the introspective reports that it relied upon to explain perception were too subjective to be of much scientific value. Furthermore the 'innate physiological processes' to which the Gestaltists attributed their laws of organization have been largely discredited.

Gestalt phenomena have long resisted quantification, which also mitigated their effect. Even so, today's computational technology has facilitated a revival of interest in developing a theoretical account of the organizational tendencies outlined by the Gestalt psychologists.

Underspecification: the 3D, 2D, 3D problem

By way of a simple example, six plus three is a forward problem, which is readily resolved to equal nine. But to simply say 'nine' and then ask, 'how was it produced by summation?' raises an inverse problem for which an infinite number of solutions exist, although constrained to equal nine. The problem is underspecified.

If only two numbers consisting of single or double digits are involved in the summation and the absolute value of one is twice as large as that of the other, for example, then such constraints allow some finite and reasonable interpretation. Six plus three, or eighteen minus nine, equals nine. Specification of constraints allows solutions to inverse problems.

In mathematical terms, the process of image formation is a forward problem and is well understood in terms of optical physics (akin to camera). The process of vision is an inverse problem, particularly at distances beyond effective binocular disparity, or for viewing through just one eye, and the solution is underdetermined. Visual information is made available at the retina in the form of a two-dimensional punctate array, corresponding to intensity values for separate points of light, encoded in the rods and cones. Yet, given this surface representation, the perceptual system can create a three-dimensional representation of the environment.

Since the inverse situation is underspecified, an infinite number of appropriate physical scenes could give rise to identical retinal images. See Figure 1.

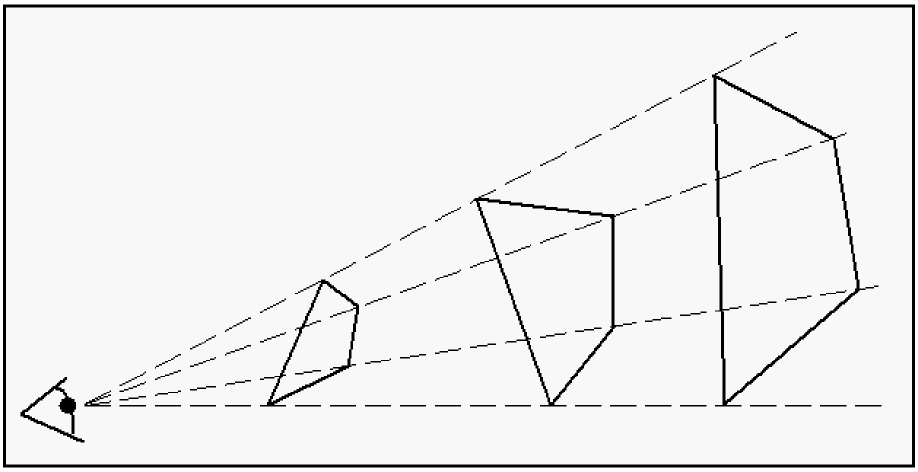


Figure 1: Different forms, same projection onto retina.

Hence recovery of shape and structure is a problem for which it appears not possible to devise an exact algorithm. It is generally regarded as an intractable, combinatorial optimisation problem. The problem becomes one of recovering near optimal structure against sensory constraints. (In mathematical terms, some function of some number of variables is minimized or maximized subject to constraints on the variables.) If the number of possibilities is discreetly constrained by one consideration or another, then solutions to the inverse problem are limited: it becomes manageable.

For some (Gibson, 1979; Marr, 1982) constraints are neither lodged in the mind nor are they active constituents in the perceptual process. They are conditions that the world must satisfy if a computational approach is to work. Constraints are passive guarantors that are external to the perceptual process.

For others (Rock, 1979, 1983, for example, and the transactionalists) assumptions and knowledge are mental contents that are active in constraining perceptual processes about regions of legitimate solutions. Assumptions and knowledge legislate selection of the most likely distal attributions. Constraints are active legislators that are internal to the perceptual process.

Approaches to perceptual organization

Approaches to perceptual organization conveniently reduce to four categories. The first contends that higher level organizational phenomena can be explained in terms of neurophysiological processes associated with the sensory level (Marr, 1982). The second category includes 'ecological' theories, as per Gibson (1979), whereupon perception is simply a process of 'resonating' to information in the stimulus array provided by the environment. All relevant visual information is inherent in the structure of ambient light. The third category, which derives from Helmholtz (1867-1925), is the 'likelihood' approach. Proponents of this approach contend that perception involves interaction between sensory and cognitive processes. Sensory elements are supposed to be organised by unconscious inference into the most likely hypotheses concerning their source (Albert & Hoffman, 1995; Brunswik, 1956; Gregory, 1970). Finally, proponents of the 'simplicity' approach stress the autonomy of perceptual processes, and, like the Gestaltists, contend that perception tends to follow some form of minimum principle. Stimulus elements are supposed to be organised in such a way as to provide the simplest description of the source of stimulation (Attneave, 1982; Restle, 1979; van der Helm & Leeuwenberg, 1996).

