UNIVERSITA' DEGLI STUDI DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di PSICOLOGIA GENERALE

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE PSICOLOGICHE
INDIRIZZO PERCEZIONE E PSICOFISICA
CICLO XXI

INTRINSIC AND EXTRINSIC STIMULUS FACTORS OF PERCEPTUAL SLANT IN DEPTH

Direttore della Scuola : Ch.mo Prof. Luciano Stegagno

Supervisore : Ch.mo Prof. Luigi Burigana

Dottorando : Francesco Martino

Index

Introduction

One of the main concepts in the study of depth perception is that of a “depth-cue”. This expression is used to indicate any property of the proximal stimulation that gives information about some 3-D characteristics of the physical world that we inhabit. Several studies have been focused on understanding the mechanisms behind depth-cues integration, i.e., how the information provided by each one of these different sources is combined, in determining 3-D properties of the visual world. Several authors underline that the information available at the level of proximal stimulation does not suffice to specify stable and reliable 3-D percepts, so that the additional concept of “prior constraints” has been introduced. “Prior constraints” constitute assumptions presumed to be carried out by the visual system in order to supply additional information that could enrich the one already available in the optic array.

These problems constitute the main theme of my research work, as I will present it in this dissertation. Compared with other studies, what distinguishes my contribution is the type of perspective adopted when considering information about depth. Depth cues are often distinguished using “qualitative-categorical” criteria. On the contrary, I will use “spatial-topological” criteria, as defined below.

My thesis is organised in five chapters, plus a summary conclusion section. The first two chapters present a general framework for describing the problems of depth perception, and try to clarify what I mean by “spatial-topological” criteria and integration between cues residing on different locations. The three remaining chapters describe six experiments aimed at an in-depth examination of some of the ideas presented in the first two chapters.

Chapter 1 addresses some general questions about depth-cues and related depth-percepts. The components of a “psychophysical inference rule”, i.e., a rule specifying how a depth-cue may inform about a depth-percept, are defined, and peculiar aspects of its basic components are discussed. Then, I consider the integration of information from several depth-cues. Main definitions are expressed by symbolic formulas, in set-theoretic terms.

The set-theoretic description is used again in Chapter 2 to introduce the problem of integration between cues which may be located on different units in the optical stimulus. In particular, I specify the distinction between “intrinsic” and “extrinsic” properties of the optical stimulus. A generic property *P* carried by object *A* is intrinsic if, after removing every other object constituting, respectively, the visual stimulus or the visual scene, *P* maintains on *A*. A generic property *P* is extrinsic if, by removing every other object constituting, respectively, the visual stimulus or the visual scene, that property does not maintain on *A*. In the last section of Chapter 2, I provide a description of the concept “optical contact”, which constitutes an important example of “intrinsic depth-related property” (relative to the optical stimulus). The role of optical contact in determining some depth-percepts is explored in four experiments, described in Chapters 3 and 4.

In Chapter 3, two experiments that analyse the role of optical contact as a factor capable of inducing perceptual slant are presented. Using computer graphics, I simulated a corridor with a pole inside it. The position of the ends of the pole within the ceiling and the ground of the simulated corridor are the “optical contact” of concern. The experimental hypothesis was that a discrepancy between these optical contacts might influence the apparent slant in depth of the pole. Results supported the hypothesized dependence of apparent slant on discrepancy of contact. They also revealed two other effects on the apparent slant, involving the pictorial tilt of the pole and its position on the right or the left of the vertical median through the stimulus.

In Chapter 4, two experiments that analyse how information from “optical contact” interacts with other stimulus properties are presented. The first experiment investigated the role of optical contact in inducing the slant of a target-stimulus with an irregular shape, whereas the second experiment investigated how optical contact affects the slant of a regular shape target, with strong monocular depth-cues (namely, texture and linear perspective) residing on it. The hypothesis behind these two experiments was that, given the highest degree of internal coherence of two-dimensional stimuli with respect to one-dimensional stimuli like those used in the experiments described in Chapter 3, there should be a stronger resistance in seeing the former ones as perceptually slanted. Results showed that optical contact could induce perceptual slant when no strong depth-cues are present on the target-stimulus. When this is not the case, then its effect appears to be vetoed by other depth-cues.

Chapter 5 describes two further experiments about the role and the integration of peripheral and foveal information in determining the slant of textured surfaces. These experiments have been made as a part of my doctorate studies when I was a visiting Ph.D. student at the Laboratory of Experimental Psychology at the Catholic University of Leuven, where I collaborated with Professor Johan Wagemans and Doctor Pedro Rosas in a project which aims at studying the role of eye movements in the general problem of “slant-from-texture”. Using the Gaze-Contingent Window Technique (cf. Rayner, 1998), I projected discrepant information in the foveal and peripheral portion of the visual field. Using a constant stimuli procedure with three different types of textures and three different sizes of the Gaze-Contingent Window, I found that subjects rely on foveal information especially when irregular/noisy patterned textures are presented, thus suggesting a process of integration between foveal and peripheral information in complex stimuli. Indeed, integration heavily relies on attentional factors and the different strategies used by subjects.

In the concluding section, I make some further comments on the theoretical analysis described in Chapter 1 and 2. Furthermore, I draw some general conclusions on the results found in my experiments. Some possible developments for future researches are also proposed.

Chapter 1

Sources of Information for Depth-Perception

The capability of the human visual system to inform us about the surrounding environment and to permit a successful interaction with it depends on the ability to gather information from different relevant sources and successfully integrate them. One of the puzzling questions of vision science is how all this information, available at the retinal level and two-dimensional in its nature, is integrated and how this integration allows the full phenomenal reconstruction of the outside world in its three-dimensional structure.

The perceptual contents relating to three-dimensional spatial organisation of the phenomenal scene, which, for the sake of simplicity, I will refer to as *depth-percepts*[1], involve properties such as the distance between the observer (the phenomenal ego) and a single perceptual object (*egocentric distance*); the distance between two distinct objects in the perceptual scene (*exocentric distance*); curvature and orientation of visible surfaces; shapes of solid objects; and so on.

In the psychophysical perspective, the aspects in the proximal stimulus specifically responsible for the emergence of depth-related properties in the phenomenal scene are often called "*depth-cues*". Thus, depth-cues are properties of definite parts of the visual stimulus. Each depth-cue contributes differently and in different amounts to specifying the values of some *depth-percepts* (see Figure 1.1 for a synthetic representation of this idea).

---

[1] Bülthoff and Mallot propose the following definition of a depth-percept or, as they propose to call them, *3-D descriptors*: "At least at a low level of abstraction, multiple representations of three-dimensional structure exist, which will be called 3-D descriptors [...]. These 3-D descriptors are sufficient for simple behaviour and it is unclear whether a single complete representation of visible surfaces exists at all […] Raw data from depth-cues such as shading, texture or disparity can be thought of as a trivial, or zero-order, representations of the spatial structure of a scene. Based on these data, higher-order descriptors are derived that make interesting spatial properties of the viewed scene explicit" (Bülthoff & Mallot, 1990, p. 121).

Some authors noticed that the use of the term “cue” is strongly associated with an empiricist position inside the long-term debate nativist/empiricist[2]. These authors suggest using the more neutral term “*source of information”* (Cutting & Vishton, 1995). For the sake of consistency with previous literature on depth perception, in this dissertation I will use both expressions, keeping in mind that the second one would be preferable since it does not require any strong theoretical assumptions. I would also to underline that not every vision scientist accepts the general idea of a simultaneous use of several depth-cues for recovering contents relating to properties of the physical world (Burton & Turvey, 1990).

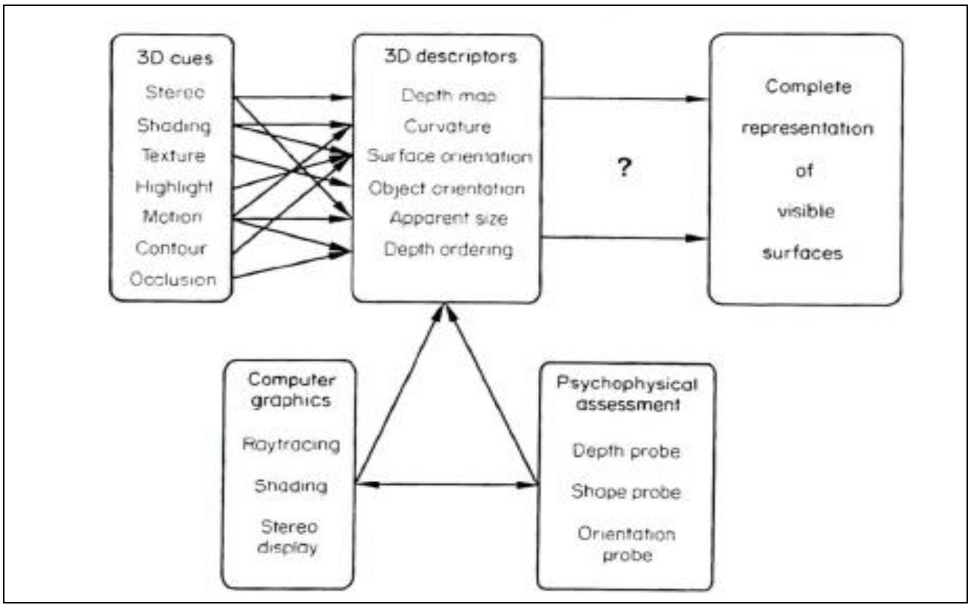


Figure 1.1. This figure is taken from Bülthoff (1991). It expresses the relationship between depth-cues, which jointly participate in estimating depth-percepts (or, as they refer to, 3-D descriptors). The existence of a complete representation of visible surfaces, if any exists, is not necessary. In the lower part of the figure, the relation of this very general model of 3-D perception and instruments used by psychophysicists to understand how it works are shown.

---

[2] The term “cue” comes from theatrical jargon, and implies “a knowledgeable observer, one with foreknowledge about when and how to act on layout information” (Cutting & Vishton, 1995, p. 70).

Depth-percepts are often studied in isolation, meaning that experiments are mainly focused on the ways in which one or more depth-cues are able to provide information about one single depth-percept (see, for instance, the idea of “slant from texture”, “shape from motion and shading” and similar expressions to be found in the depth-perception literature). The distinction among different depth-percepts is mainly a methodological one, and “the question of what an interesting … [depth-percept] is, has to be answered in the light of the action that it should subserve […]. Eventually, this process may or may not lead to a single complete representation of visible surfaces as was proposed by Marr & Nishihara (1978)” (Bülthoff and Mallot, 1990, p. 121).

1.1 Depth-cues: classifications and definitions

The interest in identifying and classifying which sources of information enable the visual system to recover depth-related properties of the visual world has a very ancient history. Some of the first results can be dated back to the IV century B.C., when Euclid introduced and studied from a geometrical point of view how the different heights in the visual field of images of different objects could provide information about the spatial position of the objects themselves (Burton, 1945). Several centuries after Euclid, Ibn al-Haytham (Latinized name: Alhacen, 965-1039) resumed investigation on the properties of the rays of light and on linear perspective. Among others, important studies on depth-perception were contributed by Leonardo Da Vinci (1452-1509) (Taylor, 1960) about the role of relative size and aerial perspective, and later by George Berkeley (1685-1763), especially concerning the role of physiological cues (Berkeley, 1709). A systematic categorization of depth-cues and their possible forms of integration is that defined by Hermann von Helmholtz (Helmholtz, 1866).

Among the depth-cues most extensively studied, we can mention binocular disparity, motion parallax, eyes convergence and accommodation, linear perspective, texture gradients, occlusion, relative size, height in the visual field, shading and kinetic depth. As the names suggest, these cues are very different from one another: the list provided is not exhaustive, and along the history of psychology, several analogous lists have been proposed (Helmholtz, 1866; Carr, 1935; Boring, 1942; Gibson, 1950a; Cutting & Vishton, 1995; Bruce, Green & Georgeson, 1996; Howard & Rogers, 2002 among the others). In recent years, some authors have also tried to discover and have proposed new potential cues for depth perception (e.g., gravity; cf. Watson, Banks, van Hofsten and Royden, 1992).

An interesting issue related to this topic is which sources of information can be really considered as depth-cues. Consider, for instance, the case of linear perspective. Some authors do not consider linear perspective as a specific cue to depth. They prefer to define it as the combination of separate sources of information (occlusion, texture gradients, relative size) (Cutting, 1997). On the contrary, other authors treat linear perspective as a single depth-cue, in spite of the fact that other depth-cues are related to its effect (Gibson, 1950a; Kubovy, 1986). These authors refer to linear perspective as the phenomenon of perceptual convergence of physical parallel lines while they recede into distance.

I report these issues to show the theoretical complexity in providing a clear definition of depth-cues. In the next pages, I will describe a set-theoretical framework that aims at illustrating the way in which depth-cues determine or constrain corresponding depth-percepts. This theoretical framework has been proposed by Burigana & Martino (2009), with the intention to provide some results that might be useful premises for formal treatment of problems in psychophysics of depth perception. The principal goal of this analysis is to show how concepts involved in it may be profitable for expressing several questions and hypotheses currently debated in this field.

1.2 Depth-cues and depth-percepts: A set-theoretic formal description

One special characteristic of the analysis I am going to propose is the epistemological standpoint taken: the perspective is centred on the scientist (the psychophysicist) rather than the perceiver. In the following pages, I will refer to hypothetical computations involving different pieces of information, with information concerning a depth-cue as the input of computation, and information concerning a depth-percept as the output. I would like to stress that, when using the expression "computation", I do not refer to inferences or calculation performed by the visual systems, according to an empiristic or Helmoltzian paradigm[3], but to logical operations, carried out by a psychophysicist, who aims at obtaining rules allowing her/him to discover and express possible regularities in the relationship between optical stimuli and perceptual results. Of course, the perceiver is a fundamental component of this system, because s/he is the source of the components concerning perceptual results, thus playing a fundamental role in such an "information game". Then, my discussion will be limited to describe the association between stimulus inputs and perceptual outputs, and no conclusion will be drawn about what is going on in the perceiver during visual observation (according to the classical formulation of "external psychophysics", as intended by Fechner, 1860/1966). Placing the psychophysicist at the core of the discussion, both internal psychophysics and external psychophysics views are equally admissible, with the first one more general and theoretically free. Further consideration on this point will be provided in the next pages.

1.2.1 Structures of data

The fundamental unit of analysis I consider is the *observational case*. I define it as any concrete episode in which one observer performs an observation directing

---

[3] In particular, I am referring to the theories that interpret perceptual outcomes as the results of "unconscious inferences" (Helmholtz, 1866; Brunswick, 1966) or "reasoning-like operations" (Gregory, 1980).

her/his gaze and attention toward a certain part of the physical environment. In any observational case, three kinds of organised *structures of data* can be distinguished, each corresponding to an information pattern concerning the observational case itself: the first structure corresponds to the optically relevant information of the physical context (the ‘distal stimulus’, in the standard terminology of psychophysics), the second structure corresponds to the information characterising the optical image projected on the retina (the ‘proximal stimulus’), and the third structure to the set of data regarding the phenomenal results experienced by the observer.

In the following pages, I will denote any single observational case by *o*. I will also denote each one of the three levels of the classic psychophysical distinction into distal and proximal stimuli and perceptual experience using three different prefixes. In particular, the three kinds of data structures will be respectively marked as S-structure (or structure of S-data, where S stands for physical Scene), I-structure (or structure of I-data, where I stands for retinal Image), and P-structure (or structure of P-data, where P stands for Phenomenal scene). I will use these prefixes in order to quickly, coherently and synthetically identify the characteristics of data belonging to distinct data structures.

The classical distinction into three layers of data can be explained from two different standpoints. The three structures can be seen in *causal order*, with data in the S-structure being concomitant causes of data of the I-structure through the “image formation process”, and data in the I-structure seen as concomitant causes of the data in P-structure through “percept formation process”. However, the same structures allow us to distinguish data in *epistemological terms*, without considering a definite causal order but focusing on how data are generally acquired as contents of knowledge concerning the same observational case. Thus, contents in the S-structure are acquired by the psychophysicist through optical and geometrical analysis of the environment; contents in the I-structure are acquired through optical and projective geometrical computations; contents in the P-structure are acquired from responses and descriptions provided by the observer.

1.2.2 Preliminary concepts

Considering the observational case *o*, a *depth-cue* is defined as a *property*[4] *X*(*o*) of a definite part of the visual stimulus, a part which I will generally conceive as an "aggregate" *A*(*o*) within the *I*-structure *I*(*o*) which inheres in the observational case *o*. I choose the term "aggregate" as a general form which allows for varying "numerical complexity" (the number of constituent units) of the parts carrying the properties to be considered. A depth-cue may be a property of one single unit or a pair of units in the I-structure[5], of one unit and an extended and undefined I-region (e.g., height in the field cue) or the collective property of sets of many units[6]. Generally, the higher the numerical complexity of the bearer of the depth-cue, the greater the cost in describing that depth-cue.

A depth-cue *X*(*o*) is generally considered in connection with a depth-percept *Y*(*o*); depth-percept *Y*(*o*) can be conceived as a *property* of a certain aggregate *B*(*o*) within the P-structure *P*(*o*) which inheres in the observational case *o*. When a connection between *X*(*o*) and *Y*(*o*) in a determined psychophysical problem is established, it means that these two contents correspond in position with each other, i.e., the part of space-time in structure *I*(o) occupied by aggregate *A*(*o*) is the counterpart of the part of space-time in structure *P*(*o*) occupied by aggregate *B*(*o*). The definition of a criterion of correspondence between the aggregates in the two structures is an open and intricate issue: one of the reasons is that the definition of aggregates pertaining to the I-structure is problematic and often there is no unique description of aggregates in the *I*-structure of the problem, since "stimulus definition is theory dependent" (Gordon, 1989, p. 234).

The core concept of this analysis is the definition of a *psychophysical inference rule*. With this term, I refer to a rule that allows predicting certain contents in the

---

[4] The term property is used here in its widest meaning, "as a generic term to cover both monadic (one-place, nonrelational) properties and (polyadic, multi-place) relations" (cf. Swoyer, 2000, §1.1).

[5] For instance, motion parallax X_μ can be computed as the angular difference of the position of two units A_1 and A_2, corresponding to the two images of the same object obtained at time t_1 and t_2. The aggregate carrying this depth-cue can be described as $A = (A_1, A_2)$.

[6] For instance, texture gradient of a textured I-region, composed of sub-units generally referred as texels.

perceptual result given certain contents in the optical stimulus. My immediate purpose is clarify this concept. The term $Y(o)$ corresponds to a P-content directly experienced by the observer during case *o*. From the standpoint of the psychophysicist, the P-content $Y(o)$ is an unknown term. By formulating a *psychophysical inference rule*, the psychophysicist tries to infer the exact value of $Y(o)$ on the basis of certain items of available information. In order to distinguish between both acceptations of the P-datum, I refer to them using the symbol $Y(o)$ for the P-datum when it is experienced by the observer, and the symbol $Y^*(o)$ for the P-unknown datum for the psychophysicist[7]. Briefly stated, a *psychophysical inference rule* is a rule enabling the scientist to make predictions about P-unknown Y^* on the basis of the I-datum X.

1.2.3 Systems of preliminary conditions

The terms constituting this definition are not exhaustive to express the rule as it is: in order to completely define a psychophysical inference rule, it is necessary to involve a "system of preliminary conditions" belonging to the I-side and the P-side of the psychophysical schema. Preliminary conditions can be considered as a definite set of *properties*, residing respectively on I-aggregate A and P-aggregate B, generally involved in the *logical construction* of the *psychophysical inference rule*. These conditions are necessary to establish the *existence and the uniqueness* of terms X and Y, and to constitute a *platform* on which the *information transfer* from X to Y is made possible. For the sake of simplicity, I will first consider the P-side of the problem, and then I will analyse the I-side of the same problem.

---

[7] From the standpoint of the perceiver, thematic P-content $Y(o)$ is a depth-related property in phenomenal scene $P(o)$, directly experienced by her/him during episode *o*. Instead, from the standpoint of the psychophysicist – when presumed to be distinct from the perceiver – the thematic P-content is an unknown term, the exact value of which s/he tries to 'infer' on the basis of certain items of information – among others, thematic I-content $X(o)$, as a datum in the optical stimulus. As I explained in the text, in order to better distinguish between both the acceptations of the thematic P-content in a psychophysical problem – that of a P-datum for the perceiver and that of a P-unknown for the psychophysicist – we refer to them with different symbols: $Y(o)$ for the former and $Y^*(o)$ for the latter.

In the previous paragraph, I stated that other properties beside property Y residing on some selected aggregate B are *explicitly* or *implicitly* involved in the logical construction of the problem. These properties are specified as basic requirements of aggregate B to ensure the existence of term Y, through a set Ψ of conditions, which I refer to as *preliminary P-conditions.* Furthermore, two or more distinct aggregates bearing the same property Y may be present in the same P-structure. We may presume that the system Ψ contains *discriminating conditions* that enable the identification of which one of these aggregates is the intended bearer of the thematic property Y.

As an example, let us consider the psychophysical problem of relative-size as a cue to depth. Thematic P-content Y, in the simplest versions of the problem, is the separation in depth between two P-units, B_1 and B_2, which are equal in form with each other (isomorphic P-units); more precisely, it is the ratio between the egocentric P-distances of units B_1 and B_2. Thus, a completely explicit statement of the corresponding inference rule should comprise a set of conditions ensuring that the P-structure of any observational case to be considered contains a pair of isomorphic units, as well as conditions specifying which pair of isomorphic units is intended, in case there were more than one pair of such units within the same P-structure. The intended pair $B=(B_1,B_2)$ of P-units is the thematic P-aggregate in the rule, and the set of conditions qualifying and identifying that pair is the system Ψ of preliminary P-conditions, which ensures the existence on B of thematic P-property Y (the separation in depth between units B_1 and B_2) but does not include Y; nor is it a sufficient basis for deriving Y through computation. Figure 1.2 shows a pictorial example illustrating the situation. The left B_1 and right B_2 rectangular objects, which are equal in form, constitute the thematic aggregate $B=(B_1,B_2)$ in the P-structure; the apparent separation in depth between both objects (which, as units in the phenomenal scene, are presumed to appear equal in size) is the thematic P-content Y in the psychophysical problem; the ratio between the height of object B_1 and that of object B_2 (now considered as parts of the optical stimulus, i.e., as I-units) is the thematic I-content X in the problem, to be used as a source of information for inference about Y.

Note that, in order to state that units B_1 and B_2 are 'equal in form', we must refer to the inner figural organisation, and the most natural way of doing this is to compare their perceptual organisations, i.e., the two rectangular objects as they are directly seen. Considering this, we state that the preliminary conditions qualifying the context for the thematic P-content of a psychophysical problem belong to the P-side of the psychophysical scheme, i.e., they are requirements directly concerning the P-structure (rather than the I-structure) of an observational case.

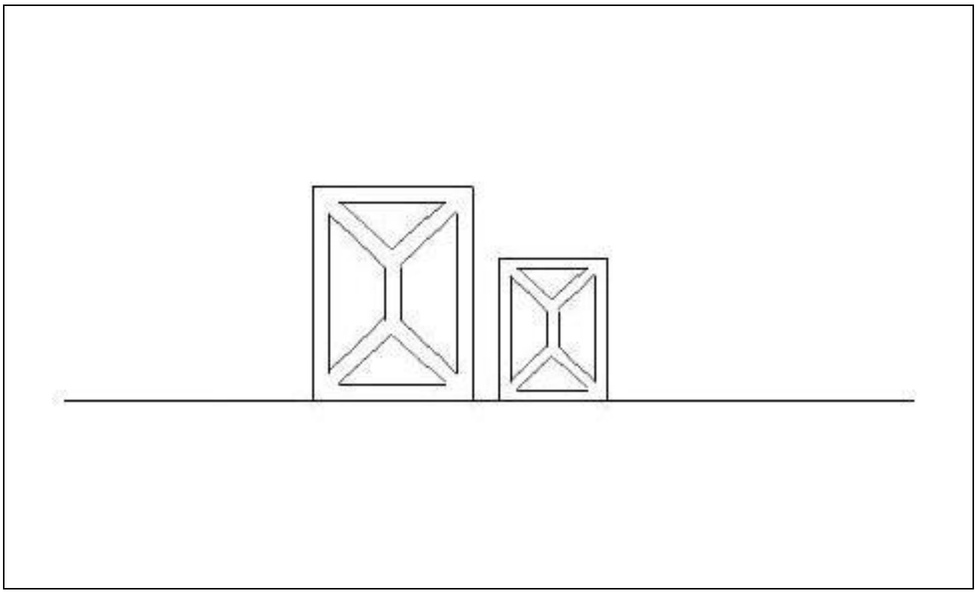


Figure 1.2. An example illustrating the relative-size cue for exocentric distance in depth.

With reference to system Ψ of preliminary P-conditions, we may abstractly consider the universe O_Ψ of observational cases such that, for each $o \in O_\Psi$, there exists one single aggregate $B(o)$ complying with the conditions in Ψ. For each o, o' $\in O_\Psi$, the conditions specified by Ψ are equivalent to each other in the aggregate $B(o)$ in $P(o)$ and aggregate $B'(o)$ in $P'(o)$, i.e., conditions in Ψ behave like *constants* across universe O_Ψ. Since P-content Y is not uniquely specified or deducible by Ψ, for some $o \neq o' \in O_\Psi$, it may happen that $Y(o) \neq Y'(o)$. Thus, content Y qualifies as a *variable* when it is referred to universe O_Ψ of observational cases. In set- theoretic terms, variable Y is a function Y: $O_\Psi \rightarrow \Upsilon$, where Υ is a homogeneous class of comparable *P*-properties. The complexity and number of elements of $\mathcal{Y}$ may be remarkably different for different *P*-properties: for instance, ordinal information (which one of two different objects is farther from the observer) can be indicated by a simple discrete dichotomous variable, whereas absolute distance needs a variable with a wide and continuous range.