Unlike the first two approaches, the last two attach importance to perceptual processes. Nonetheless each has attracted criticism (Leeuwenberg & Boselie, 1988; Pomerantz & Kubovy, 1986). Leeuwenberg and Boselie, for example, argue that likelihood starts from interpretations, whereas it is supposed to explain the existence of interpretations. They show that the likelihood of an interpretation is a consequence of its simplicity. Once an interpretation is taken for reality, then its likelihood is self-evident. In other words, likelihood is a secondary implication of the perceptual process.

More generally, for the likelihood approach, assumptions required to assign likelihoods to even simple hypothetical situations associated with a pattern of stimulation are elaborate and frequently tenuous (cf. Feldman, 1996). And the simplicity approach requires debatable assumptions concerning what constitutes the basic elements of a description and the process by which descriptions are supposed to be constructed. Besides these difficulties, both approaches fail to consider any active contribution by the perceiving organism towards the process of organization (Dodwell, 1984).

There is obviously some overlap in at least some of the four approaches: likelihood of an interpretation as a consequence of simplicity, for example. Moreover, approaches to cortical functioning such as sparse coding (a compromise between local and distributed representation) and Bayesian methods (allied to the likelihood approach in any event) suggest the four approaches are not mutually exclusive. However, these divisions of propositions span a considerable research period with comparatively recent citations, and their considered differences still grant them currency as bases for perceptual organization.

Transformational approaches to visual perception

A further division of propositions, although not as well elaborated, avoids some difficulties of the other approaches. Proponents of this division take a 'transformational' approach. Although interpretations vary, most researchers in brain science consider that a major function of the human brain is to provide a model of the environment. Hence it might be expected that some of the general physical principles operating in nature be reflected in the way the brain constructs such a model. For example, in a series of articles, Shepard (1984, 2001) argues that the external constraints that have been most invariant throughout evolution have become internalized in the perceptual system. In Shepard's view, the principles of kinematic geometry provide an important set of such constraints. A general conceptual framework for dealing with invariant constraints is that of symmetry under a (mathematically defined) group of transformations.[1]

Shepard also includes a form of 'minimum principle', which governs the motion of rigid objects. He shows that when presented with alternating views of an object in two different positions, an observer perceives a single object that appears to move between the two positions along the path of simplest rigid transformation.

Shepard's transformational approach appears to overcome some limitations of the likelihood and simplicity approaches. However, the notion of 'internalization' has been criticised as misleading and unhelpful (Schwartz, 2001). Schwartz notes, for example, that Shepard's evidence for internalization of kinematic geometry largely comes from studies of apparent motion, which is not ecologically representative.[2] So how might the principles of kinematic geometry have become internalized on the basis of such evidence? And when presented with instances of real motion that disregard kinematic principles, we have no problem in seeing the motion veridically. Hence principles established by cases of apparent motion that are readily voided by cases of real motion, which is more ecologically representative, must be of little consequence.

---

[1] In mathematics, symmetry is framed in terms of invariance under transformation. A figure, for example, is symmetric with respect to a transformation if the figure comes out unchanged after applying the transformation.

[2] Apparent motion, as opposed to real motion, involves discrete changes in position, such as that experienced between a pair of lights that flash in alternating fashion.

More generally, the likelihood, simplicity, and transformational approaches are chiefly descriptive, rather than process oriented. They lack a computational formulation regarding processing steps.

Low-level grouping

Image elements have properties such as shape, size, orientation, colour, brightness, and position; any combination of which can define groupings. Dealing with just one property at a time eliminates interactions between properties that can result in conflicting groupings. Monochromic dots, or small monochromic circles, effectively have just one property, which is position. Patterns constituted of such elements allow determinations of grouping characterized by position: arguably the most rudimentary grouping.

Segments defined by such grouping have spatial extent, and hence properties such as shape, size, and orientation. For example, dots grouped as orientated pairs form virtual lines that produce textures with clear structure. Glass patterns (Glass, 1969) have this kind of grouping. And for more complex patterns, orientated virtual lines can be grouped still further, and so on.

Functionally, the first section of this thesis presents a computational approach pitched more to low-level grouping, associated with the sensory level, in point patterns. It seeks to exploit the proximity principle of Gestalt psychology in a comprehensive way, not only defining relative distance measures but other measures related to proximity as well. Structurally, each chapter belonging to the first section outlines various theoretical devices, and for the purpose of prompt consolidation each device is followed by application to low-level grouping.