Analogous considerations can be made relating to the I-side of the problem: considering the I-property *X* of the aggregate A in the I-structure, other properties of the aggregate *A* must be presumed in the logical construction of the psychophysical problem to specify *X* as a depth-cue. For example, if an I-unit does not have in the I-structure any other I-unit equal in form to it, then a relative-size cue involving that I-unit cannot be computed; if an I-structure corresponds to an observational case without any motion (the perceiver is not moving with respect to the observed object, nor is the object moving with respect to the perceiver), then no motion-related cue to depth (motion parallax, optic flow, etc.) can be taken into consideration; and so on. This set of preliminary I-conditions is here referred to as Φ. As in the previous case, the system of preliminary I-conditions is specifically referred to the intended aggregate A; it ensures the existence and computability of the property X on that aggregate, but does not allow to determine the value of *X* by mere computation.

Analogously to some previous considerations, we may refer to some universe O_Φ of observational cases such that, for each $o \in O_\Phi$, there is one single aggregate $A(o)$ satisfying every condition in Φ. For some $o \neq \mathrm{o}' \in O_\Phi$, it could happen that $X(o) \neq X'(o)$, so that *I*-contents (i.e., depth-cues) are qualified as *variables*, while conditions in Φ are *constants*. In set-theoretic terms, *X* is a function $X\colon O_\Sigma \rightarrow \mathcal{X}$, where $\mathcal{X}$ is a homogeneous class of comparable *I*-properties. The complexity and number of elements of $\mathcal{X}$ may be remarkably various for different *I*-properties: for instance, occlusion as a cue to depth has a very restricted range (when comparing two different objects, a simple discrete dichotomous variable suffices to indicate which object is occluding and which object is occluded), while cues like motion parallax have a wider and continuous range.

1.2.4 The constraining effects of stimulus factors

The last component of a psychophysical inference rule is the *constraining effect* F of I-content X on P-content Y. The term ‘constraining’ indicates that I-content X is able to reduce the uncertainty about the P-content Y, by determining a subset of compatible values for Y, within the range $\mathcal{Y}$ of possible values for Y[8]. This effect could be *punctual* or *limiting*.

The constraining effect is *punctual* if, given any observational case o and the I-content (depth-cue) $X(o)$, a single solution $F(X(o)$ is determined for unknown $Y^*(o)$ (thus, the inferential result can be expressed by equation $F(X(o)) = Y^*(o)$). Solution $F(X(o))$ is one single element of the homogeneous class $\mathcal{Y}$; F is valid on o if and only if $Y(o) = F(X(o))$ (i.e. the real value $Y(o)$ is equal to the predicted value $Y^*(o)$).

To illustrate this point, let us consider occlusion as a cue to depth. For the sake of simplicity, we will refer to the simple stimulus configuration shown in Figure 1.3, corresponding to the observational case o. The picture is composed by two I-units, A_1 and A_2, which compose the aggregate A. Let us consider the P-units B_1 and B_2, corresponding to the perceptual rendering of I-units A_1 and A_2, and corresponding to the P-aggregate. Two perceptual solutions are possible: in the first one, referred to as “mosaic interpretation”, the two P-units are seen respectively as a rectangle and an L-shaped form; in the second one, referred to as “occlusion interpretation”, the two P-units are seen as two rectangles differently positioned in depth.

Occlusion as a cue to depth is defined on an I-aggregate A composed by (at least) two adjacent I-units, A_1 and A_2. The set of preliminary I-conditions Φ for this cue are: aggregate A carrying the depth-cue $X(o)$ is composed by two I-units, A_1 and A_2; the two I-units are contiguous; the adjacent contour of the two I-units are T-

---

[8] The term “constraint” is widely used in the computational literature, especially in the paradigm known as Constraint Satisfaction Problem (CSP) or Constraint Network Theory (Montanari, 1974; Dechter, 2003). The analogy with our use of the word ‘constraint’ is close: finding a solution to a CSP is defined as the finding of ”an assignment of values to a given set of variables subject to constraints on the values which can be assigned simultaneously to certain specified subset of variables” (Jeavons, Cohen & Cooper, 1998, p. 251). It is important to stress that constraints here considered are relational ones, and are different from the concept of “prior constraints” in vision science literature, as I shall discuss later.

junctions or Y-junctions. Depth-cue $X(o)$ is a variable that may take value 0 if A_1 is the occluding object, and 1 if A_2 is the occluding object[9]. In the specified case o, $X(o) = 1$[10].

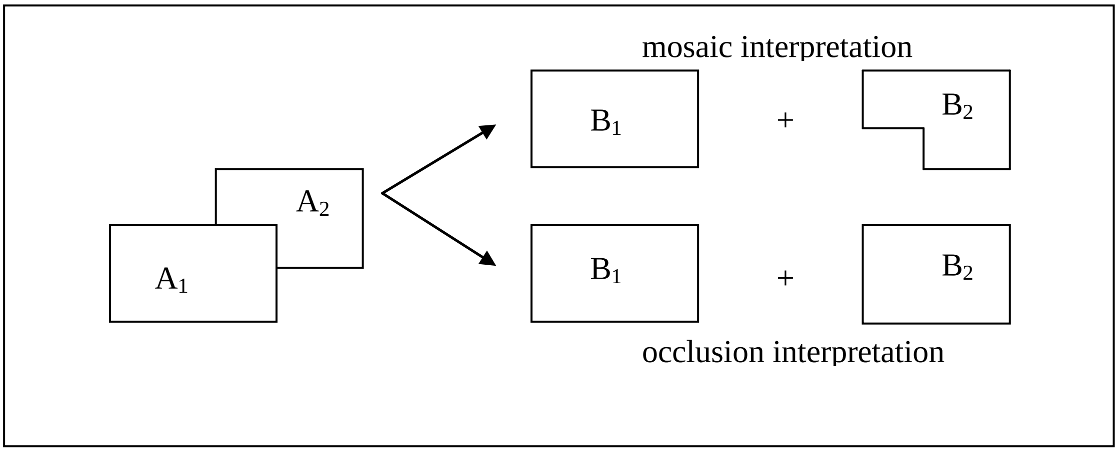


Figure 1.3. Image adapted from Van Lier, van der Helm, Leeuwwenberg (1994). The stimulus pattern discussed in the text is shown on the left. The two possible perceptual solutions are shown on the right.

The P-content $Y(o)$ is carried by the aggregate B. P-content $Y(o)$ can take two values[11]: 0 if P-unit B_1 is seen as farther than P-unit B_2, 1 if P-unit B_2 is seen as farther than P-unit B_1. The inferential result for the case o and the constraining effect F can be described by the following equation:

$$X(o) = 1 \Rightarrow Y^*(o) = F(X(o)) = 1.$$

More generally, the constraining effect may be simply *limiting* in character. A psychophysical inference rule is *limiting* if, given any observational case o and the I-content $X(o)$, a subset of elements of the class $\mathcal{Y}$ is the solution, so that $Y^*(o) \in$

---

[9] The attribution of alternative values 0 and 1 to the two possible alternatives is arbitrary.

[10] There is no definitive explanation on which factors contributes to the characterization of an I-unit as occluding or occluded. For some possible hypotheses, cf. the review in Van Lier, van der Heelm & Leeuwenberg (1995).

[11] The attribution of alternative values -1, 0 and 1 to the two possible alternatives is arbitrary.

$F(X(o))$. F is valid if and only if $Y(o) \in F(X(o)$. The smaller set $F(X(o))$, the more precise the inference allowed by the rule. In the extreme, set $F(X(o))$ may be a singleton, which is the case of a *punctual* constraining effect by the depth-cue through the rule.

Let us consider the previous example; if the P-content is the pair $Y = (y_1, y_2)$ of egocentric distances of the two objects corresponding to P-units B_1 and B_2, the constraining effect F can be described by the following equation:

$$X(o) = 1 \Rightarrow Y^*(o) \in F(X(o)) = \{(\mathrm{y1}, \mathrm{y2}) \in \mathfrak{R}^+ \times \mathfrak{R}^+ : y_2 > y_1\},$$

where $\mathfrak{R}^+$ is the set of positive real numbers.

As shown by these examples, the strength of the constraining effect acted upon by a depth-cue depends on which depth-percept is considered for inference. In principle, it may be that a depth-cue does not suffice to constrain a punctual solution for a specific depth-percept if considered singly, as this constraint is a limiting one. Yet, it is possible for two limiting constraining effects (acted upon by distinct depth-cues) to determine a punctual solution by combining their effects. A discussion about cues combination will follow in the next pages.

1.2.5 Summary

In summary, we may state that a psychophysical inference rule can be described by this list of components: (Φ, Ψ, X, Y, F) (Figure 1.4). The rule can be referred to the universe $O_{\Phi\Psi} = O_\Phi \cap O_\Psi$ of possible observational cases so that for each $o \in O_{\Phi\Psi}$, there is one single aggregate A in $I(o)$ satisfying conditions in Φ, one single B aggregate in $P(o)$ satisfying all conditions in Ψ, and a correspondence in space and time holds true between the two aggregates. Since I-content X and P-content Y are functions having universe $O_{\Phi\Psi}$ as their domain and respectively class $\mathcal{X}$ of possible I-properties and class $\mathcal{Y}$ of possible P-properties as their range, the constraining effect F in the inference rule may be conceived as:

- a function from class $\mathcal{X}$ to the power set $\mathcal{P}(\mathcal{Y})$ of class $\mathcal{Y}$, if the constraining effect F is limiting;

- a function from class $\mathcal{X}$ to $\mathcal{Y}$, if the constraining effect F is punctual.

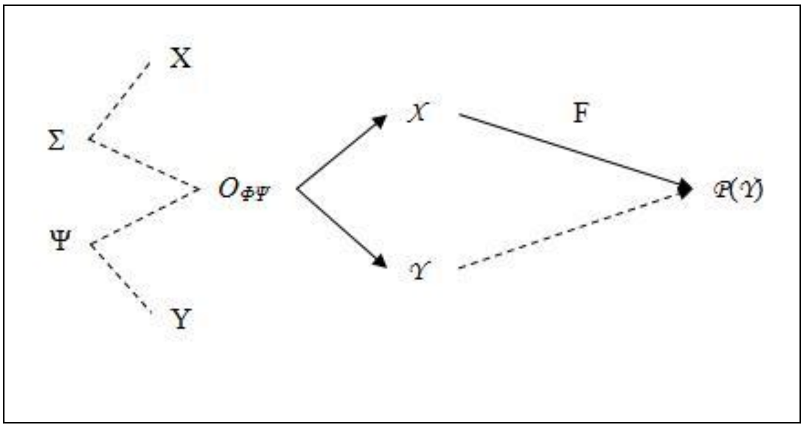


Figure 1.4. The components of a psychophysical Inference Rule

1.3 Prior constraints as additional sources of information

The approach outlined so far allows us to describe the concept of "prior constraint", which is typical in the modern computational approach to vision science. *Prior constraints* corresponds to a-priori assumptions on which human observer rely, used when information is ambiguous (Gregory, 1980; Rock, 1983)[12].

The term *constraint* is used in computer science to indicate a rule that reduces the set of admissible solutions to a problem, when this is judged to be an "ill-posed problem" [13]. Resting upon the idea that "vision is inverse optics" (in some respects), some authors stated that most of the problems in vision are substantially ill-posed problems (Poggio, Torre & Koch, 1985, p. 314). In computational vision, constraints are often in their original sense, i.e., they are used to find unique solutions to ill-posed problems. Some authors expressed the idea that "rather than considering visual

---

[12] The need for natural and artificial constraints is questioned by Gibson (1979) who proposed a theory (referred to as "direct vision theory" or "ecological optics") in which visual information is completely specified in the optical array and its invariant properties, with no need for further assumptions about the environment.

[13] According to Hadamard (1923), a problem is "well defined" when it has a solution, this solution is unique and depends continuously on the initial data. "Ill posed" problems fail to satisfy one of these criteria.

constraints as merely a technique to render a problem well-posed, one can treat [them] on an equal basis with sensory data“(Mamassian & Landy, 2001, p. 2653).

Clark and Yuille (1990, p. 6) classify prior-constraints in three different categories:

- *Physical constraints*: constraints based on laws of geometry and physics, which are generally valid. Examples of this category are some elementary regularities defined in Euclidean geometry and the Newton’s laws of motions;
- *Natural constraints*: constraints derived from everyday observation of the environment and which are usually valid. Examples of these constraints are surface smoothness, object rigidity, but also higher level heuristics such as the “generic viewpoint” constraint (Freeman, 1996);
- *Artificial constraints*: constraints based on high level knowledge or expectations about the environment are incorporated into sensory process; one possible example is the assumption that the ceilings of a room have uniform height, leading to a wrong perception in the well-known illustration of “Ames Room”.

As an example, I mention the so called “shading-constraint”, which states that the position of a source of light is usually assumed to be above the observed object, with a bias to the left between 20° and 30° off the vertical (Mamassian & Goutcher, 2001). Todd & Mingolla (1983) and Ramachandran (1988) showed that, given a shading pattern for which two possible interpretations are possible (for instance, a certain pattern is produced by a convex object if light comes from below or by a concave object if light comes from above), then the unambiguous location of the source of light can provide an unambiguous interpretation of the depth of the visual stimuli (source of light is located above the stimulus, so the pattern is seen as a concave object). Mamassian and Landy (2001) provide an example of the intervention of this constraint, reported in Figure 1.5.

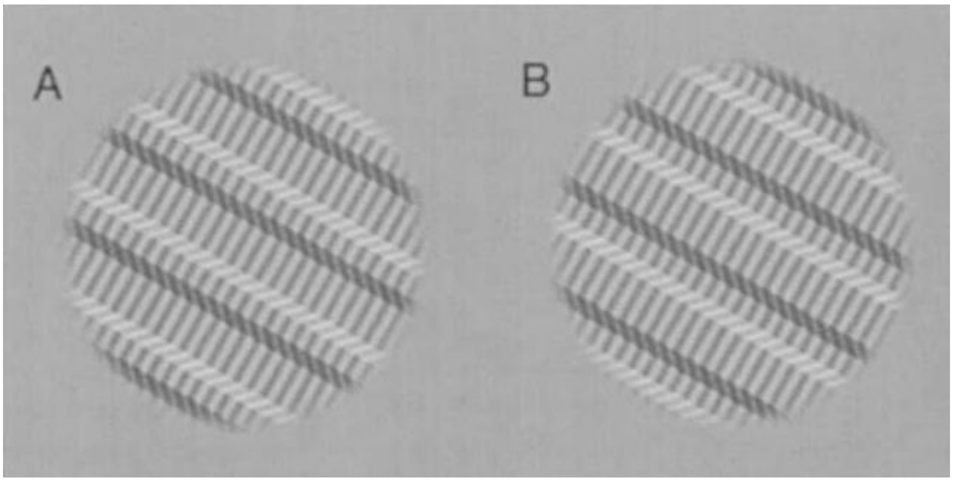


Figure 1.5. An example from Mamassian and Landy (2001). It represents two patches of surfaces that appear to be formed of a series of raised strips, running from left to right. Observers usually describe patch A as having narrow ridges and wide valleys. If the patch is rotated by 180° (patch B), then the interpretation that observers report consists in narrow strips as indented and wider strips as emerging. The different interpretations can be explained by considering the different reflectance of the bevelled regions between the strips, which are seen as illuminated by a single source of light located above the patch, but both solutions are plausible, at least theoretically.

The set-theoretic approach proposed in Section 1.2 can be extended in order to frame the concept of "prior constraint". Consider a generic P-aggregate B, on which two generic perceptual properties are residing: the first one is some depth-percept Y, and the second one is some perceptual property W, not necessarily depth-related. At the same time, consider the I-aggregate A, corresponding in position with B, on which a depth-cue X is residing. For an observational case o, it may occur that $X(o)$ has no constraining effect on $Y(o)$ (i.e., set $F(X(o))$ equals the whole set $\mathcal{Y}$ of admissible solutions to P-unknown Y^*), but it has a constraining effect on (Y, W)(i.e., set $K(X(o))$ is a proper subset of Cartesian product $\mathcal{Y} \times \mathcal{W}$, which is the set of admissible solutions to P-unknown W^*). Set $K(X(o)) \subset \mathcal{Y} \times \mathcal{W}$ may be a function K_0 from $\mathcal{W}$ to $\mathcal{Y}$. Thus, if there is another source of information V beside $X(o)$, and this source allows the determination of a solution $w \in \mathcal{W}$ to unknown W*, then one solution $y = K_0(w) \in \mathcal{Y}$ to unknown Y^* becomes also determined (see also Figure 1.6).

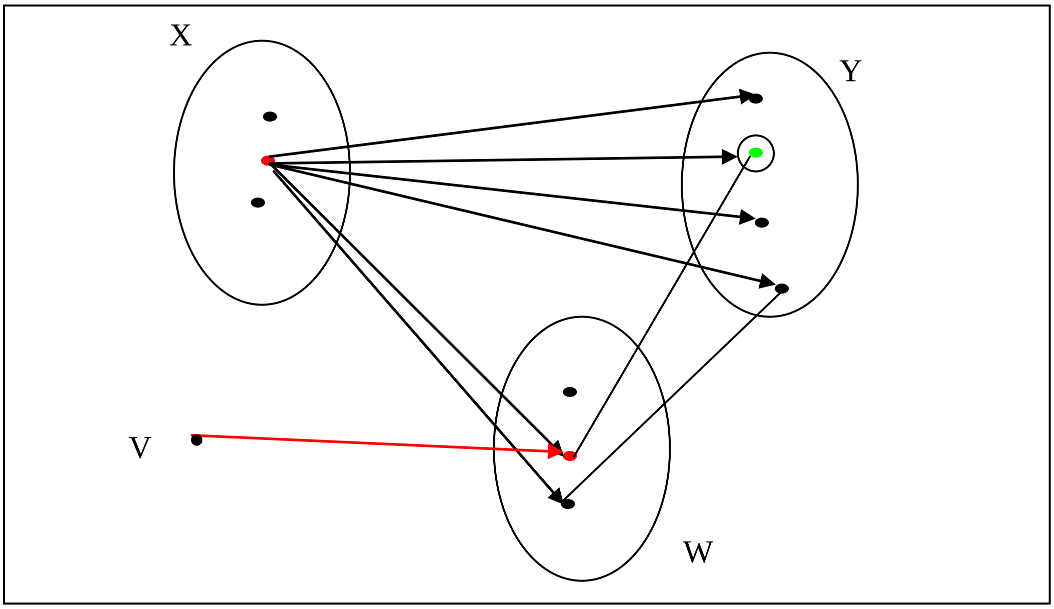


Figure 1.6. Graphical representation of the constraining effect of a prior constraint. Depth-cue X(*o)* takes the value *x* (represented by the red dot in the set *X*), which alone is not enough to uniquely specify *Y*. The intervention of an external additional source of information, *V*, promotes value *w* (represented by the red dot in the set *W*), constraining the possible solutions of *Y*, so that a unique solution *y* (represented by the green dot surrounded by a circle) in the observational case *o* becomes determined.

In the presumed terms, a "prior constraint" can be defined as an intrinsic tendency of the visual system to favour a certain solution P-unknown *W** which allows, through the cooperation with the I-content *X*, to determine a unique solution to *P*-unknown *Y*.

As an example, let us consider the stimulus situation described in Figure 1.5. Let us refer to an observational case *o* in which an observer is looking at part A of the figure. One possible segmentation of the stimulus is into 15 I-units, corresponding to the raised strips running from left to right (see Figure 1.7). I will refer to strips as a_1, a_2, ..., a_{15}, going from the upper strip to the lower one. In order to simplify the analysis of this case, I will focus only on the aggregate A = (a_6, a_7, a_8, a_9), since the same pattern is repeated continuously along the vertical axis.

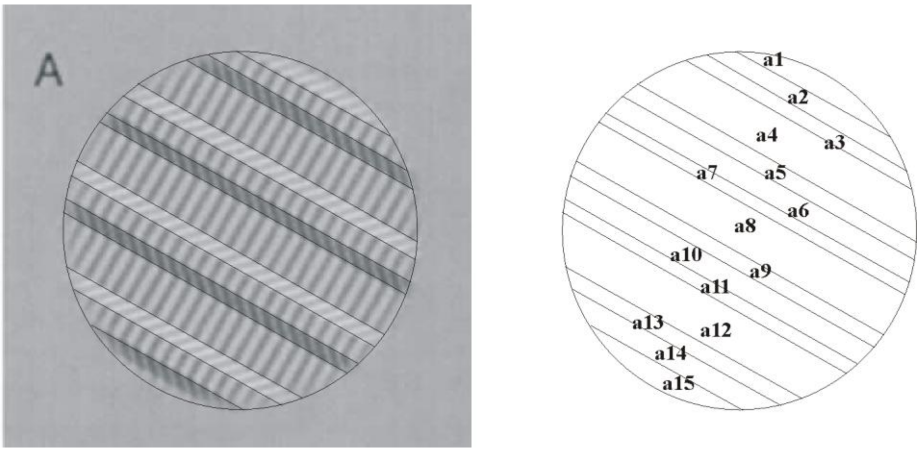


Figure 1.7. Graphical representation of the segmentation of Figure 1-3 (A) proposed in the text.

I-aggregate A corresponds in position to the P-aggregate B, which can be conceived as a three dimensional structure composed by surfaces b_6, b_7, b_8, b_9, with their own orientation in space. We can identify two terms on the P-structure considered: P-unknown depth-related property Y^*, admitting alternative solutions “appearing convex” (denoted by y_1) and “appearing concave” (denoted by y_1); P-unknown depth-related property W^*, concerning the grey colour of the surfaces composing aggregate B, admitting 15 solutions (i.e., the number of partition of a set composed by four elements), from w_1 (all the P-surfaces have the same colour) to w_{15} (no P-surface have the same colour of another P-surface). Depth-cue (X) has no constraining effect on depth-percept Y taken alone, since both solutions are compatible with it, but a single solution can be derived by considering the intervention of an unknown light-related P-content V^*, admitting solutions “the light comes from above” and “the light comes from below” (denoted by v_1 and v_2 respectively). Thus, in the stimulus condition described in Figure 1.7, a definite relational constraint on set of variables (Y^*, W^*, V^*) may be presumed, expressed by the following set of triples:

$$K(X(o) = \{(y_1,w_1,v_1), (y_1,w_2,v_1), (y_1,w_2,v_2), (y_1,w_3,v_1), (y_1,w_3,v_2), (y_1,w_4,v_1),$$
$$(y_1,w_4,v_2), (y_1,w_5,v_1), (y_1,w_5,v_2), (y_1,w_6,v_1), (y_1,w_6,v_2), \ldots, (y_1,w_{16},v_1), (y_1,w_{16},v_2),$$

(y_2,w_1,v_2), (y_2,w_2,v_1), (y_2,w_2,v_1), (y_2,w_2,v_2), (y_2,w_3,v_1), (y_2,w_3,v_2), (y_2,w_4,v_1), (y_2,w_4,v_2), (y_2,w_5,v_1), (y_2,w_5,v_2), …, (y_2,w_{15},v_1), $(y_2,w_{15},v_2)\}$.

The so-called light constraint favours the solution v_1 ("light come from above") for unknown V^*; further, the visual system would also prefer a solution w_1 for unknown W^* (i.e., seeing the four *P*-units as equal in texture colour). This is an additional prior constraint, which may be related to the so-called "minimum principle" in visual perception (cf. Hatfield and Epstein, 1985). This principle, when referred to the present case, implies the rule that perceiving a uniform texture on a folded surface is structurally simpler than perceiving four surfaces with different textures. These additional sources of information determine y_1 as the only admissible solution for unknown Y^*[14]. A prior constraint is then supposed to intervene as a "supplementary source of information", the root of which, however, is in the visual system, not in the optical stimuli.

1.4 Integration of depth information

Cues may differ from each other as regard two aspects: *quality*, i.e., the type of cue, and *position*, i.e., the aggregates on which cues are residing. Given any observational case *o*, it may occur that several and different depth-cues are available in the I-structure. For any couple of cues belonging to the I-structure of the observational case *o*, three cases are possible:

- Cues may differ only in quality: they are of different type but reside on the same I- aggregate;
- Cues may differ only in position: they are of the same type but reside on different I-aggregates;
- Cues may differ both in quality and in position.

---

[14] Constraint Network Theory (see note 7) provides a theoretical framework and efficient algorithms for solving problems involving relational constraints like the one described in this example.

The general problem of “integration of depth information” concerns the modalities in which several depth cues may interact with one another in determining the P-structure in one observational case *o*.

A conspicuous part of contemporary research about depth perception is specifically aimed at understanding how two or more cues of different types and residing on the same I-aggregate determine the depth-related properties of the corresponding P-aggregate. In the next section, I will provide a preliminary discussion of this problem, using the concepts previously introduced. In the next chapter, I will analyse the case which constitutes the central theme of this dissertation, i.e., the case in which cues residing on different aggregates concur to determine depth-related properties of the P-structure,

1.4.1 Modalities of integration of depth cue information

For any observational case *o* we may consider, multiple sources of depth-related information are simultaneously available. These sources of information are often redundant (to some extent), and support similar solutions for the same or closely related depth percepts. However, abundant evidence showed that depth cues differ widely in their perceptual effectiveness (Cutting & Vishton, 1995; Sedgwick, 2001). In particular, Sedgwick (2001) distinguishes between two types of differences in effectiveness among cues: the first one is based upon *situational* differences, and the second one upon the *sensitivity* of the visual system.

The first type of differences is related to the limits of resolution of different cues for different distances: for instance, occlusion as a cue to depth limitations are related to the visual acuity of the observer, whereas stereopsy is effective only at short distances (cf. Cutting & Vishton, 1995).

The second type of differences is related to the varying sensitivity of the visual system for different features of the information: for instance, Cutting & Millard (1984), studying the effectiveness of different texture gradients on different depth percepts, found that perspective gradient is relevant in determining depth-related perceptual properties of flat surfaces, but it is quite irrelevant in determining the curvature of a surface. On the contrary, compression gradient is relevant in determining depth-related perceptual properties of a curved surface, but quite irrelevant in determining depth-related properties of flat surfaces.

Indeed, the integration of different sources of information is necessary in order to obtain a stable and reliable perception of the characteristics of the visual world. As Todd (1985) pointed out: ”Objects and events in a natural environment can be multiply specified by many different sources of information, each of which is detected by a specialized processing module with its own individual limitations. In any given situation, we would expect to obtain erroneous outputs from some of these modules because of inappropriate viewing conditions, but it would be most unlikely for two or more of them to fail in exactly the same way” (p. 708).

Different modalities of depth-cues combination have been conjectured and discussed in the recent years. For instance, Howard and Rogers (2002) proposed a list of 11 different modalities. Bülthoff and Mallot (1990) restricted the description of possible interactions to 4 principal modalities: *accumulation* (information carried out by each depth cues are combined to specify a more stable and reliable depth-percept); *cooperation* (estimate offered by each source of information are insufficient to specify a depth percept, but sources joint effect is able to support a definite one); *disambiguation* (a cue may resolve the ambiguity inherent in the information provided by another cue) and *veto* (in case the information offered by two cues has a large discrepancy, one of the two sources of information may be ignored, so that the final estimate relies on one of them only).

Clark and Yuille (1990) classify hypotheses concerning depth-cues integration (or fusion, according to their glossary) into two major categories, referred to as “weakly” and “strongly coupled data fusion” mechanisms[15]. In the following years, this distinction has been adapted to human vision (cf. Landy, Johnston, Maloney & Young, 1995).

---

[15] The classification of fusion models defined by Clark and Yuille (1990) is a part of the discussion on the machine vision problem. Actually, the authors refer to fusion algorithms in machine vision, not to human vision mechanisms. As the authors underline in their book, this classification can be easily applied to general mechanisms of the human visual system.

In “Weak Fusion Models” (WF models), each depth cue is presumed to be processed independently and separately from concomitant depth cues. Single estimates are presumed to combine only at the end of the elaboration processes for separate sources of information. Examples of WF models are: linear models including a weighted combination of the outputs, with weights dynamically assigned on the basis of the reliability of the sources of information (Dosher, Sperling & Wurst, 1986; Bruno & Cutting, 1988), algebraic fusion models as those proposed by Alomoinos and Shulman (1989) or a combination of these two modalities.

In “Strong Fusion Models” (SF models), the operation performed by the visual system to elaborate a single source of information may be affected by the output of another operation or another source of information, so that data are not processed independently. Typical examples of this kind of models use feedback loops, feed forward adaptation (Clark and Yuille, 1990), constraint network based procedures (Burigana, 1999).

The concepts presented in Section 1.2 may be applied to the general problem of depth cues integration on one I-aggregate. For the sake of simplicity, I will refer to cases in which only two depth-cues are available, but the considerations presented below can be easily extended to more complex cases.

The terms involved in the analysis of depth cues integration the depth percept Y, the P-aggregate B on which it is residing, the depth-cues X_1 and X_2, the I-aggregate A corresponding in position with B on which the two depth cues are residing, the systems of preliminary I-conditions Φ and preliminary P-conditions Ψ, and the components F1 and F2, expressing the constraining effect respectively of X_1 and X_2 on Y. $O_{\Pi\Sigma} = O_{\Pi} \cap O_{\Sigma}$ is the universe of observational cases satisfying all conditions in $\Sigma \cup \Pi$. Following the discussion in Section 1.2, with reference to the universe $O_{\Pi\Sigma}$, I-properties X_1 and X_2 and P-property Y are functions having universe $O_{\Pi\Sigma}$ as their domain, and respectively classes $\mathcal{X}_1$, $\mathcal{X}_2$ and $\mathcal{Y}$ as their ranges. Thus, F_1 and F_2 become functions with classes $\mathcal{X}_1$ and $\mathcal{X}_2$ as their domains, and power set $\mathcal{P}(\mathcal{Y})$ of class $\mathcal{Y}$ as their codomain.

On considering any generic observational case $o \in O_{\Pi\Sigma}$, it may occur that both constraining effects F_1 and F_2 are *punctual* and *concordant*, so that $F(X_1(o)) = F(X_2(o)) = \{y\}$. This situation corresponds to the one generally referred to as "cue accumulation". In these conditions, the inference based on depth-cues X_1 and X_2 for P-unknown Y^* is that y is its solution, i.e., $Y^* = \{y\}$.

A different case would be the one in which constraining effects F_1 and F_2 are *limiting* and *compatible*. In set-theoretic terms, this could be expressed by $F_1(X_1(o)) \cap F_2(X_2(o)) \neq \varnothing$, i.e., given the observation case o, there are solution in the set Υ which are compatible with both constraining effects of X_1 and X_2 on Y. It may occur that

$$E(F_1(X_1(o)) \cap F_2(X_2(o))) < \min(E(E(F_1(X_1(o)), E(F_2(X_2(o)))$$

where, for any $Q \subseteq \Upsilon$, $Q(E)$ is a measure of set Q in its extent. This situation would correspond to the one defined as *cooperation* (but it may be easily adapted to the case of *disambiguation*). This would mean that the combined effects of cues X_1 and X_2 are suitable for specifying a set of solutions to P-unknown Y^* that has reduced extension compared to the one specified by the two depth-cues singly considered. In the extreme case, it may happen that $F_1(X_1(o)) \cap F_2(X_2(o)) =\{y\}$, so that only one solution to P-unknown Y^* is specified by the combined effect of these depth cues.

The last case presented in this analysis is the one in which constraining effects F_1 and F_2 are *limiting* and *incompatible* with each other. In set-theoretic terms, this could be expressed by $F_1(X_1(o)) \cap F_2(X_2(o)) =\varnothing$, so that no solution to P-unknown Y^* consistent with both cues X_1 and X_2 would exist. This case is often referred as "cue conflict", and rarely occurs in ecological context. Furthermore, "with large discrepancies between information sources, the nervous system may exhibit robust behaviour, in which a discrepant source is discounted" (Ernst and Banks, 2002, p. 432). Indeed, such a situation is quite common in the psychophysics of vision, and can be considered as experimental artefact. Researchers have often constructed visual situations in which two or more depth cues located on the same object convey discrepant information, and the subjects' responses are used to analyse the role or the weight of each depth cue in contributing to the final estimate of depth (Gogel, 1972; Cutting and Millard, 1984; Bülthoff and Mallot, 1990). The usefulness of such a paradigm has often been proved: however, it cannot be applied without paying attention to the results treatment.

Chapter 2

Intrinsic and Extrinsic Properties of Stimulus Factors

The study of the different variables concurring to the perception of objects within a scene can be carried out at different levels. According to Andersen, Braunstein and Saidpour (1998), three main separate levels can be distinguished:

- The first level corresponds to the scene, i.e., the entire visible 3-D environment; the variables of interest are those that affect the overall perceived depth of a scene. Studies belonging to this category are those analysing how the perceived depth of an image is influenced when viewed within a frame, such as a looking glass, a window or a mirror (Schlosberg, 1941; Goldstein, 1987; Reinhardt-Rutland, 1999; Lawson, Bertamini and Liu, 2007) or the effect of viewpoint on scene perception (Sedgwick, 1991).
- The second level corresponds to the layout, i.e., the relative positions of surfaces and objects within the scene. Andersen et al. (1998) include in this category those studies addressing the degree of veridical perception of layout (Gibson, 1950b; Da Silva, 1985), but even those studies focused on the effect of separate and combined depth-cues and how they affect depth perception.
- The third level corresponds to single characteristics of surfaces and objects. This category comprises those studies coping with problems like orientation of a single surface, or objects shape.

The distinction here described is notable as it shows how “the perception of 3-D scenes usually involves an integration of information about overall scene depth, the layout of relative position of objects in the scene, and the properties of the objects within the scene” (Bian, Braunstein & Andersen, 2005, p. 802). Unfortunately, variables belonging to each one of these levels have mostly been studied in isolation, and “relatively little is known about how information about scene depth, layout, and object properties from viewer-centred and object-centred sources is integrated to provide a perception of a 3-D scene” (Bian et al., 2005, p. 802).

In the present chapter, I will provide some considerations on a possible distinction between factors concurring to depth perception, factors which can be either extrinsic or intrinsic, and how they interact with one another. Four experiments aiming at the analysis of how intrinsic and extrinsic factors jointly concur in determining the perception of slant of surfaces will be present in Chapters 3 and 4.

2.1 Integration among depth cues in different positions

In Chapter 1, I described the general problem of integration of depth information, i.e., how several depth-cues in one observational case may interact with one another in determining the perceptual result. For any observational case, I distinguished among three different potential situations, based on the idea that, for any two cues in one observational case, these cues may be equivalent in type (but not in position), in position but not in type, of both in type and in position (cf. p. 26). In the same chapter, I focused on the analysis of one of these situations, which is also the most common in the current depth perception literature, i.e., the case in which two (or more) depth cues have a different type but are residing on the same I-aggregate.

In this chapter, I will focus on the situation in which two cues are not residing on the same I-aggregate. In general, we may describe the situation as composed by a depth-percept Y residing on a P-aggregate B, and a number of depth cues $X_1, ..., X_m$ residing on I-aggregates $A_1, ..., A_m$, some of which may not correspond in position with B. Let $Y_1, ..., Y_m$ be depth-percepts residing on I-aggregates $B_1, ..., B_m$ and *directly conditional*[1] on depth-cues $X_1, ..., X_m$ (so that B_i corresponds in position with A_i, for $i = 1, ..., m$). Let depth-percepts $Y_1, ..., Y_m$, and Y be involved in a $(m+1)$-ary relational constraint C (constraint C may be represented as a subset of $\mathcal{Y}_1 \times ... \times \mathcal{Y}_m \times \mathcal{Y}$, i.e., the Cartesian product of the possible values of the P-properties considered). In these conditions, given the direct constraining effect of $X_1, ..., X_m$, on depth-percepts

---

[1] A "direct conditioning" of a generic depth cue X_i on a generic depth percept Y_i corresponds to the constraining effect of depth cue X_i residing on the aggregate A_i on a depth percept Y_i residing on the aggregate B_i, when A_i and B_i correspond to each other in position.

$Y_1, ..., Y_m$, and given the relational constraint C, depth cues $X_1, ..., X_m$, may induce an *indirect constraining effect* on depth-percept Y.

To exemplify, let us compare the pictures presented in the two parts of Figure 2.1, denoted by (i) and (ii).

The picture in Figure 2.1(i), presumed to be the I-structure of an observational case o_1, is composed of 4 I-units, denoted by a_1, a_2, a_3, and a_4, which are the upper circle, the right circle, the lower circle and the left circle. Let us consider the P-units b_1, b_2, b_3 and b_4, corresponding to the perceptual rendering within o_1 of I-units a_1, a_2, a_3, and a_4. Let us indicate by $y(b_i)$ the egocentric perceptual distances of the P-units b_i, for $i = 1, ..., 4$. We may be interested in determining the relative (ordinal) position in depth of the four circles with respect to one another. We denote the P-unknown depth-related property as R^*. The set of alternative possible solutions to P-unknown R^* is denoted as $\mathcal{R}$, and is a set of 77 solutions[2]. Among others, one possibility is that circles have the same perceptual egocentric distance from the observer, i.e., $y(b_1) = y(b_2) = y(b_3) = y(b_4)$, but it may also happen that the circle referred to as b_1 appears to be farther than the others, which, in turn, have the same egocentric distance, i.e., $y(b_1) > y(b_2) = y(b_3) = y(b_4)$, and so on.

In principle, since no definite depth-cue is acting on the I-aggregate $A_1 = (a_1, a_2, a_3, a_4)$, there is no constraining effect on any possible solution to P-unknown $R^*(o_1)$, so that any solution is plausible. Actually, subjects report that circles are lying on the same frontoparallel plane (corresponding to the solution $y(b_1) = y(b_2) = y(b_3) = y(b_4)$). A possible explanation hypothesizes the intervention of a prior-constraint, “equidistance tendency”, which reflects “the tendency for objects or parts of objects, in the absence of effective distance cues, to appear visually at the same distance as each other with the strength of this tendency being inversely related to the directional separation of the objects or part”[3] (Gogel, 1965, p. 153).

---

[2] The number of possible solutions is calculated multiplying the number of partition of a set of four elements respectively in 1, 2, 3 and 4 subsets by the number of possible permutations of elements in a set with cardinality respectively of 1, 2, 3 and 4.

[3] The change in effectiveness of perceptual interactions as a function of object separation is strictly connected to the “adjacency principle”, which affirms that the “effectiveness of cues between objects

Let us now consider Figure 2.1 (ii), presumed to be the I-structure of another hypothetical observational case o_2. The picture contains the same four I-units of Figure 2.1(i), a_1, a_2, a_3, and a_4, constituting the I-aggregate $A_1 = (a_1, a_2, a_3, a_4)$, and nine further I-units, constituting the I-aggregate $A_2 = (a_5, a_6, a_7\ a_8, a_9, a_{10}, a_{11}, a_{12}, a_{13})$.

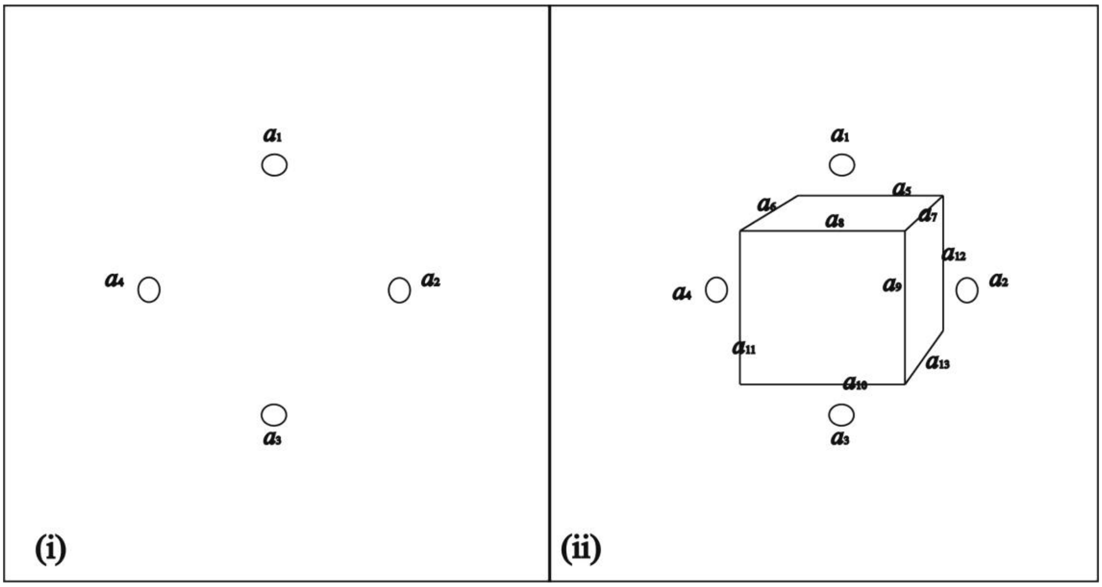


Figure 2.1. In Panel (i), the four circles appear to be not displaced in depth, i.e., they appear to lie on the same frontoparallel plane. In panel (ii), the four circles appear to have different position in depth, induced through adjacency by the edges of a phenomenal cube.

I-aggregate A_2 is carrying a definite depth cue $X(o_2)$, the perspective cue, which supports a 3-D perceptual organisation of the picture, so that the corresponding P-aggregate B_2 is seen approximately as a cube, with a well-defined extension in depth. Aggregate B_2 has the following visible edges: b_5, b_6, b_7 b_8, b_9, b_{10}, b_{11}, b_{12}, b_{13}, corresponding in position with I-units a_5, a_6, a_7 a_8, a_9, a_{10}, a_{11}, a_{12}, a_{13} respectively.

If we specifically consider the four circles, now they presumably appear to be differently located in depth: in particular, subjects usually refer to prefer the solution $y(b_1) = y(b_2) > y(b_3) = y(b_4)$: the circles that are adjacent to an edge of the perceptual cube appear to be coplanar with the frontal face to which the edge is belonging.

---

in determining perceived object characteristics […] is inversely related to the perceived separation of the objects" (Gogel, 1976, p.839).

In order to explain this effect, let us consider the subset $B_3 = (b_5, b_{10}, b_{11}, b_{12}) \subseteq B_2$. Depth cue X (linear perspective) supports the following solution: $y(b_5) = y(b_{12}) > y(b_{11}) = y(b_{10})$. Let us also introduce the relation N, which expresses the I-adjacency between two I-units. In case o_2, $N = \{(a_5, a_1), (a_{12}, a_2), (a_{10}, a_3), (a_{11}, a_4)\}$.

According to the "equidistance tendency", in absence of effective distance cues, if two I-units are proximally close to one another in the proximal stimulus, then the corresponding P-units appear to be at approximately the same distance.

Now let us presume the following entities: an observational case o formed of two generic I-units a_j and a_k, adjacent to each other; two P-units corresponding in position with a_j and a_k, denoted by b_j and b_k; the egocentric distance of P-units b_j and b_k, denoted by $Y(b_j)$ and $Y(b_k)$; depth cue X_k residing on I-unit a_k; the constraining effect of depth cues X_k on thematic P-content $Y(b_k)$, denoted by F. Then, we can formulate the constraint C, expressing the "equidistance tendency", through the following rule, expressed using logical terms:

IF

- two generic I-units a_j and a_k are adjacent to each other;

and

- there are no depth-cues on a_j with a constraining effect on the set of possible solutions for depth-percept $Y(b_k)$ carried by P-unit b_k,

and

- the constraining effect F of depth cue X_k has a punctual or limiting effect on the set of possible solutions for depth-percept $Y(b_k)$ carried by P-unit b_k,

THEN

- the set of possible solutions for depth-percept $Y(b_j)$ is equal to the set of possible solutions for depth-percept $Y(b_k)$

Considering the joint effects of constraint *C*, relation *N* and the solution supported by depth cue *X* in o_2, the only possible solution to P-unknown $R^*(o_2)$ is $y(b_1) = y(b_2) > y(b_3) = y(b_4)$.

Two short comments follow. The first is about the nature of constraint C. We may say that it expresses a constraining interaction between perceptual properties. In general, any theorisation about indirect constraining effects depends on connections between perceptual properties, which serve as the mediators in those indirect effects, and fall within the category of "perceptual interactions" (cf. Rock, 1983, Chapter 10).

The second comment is about the general tendency that enables constraint *C* to be effective. This phenomenon can be framed within the general concept of "depth induction effect"[4], and it is one of the main topics addressed in this chapter and in the following two.

2.2 Intrinsic and extrinsic properties of stimulus factors

In order to clarify and conceptualise the idea of 'depth induction effect', we need to introduce some general concepts related to the depth properties of a generic aggregate, and in particular, to present the distinction between " being intrinsic" and "being extrinsic" when referring to depth-related properties.

Properties are intrinsic or extrinsic *per se* and these characteristics are rooted in the property definition itself. This means that, considering any possible case in which a property *X* is defined, either this property is always intrinsic, or it is always extrinsic. Ellis (1991) defines 'intrinsic properties' as those that are possessed by objects independently of any outside force acting on them, whereas 'extrinsic properties' are those possessed by objects only in virtue of some outside forces. For instance, in physics, the most common example of such a distinction is the one between *mass*, which is an intrinsic property, and *weight*, which is an extrinsic property. Considering that the notion of intrinsicness is a matter of debate in philosophy, and considering that this concept is indeed central in my present

---

[4] I use term *induction effect* in the sense meant, for instance, by Gogel (1972).

formulation, I will try to clarify the sense in which I use this distinction, in order to avoid possible confusion. For a general discussion about the distinction between intrinsic and extrinsic properties, cf. Weatherson (2006). In defining intrinsic and extrinsic properties of stimulus factors, I will use a distinction similar to the one proposed by Vallentyne (1997)[5].

For any observational case *o*, we may consider the set of all I-aggregates (or parts) A_1, A_2, …, A_n identifiable in it, which I denote by $\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$. In the presumed observational case *o*, a I-aggregate $A_i \in \mathcal{A}$ may exist such that A_i carries a well-defined depth property X_i that is *intrinsic* to it, i.e., if we suppose another observational case *o*’ in which other aggregates or parts different from A_i are eliminated[6], still property X_i carried by *A* would maintain in *o*’.

In the presumed observational case *o*, another I-aggregate $A_j \in \mathcal{A}$ may exist such that A_i carries a depth property X_j that is *extrinsic* to it, i.e., if we suppose another observational case *o’’* in which other aggregates or parts different from X_j are eliminated, property X_j carried out by A_j would not maintain in *o*’’.

The most common cues to depth (e.g., texture, linear perspective) are usually considered as intrinsic factors. A possible claim could be that some commonly considered depth-cues are not intrinsic, as they do not maintain if some aggregates of the I-structure are removed. For instance, occlusion as a cue to depth requires at least two units in order to be effective. If we remove one of these two units, then we cannot consider occlusion as a well-defined cue to depth. As an answer to this observation, I underline that occlusion is not defined on a unit, but on an aggregate composed of (at least) two units[7]. We can state that occlusion is a relational property, in the sense that it is a relation between units in one aggregate. As Weatherson

---

[5] Vallentyne (1991) defines a *contraction* of a world *w* as “a world obtainable from the original one solely by removing objects from it” (p.211). A “*x*-*t* contraction” for the object *x* and time *t* is defined as the world “obtainable from the original one by removing , to the greatest extent possible, all objects wholly distinct from *x*, all spatial locations not occupied by *x*, and all times (temporal states of the world) except *t*, from the world” (p. 211). A property *P* is intrinsic “if and only if for any world *w*, any time *t*, and any object *x*: if *Px* at *t* in *w*, then *Px* at *t* in each *x*-*t* contraction of *w*” (p. 212).

[6] This procedure, referred to as ‘elimination’, corresponds to the operation of *contraction of a world* (where the world corresponds to observational cases), as defined by Vallentyne (1991).

[7] According to the type of information in which a psychophysicist is usually interested, any I-unit could be considered as an I-aggregate.

(2006) explains, there is a neat distinction between extrinsic and “relational” properties: “being relational is not a property of *properties* [italics of the author], but a property of *concepts* […] then relational/non-relational and intrinsic/extrinsic are quite different, for they are distinctions between different kinds of things” (p. 6).

A second type of objection which may be raised to my argument is that any depth-cue or depth-percept is an extrinsic property, since they require, to be well-defined, not only the objects carrying them, but also an observer. In my opinion, this objection is not so relevant, since it does not add anything useful to this discussion. Every perceptual property requires an observer, so that it is meaningless to discuss about an observational case without any reference to an observer: the properties I am considering do not refer to the physical world, but to the retinal image or the perceived world.

In order to clarify the concept of extrinsic I-properties, we may refer again to the example in Figure 2.1 (ii). The stimulus factors concurring to the determination of the depth-related properties of aggregate A_1 are two: the first is depth cue X (linear perspective), carried by aggregate A_2, the second is the I-adjacency between some units composing aggregate A_1 and some units composing aggregate A_2. This second factor is an extrinsic I-property, because it is an I-property directly specifying a depth percept on P-aggregate A_1 but not on P-aggregate A_2: as a source of information about depth, it is carried only by A_1 (not by A_2). Another example of extrinsic depth-cue is the so-called “optical contact” (Gibson, 1950a). Due to the role played by this concept in my research, I will explain it in greater detail in the next section.

One general comment follows, related to the ambiguity in the location of extrinsic stimulus properties. We claimed that such properties belong to the I-side of the depth perception problem, but indeed, their action in supporting a depth-related P-content is connected to the idea of a “prior constraint”: their effectiveness on determining depth-related P-properties is based on some a-priori assumption made by the visual system about the visual world. The “extrinsic property” of picture in Figure 2.1(ii) is supposed to intervene as a “supplementary” source of information: from an ecological point of view, it is not incorrect to assume that, if the projections of two objects are close in the retinal projection, then the corresponding phenomenal objects are likely to be close in the visual scene. But this assumption uses an information whose roots reside in the visual system, not in the optical stimulus. Once again, this reveals the subtleness and intricateness of problems of vision science.

2.3 Optical contact

Among the various possible depth-related extrinsic I-properties, the concept of ‘optical contact’ constitutes one of the most important examples. In this section, I will provide a definition of the concept, and a summary description of some researches in which this concept has played a central role. The interaction of optical contact information with other depth-related information is the main theme explored in the experiments described in Chapters 3 and 4.

The concept “optical contact” refers to the topological contact between the 2-D proximal image of an object and a more extended surface in the 2-D proximal stimuli (as opposed to contact between an object and a background surface in the physical world, which may be called “physical contact”). An object realising an optical contact with a background surface, well extended in depth by virtue of certain depth cues (e.g., texture) is generally seen as lying on the surface, in the absence of strong cues specifying the depth-related properties on it. Due to this, the optical contact could be seen as perceptual contact between the object and the background surface, and this would determine the egocentric perceptual distance of the object in relation to the background surface. A convincing demonstration of this effect is provided by a classical stimulus situation created by James Gibson (see Figure 2.3), in which the importance of underlying background surfaces and their effectiveness in determining spatial layout is clearly shown.