A generative transformational approach

The second section of the thesis presents a computational approach pitched more to the perceptual level. This advances the generative transformational approach mooted by Vickers (1996, 1997, 2001), and furthered by Vickers, Navarro, and Lee (2000), and also mentioned by Vickers and Preiss (2000). It has much in common with other transformational approaches, but has some distinguishing features. Like Shepard's, this approach assumes that perception is guided by the internalization of geometric constraints in the external world.[3] It differs in that it proposes a set of operations by which such constraints might be implemented.

The locus of internalization is conceived in more literal terms than those employed by Shepard. The principles internalized are not just those that apply to the external world, but ones that also govern *action* in the world. Transformations on stimulus elements performed by the perceptual system are considered internal versions of actions that can be performed in association with corresponding physical objects; and the principles and constraints invoked are subordinate to *generating* a perceptual result.

In accord with this proposition, a number of people (e.g., Kosslyn, 1994) have suggested that the mental rotation of visual stimuli in experiments by Shepard and others (e.g., Shepard & Cooper, 1982) is closely associated with pre-motor processes. For example, Wohlschläger and Wohlschläger (1998) found that performing rotational hand movements interfered with simultaneously executed mental object rotation, but only if the axes of rotation coincided in space. On this basis, Wohlschläger (2001) suggested that mental object rotation is an imagined (covert) action, rather than a pure visual-spatial imagery task. Consistent with this view, these authors showed that the mere planning of hand movements interfered with mental object rotation.

---

[3] Vickers' (2001) 'concrete instantiation' of the most general physical principles identifies *rules* of perception, hence a formulation of visual abilities simply in terms of rules for prediction can be used to circumvent 'internalization' (Hecht, 2001, p. 755).

These findings underline a relationship between mental object rotation and anticipation of physical movements. Hence it is plausible to regard transformations of stimulus elements, envisaged by the generative transformation model of perceptual processing, as having evolved from the internal direction of what were originally overt actions.

The generative transformational approach also cites objects in terms of process history (Leyton, 1992). Leyton describes a perceptual system that incorporates action in terms of reconstruction of original symmetry by reversal of the process that generated asymmetry in a currently observed object.

Functionally, the second section of this thesis presents a computational approach oriented toward the perceptual level. Again, it seeks to exploit distance measuring devices, based on proximity, in a comprehensive way. These are applied to static and dynamic two- and three-dimensional point patterns. Structurally, each chapter belonging to the second section outlines various computational devices that achieve desired outcomes.

Before beginning in earnest

As stated in the preface, the significant element common to the two sections is the nearest neighbour metric. While being on the lookout for attendant limitations, this thesis seeks to thoroughly explore nearest neighbour along with its encompassing family of proximity measures. Even so, it is well appreciated that proximity is just one of several Gestalt principles of grouping; the others being similarity, closure, common fate, and continuity. However, there are interactions between the Gestalt principles: continuity can be on the basis of proximity or similarity, for example. In fact, proximity has been considered a particular case of grouping by similarity: the similarity of position of image elements. That is, grouping by regularity, or consistency, of proximal spatial arrangement. Proximity, along with regularity, is fundamental to the thesis, and inasmuch as this is implied by similarity, then similarity is fundamental to the thesis.

Another element that persists across the two sections is that of symmetry, which is commonly classified as a simple harmony of proportions, or balance (Jenkins, 1982). Reflection, translation, and rotation are basic forms of symmetry, which in its general form requires a configuration of components to show invariance under a group of automorphic transformations (Weyl, 1952).

Many theoretical approaches to low-level vision not unreasonably stem from research in neurophysiology, which suggests a theory in which feature detector cells are tuned to respond to fixed attributes, not attributes scaled to pattern configuration. However, the human visual system is demonstrably sensitive to Glass pattern structure that varies over a wide range of scales from one stimulus to the next or even within a single stimulus. Hence this thesis implicitly deals with proportions among attributes. For a measure such as distance, for example, proportions among distances are important, not the absolute distances themselves.[4]

Although various cognitive activities may be distinguished by invariants associated with different transformation groups and related symmetries, this thesis employs affine transformations by way of example.[5] Affine transformations describe some ways in which an

---

[4] Distances throughout the thesis are calculated in screen units, which depend on the scale mode (unit of measurement for coordinates) chosen for a set of graphic displays. A value given in screen units, by itself, has no absolute significance, only a significance relative to another value or other values given in the same screen units.

[5] **Affine:** translation, rotation, uniform stretching, that carries straight lines into straight lines and parallel lines into parallel lines, but may alter distance between points and angles between lines. (After Dictionary entry in Encyclopaedia Britannica, Standard Edition, 2002.) Affine transformations include translations, dilations and contractions, rotations in depth and in the plane, shears, and reflections.

image can change on the retina; although it is recognised that non-uniform transformations, too, are essential. The question of which geometry best qualifies human visual perception continues to be debated. In any event there appears to be a hierarchy of transformations in which subordinate sets of transformations may be considered subsets, or special cases, of more inclusive, or superordinate, sets.