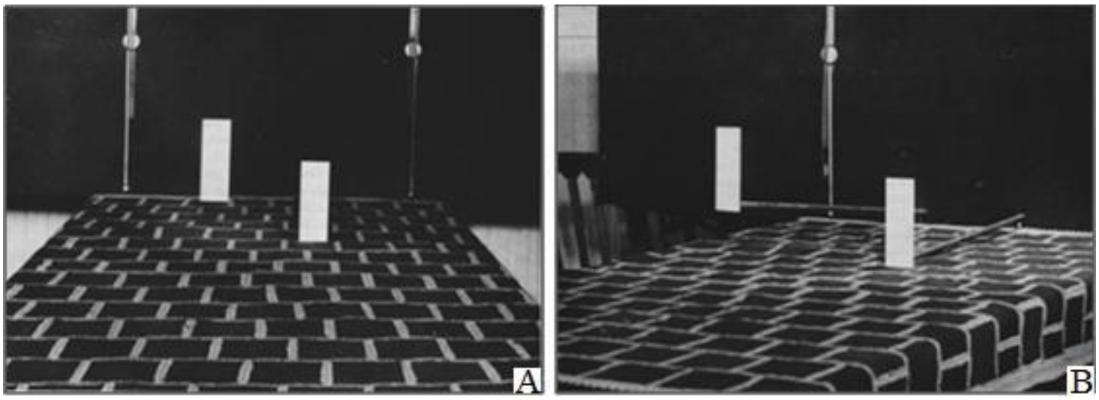


Figure 2.2. Images taken from Gibson, 1950a. Two card-board rectangles and a textured ground surface constitute the main elements of the visual scene. Panel A shows the front view of the scene, in which the left rectangle seems to be more distant than the right rectangle. Panel B shows a lateral view of the scene, in which we can clearly see that the rectangles are actually at the same egocentric distances, but the left one is suspended above the surface. The different apparent locations in depth of the two rectangles can be explained by referring to the optical contact between left rectangle and the ground surface, which determines the egocentric distance of the rectangle itself.

This hypothetical principle is fundamental in explaining depth-related perceptual properties, mainly, egocentric and exocentric distances of the objects in the visual scene. In his *ground theory of space perception*, Gibson (1950a) emphasizes the role of the ground plane in distance perception: objects are generally perceived relative to a background surface, so that the ground surface becomes of special importance. The interpretation of this phenomenon suggested by the author is that human beings are *terrestrial* animals, and the environment in which they live is a continuous, more or less horizontal, ground surface, on which objects lie, and by means of which their position becomes related to each other. Precisely for this reason, the terrain (ground surface) has a special role in determining the spatial layout. Gibson maintained that “the problem of three-dimensional vision, or distance perception, is basically a problem of the perception of a continuous surface which is seen to extend away from the observer [...] the ground is the basis of visual space perception” (Gibson, 1946, p. 420; quoted by Bian, Braustein & Andersen, 2006).

Some remarkable developments of this idea have been proposed recently by Meng & Sedgwick (2001, 2002), who found that information about objects lying on different surfaces can propagate from the ground to other surfaces through “nested contact relations”. Using computer generated images, these authors asked their subjects to adjust the position of a probe lying on a vertical track on the ground plane to match the distance of a cube which could be seen, through optical contact, as lying on different blocks having contact relations with surfaces other than the ground. They draw the conclusion that “local spatial relations between objects and their platforms are only partially integrated with more global spatial relations between the discontinuous surfaces of the platforms” (Meng & Sedgwick, 2002, p. 1). In general, the presence of discontinuities among surfaces, or discontinuities on the ground plane itself, are able to affect the accuracy and effectiveness of judgements based on the “ground theory of perception” (Gibson, 1950; Sinai, Ooi & He, 1998; Feria, Braunstein & Andersen, 2003; Wu, Ooi & He, 2004).

Bian et al., (2005, 2006) tested the relative effectiveness of the ground surface and other environmental surfaces in determining the perceived layout of three-dimensional scenes, by presenting to their subjects two posts (or two oblong ellipses), in optical contact with both the ground surface and the ceiling surface, or with both sidewalls, thus producing a “contradictory layout information”. They reported a dominance effect of the ground plane over other environmental surfaces. This dominance effect is judged by the authors to be ecologically plausible since “ceilings and walls [...] are usually artificial surfaces and are not universal” (Bian, et al., 2005, p. 803).

Madison, Thompson, Kersten, Shirley & Smits (2001); Sauer, Braunstein, Saidpour & Andersen (2002); Koning & Van Lier (2003); Ni, Braunstein & Andersen (2004, 2005, 2007), studied how the information conveyed by the optical contact interacts with other depth-related stimulus factors. These researches generally showed that shading is able to suppress the role of the optical contact. Ni et al., (2005) found that, when more than one potential shadow is present in the scene (a situation that is very common in the natural environment), the role of optical contact can still be relevant.

In the following two chapters, I will present four experiments where the optical contact and its interaction with other stimulus factors play a central role.

Chapter 3

Perceptual slant induced through the optical contact

3.1 Introduction

The idea for this research originated from the comments made by Bian, Braunstein & Andersen (2005) in presenting their Experiment 4. The comments are about the possibility that a discrepancy in “optical contact” (“contradictory layout information”) between the upper and lower ends of an elongated object may turn into an apparent “slant in depth” of the object itself in the perceived scene.

In these experiments, pictorial stimuli of the kind illustrated in Figure 3.1 were used. They differed from those of Bian et al. in some respects, which are specified in introducing Experiment 1. The stimulus is formed of four textured regions, called the (pictorial) frame, and an oblique rectilinear segment, called the (pictorial) target. The frame is so constructed as to simulate a corridor, the ground, ceiling, and side walls of which are covered by a random checkerboard texture. There is a difference in “optical contact” between the ends of the target, in that the distance (on the pictorial plane) of the upper end from the upper border of the frame is larger than the distance of the lower end from the lower border of the frame. On account of this difference, if we presume that the frame actually appears (in the perceived scene) as a corridor, and that (in the perceived scene) the upper and lower ends of the target preserve their contact with the ceiling and ground of the corridor, then we predict that the two ends will differ from each other in their apparent distance from the observer: the upper end appears farther than the lower end. As a consequence, if we further presume that the target perceptually preserves its rectilinear form, then we predict that it will appear as a pole “slanted in depth” inside the corridor: more precisely, a pole slanting upward. Similar reasoning allows us to infer that, for a stimulus in which the distance of the upper end of the target from the upper border of the frame is smaller than the distance of the lower end of the target from the lower border of the frame, the target itself should appear as a pole slanting downward. This is the kind of prediction we planned to test in these experiments. As may be seen in the above argument, these predictions rest on a system of assumptions, of which the most important (for our purposes) is that the "optical contact" of the ends of the target – i.e., their local coincidence with points in the upper and lower regions of the frame in the 2-D pictorial stimulus – becomes preserved as a "perceptual contact" – i.e., the ends of the target keep touching the ceiling and ground of the apparent corridor in the 3-D perceptual scene (cf. Section 2.3).

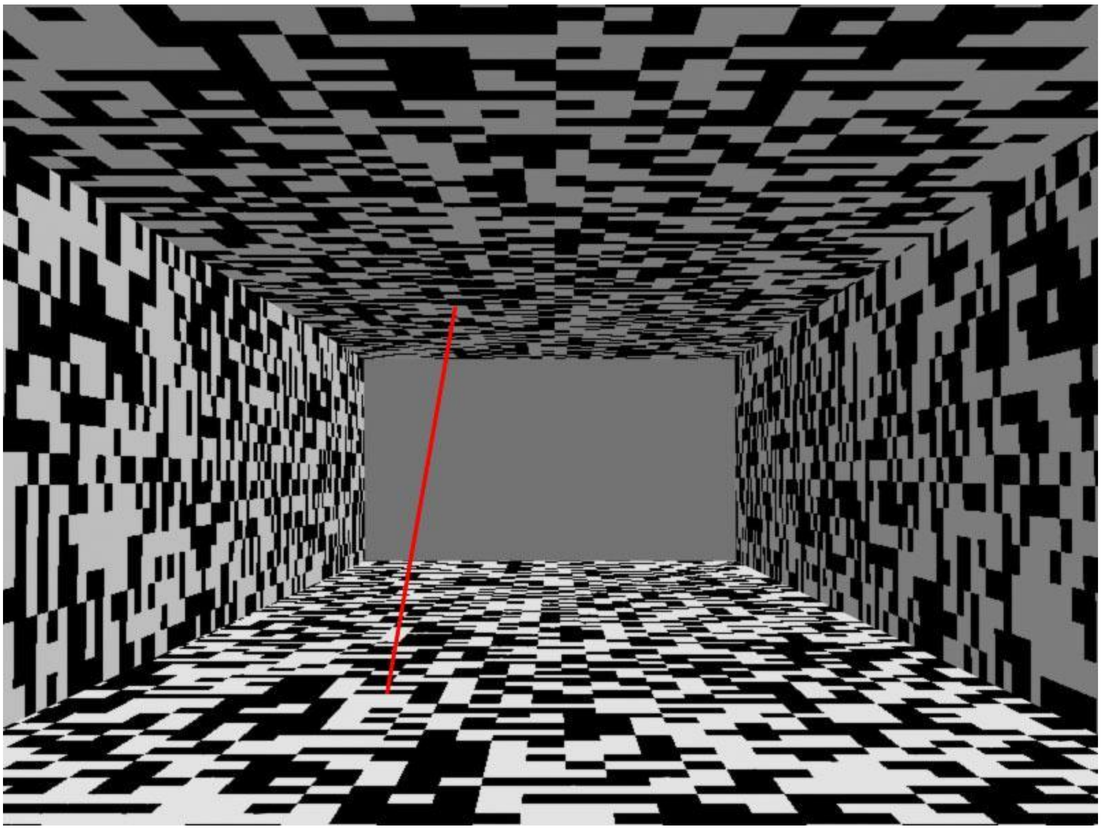


Figure 3.1. One of the pictorial stimuli (stimuli were shown on a monitor).

As illustrated in this example, the key perceptual property in our study is slant in depth, more specifically, the slant of an elongated, rectilinear, thin object, to which we briefly refer as a linear object – a pole in its phenomenal appearance. I choose the vertical in the 3-D space as the basic reference direction, and define the slant of any linear object as the acute angle between the sagittal projection of the object – i.e., its orthogonal projection on the sagittal plane – and the vertical in the space. More precisely, the slant in depth (henceforth denoted by S) is measured by a positive value ($S>0$) if, relative to the vantage point, the upper end of the linear object is farther than the lower end (upward slant), a negative value ($S<0$) if the upper end is closer than the lower end (downward slant), and the zero value ($S=0$) if both ends are at the same distance (vertical slant). For example, by applying this criterion and following the above argument, we may state that the perceptual slant of the linear target in Figure 3.1 is expected to be positive in value – an upward perceptual slant[1].

The other key concept in this study is that of “optical contact” (also called “image-based connectedness”; cf. Koning & van Lier, 2003). As explained in chapter 2, it presumes reference to some 2-D structure – a pictorial stimulus, or the optical image of a 3-D physical scene – and expresses the relation holding between any two components in the structure when a certain portion of one component is locally coincident with a certain portion of the other. In this sense, for example, I stated above that the upper end of the linear target in Figure 3.1 is in optical contact with the upper region of the frame, and the lower end with the lower region. This concept is important because it is involved in the following hypothetical rule of psychophysical inference: in certain conditions, any two components that are in optical contact as parts of a 2-D optical stimulus tend to appear in perceptual contact as parts of the resulting 3-D perceptual scene. This is a hypothetical law which, combined with other principles and assumptions, allows us to make definite predictions concerning perceptual results. It asserts, in substance, that there is a tendency in the percept formation process to preserve certain optical contacts as perceptual contacts, i.e., a tendency to contact preservation.

As I commented in Chapter 2, the concept of “optical contact” occurs in the classic works of James J. Gibson, and the hypothesis we call “tendency to contact

---

[1] This use of the term “slant” does not exactly coincide with the prevailing use of the term in the literature on depth vision, in which surfaces rather than linear objects are mainly examined, and the slant of a plane surface is defined as the angle between the line of sight and the surface normal. Cf. Stevens (1981).

preservation" plays a crucial role in his "ground theory" of space perception (cf. Gibson, 1950a, pp. 6-7, 176-180; also Sedgwick, 1986). The contacts of main concern for Gibson's theory are those between the optical image of the ground surface – the terrain – in an outdoor environment, and the optical images of objects – trees, buildings, etc. – variously located on that surface. The basic claim is that the perceptual extension in depth of the ground surface (due to stimulus information, like texture gradients, which are inherent in its optical image), and the optical contacts between images of the objects and the image of the ground surface, are determining factors for the perceptual distribution in depth of the objects themselves – the spatial layout of the perceived scene. This connection may be conceptualized as a "depth induction effect", i.e., an effect directed from the ground surface (which is endowed with an "endogenous" 3-D spatial organization, due to depth cues residing on its optical image) to the several objects resting on the surface (the perceptual location of which, insofar as it depends on contact relations with the ground surface, is "exogenous" in character). In recent years, this classic theoretical paradigm has inspired a fresh stream of studies on depth vision (cf. Section 2.3). In some of these studies, the paradigm is applied in a generalized form, considering that not only a ground surface but also any spatially organized part of the perceptual scene may play the role of an "inducing factor", in a depth induction effect mediated by contact relations (cf. the concepts of "propagation of depth information" in Sauer, Braunstein, Saidpour, & Andersen, 2002, and "sequential surface integration" in He, Wu, Ooi, Yarbrough & Wu, 2004).

In summary, elements involved in the depth induction effect considered are: the "inducing component", a frame perceptually appearing as a corridor; the "induced component", a target perceptually resembling a pole in the corridor; and the depth property to be judged, the apparent slant of the target, as it is induced by the frame through contact relations.

3.2 Experiment 1

The main differences between the stimuli in these experiments and those in Experiments 4 and 5 of Bian et al. (2005) are the following. First, the frame in each stimulus is formed of four regions, which are circularly adjacent so as to simulate a corridor (the frame in the stimuli of Bian et al. consists of two regions, which are separated from each other and correspond to the upper and lower regions in our situation). With this choice I intended to strengthen the unitary aspect of the perceptual frame (the corridor) and to support a precise correspondence between points in the upper and lower regions, as the difference in the simulated egocentric distance of points in the two regions was the main factor in our experimental design. Second, the target in the stimulus is a thin linear segment, i.e., a pictorial component which in its appearance is one-dimensional and devoid of internal structure. By this I intended to exclude stimulus factors which might condition the perceptual slant of the target (in an “intrinsic” way) and thus interfere with the depth induction effect I planned to examine (such an interference was explicitly considered by Bian et al. who, for this reason, substituted the rectangular striped targets in their Experiment 4 by elliptical uniform grey targets in Experiment 5). Third, all our stimuli were presented to the subjects so that the intrinsic vertical of the frame coincided with the vertical in the physical environment, but in part (two-thirds) of our stimuli, the target was shown tilted, to the right or left (in the stimuli of Bian et al., the targets were all upright, whereas the regions forming the frame were tilted, in either direction). The reason for this choice was that I surmised that the vertical position of the target (in the stimulus) might directly favour the response of a vertical position of the target (in the perceptual scene), i.e., the response of null apparent slant, thus obscuring the depth induction effect I aimed to study.

There are two characteristics of the stimulus which enter as the main independent variables in our experiments, and they are related to the first and third comments in the previous paragraph. I call one of these characteristics the *vertical unbalance* of the target, denoted by V. In direct terms, variable V represents the disagreement between the distance of the upper end of the target from the upper border of the frame (a distance on the 2-D stimulus, denoted by D_u), and the distance of the lower end of the target from the lower border of the frame (denoted by D_l). More precisely, variable V represents the difference $c(D_u)–g(D_l)$, where c and g are functions transforming pictorial distances of points in the upper and, respectively, the lower region of the 2-D frame into simulated egocentric distances of corresponding points in the ceiling and, respectively, the ground of the simulated 3-D corridor. Ceiling and ground are simulated to be horizontal surfaces, which implies that the simulated slant of the target is $\tan^{-1}((c(D_u)–g(D_l))/h)$, where h is the simulated height of the corridor. Thus, the simulated slant of the target is positive (upward), null (vertical) or negative (downward), depending on whether the value of variable V is positive, null or negative. This argument concerns the simulation process and rests on some assumptions, one of which is that, in the simulated scene, the ends of the target are in contact with the upper and lower surfaces of the corridor. For this very reason, the argument has a bearing on our perceptual problem. That is, if by experiment I find that there is a consistent relationship between vertical unbalance V and apparent slant S of the target – i.e., as stimulus variable V passes from negative to positive values, there is a corresponding passage of response variable S from the “downward” to the “vertical” and “upward” alternatives – then we may conclude that the assumptions implicit in the simulation process also hold true – within limits – of the percept formation process. In particular, we may conclude that, in the perceptual process, there is a tendency to contact preservation, which is the key condition for the slant induction effect in our situations.

The other stimulus characteristic entering as a main independent variable in our experiments is the *tilt* of the target, denoted by T. It represents the acute angle (on the pictorial plane) formed by the target and the intrinsic vertical of the frame – which, as I have stated, coincides with the vertical in the environment. The angular measure is taken as positive or negative, depending on whether the target leans to the right or the left. The main reason for introducing this variable was that stated above: I surmised that there is a privileged association between the vertical position of the target in the stimulus (condition T=0) and the judgment of a vertical slant of the target in the perceived scene (response S=0), and that, for this reason, the hypothesized depth induction effect may be better appreciated on targets which are not vertical in the stimulus (condition T≠0). In other words, I expected an interaction between factors T and V on response variable S, in that the dependence of S on V is stronger in conditions T<0 and T>0 than in condition T=0.

The stimuli in Experiment 1 differed from one another in two other respects: the *length L* of the target, and the *horizontal position H* of the target itself (i.e., its position on the left or right of the median vertical of the frame). When planning the experiment, I did not come to definite expectations regarding the effects of these stimulus variables – I introduced them simply to create a sufficiently large variety of stimuli on which to test the basic hypotheses of our study involving variables V, T and S. However, as we shall see, the data bring to light a notable effect of variable H (interacting with T), which substantially enriches our initial theoretical framework. Also note that variable L, which represents the length of the target on the pictorial plane, may equally be described in terms of the simulated egocentric distance of the target. In fact, if we presume that the upper and lower regions of the frame simulate the ceiling and ground surfaces of a corridor, and that in the simulation both ends of the target are in contact with these surfaces, then the smaller the length of the target (in the 2-D stimulus), the larger its egocentric distance (in the 3-D simulated corridor).

3.2.1 Experimental setup.

Stimuli. 90 stimuli were used, obtained by combining 5 levels of factor V, 3 of T, 3 of L, and 2 of H. The five levels of factor V (vertical unbalance of the target) are denoted by −2, −1, 0, 1 and 2, and were specified so as to produce values of −65°, −45°, 0°, 45° and 65° as the simulated slant of the target. Note that to each one of the five levels of V there corresponds one difference $c(D_u)–g(D_l)$ which is constant among the 18 stimuli having V at that level, but not only one difference $D_u–D_l$

sharing the same constancy. For example, to produce the 18 stimuli in which $V=-2$ (the simulated slant of the target is −65°), I had to adapt distances D_u and D_l, and their difference D_u-D_l, on account of differences in variables T and L among the 18 stimuli. The three levels of factor T (tilt of target) are denoted by −1, 0 and 1, and correspond to tilts −10° (10° to the left), 0° (vertical position) and 10° (10° to the right). The three levels of factor L (length of target) are denoted by 1, 2 and 3, and correspond to 12.8, 14.2 and 15.6 cm as measured on the pictorial plane (the screen on which the stimuli were shown). Lastly, the two levels of factor H (horizontal position of target) are denoted by −1 and 1, meaning that the centre of the target is 5 cm on the left or right of the median vertical of the stimulus. For example, that shown in Figure 1 is the stimulus in which $V=2$, $T=1$, $L=2$ and $H=-1$.

The frame was the same in all 90 stimuli, and was that illustrated in Figure 1. It was generated by computer graphics (using a 3-D modelling software package, Autodesk 3DS Max 8.0), and programmed to simulate a corridor 550 cm wide, 300 cm high, and 2120 cm deep. The simulated vantage point was at height 130 cm from the ground of the corridor and at distance 120 cm from its opening. On such a frame, there are three kinds of depth cues which support its perceptual rendering as a 3-D structure: texture (a random checkerboard texture covering the four regions), linear perspective (which involves, in particular, the adjacency lines between regions), and illumination perspective (in the simulation, an invisible light source was located next to the front-top-right corner of the corridor).

The stimuli were displayed on a 22-inch CRT monitor with a pixel resolution of 800×600, controlled by a Windows 1998 workstation. On the monitor, the width and height of the contour of the frame were 40 and 30 cm, and the width and height of its central region were 13.8 and 7.4 cm. The light and dark components of the texture and the central region of the frame were of different shades of grey (the mean luminances of light components in the top, bottom, left, and right textured regions were 27.42, 102.86, 68.08, and 31.45 cd/m^2; the mean luminances of the corresponding dark components were 1.07, 1.15, 0.96, and 0.99 cd/m^2; the luminance of the central region was 28.12 cd/m^2). The target was red (luminance 43.47 cd/m^2).

The distance from the eyes to the monitor was 50 cm. A chinrest was mounted at a position appropriate to this viewing distance. At this distance, the two external sides of the frame subtended visual angles of 43.6 and 33.4 degrees, and the target in its three possible lengths subtended visual angles of 14.59, 16.16, and 17.73 degrees. Vision was binocular. The experiment was run in a darkened room.

Procedure. For each participant, the experimental session began with a practice phase, the aim of which was to illustrate the kind of stimuli to be observed and how to express responses. The main point was to explain to the participants the spatial property they were asked to judge, i.e., the apparent slant in depth of the red pole in the scene, which is the inclination in the forward/backward direction, to be distinguished from the apparent tilt (inclination in the rightward/leftward direction), or other types of inclination. For this purpose, not only some images on the monitor generally similar to those to be shown in the main phase were presented, but also a simple mechanical device in which a suspended pole was rotated in various directions by the experimenter.

In the main phase of the session, all 90 stimuli were separately presented, in an order independently randomized for each participant. The sequence of 90 trials was split into three blocks of equal size and separated by pauses, the duration of which was chosen by participants. In each trial, a white fixation cross (0.5 sec) first appeared in the centre of the screen, and then one of the 90 stimuli was shown for 3 sec, this time being sufficiently short to prevent observers from exercising any “imaginative geometry” on the stimulus. The observers’ task was to judge whether the red pole in the scene was slanted upward, downward, or neither (it appeared lying on a frontal plane), and to express their judgment, after the image had disappeared, by pressing one of three buttons of a response box. The central button was associated with judgment “no slant” for all participants, whereas the left and right buttons were balanced for “upward slant” and “downward slant” across participants. The experimental session, including both practice and main phases, took from 14 to 17 min to complete.

Participants. The experiment was carried out by 30 participants, who took part in it voluntarily. They were undergraduate or graduate students of several faculties in the University of Padova (Italy), ranging in age from 21 to 30 (mean 23.91), 16 women and 14 men. All had normal or corrected-to-normal visual acuity, and were unaware of the hypotheses of our study.

3.2.2 Results and discussion

The experiment comprised four factors (V,T,L,H), response variable S (which is categorical), and repeated measures (each of the 30 participants gave a response to each of the 90=5×3×3×2 experimental stimuli). The data were analysed by fitting and interpreting a log-linear model (Agresti, 2002). First, from the data-file I obtained a 5-dimensional frequency table (m_{vtlhs}) where, for $v\in\{-2,-1,0,1,2\}$, $t\in\{-1,0,1\}$, $l\in\{1,2,3\}$, $h\in\{-1,1\}$, and $s\in\{-1,0,1\}$, frequency m_{vtlhs} was the number of participants who gave response $S=s$ when presented with the stimulus in which $V=v$, $T=t$, $L=l$, and $H=h$. Then I searched for a log-linear model which was optimal in simplicity and goodness-of-fit relative to table (m_{vtlhs}) of observed frequencies. The search was carried out in two stages. First, I searched for a *hierarchical* log-linear model which was optimal in the stated sense (to be “hierarchical” for a log-linear model means that, e.g., if the model comprises a component representing the interaction of two factors A and B, then it also comprises components representing the individual actions of A and B). Second, I tried to simplify the optimal hierarchical log-linear model further by omitting some of its components, so far as such omissions did not entail any substantial increase in residual deviance (i.e., a substantial loss of goodness-of-fit). If omissions are actually made in this stage, then the resulting log-linear model is *non-hierarchical*, and optimal in simplicity and goodness-of-fit to the data.

The log-linear model resulting from this search is given by this equation:

$$\log(m_{vtlhs})=\mu+\lambda_s^{S}+\lambda_{vs}^{VS}+\lambda_{ts}^{TS}+\lambda_{ths}^{THS}.$$

It expresses the logarithm of the observed frequencies as the sum of five components, which are intercept μ (with no useful meaning, in the present context); component λ^{S}, directly associated with response variable S; components λ^{VS} and λ^{TS}, specifying the separate actions of factors V and T on response S; and component λ^{THS}, specifying the interaction of factors T and H on response S. The model is non-hierarchical (e.g., it has component λ^{THS} but not component λ^{HS}). Its residual deviance (likelihood-ratio statistic) is G^2=200.28, with df=251 (=270−19, 270 being the size of the frequency table, 19 the number of free parameters in the model), and p=0.991 (relative to distribution χ^2_{251}). The low value of residual deviance (when compared with the "null deviance" of our data, which is G^2=1557.86) and the corresponding high value of p are proof of a very good fit of the model to the data. Table 3.1 lists the maximum-likelihood estimates of the four classes of parameters forming the model (the estimate of the intercept is μ=2.029).

Component λ^{S} in the model (specific deviance G^2=95.236, df=2, p<0.001) means that, in the whole set of 2700=90×30 responses by participants, alternatives −1, 0 and 1 were not of equivalent presence. More precisely, part (i) of Table 3.1 shows that the most preferred response was S=0 (no slant in depth, the pole appears vertical), then response S=1 (the pole appears slanted upward), and lastly response S=−1 (the pole appears slanted downward). This order can also be directly seen on the collapsed frequencies associated with the three alternatives, which are 1051 for S=0, 899 for S=1, and 750 for S=−1. Unexpected, in particular, is the difference in frequency between alternatives S=1 and S=−1 (significant to the binomial test with p<0.001), considering that, in the plan of our experiment, there was an exact balance between conditions which – according to our hypotheses – should favour the one and those which should favour the other of the two alternatives. The difference reveals a bias in favour of response "upward slant". I shall return to this result when discussing the data from Experiment 2.

Component λ^{VS} in the model (specific deviance G^2=568.225, df=8, p<0.001) means that there is a strong association between stimulus variable V (the vertical unbalance of the target) and response variable *S*. Table 1.ii shows that this is an increasing monotone association, when ranges {–2,–1,0,1,2} and {–1,0,1} of the two variables are ordered in the natural way. In more detail, the table shows that, as variable *V* increases over its range, the chance of response *S*=1 increases, that of response *S*=–1 decreases, and that of response *S*=0 first increases and then decreases.

(i)

λ^{S}	*S*=–1	*S*=0	*S*=1
	–0.153	0.163	–0.010

(ii)

λ^{VS}	*V*=–2	*V*=–1	*V*=0	*V*=1	*V*=2
S=1	–0.456	–0.367	–0.045	0.267	0.601
S=0	–0.148	0.054	0.177	0.113	–0.196
S=–1	0.604	0.313	–0.132	–0.380	–0.405

(iii)

λ^{TS}	*T*=–1	*T*=0	*T*=1
S=1	0.279	–0.685	0.405
S=0	–0.562	1.216	–0.653
S=–1	0.282	–0.531	0.248

(iv)

λ^{THS}		*H*=–1			*H*=1	
	T=–1	*T*=0	*T*=1	*T*=–1	*T*=0	*T*=1
S=1	–0.338	0.066	0.272	0.338	–0.066	–0.272
S=0	0.129	–0.022	–0.107	–0.129	0.022	0.107
S=–1	0.209	–0.044	–0.165	–0.209	0.044	0.165

Table 3.1. Maximum-likelihood estimates of parameters in log-linear model accepted for data of Experiment 1.

This sequence was precisely what I expected on the basis of the main hypothesis, concerning the dependence of the apparent slant of the target on the discrepancy between optical contacts at its ends. The same course can be directly seen in Table 3.2, which shows the collapsed frequencies on the pair of dimensions (*V*,*S*). It is also illustrated in Figure 3.2, in which three logit functions computed on the frequencies in Table 3.2 are shown (note, in particular, the regular trend of logit(S=1,S=−1)).

	V=−2	V=−1	V=0	V=1	V=2
S=1	107	120	164	217	291
S=0	180	226	253	230	162
S=−1	253	194	123	93	87

Table 3.2. Frequencies of responses *S*=−1, *S*=0 and *S*=1 to stimuli with different values of property *V* (vertical unbalance of target) from Experiment 1.

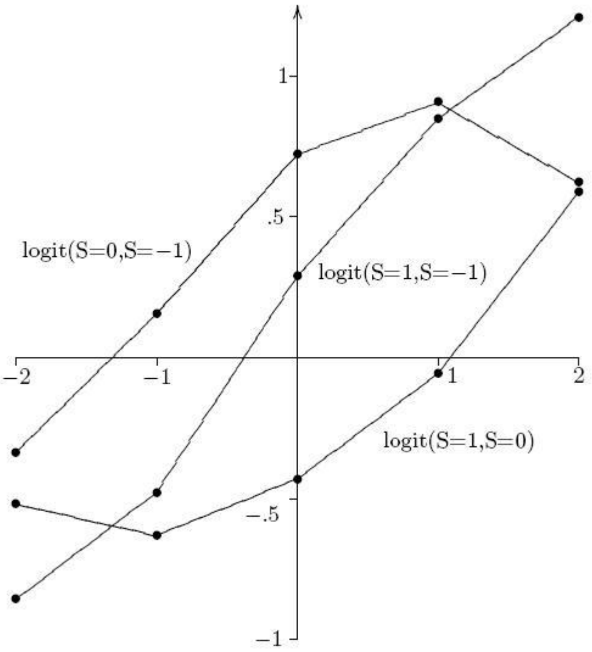


Figure 3.2. Logit functions for comparing frequencies on alternative values of variable *S*, in relation to variable *V*. E.g., logit(S=1,S=−1)=log($m_{v\ldots1}/m_{v\ldots-1}$), for $v\in\{-2,-1,0,1,2\}$, where $m_{v\ldots1}$ and $m_{v\ldots-1}$ are frequencies in first and third rows of Table 3.2.

Component λ^{TS} in the model (specific deviance G^2=1975.407, df=4, p<0.001) signifies quite a strong connection between stimulus variable T (the pictorial tilt of the target) and response variable S. The scheme, as revealed by the values in Table 3.1.iii, is as follows: in condition T=0 (the target was upright in the 2-D stimulus) there was a sharp dominance of response S=0 (the target was judged to have zero slant in the 3-D perceptual scene), whereas in conditions T=–1 and T=1 the chance of response S=0 was greatly diminished, and correspondingly there was an increase in the chance of responses S=–1 and S=1. The same scheme is directly seen in Table 3.3, which shows the frequencies collapsed on the pair of dimensions (T,S). On this table we compute 708/192=3.6875 as the odds of response S=0 in condition T=0, whereas 181/719=0.2517 and 162/738=0.2195 are the odds of that response in conditions T=–1 and T=1. Thus, there was a strong privileged association between stimulus condition T=0 and response S=0. This response prevailed in general, as pointed out when discussing component λ^{S}, but its prevalence became much greater in stimulus condition T=0. Note that, relative to variable V, the effect of condition T=0 on variable S has something of a “veto effect” or “cue dominance” (Bülthoff & Mallot, 1990, p. 123; Howard & Rogers, 2002, p. 470): condition T=0 makes response S=0 almost mandatory, so that, in that condition, variable S is largely unaffected by V. For an explanation of this I appeal to the “generic viewpoint assumption” (Freeman, 1994) by considering that, if a linear object is not upright in the 3-D space, then there are only two viewing directions from which its optical image is upright (for an observer standing ordinarily), whereas if the object is upright in the 3-D space, then its optical image is upright from every viewing direction. Thus, the “generic view principle” predicts that, when the 2-D image of a linear object is upright (i.e., condition T=0 is true), then the impression of the object standing vertically in the 3-D space will prevail (i.e., response S=0 becomes highly probable).

Component λ^{THS} in the model (specific deviance G^2=195.022, df=4, p<0.001) proves the existence of an interaction of stimulus variables T and H (pictorial tilt and horizontal position of target) on response variable S. Actually, the values in Table 1.iv show that, when H=–1 (target in left visual hemifield), condition T=–1 favoured response S=–1 and condition T=1 response S=1, whereas when H=1 (target in right visual hemifield) the opposite was the case, i.e., condition T=–1 favoured response S=1 and condition T=1 response S=–1. The same configuration can be seen in Table

4, which lists the frequencies collapsed on the triple of dimensions (T,H,S). On the table we find, for example, that in condition H=–1 the odds ratio for T=–1 or T=1, and S=–1 or S=1 is 142×121/205×263=0.3186, whereas in condition H=1 the odds ratio for the same four positions is 241×196/50×158=5.9792 – quite the opposite.

	T=–1	T=0	T=1
S=1	383	95	421
S=0	181	708	162
S=–1	336	97	317

Table 3.3. Frequencies of responses *S*=–1, *S*=0 and *S*=1 to stimuli with different values of property *T* (pictorial tilt of target) from Experiment 1.

This interaction does not correspond to a hypothesis conceived in planning the experiment, yet it may be given a plausible interpretation in geometric-optical terms. Indeed, let us presume that a linear object (a pole) is given in the physical 3-D space so that its tilt relative to the vertical is null (i.e., the pole is parallel to the sagittal plane). Then a simple argument shows that, when the object lies in the left hemifield (i.e., H=–1), then its polar projection on the frontal plane passing through the vantage point is tilted left (T<0) or right (T>0), depending on whether the object is physically slanted downward (S<0) or upward (S>0). The same argument shows that, when instead the object lies in the right hemifield (i.e., H=1), then its polar projection is tilted left (T<0) or right (T>0), depending on whether the object is physically slanted upward (S>0) or downward (S<0). Thus, the following relational constraint on variables T,H,S is true at the optical level (the image formation process):

if H=–1, then T<0 or T>0, depending on whether S<0 or S>0;

if H=1, then T<0 or T>0, depending on whether S>0 or S<0.

Now, let us turn this optical constraint into a (conjectural) psychophysical constraint, simply by interpreting S as the perceived (rather than the physical) slant of the object. The rule then allows us to predict that, if H=–1, then stimulus conditions T<0 and T>0 favour responses S<0 and S>0, respectively, whereas if H=1, then stimulus conditions $T<0$ and $T>0$ favour responses $S>0$ and $S<0$, respectively. But this prediction exactly fits the configuration shown by the parameter estimates in Table 1.iv and the response frequencies in Table 3.4. This argument is in line with the heuristics “vision is inverse optics” (cf. Longuet-Higgins, 1986; Poggio, Torre & Koch, 1985). For this reason, I refer to the effect revealed by component λ^{THS} of the model as the “process by inverse optics”, distinct from the “process by optical contact”, which is the basic hypothesis of our study and corresponds to component λ^{VS} of the model.

	$H=-1$			$H=1$		
	$T=-1$	$T=0$	$T=1$	$T=-1$	$T=0$	$T=1$
$S=1$	142	54	263	241	41	158
$S=0$	103	349	66	78	359	96
$S=-1$	205	47	121	50	50	196

Table 3.4. Frequencies of responses $S=-1$, $S=0$ and $S=1$ to stimuli with different values of properties T and H (pictorial tilt and horizontal position of target) from Experiment 1.

There are 5 variables in the dataset, so that the saturated log-linear model fitting the data has $2^5=32$ components. I found a well-fitting model comprising only 5 of these components, and this may be viewed as a success in simplicity. To conclude, I add comments on some of the possible components which fail to be present in the accepted model. The absence of all components that only involve stimulus variables V,T,L,H – or some of them – is no surprise, as it is simply due to the design of our experiment, which is completely balanced. All components involving variable L are excluded from the model, which means that the pictorial length of the target has no effect – either singly or in a combined way – on the apparent slant of the target, within the limits of the experiment. The same may be said of the simulated egocentric distance of the target, which is intrinsically related to variable L. Variable H (the horizontal position of the target) plays a special role: it influences the apparent slant by interacting with pictorial tilt T (component λ^{THS} discussed above), but does not, alone, condition the apparent slant (component λ^{HS} is not in the model). In introducing the experiment, I stated that an interaction of stimulus variables *V* and *T* on response variable *S* was expected, as I surmised a privileged association between condition *T*=0 and response *S*=0, so that the dependence of *S* on *V* would be stricter and clearer in conditions *T*=–1 and *T*=1 than in condition *T*=0. By submitting this expectation to statistical testing, I find that the decrease in residual deviance which results from adding component λ^{VTS} to the model is G^2=18.671, with df=16 (so that *p*=0.2861). Thus, the gain in goodness-of-fit is negligible, and component λ^{VTS} is not included in the model. I may conclude that there is an influence of pictorial tilt *T* on perceived slant *S*, but this is adequately expressed as the direct action of *T* on *S*, rather than as the interaction of *T* and *V* on *S* (as initially surmised).

3.3 Experiment 2

The stimuli in Experiment 2 were similar to those in Experiment 1 (cf. Figure 3.1). Compared with the first experiment, the plan of the second differed in three respects, two of which concerned the experimental factors and one the experimental task. On one hand, variable L (the pictorial length of the target) was removed from the set of experimental factors, which means that the stimuli in the new experiment were all equivalent in that characteristic. No loss of import was expected, also considering that the data from Experiment 1 showed no involvement of variable L in any significant effect on the perceptual property under study. On the other hand, the number of levels of factor T (the pictorial tilt of the target) was increased, from three to five. The data from Experiment 1 showed that factor T played important roles in conditioning the response variable, so that I judged it worthy of closer scrutiny. The task for participants in Experiment 2 was “matching by adjustment” rather than a categorical judgment as in Experiment 1. I thus aimed at obtaining more precise information – quantitative rather than categorical – concerning the apparent slant of the target in the scene.

Based on measures produced through “matching by adjustment”, the aims of Experiment 2 were as follows: to confirm and specify better the dependence of apparent slant *S* on vertical unbalance *V* of the target (what I call the “optical contact process”, as a case of the depth induction effect, which represents the chief issue of this research); to confirm and specify better the dependence of *S* on pictorial tilt *T* of the target, both in terms of main effects of *T* and in terms of interaction effects of *T* and *H* (what I call the “inverse optics process”, revealed by the data of Experiment 1); to explore the interaction between both processes, in particular, to appreciate their relative importance in conditioning the apparent slant of the target.

3.3.1 Experimental setup

Stimuli. Fifty stimuli were produced by combining 5 levels of factor *V*, 5 of factor *T*, and 2 of factor *H* (the horizontal position of the target). The levels of *V* and *H* are labeled as in Experiment 1, and correspond to the same measures defined above. The levels of *T* are denoted by –2, –1, 0, 1 and 2, and correspond to measures –15°, –5° (tilts of the target to the left), 0° (no tilt), and 5° and 15° (tilts of the target to the right). The pictorial length of the target, as measured on the monitor, was kept constant (14.2 cm) across the stimuli. The distance from the eyes to the monitor was 74 cm. At this distance, the visual angle subtended by the target was 10.96 degrees.

All other conditions – the geometric, photometric and simulated properties of the frame, colour of the target, properties of the monitor, position of the observer relative to the monitor, etc. – were the same as in Experiment 1.

Procedure. The mechanical device for the adjustment task was a ball-and-socket joint. The socket – a truncated hollow sphere – was fixed through a pedestal to the table at which the participant was sitting. A pole 5 mm in diameter, 20 cm in length, and red in colour was joined to the ball, and could be rotated by hand in all directions (max 86° from the vertical). At the base and vertex of the pole were two markers which reflected rays coming from three infrared cameras (ELITE motion analysis system, Bioengineering Technology & Systems, BTS), these being placed at three upper corners of the experiment room. The cameras were linked to the workstation used to control the experiment, and the integrated system was equipped so as to compute and record the 3-D orientation of the adjustable pole at any given moment in the experimental session. The orientation was expressed as a pair of angular measures, which were slant (with respect to the vertical, as defined in the Introduction) and tilt (i.e., the tilt of the orthogonal projection of the pole on the frontal plane). The device for adjustment was placed in front of the participant, 24 cm from the chin-rest and at a height so that the vertex of the pole (when it stood vertical) was 17 cm below the line of sight (from the eye to the centre of the monitor). On both sides of the device, at a distance of 45 cm, were two lamps which selectively lighted up the pole when the participant was performing the adjustment task. Apart from these lamps and the monitor, there was no other source of light to the participant in the experiment room. The set-up is sketched in Figure 3.3.

The organization of the experimental session – practice and main phases, randomized order of stimuli, a pause half way through the series of stimuli, etc. – was similar to that of Experiment 1. For each participant, the main phase was a series of 50 trials, one for each of the stimuli. In each trial, after a signal on the monitor (a white fixation cross, 0.5 sec), a stimulus was shown for 5 sec. As the stimulus disappeared, the side lamps were automatically turned on, and the participant had to adjust the pole of the device to match the orientation of the pole in the scene he or she had just seen on the monitor – i.e., to adjust the direction of the mechanical pole so that it was 3-D parallel to the pole in the scene. This done, the participant pressed the space-bar on a computer keyboard, so that the orientation of the adjusted pole could be computed and recorded. The session took from 40 to 45 min to complete.

Participants. The experiment was carried out by 20 subjects (9 women and 11 men), whose general characteristics were similar to those of participants in Experiment 1. Eight of them had also taken part in Experiment 1, about three months before.

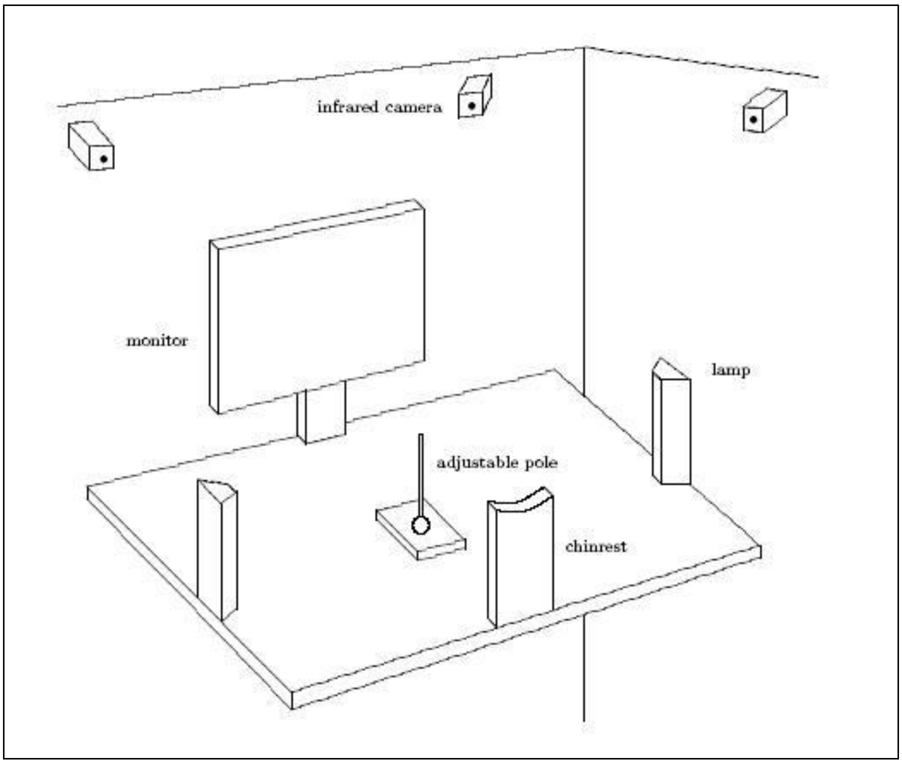


Figure 3.3. Set-up of Experiment 2, with monitor, device for adjustment, and infrared cameras for capturing orientations of adjustable pole.

3.3.2 Results and discussion

The data considered were those on stimulus variables *V*,*T*,*H* and response variable *S* which, in Experiment 2, was the *matched slant* of the adjustable pole, i.e., the slant of the pole of the mechanical device as it was computed at the end of each adjustment operation. On these data, I ran a repeated-measures three-way ANOVA, which yielded significant results for the main effects of factor *V* (*F*(4,76)=27.568, *p*<0.001), main effects of factor *T* (*F*(4,76)=7.771, *p*<0.001), interaction effects of factors *V* and *T* (*F*(16,304)=4.299, *p*<0.001) and interaction effects of factors *T* and *H* (*F*(4,76)=16.375, *p*<0.001). No other significant effects were obtained (level of significance 0.001).

To interpret these results, I refer to Figure 3.4, which was obtained by plotting the specific means of variable *S* against the levels of factor *V* (part (i)), the levels of factor T (part (ii)), the combined levels of factors V and T (part (iii)), and the combined levels of factors T and H (part (iv)). Figure 3.4.i shows that, between response variable S and stimulus variable V, there is an increasing monotone dependence: as V increases over its range $\{-2,-1,0,1,2\}$ there is a corresponding increase in S, from negative (downward slant) to positive values (upward slant). This dependence fits that found in the data of Experiment 1 (cf. Figure 2), and both of them support the main claim of the study: the vertical unbalance of the target in the stimulus does influence the apparent slant of the target itself, in a way which accords with a tendency to contact preservation. Figure 3.4.ii shows two special aspects of the relation between stimulus variable T and response variable S. One aspect is that, among the values of T, value $T=0$ was the one which mostly favoured a response close to $S=0$, i.e., the adjustable pole was set perpendicular to the line of sight. This finding corresponds to the privileged association between condition $T=0$ and response $S=0$ noted in discussing component λ^{TS} of the log-linear model for Experiment 1. The other aspect is that, when $T\neq0$, the matched slant moved toward positive (rather than negative) values, so that, on the whole, the upward direction of matched slant prevailed. This result is commented on in the next paragraph. Figure 3.4.iii represents 15 means of variable S, corresponding to subsets of stimuli distinguished by the levels of factor V and by conditions $T=0$, ($T=-1$ or $T=1$), and ($T=-2$ or $T=2$) on factor T (I apply this simplification, because diagrams of means for $T=-1$ and $T=1$ run quite close to each other, and the same is true of $T=-2$ and $T=2$). Figure 4.iii shows that the diagram relating the means of S to the levels of V is steeper in condition ($T=-2$ or $T=2$) than in condition $T=0$, which plausibly signifies that the dependence of S on V is stronger in the former condition than in the latter. Thus, Experiment 2 gives a positive answer to one of the initial questions concerning the interaction of stimulus variables V and T on response variable S (the answer from Experiment 1 was not convincing, as I noted in commenting on the absence of component λ^{VTS} in the accepted log-linear model). Lastly, Figure 3.4.iv shows that there were two circumstances in which the tendency to set the pole in the upward direction became reinforced: when $H=-1$ and $T>0$ (i.e., the target was in the left visual hemifield and tilted to the right in the stimulus) and when *H*=1 and *T*<0 (i.e., the target was in the right hemifield and tilted to the left). These associations fit the psychophysical relational constraint described in commenting on component λ^{THS} in the log-linear model, and may be taken as a further proof of the intervention of a "process by inverse optics" conditioning perceptual slant.

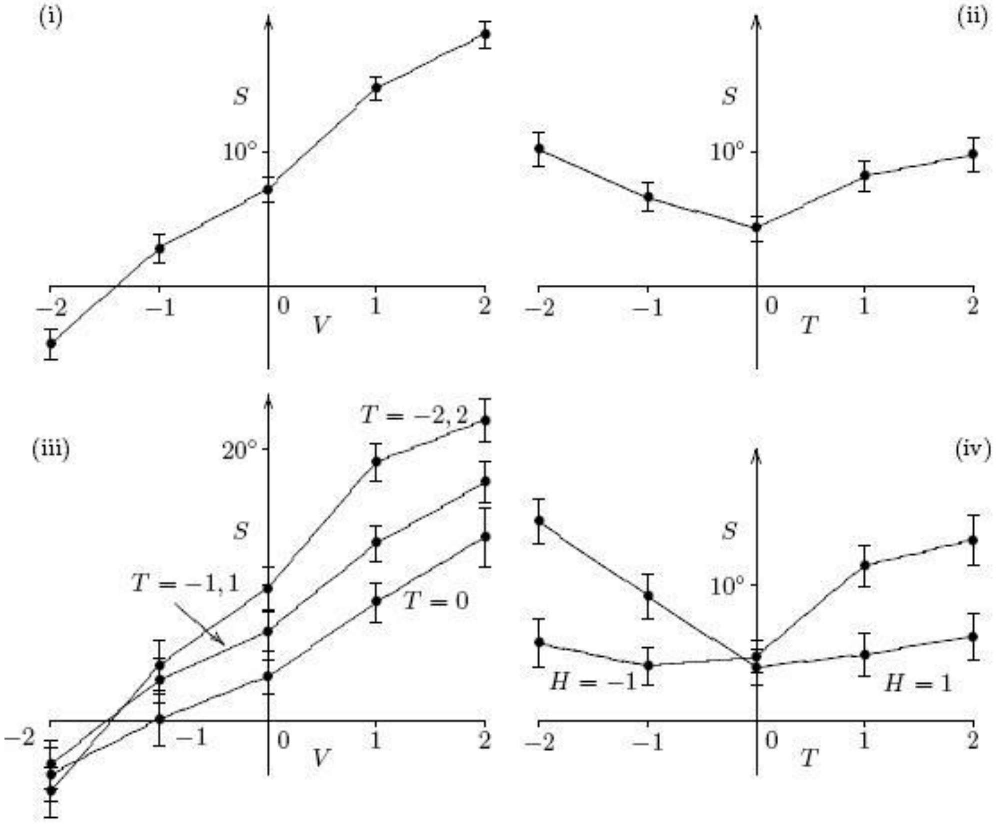


Figure 3.4. Plots of means of matched slant *S* corresponding to levels of *V* (part (i)), levels of *T* (part (ii)), combined levels of *V* and *T* (part (iii)), and combined levels of *T* and *H* (part (iv)), from Experiment 2. Matched slant (ordinate) is measured in degrees. Positive vs negative values for upward vs downward slant. Length of vertical bars is twice standard error of corresponding means.

Figure 3.5 plots the means of the matched slant for different levels of factor *V* against the simulated slants corresponding to those levels. Two aspects of the diagram are conspicuous: its greatly diminished gradient when compared with the identity function (0.161 is the slope of the least-squares fitted line) and its displacement on the upper half-plane (7.155 is the mean matched slant in condition V=0, and 7.783 is the intercept of the least-squares fitted line). The first aspect means that, for any two stimuli, the difference between the associated matched slants – and, presumably, between the perceived slants, which are estimated through the adjustment operation – was much smaller than the difference in slant as it was simulated by computer graphics. There were probably several factors concurring to produce this effect. One is perceptual underestimation of the depth of the frame (the apparent corridor), when compared with the simulated depth, which was set at 2120 cm in the computer graphics program. Actually, for the stimuli, the apparent extent in depth of the frame is only supported by pictorial cues of texture, linear perspective, and illumination perspective (whereas other stimulus circumstances, like binocular vision in the absence of binocular disparity, signal the flatness of the scene), so that it is not surprising that the truly perceived depth of the corridor was definitely smaller than that theoretically fixed by simulation. But a smaller perceived depth of the frame obviously implies a smaller perceived slant of the target, as it is induced through optical contacts (on discrepancies between depth simulated in virtual environments, and depth actually seen in them, cf. Loomis & Knapp, 2003). Another factor of slant flattening may be the general “frontal tendency” or “tendency to the frontal-parallel plane”, which is a classic hypothesis in the study of space perception (cf. Gibson, 1950b, p.381; Koffka, 1935, pp. 231-232). The other conspicuous aspect of the diagram in Figure 3.5, i.e., its displacement on the upper half-plane, is also apparent in Figure 3.4.ii, which shows the subjects’ general tendency to set the adjustable pole in upward (positive) directions – in spite of the exact balance within the experiment between conditions favouring an upward and those favouring a downward apparent slant of the target. A general prevalence of upward over downward responses was also found in Experiment 1 (cf. component λ^S of the log-linear model), and this may be interpreted as a perceptual bias – contingent on our variety of stimuli – for an upward apparent slant of the target. But the large effect found in Experiment 2 may also be due to other factors specifically related to the response procedure. For evidence on this point, I performed one additional experiment, in which 18 of the 20 subjects who participated in Experiment 2 were set in the same conditions, except that the monitor showed no structured stimulus (it was illuminated but empty). The task was to adjust the pole of the response device so as to appear vertical. The values I obtained ranged from –2.8° to 9.01° (on the matched slant scale, measured by the ELITE system), with M=2.47° and SD=2.92°. Thus, I may conclude that there was also an effect conditional on the special set-up of Experiment 2 which concurred to determine the upward bias observed in the data.

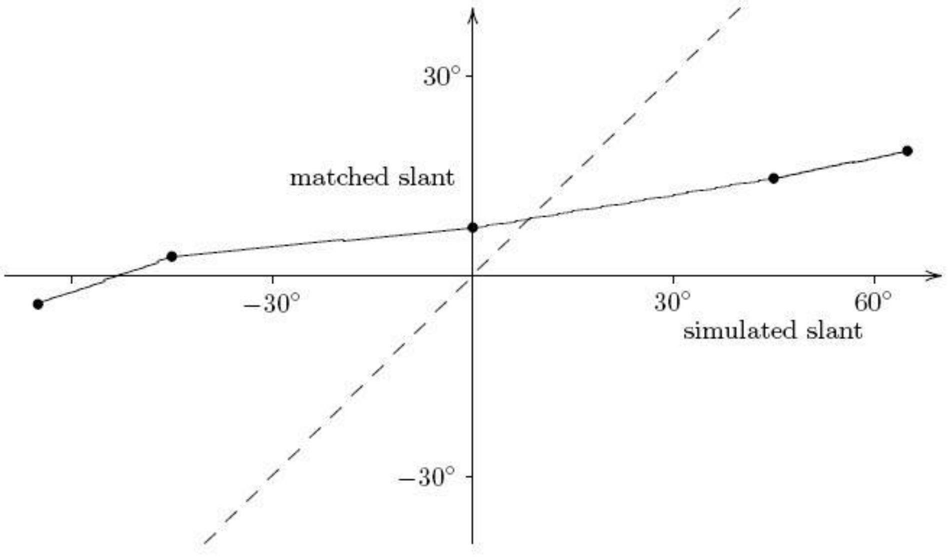


Figure 3.5. Mean matched slants from Experiment 2, plotted against simulated slants of target inside simulated corridor. Matched and simulated slants are measured in degrees.

In this and the next two paragraphs, I compare the “process by optical contact”, which represents the main hypothesis of our study and has received support from both experiments, and the “process by inverse optics”, which has emerged in discussing component λ^{THS} of the log-linear model fitted to the data of Experiment 1 and has been confirmed by the interaction of factors T and H on variable S in Experiment 2. The set of 50 stimuli in Experiment 2 may be partitioned into set A_d of 20 stimuli, in which V<0, set A_v of 10 stimuli, in which V=0, and set A_u of 20 stimuli, in which $V>0$. The same set of 50 stimuli may also be partitioned into set B_d of 20 stimuli, in which either ($H=-1$ and $T<0$) or ($H=1$ and $T>0$), set B_v of 10 stimuli, in which $T=0$, and set B_u of 20 stimuli, in which either ($H=-1$ and $T>0$) or ($H=1$ and $T<0$). Now, stimuli in A_d, A_v and A_u are those in which, by the optical-contact process, downward ($S<0$), vertical ($S=0$) and upward ($S>0$) responses are favoured, respectively. The same may be said of stimuli in B_d, B_v and B_u, but with reference to the inverse-optics process. Of special interest are some intersections between these sets: stimuli in $A_d \cap B_d$ (resp., in $A_u \cap B_u$) are those in which the two processes both support a downward (resp., upward) response, whereas stimuli in $A_d \cap B_u$ and $A_u \cap B_d$ are those in which the two processes disagree, as for any stimulus in $A_d \cap B_u$ the optical-contact process supports a downward response and the inverse-optics process an upward one, and the opposite is true for any stimulus in $A_u \cap B_d$. These combinations form a suitable base in order to examine how the two processes interact with each other, and to compare their effectiveness in conditioning the same perceptual property (the apparent slant of the target).

Table 3.5 lists the means and standard deviations of the response variable (matched slant) for 8 of the 50 stimuli in Experiment 2 (each statistic is computed on 20 measures, which are the slants matched by the participants for one of the stimuli). The 8 stimuli are representative, as they are extreme in the four intersected blocks. Specifically, stimuli ($H=-1,T=-2,V=-2$) and ($H=1,T=2,V=-2$) are extreme in block $A_d \cap B_d$, stimuli ($H=-1,T=2,V=-2$) and ($H=1,T=-2,V=-2$) in block $A_d \cap B_u$, stimuli ($H=-1,T=-2,V=2$) and ($H=1,T=2,V=2$) in block $A_u \cap B_d$, and stimuli ($H=-1,T=2,V=2$) and ($H=1,T=-2,V=2$) in block $A_u \cap B_u$. Two aspects of the table deserve comment. One is that, on both left and right parts of the table, the means which are extreme (either negative or positive) correspond to stimuli in which the optical-contact and the inverse-optic processes agree with each other in supporting a certain direction of perceptual slant. For example, on the left part, the extreme negative mean is that for stimulus ($H=-1,T=-2,V=-2$), which belongs to block $A_d \cap B_d$ (both processes support a downward perceptual slant), and the extreme positive mean is that for stimulus

(H=−1,T=2,V=2), which belongs to block $A_u \cap B_u$ (both processes support an upward perceptual slant). A similar scheme may be seen on the right part of the table. This configuration may be interpreted as proof that the two processes interact by *accumulating* their effects on the resulting perceptual property (the apparent slant of the target): when the predicted effects of both processes are in agreement, then the resulting property is more extreme in value (than when their predicted effects disagree), and the value accords in direction with the predicted effects (see Section 1.3.2 for a definition of accumulation process among depth cues). The other aspect deserving comment regards differences between means in the table. Specifically, I find that the two differences of means in the same *row* of the left part of the table (i.e., 24.308−(−6.053)=30.361 and 16.734−(−9.321)=26.055) are much larger than the two differences of means in the same *column* of that part (i.e., −6.053−(−9.321)=3.268 and 24.308−16.734=7.574). This discrepancy is corroborated by the t-test, which yields significant results for both differences within row (p<0.001), but for no difference within column (p=0.575 and p=0.127). The same scheme holds true on the right part of the table. Now, differences within rows are those due to V (the vertical unbalance of the target), which is the key factor in the optical-contact process, whereas differences within columns are those due to T (the pictorial tilt of the target), which is the key factor in the inverse-optics process. Thus, the observed discrepancy between differences may be interpreted as proof that the optical-contact process is *stronger* than the inverse-optics process in conditioning the perceptual slant of the target.

Presuming that the effects of both processes combine in a linear way (cf. Dosher, Sperling & Wurst, 1986; Meese & Holmes, 2004), let us fit a linear regression model to the data. The response variable is S (the matched slant), the explanatory variables are V (the vertical unbalance of the target) and a new variable W defined as $-(T \times H)$, which both have $\{-2,-1,0,1,2\}$ as the set of possible values.

	$H=-1$	
	$V=-2$	$V=2$
$T=2$	$M=-6.053$ $SD=20.604$	$M=24.308$ $SD=14.065$
$T=-2$	$M=-9.321$ $SD=16.256$	$M=16.734$ $SD=16.569$

	$H=1$	
	$V=-2$	$V=2$
$T=2$	$M=-4.991$ $SD=16.043$	$M=20.642$ $SD=13.206$
$T=-2$	$M=-0.774$ $SD=14.046$	$M=26.847$ $SD=12.599$

Table 3.5. Means and standard deviations of matched slants (measured in degrees) for 8 of 50 stimuli in Experiment 2.

Note that the hypothesis that the inverse-optics process is effectual on variable S (according to the psychophysical relational constraint introduced in discussing data from Experiment 1) is tantamount to the hypothesis that variable S depends in an increasing monotone way on new variable W. The least-squares linear equation for our 50×20=1000 data is the following:

$S=7.784+5.738\times V+2.196\times W$.

The residual standard error of the data is 13.67 (df=1000–3=997) and the coefficient of determination is 0.288. All three estimated parameters ($\alpha=7.784$, $\beta^V=5.738$, $\beta^W=2.196$) are significantly greater than zero (t-test, $p<0.001$), and estimate β^V is significantly greater than estimate β^W (F-test, $p<0.001$). Now, the positive value of α – the intercept of the linear regression – corresponds to the prevalence of positive matched slants discussed when referring to Figures 3.4 and 3.5. The positive values of β^V and β^W – the coefficients of the linear regression – correspond to the monotone way in which the optical-contact and the inverse-optics processes are presumed to condition the apparent slant of the target. Lastly, the fact that β^V is significantly greater than β^W corroborates the conclusions reached in the previous paragraph, concerning the difference in strength between the two processes in conditioning the perceptual slant of the target.

I computed the linear equation as a supplement to our discussion of Table 3.5. The equation cannot be proposed as a formula describing the real way in which response variable S is formed, for at least two reasons. One is that coefficients β^V and β^W in the equation are conditional on values $\{-2,-1,0,1,2\}$ of stimulus variables V

and *W*, relative to which they are computed, and these values are (partly) arbitrary. We have no explicit measures of the slants specified separately by the optical-contact and inverse-optics processes: we only presume that these slants are monotonically related to the values of *V* and *W* applied in the computation. Thus, coefficients β^V and β^W cannot be directly interpreted as the “weights” of the two processes in a (presumed) linear combination giving rise to the perceived slant of the target. The other reason is that the linear scheme itself presumably is not well suited to represent the way in which several factors interact in determining response variable *S*. One sign to this effect is the relatively low coefficient of determination (R^2=0.288) associated with the linear equation (however, when the regression is computed over the matched slants averaged across participants, then R^2=0.845). Another sign is the veto-like effect of condition *T*=0 on variable *S*, revealed by Experiment 1 (Tables 3.1.iii and 3.3) and corroborated by Experiment 2 (Figure 3.4.ii).

3.4 General discussion

Perceptual slant is a notable aspect of vision in depth, and has been the subject of several studies in the psychophysics of visual perception (representative recent contributions are Creem, Regehr, Gooch, Sahm, & Thompson, 2004; Grove, Ono, & Kaneko, 2004; Hillis, Watt, Landy, & Banks, 2004; Knill & Saunders, 2003; Oruç, Maloney, & Landy, 2003; Rosas, Wichmann, & Wagemans, 2004, 2007). I conceived present work as a contribution to this subject, and directed research following three methodological choices. First, I limited myself to studying the apparent slant of a *linear* object (its appearance is that of a pole), rather than the slant of a surface, or of objects of a more complex 3-D shape. Second, I treated apparent slant as an endogenous perceptual property, which means a property induced on the carrier object by a structured frame external to it (rather than slant as an exogenous property, directly supported by depth cues residing on the carrier, like texture gradients on a surface, shading, linear perspective, etc.). Third, I considered cases of slant induction in which a key role was played by *contact* relations between the inducing component (the frame/corridor) and the induced component (the target/pole), according to a hypothetical "tendency to contact preservation" in the passage from the 2-D pictorial stimulus to the 3-D perceptual scene.

The main prediction of this research was that there is a consistent association – specifically, a monotone dependence – between the stimulus property called "vertical unbalance" of the target, denoted by V, and the perceived slant of the target, denoted by S. The argument leading to this prediction rests on the hypothesis of contact preservation, and the hypothesis that, through contact, the intrinsic 3-D perceptual organization of the frame may condition the apparent 3-D orientation of the target. The experimental data lent concordant support to the prediction, in spite of differences in the experimental procedures (a categorical task in Experiment 1, a matching by adjustment task in Experiment 2). The expected monotone dependence of response variable S on stimulus variable V is given by component λ^{VS} of the log-linear model fitted to the data of Experiment 1, and by the main effects of factor V on variable S in the ANOVA run on the data of Experiment 2. Evidence in favour of the prediction is also evidence in favour of both the hypotheses from which that prediction derives.

Besides supporting main expectation, the data also revealed two other processes which involve stimulus variable T and affect responses about the apparent slant of the target. One process is a privileged association between condition T=0 (the target is pictorially vertical) and response S=0 (the target is judged to be vertical in slant in the 3-D perceived scene); the other is a combined effect acted upon response variable S by stimulus variables T and H – an effect I interpreted following the heuristics "vision is inverse optics". The two processes are testified by components λ^{TS} and λ^{THS} in the log-linear model fitted to the data of Experiment 1, and by the main effects of factor T and the interaction effects of factors T and H on response variable S in the ANOVA run on the data of Experiment 2. The emergence of these processes is proof that, even in the simplified observational contexts considered in our experiments, the real process of formation of a response concerning apparent slant is presumably at a higher level of complexity than the basic paradigm of a "depth induction effect". The evidence also calls for an examination of the way in which the distinct processes interact in conditioning responses of apparent slant. The interaction of effects determined by "optical contact" and other factors of perceptual depth – like occlusion, shadows, or motion parallax – has been examined in some recent studies on depth vision (cf. He et al., 2004; Madison, Thompson, Kersten, Shirley, & Smits, 2001; Ni, Braunstein & Andersen, 2004, 2005, 2007). The comparison I made between the "process by optical contact" and the "process by inverse optics", using the results of Experiment 2, is a step in this direction.

The paper mentioned at the beginning of this work (Bian et al., 2005) is specifically concerned with the "ground dominance effect". This is the name of a psychophysical hypothesis according to which, for a stimulus comprising a bottom region and a top region, which become a ground surface and a ceiling surface in perceptual rendering, the optical contact of an object with the former region is more strongly binding than that with the latter, in the sense that the hypothesized tendency to "contact preservation" is stronger relative to the ground surface. If this is true, it implies that, when the stimulus is such that bilateral contact of the object is "phenomenally impossible", i.e., it is structurally unlikely that the object will be perceived as being in contact with both ground and ceiling surfaces (in spite of bilateral contact in the stimulus), then the former alternative should prevail, i.e., the object is perceived as being in contact with the ground only (unilateral contact). This is the prevalence named "ground dominance". There is an obvious connection between this hypothesis and Gibson's "ground theory" of space perception. The hypothesis lends itself to an ecological and adaptive interpretation, as the behavioural environment is thoroughly conditioned by the force of gravity, so that, in natural contexts, static objects are and appear to be resting on the ground surface (the terrain), not suspended or floating above it (cf. Nakayama, 1994). In Bian et al. (2005) and other experimental studies (Bian et al., 2006; McCarley & He, 2000), evidence has been gathered in favour of the hypothesis.

Is there anything I can add to debate this problem, based on the data from our experiments? I prepare an answer by the following argument, concerning a relation between stimulus variable V, response alternative $S=0$, and the perceptual contacts of the target. If $V>0$ (as in Figure 3.1), then response $S=0$ is phenomenally compatible with either no perceptual contact of the target (the pole appears to be floating inside the corridor), or with unilateral perceptual contact of the target with the *ground* (a unilateral perceptual contact with the ceiling is not compatible with percept $S=0$, since the two circumstances would imply that the pole appears to "pierce" the ground surface and continue beneath it, which is impossible, given the perceptual opacity of the surface). A symmetric argument leads us to conjecture that, if $V<0$, then response $S=0$ is phenomenally compatible with either no perceptual contact of the target, or with unilateral perceptual contact of the target with the *ceiling*. Now, let us interpret the hypothesis of "ground dominance" as stating that to renounce perceptual contact with the ground is more difficult/unlikely than to renounce perceptual contact with the ceiling. Based on this notion and the above argument, we would then predict that response $S=0$ is more likely in condition $V>0$ than in condition $V<0$, as – for *unilateral* perceptual contact – the former implies loss of contact with the ceiling and the latter with the ground. I compare this prediction with the data of Experiment 1, considering only data with $T=0$ (to ensure a simpler context, as $T\neq0$ entails the intervention of the "process by inverse optics"). The results are as follows: in the 2×3×2×30=360 responses to stimuli in which $V>0$ and $T=0$, alternative $S=-1$ has frequency 15, alternative $S=0$ frequency 279, and alternative $S=1$ frequency 66; the corresponding frequencies for stimuli in which $V<0$ and $T=0$ are 73, 271 and 16. Thus, this section of the data does not support the stated prediction. I must add, however, that the stimuli used in our experiments are not very suitable for testing the "ground dominance" hypothesis. Indeed, the target in the stimuli has no inner structure constraining it to a certain orientation in the perceptual 3-D space: in this regard, the target is highly "flexible" in spatial orientation (especially when $T\neq0$). *Bilateral* perceptual contact of the target with the frame is phenomenally quite plausible, and through bilateral perceptual contact any discrepancy in optical contact of the two ends of the target may turn into the apparent slant of the target itself.

Chapter 4

Interaction between Optical Contact and Objects Shape

4.1 Introduction

The two experiments described in Chapter 3 showed an effect of the optical contact in determining the perceptual slant of a pole. Furthermore, it has been found that this effect interacts with concomitant intrinsic factors. In particular, I found that if a pole has zero tilt, the chance of perceiving it as having zero slant increases, even when optical contact indicates a positive or negative slant, and that slant is affected by the interaction of tilt and horizontal position, as may be predicted based on a perspective analysis of the scene.

The solution to choose a pole as target was mainly due to an interest in exploring the effect of optical contact in relatively simple observational contexts. The scene analyzed in the two previous experiments was composed by two different aggregates, a frame and a target, in which the first one had a well-defined intrinsic three-dimensional structure, whereas the second missed this feature (as I found, the only intrinsic factors were those related to the tilt of the target). In principle, it is possible to conceive more complex situations in which, for instance, the aggregates are more than two (see, for instance, Bian, Braunstein & Andersen 2005), in which I may presuppose the intervention of a chain of reference links so that some aggregates can be both frame and target (see, for instance, Meng & Sedgwick, 2001), in which different types of contact between target and frame are realized (see, for instance, Koning & Van Lier, 2003), in which target has a more defined three dimensional structure as effect of more intrinsic properties residing on it. The experiments described in this chapter will specifically address this last topic.

It is well-known that orientation in depth depends on the shape of the object stimuli: for instance, Braunstein & Payne (1968) found that “ellipses provide a much weaker cue to orientation than do rectangles” (reported in Bian, Braunstein &

Andersen, 2001, p. 811). The experiments described in the current chapter address the problem of how vertical unbalance resulting from the optical contact may perceptual slant for bi-dimensional targets (plane figures). The possible presence of factors as “good form tendency” or “tendency to form regularization” (Wertheimer, 1923; Koffka, 1935; Metzeger, 1976) will be the main difference between the two experiments: in particular, the first experiment investigates the role of optical contact when irregular shape targets are used as stimuli, while the second experiment investigates the role of optical contact with regular shape targets. The main hypothesis is that the effect of the optical contact on two-dimensional stimuli should be weaker than the effect on one-dimensional targets, given the highest degree of internal coherence for the first type of stimuli. This difference between two-dimensional and one-dimensional stimuli could result in a stronger resistance in seeing the former as perceptually slanted.

A wide surface allows the introduction of a larger set of cues, residing on it, such as shading (i.e., difference in brightness among different parts of the aggregate) or texture. Then, the use of planar surfaces is suitable to provide further possibilities about the interaction between intrinsic and extrinsic factors in determining the perceptual slant. In particular, the second experiment is aimed a studying the interaction of some of these factors and the optical contact.

The two experiments described in the next pages (in particular, the second one) both deal with factors that have precise location in the stimuli figure. This fact gives the opportunity of studying which factors are sought, in the visual scene, when providing a judgment on some characteristics of it. By collecting gaze position as long as subjects visually explore the scene, it is possible to determine which kind of information they look for, and whether this information is effectively used in the perceptual judgments. Eye gaze records have been very useful in studying how observer looks at a scene, which parts of it are generally considered as more informative, and which are those that capture the observer’s attention (Loftus & Mackworth, 1978; Rayner, 1998). The areas in which optical contact is realized can be considered very high in informativeness about slant. I presume that a consistent part of the subject's gaze is directed to these areas. Eye movements are explored only in the second experiment.

4.2 Experiment 1

The first experiment had the aim to test if optical contact is crucial to determine perceptual slant also when targets are two-dimensional.

Since the main goal of this experiment was to study only the effect of vertical unbalance V on targets with higher structural complexity than those previously studied, and given the privileged association between $T = 0$ (target vertical on the image plane) and perceptual response $S = 0$ (target has no slant), considering the results described in Chapter 3, I decided to include only situations in which $T \neq 0$ (target is tilted on the image plane).

Perceptual further factor which could affect perceptual slant is the type of contact that takes place between the frame and the target. In particular, a different effect could be due to the shape of the extremity of the target in optical contact with the frame. Bian, Braunstein & Andersen (2002) did not directly address this problem in their experiment about the "ground dominance effect" but, after changing the shape of the target, they found a weaker effect when the target was elliptical. The main difference about the rectangular post and the ellipses they used is relative to the extremities, which are rectilinear in one case and curved in the other case. In order to test if this difference can be found also in the determination of slant, I introduced one further factor, the shape of the vertical extremities of the target. In particular, I distinguished between stimuli with punctual contact and stimuli with rectilinear contact (see images in Table 4.1). A different effect of contact shape could eventually be proved by comparing the amount of variation of responses among the different levels of the type of contact (under the hypothesis that the stronger is the contact, the lesser should be the variance of data).

Since the shape of target has some bearing on my experimental hypotheses, I also varied it as regards its complexity, simply measured in terms of number of edges. Initially, there was no definite expectation regarding the efficacy of this factor: this variation was introduced in order to create a sufficiently large set of stimuli on which to test the generality of the main hypothesis. Moreover, to avoid the intervention of linear perspective related factors, I used rounded edges (except for the upper and lower extremities, in which the edges were rectilinear in the case of non-punctual contact)

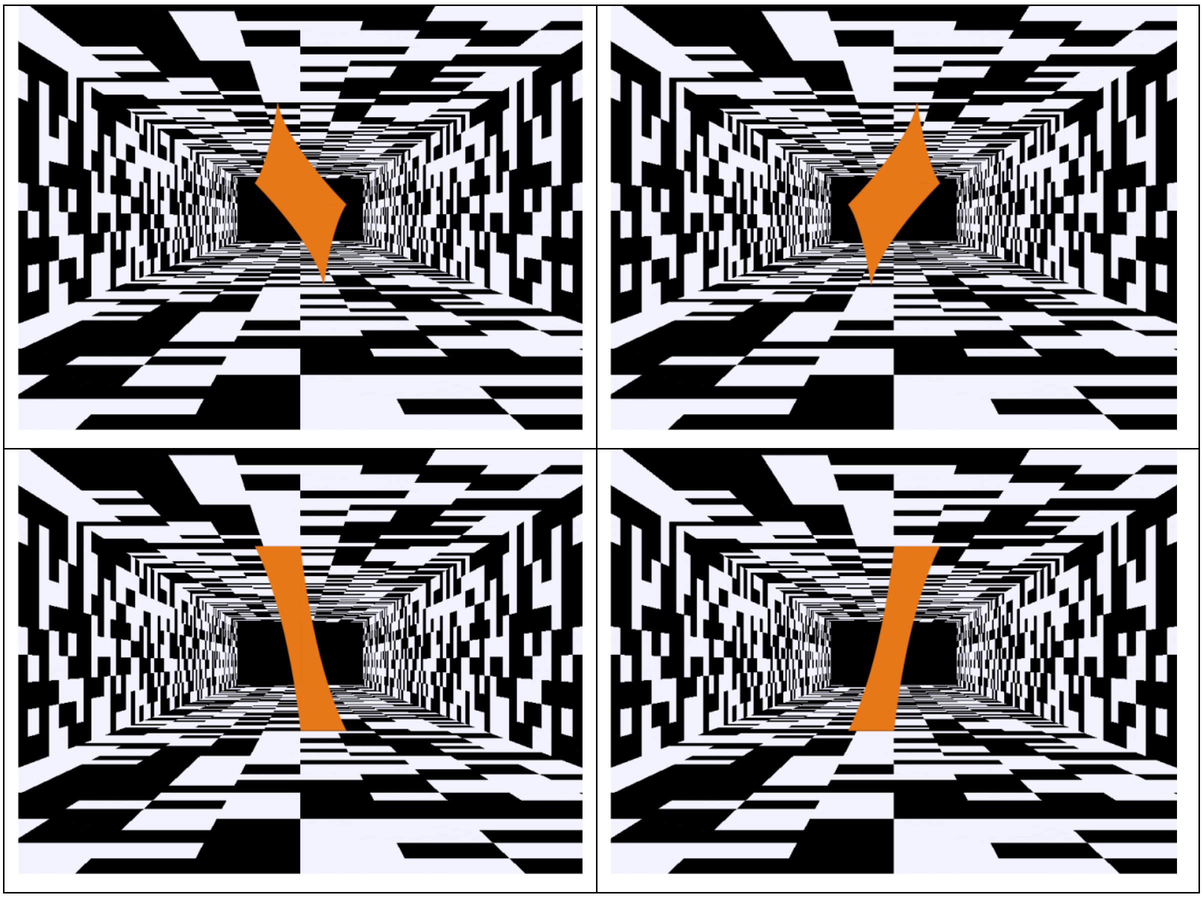
4.2.1 Experimental setup

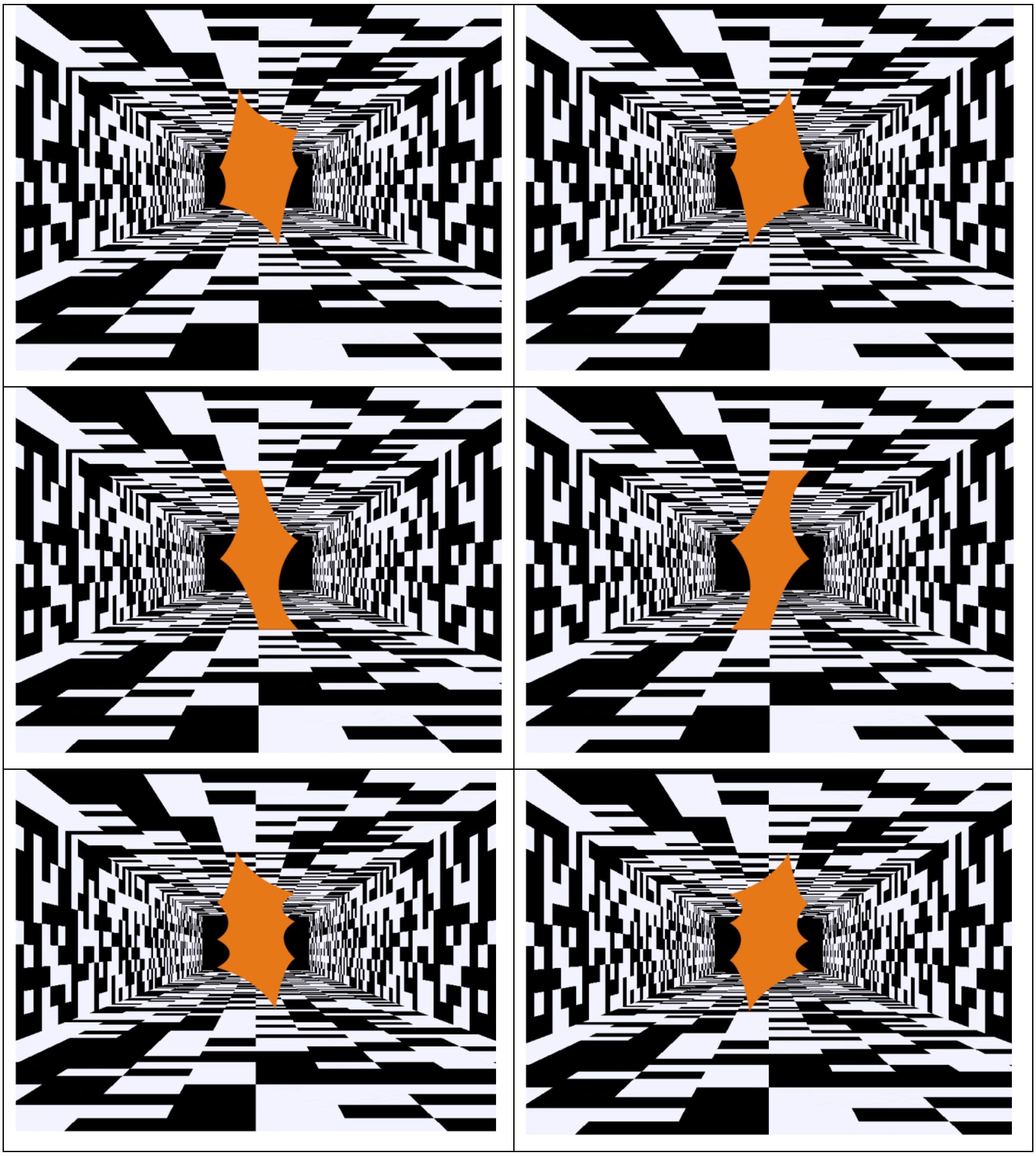
Participants. Twenty three students of the University of Padova participated in the study (mean age 24.60). All of them had normal or corrected-to-normal vision. They were naïve with respect to the purpose of the experiment. Data for three of them were discarded since the recording apparatus had some technical difficulties and data were inaccurate.

Stimuli. 84 figures served as stimuli. The tridimensional structure of the frame was created using Autodesk 3D studio Max 8.0. Target stimuli were drawn and then over-imposed on the frame using the graphic software Corel Draw X. As in the previous experiments, frame was composed of four adjacent regions, textured using a random checkerboard motif, so as to simulate a corridor extended in depth. The simulated width of the corridor was 800 cm, the simulated height was 400 cm, the simulated extension in depth 5000 cm.

The frame was the same in all the stimuli (see pictures in Figure 4.1 for an illustration of the frame). The simulated vantage point was at height 50 cm from the ground of the corridor and at a distance 0 cm from its opening. As in the experiments described in the previous chapter, three kinds of depth cues were supporting the perceptual rendering of the frame as a 3-D structure: texture, linear perspective and illumination perspective.

Target pictorial objects were obtained by combining 7 levels of factor V, 2 levels of factor C, three levels of factor E, two levels of factor T. The levels of factor V (Vertical Unbalance of the target) were denoted by -3, -2, -1, 0, 1, 2, 3 and were specified so as to produces values of -60°, -45°, -20°, 0°, +20°, +45°, +60° as the simulated slant of the target. The 2 levels of factor C (Contact Shape) were denoted as 0 or 1, corresponding to punctual contact vs. rectilinear contact. The 3 levels of factor E (number of Edges) were denoted as 4, 6 and 8, these numbers corresponding to the number of edges of the figure. The 2 levels of factor T (Tilt) were denoted by -1 and 1, and corresponded to the lateral inclination of the horizontal symmetry axis of the target for the value -10°(10° to the left) and tilt 10° (10° to the right). A sample of stimuli containing target different for number of edges, contact shape and tilt are displayed in the pictures collected in Figure 4.1.





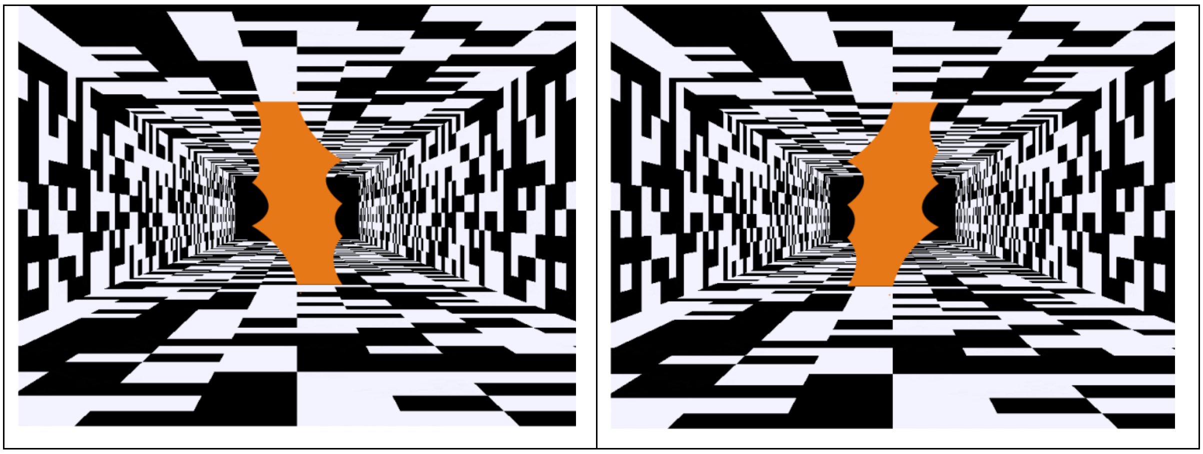


Figure 4.1. A sample of the stimuli used in the experiment. Each one of these images represents a situation in which the predicted value of the slant according to the optical contact is 0° (i.e., no slant). Pictures on the left column are left tilted, while pictures on the right column are right tilted.

Apparatus. Stimuli were displayed on a 22'-inch CRT monitor, with a pixel resolution of 800x600, controlled by a Windows XP workstation. The distance from the eye to the monitor was 57 cm. A chinrest was mounted at a position appropriate to the viewing distance. Vision was binocular. The experiment was run in a darkened room.

The mechanical device that participants were requested to adjust was a small wooden table fixed to a larger wooden table by using two hinges, positioned at the right of the participant at a distance of about 47 cm. The wooden table was 12 cm high, 20 cm wide, 1 cm thick. Participants could lean the surface forward or backward by simply pushing or pulling it manually.

Behind the wooden table, an Intersense Tracker V100 was attached in order to register in real time the slant of the table itself. Stimuli presentation and response collection were all accomplished using Dassault Systemes Virtools Dev 3.5.

Procedure. For each participant, the experiment began with a practice phase, aimed at illustrating the kind of stimuli to be observed and how to express responses. After the training phase, all 84 stimuli were separately presented in a randomized order for each participant. The sequence of 84 trials was split into two blocks of 42 trials. The two blocks were separated by a pause, the duration of which was decided by participant. Each trial had the same structure: a white fixation cross appeared and remained on the screen for 0.5 seconds. After the fixation cross disappeared, one of the stimuli was shown for 5 seconds. Then, the screen became blank. The participants' task was to adjust the wooden table to match the slant of the target as it was seen on the monitor, so that the tables was parallel to the target. After having expressed their judgment, participants were requested to press a button on a keyboard, positioned in front on them, in order for the next trial to be executed. The experimental session, including both practice and main phase, took from 8 to 15 minutes to be completed.

4.2.2 Results and discussion

The variable considered were stimulus measures V, C, E, T and response variable S, corresponding to the *matched slant* of the adjustable surface.

On these data, a repeated-measures four way ANOVA was run. The analysis yielded significant results for the main effects of factor V [$F_{(6, 144)} = 25.498$, $p < .001$, $\eta^2_p = .515$], the main effects of factor T [$F_{(1, 24)} = 9.980$, $p = .004$, $\eta^2_p = .294$], and the interaction effects of factors V, E and C [$F_{(12, 288)} = 2.629$, $p = .002$, $\eta^2_p = .099$]. No other significant effects were obtained, using a significance level of 0.05.

Results can be interpreted by referring to Figure 4.2, Figure 4.3 and Figure 4.4, in which the means of response variable are plotted against the levels of factor V, the levels of factor T and the combined levels of factor V, E and C respectively.

Figure 4.2 clearly shows that there is an increasing monotone dependence between response variable "matched slant" and stimulus variable V, as found in the two experiments described in Chapter 3: as V increases over its range, matched slant increases from negative (downward slant) to positive values (upward slant). This dependence supports our main hypothesis: the vertical unbalance of the target in the stimulus influences its apparent slant of the target, as predicted if the tendency to contact preservation holds true.

Another aspect deserving notice is that the matched slant toward positive values prevails over the matched slant toward negative values (this result can be easily noticed also in Figure 4.3). This trend has been already found in the two experiments described in Chapter 3.

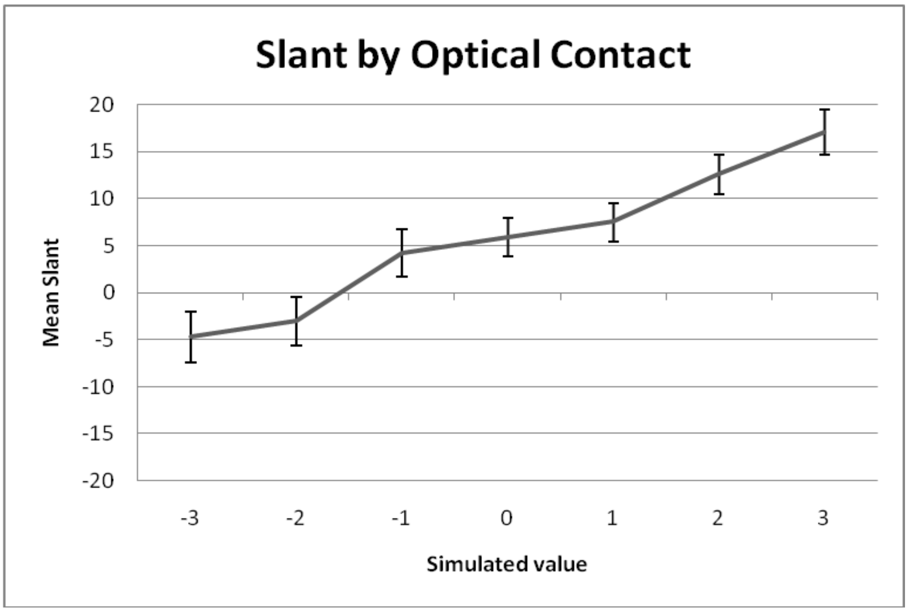


Figure 4.2. Plots of means of matched slant corresponding to levels of V. Matched slant is measured in degrees. Length of vertical bars corresponds to twice the standard error of corresponding means.

Figure 4.2 depicts means of the response variable “matched slant” corresponding to the two tilt levels used in the experiment. Unexpectedly, if the target is tilted on the right, it is more likely that it is seen as upward slanted. In the previous experiments, I did not find any significant difference in response between right and left tilt targets (excluding the interaction found with the hemifield of presentation). Indeed, an asymmetry in judgement between right and left tilted figures is not completely new in the psychological literature. For instance, Buermeister (1978) found an asymmetry in spatial perception when subject were requested to judge the lateral inclination of a left or a right tilted figure with respect to their body. However, at the present time, it is not possible to provide a full and sufficient explanation to the observed phenomenon of asymmetry. These results need further researches in order to be confirmed and explained.

No significant effects have been found for factor C (means of response variable *S* are respectively 5.89 for C = 0, and 5.40 for C = 1), neither I found any difference between data dispersion between the two levels (standard errors value are, respectively, 1.979 when C = 0 and 1.868 when C = 1).

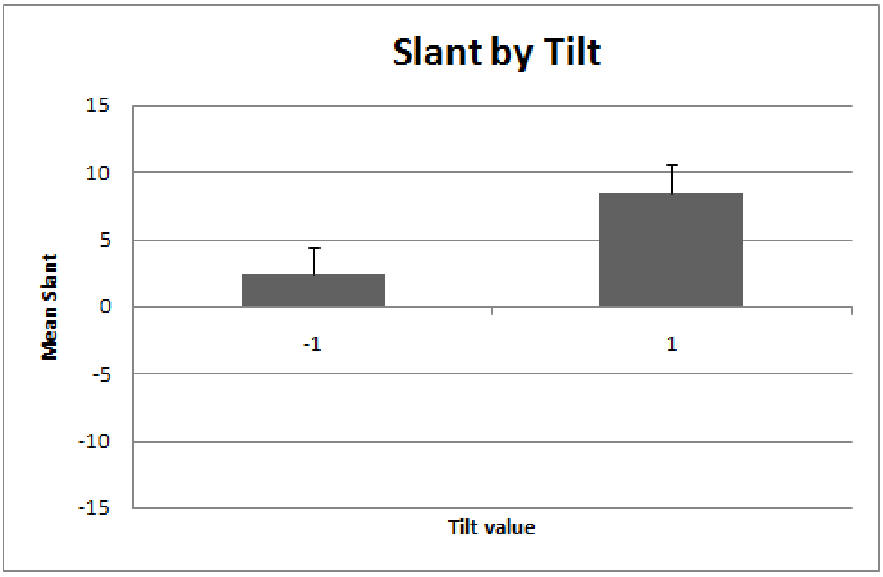
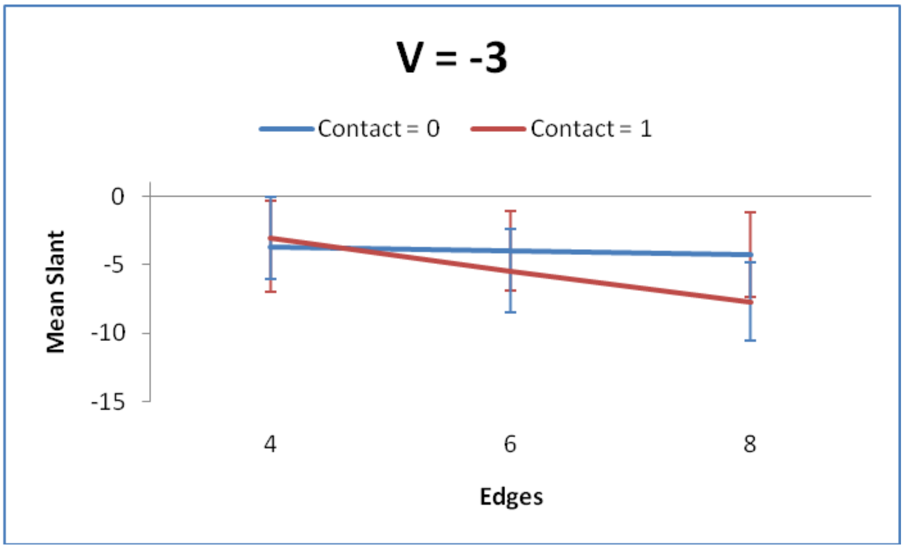
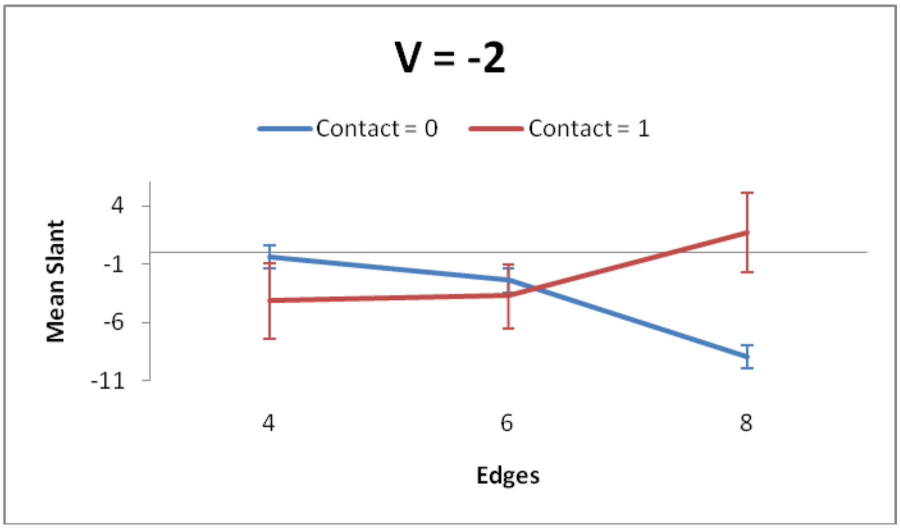


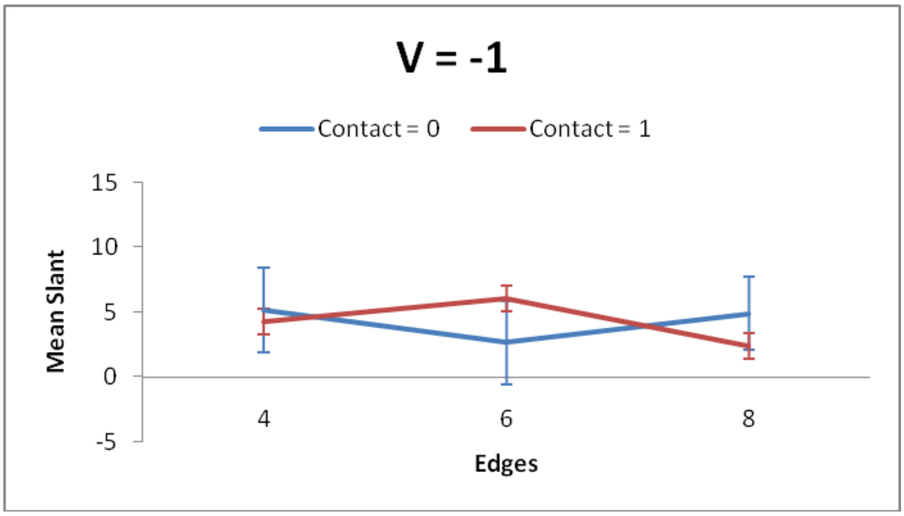
Figure 4.3. Plots of means of matched slant corresponding to levels of T (-1 corresponds to left tilt, 1 corresponds to right tilt). Matched slant is measured in degrees. Length of vertical bars corresponds to twice the standard error of corresponding means.

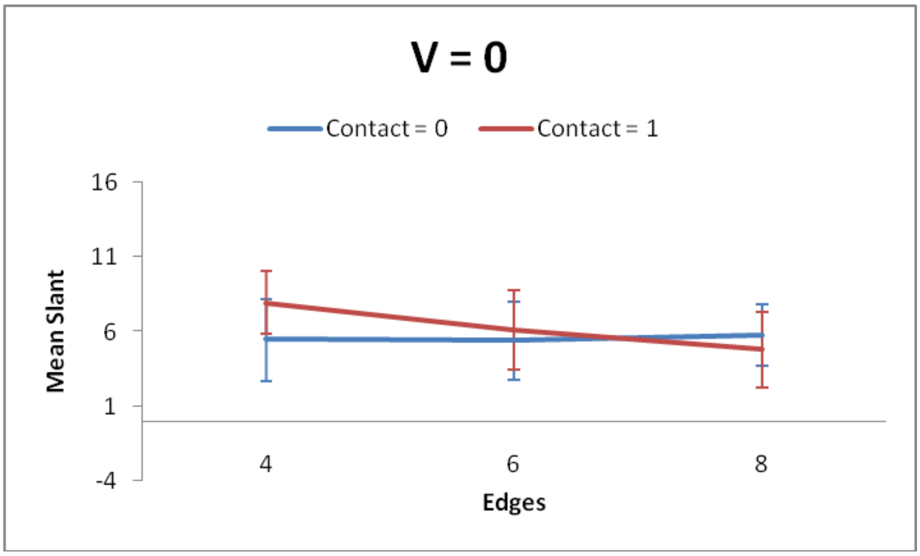
The significant interaction of factor C with other two factors (namely, E and V) suggests an effect of contact type (factor C) only emerging by interaction with other stimuli factors. Graphs shown in Figure 4.4 represent the means of the response variable "matched slant", corresponding to subsets of stimuli distinguished by the levels of factor V and condition C = 0 and C = 1 on factor E.

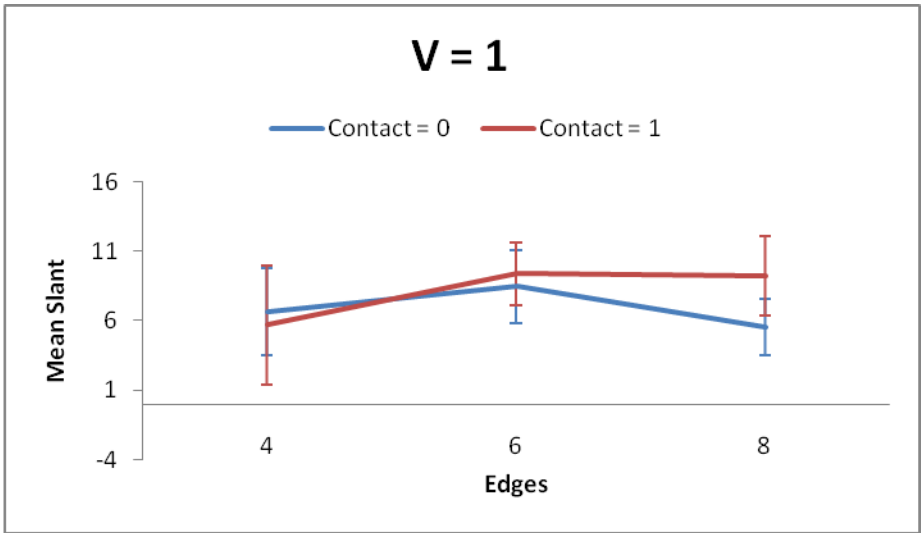
Unfortunately, results do not reveal a clear trend that could easily inform about the role of different types of contact. It is not possible, at the present time, to provide a convincing explanation of these interaction effects.

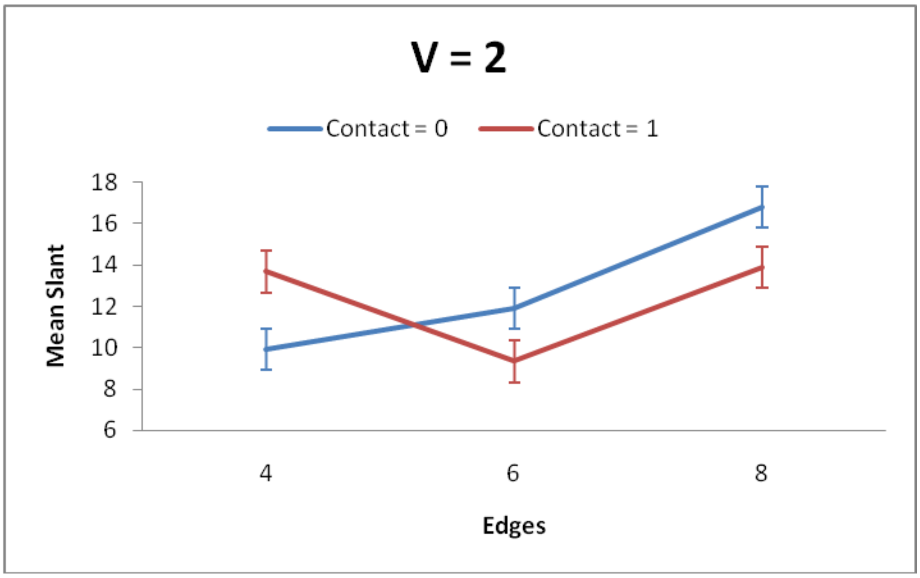

V = -3
Contact = 0
Contact = 1
Mean Slant
0
-5
-10
-15
4
6
8
Edges


V = -2
Contact = 0
Contact = 1
Mean Slant
4
-1
-6
-11
4
6
8
Edges


V = -1
Contact = 0
Contact = 1
Mean Slant
15
10
5
0
-5
4
6
8
Edges


V = 0
Contact = 0
Contact = 1
Mean Slant
16
11
6
1
-4
4
6
8
Edges


V = 1
Contact = 0
Contact = 1
Mean Slant
16
11
6
1
-4
4
6
8
Edges



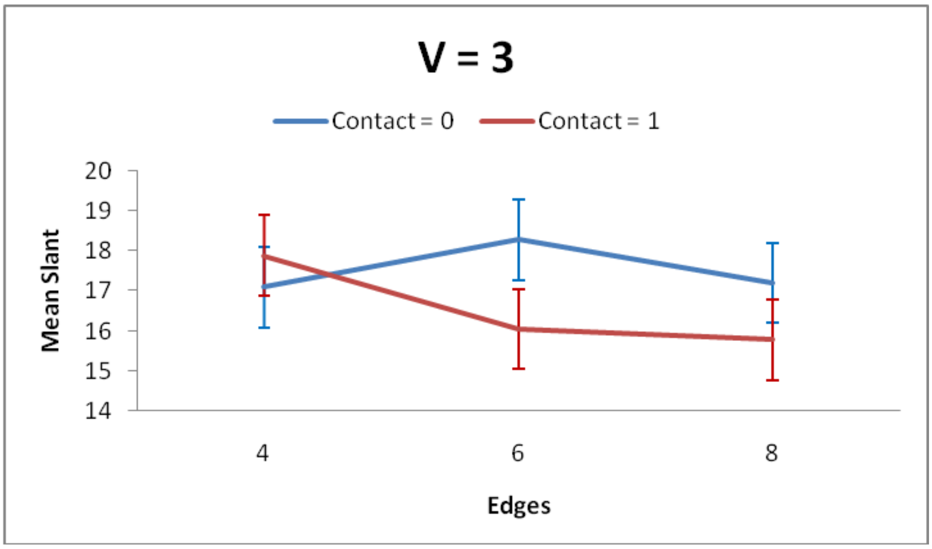


Figure 4.4. Plots of means of matched slant corresponding to combined levels of V, E and C. Matched slant is measured in degrees. Length of vertical bars is twice the standard error of corresponding means.

Comparing results from the experiments described in Chapter 3 with the current ones, it is possible to extend the role of the optical contact in inducing slant also for planar figures. This effect is strong when no other factor (e.g. slant), suggesting different perceptual solutions, is active on the same target. This suggested us to test how strong optical contact is in determining the slant of a figure if concurrent depth cues are active on the same target stimuli. In order to have an answer to this question, in the experiment described in the next pages I used the "depth cues conflict" paradigm to test the interaction between optical contact and two strong intrinsic factors, i.e., texture gradient and linear perspective.

4.3 Experiment 2

In the previous experiment, I deliberately used figures having not a regular shape in order to weaken the intervention of depth cues such as linear perspective, which refer to the geometrical shape of the structurally salient parts of the target. I made this choice since it was my intention to have a clear view on how the optical contact is able to induce slant in planar figures, without the intervention of other strong depth cues. This being proven, it is possible to introduce new stimulus factors in order to understand better the role of optical contact in complex scenes. In the new experiment described here, linear perspective factor has been introduced so that to involve the perceptual tendency to good form. This tendency states that perceptions of symmetry or rectangularity will occur more frequently than chance when the stimuli configuration allows this solution. Stimuli were built using the same frame used in the previous experiment, and a two-dimensional quadrilateral target figure. Starting from a rectangle with a simulated height of 400 and a simulated width of 200 cm, I created trapeziums which, at a given simulated distance, corresponded to how the rectangle would look like if it had the amount of slant *p* at a distance *d*. In particular, distance *d* was equal to the simulated distance of the target from the vantage point of the simulated visual scene (1670 cm), and the slant *p* corresponded to each one of the levels of the factor Linear Perspective used in the experiments (the effect has been traditionally called "contour convergence", Freeman, 1966)

Given the possibility to use a wide and regular surface as a target, I decided to test an additional factor, "texture gradient". I decided to use a "convergence gradient", as this type of gradient is more effective in specifying slant for surfaces close to frontal planar space (cf. Andersen, Sanders & Saidpour, 1998).

One of the advantages of the factors analysed in this experiment is that they have separate location in the pictorial stimulus: texture information is located within the target figure, linear perspective information (“contour convergence”) is located on the border of the figure, and optical contact information around the vertical extremities of the figure. As pointed out in the introductory section of this chapter, it is possible, through the analysis of gaze behaviour, to discover which parts forming the stimulus and in what proportion each are observed, and then on which kind of information observers pay attentions in the stimulus configuration I used. It is generally assumed that eye fixations are correlated with attention (Rayner, 1998), so that gaze analysis should enable us to understand how subjects distributed their attention when visually observing the scene.

4.3.1 Experimental setup

Participants. Thirteen students of the University of Padova participated in the study (mean age 23.31). All of them had normal or corrected-to-normal vision. They were naïve with respect to the purpose of the study.

Stimuli. 125 figures were presented as stimuli. Frame and vantage point had the same characteristics of the previous experiment.

Target figures were obtained by combining five levels of factor V, five levels of factor L, and five levels of factor X. Corresponding levels of factor V (Vertical unbalance of the target), factor P(linear Perspective) and factor X (teXture) were denoted by -2, -1, 0, 1, 2, 3 and were specified so as to produces values of -45°, -20°, 0°, +20°, +45°, as the simulated slant of the target for each one of the three factors (see figure 4.5 for an example of stimuli). The three levels of the response variable S are expressed as -1, 0 and 1, indicating downward slant, null slant or upward slant.

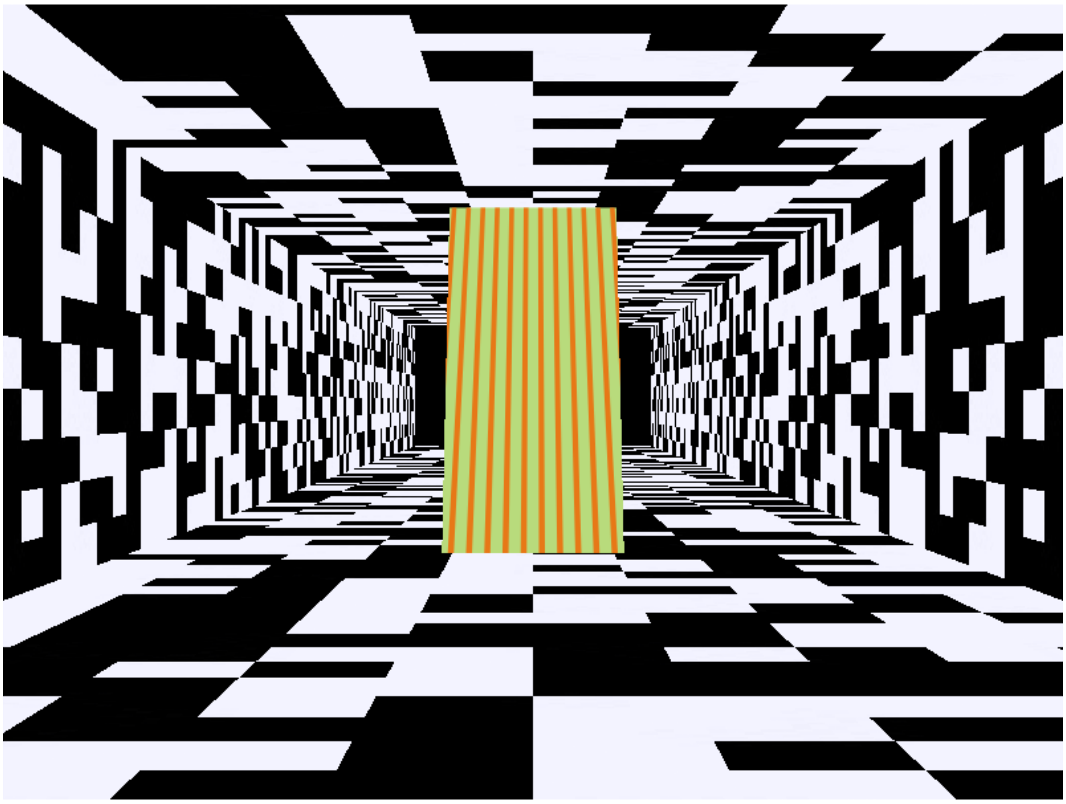


Figure 4.5. One of the stimuli used in the second experiment. Target texture gradient suggests a slant of -45°, target shape corresponds to a slant of – 20°. Optical contact as resulting from the relation between target and frame suggests a slant of 45°.

Apparatus. In order to record gaze data, I used a Tobii 1750 video-based eye-tracking system. Tobii 1750 integrates the camera and infrared lighting into a TFT 17'' monitor (1024x768 resolution). The system has an accuracy of 0.5°, a sampling frequency of 50 Hz and a reacquisition time inferior to 100 msec. It permits a relatively high freedom of movements as the camera has a recording field of 20x16x20 cm. ClearView 2.7.0 (Tobii®) has been used to record data on horizontal and vertical user's gaze screen coordinates. I considered for analysis only fixations that showed more gaze-points within an area of 1.58° visual angle, for a period of at least 200ms. Other data-points were discarded.

The centroid of the screen coordinates of gaze-points belonging to a fixation was considered as the position of the related fixation. Eye position was calibrated at the beginning of the experiment by asking the participants to follow with the eyes a moving dot on the screen. The distance from the eye to the monitor was 57 cm. Vision was binocular.

Procedure. For each participant, the experiment began with a practice phase, aimed at illustrating the kind of stimuli to be observed and how to express responses. After the training phase, all 125 stimuli were separately presented in a randomized order for each participant. The sequence of 125 trials was split into three blocks of 62, 62 and 61 trials each. The two blocks were separated by a pause, the duration of which was decided by participant. Each trial had the same structure: a white fixation cross appeared and remained on the screen for 0.5 seconds. After the fixation cross disappeared, one of the stimuli was shown for 5 seconds. Then, the screen became blank. The subjects' task was to judge whether the target was slanted upward, downward or had a null slant (i.e., it appeared to be vertical), and to express their judgment by pressing one of three keys of a computer keyboard (respectively,"r", "v", and "g"[1]). The experimental session, including both practice and main phase, took from 6 to 10 minutes to be completed.

4.3.2 Results

The experiment comprised three factors (V, L and X), a categorical response variable S, and repeated measures. Data have been analysed by fitting and interpreting a log-linear model (Agresti, 2002), as for the first experiment of Chapter 3.

Data were organised to form a 4-dimensional frequency table (m_{vlhs}), where frequency m_{vlhs} was the number of participants who gave responses S=s when presented with the stimulus in which $V = v$, $P = p$, $X = x$, for $v \in \{-2, -1, 0, 1, 2\}$, $p \in \{-2, -1, 0, 1, 2\}$, $x \in \{-2, -1, 0, 1, 2\}$ and $s \in \{-1, 0, 1\}$. Relative to this frequency table, I searched for a log-linear model which was optimal in simplicity and goodness-of-fit. The search was carried out by fitting a hierarchical log-linear model,

---

[1] This association between responses and keys on the keyboard was chosen after suggestions received by some participants in a pilot experiment, which described the choice of keys in this specific order to be "natural".

and then omitting those components which did not entail any substantial increase in residual deviance. The log-linear model resulting from this search is given by the following equation:

$$\log (m_{vlhs}) = \mu + \lambda_{ps}^{PS} + \lambda_{xs}^{XS} + \lambda_{pxs}^{PXS}$$

It expresses the logarithm of the observed frequencies as the sum of intercept μ (this component has no useful meaning in the present research); component λ_{ps}^{PS}, specifying the action of factor P on response S; component λ_{xs}^{XS}, specifying the action of factor X on response S; component λ_{pxs}^{PXS}, specifying the interaction of component P and X on response S. Model residual deviance is $G^2 = 230.9$, with df = 300, and $p = .99$ (relative to the distribution χ^2_{300}). The low value of residual deviance (when compared with the “null deviance”, which is $G^2 = 1816$) and the corresponding value of p are proof of a very good fit of the model to the data. Table 4.1 displays the maximum-likelihood estimates of the three classes of parameters forming the model (the estimate of the intercept is μ = -1.73).

(i)

λ^{PS}	P=-2	P=-1	P=0	P=1	P=2
S = -1	-34.81	14.65	15.44	-20.77	25.49
S = 0	-1.44	2.14	2.71	-0.44	-2.97
S = 1	36.25	-16.79	-18.15	21.21	-22.52

(ii)

λ^{XS}	X=-2	X=-1	X=0	X=1	X=2
S = -1	-51.37	0.89	14.20	17.94	18.35
S = 0	-0.88	0.69	1.61	-0.51	-0.92
S = 1	52.25	-1.58	-15.81	-17.43	-17.43

(iii)

Λ^{PXS}	X=-2	X=-1	X=0	X=1	X=2
			P=-2		
S = -1	-261,3	93,66	55,16	49,68	62,82
S = 0	11,85	-15,18	4,55	-1,54	0,32
S = 1	249,47	-78,48	-59,71	-48,14	-63,14
			P=-1		
S = -1	59,48	-14,4	-13,54	-16,23	-15,3
S = 0	3,06	-0,94	-1,76	0,06	-0,43
S = 1	-62,54	15,34	15,30	16,17	15,73
			P=0		
S = -1	70,38	-48,06	-12,5	2,58	-12,4
S = 0	-13,84	16,37	-0,88	-0,39	-1,26
S = 1	-56,54	31,69	13,38	-2,19	13,66
			P=1		
S = -1	62,8	-14,51	-14,12	-17,65	-16,52
S = 0	1,01	-0,50	-1,34	0,65	0,18
S = 1	-63,81	15,01	15,46	17,00	16,34
			P=2		
S = -1	68,66	-16,69	-15	-18,38	-18,6
S = 0	-2,08	0,25	-0,57	1,22	1,19
S = 1	-66,58	16,44	15,57	17,16	17,41

Table 4.1. Maximum-likelihood estimates of parameters in log-linear model accepted for data of Experiment 2.

Separate components will be commented in the next paragraphs following the same method used in the description of single components in Chapter 3.

Component λ_{ps}^{PS} in the model (specific deviance G^2 = 1447.70, df = 12, p <.001) means that there is an increasing monotone association between response variable S and stimulus variable *P* when ranges are ordered in the natural way. Figure 4.6 illustrates how, when variable P increases (resp., decreases), the chance of response variable S=1 increases (resp., decreases), and the chance of response S = -1 decreases (resp., increases). When variable P = 0, the chance of response S = 0 is maximum; the more it increases or decreases, the more the chance of response S = 0 decreases. This sequence clearly shows that the linear perspective factor strongly influences the apparent slant of the target.

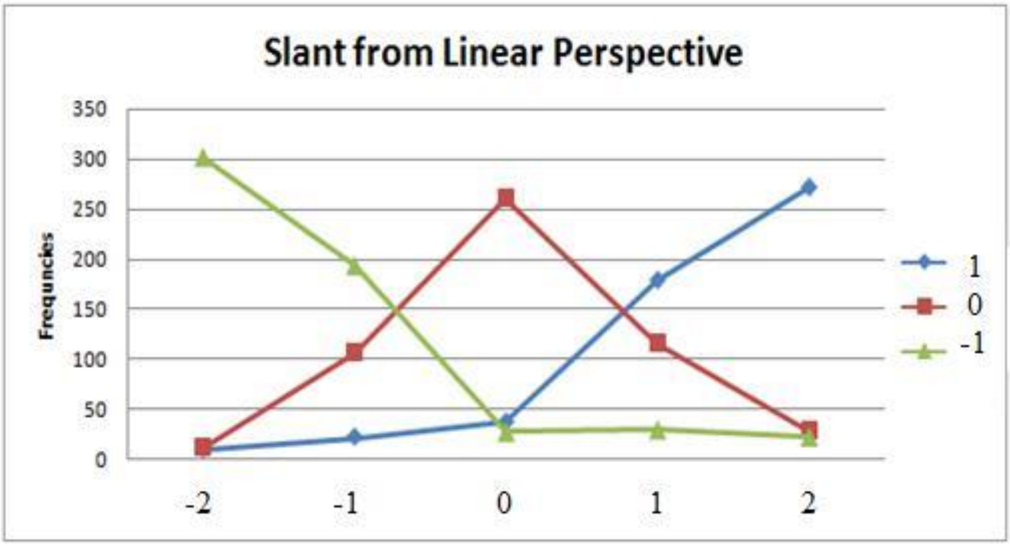


Figure 4.6. Frequencies of responses S = -1, S= 0 and S = 1 to stimuli with different values of property P (linear Perspective).

Component λ_{xs}^{XS} in the model (specific deviance $G^2 = 40.72$, df = 12, $p < .001$) means that there is an association between response variable S and stimulus variable X (texture). The general trend of the frequency of responses resembles the one already described for linear perspective (as illustrated in Figure 4.7), revealing an increasing monotone association when X range {-2, -1, 0, 1, 1} and S range {-1, 0, 1} are ordered in the natural way.

Comparing Figure 4.6 and Figure 4.7, it clearly results that the general effect of texture factor is weaker than linear perspective factor.

Component λ_{pxs}^{PXS} in the model (specific deviance $G^2 = 57.66$, df = 32, $p < .001$) means that there is an association between the interaction of stimulus variable P and X and the response variable S. Frequencies for each level of S are shown in Figure 4.8. The distribution of responses illustrates the nature of the interaction effect: when P = -2 and X = -2, the response S=-1 is prevailing over the others. When

P = -1 and X = -1 or X = -2, the chance of response S = -1 is higher than the case in which P = -1 and X = 1. When P = 0, response S = 0 is preferred in case X = 0, and it decrease monotonically. This configuration of responses may be interpreted as a proof that the two factors interaction may be described in terms of *accumulation* (Bülthoff & Mallot, 1990): if the predicted effects of both factors are congruent, then the resulting response is higher in frequency.

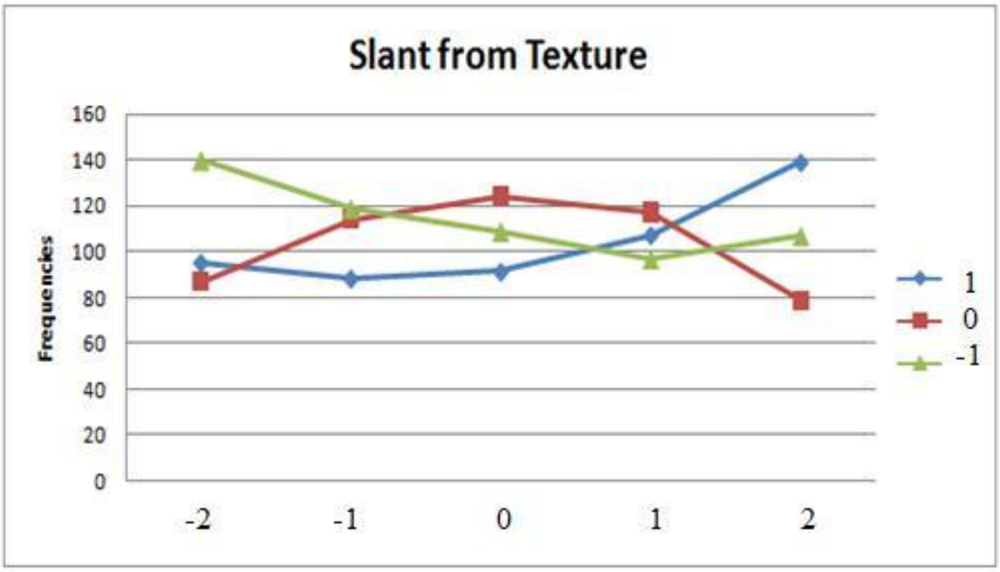


Figure 4.7. Frequencies of responses S = -1, S= 0 and S = 1 to stimuli with different values of property X (teXture).

There are some aspects deserving specific attention, regarding general differences found between these results and those from previous experiments (both those described in Chapter 3 and the first one of the current chapter). The component λ_s^S has not been accepted in the model, as it happened in the previous experiments. The bias in preferring response “upward slant” is not confirmed with the current set of stimuli. This aspect can be explained by the different sources of information used in this study: the general tendency in preferring the response “upward slant” is not present when strong intrinsic depth cues are able to fully determine slant values.

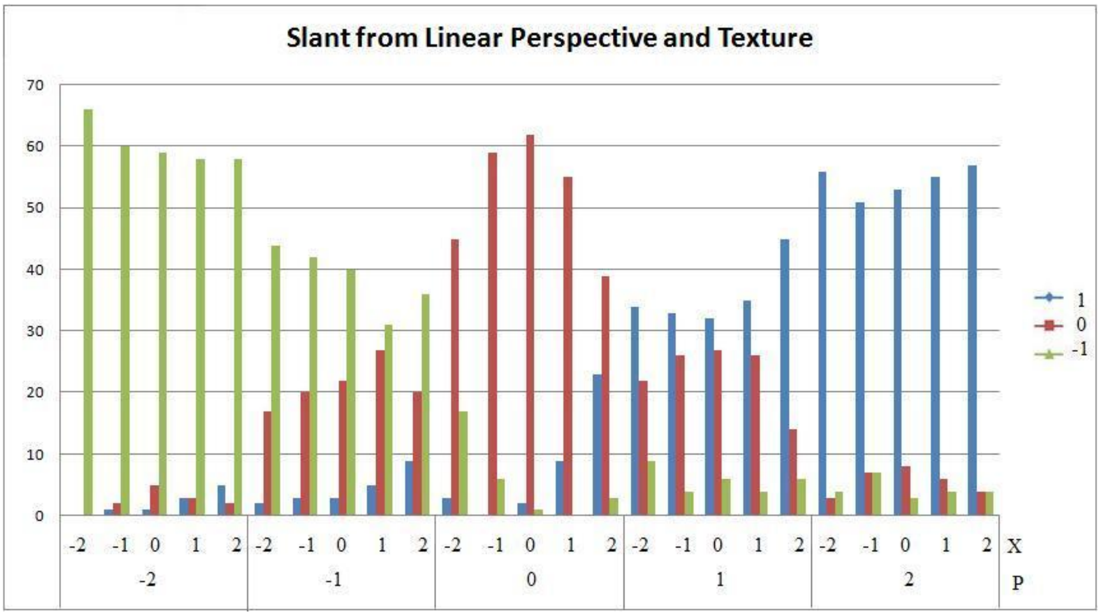


Figure 4.8. Frequencies of responses S = -1, S= 0 and S = 1 to stimuli with different values of property X (teXture) and property P (linear Perspective).

Another aspect deserving some attention is the absence of the component λ_{vs}^{VS}, which would specify the action of factor V (vertical unbalance) on response S – deviance is G^2= 6.27 (df=12, p =.90). This result show that, in presence of strong intrinsic factors, the information resulting from optical contact has no effect on the apparent slant: intrinsic factors, in this case, seem to exercise a “veto” effect (Bülthoff & Mallot, 1990).

Responses frequency analysis was followed by the analysis of the gaze behaviour of our participants during the task. The aim of this analysis, as noted above, was to understand on which kind of information observers rely on when looking at a simple stimulus configuration like the ones proposed in the current experiment. Gaze analysis was accomplished in two separate steps:

1. A part of the scene (the one surrounding the target) was organized in different Areas of Interests (AOI). AOIs are conceived as meaningful subsections of the stimuli with well-defined border. I selected three different areas: “top edge”, “bottom edge” and “body”. The first two areas were delineated by drawing a

rectangle of dimensions 22cm×7cm around the zone where the optical contact was realised. The third area, body, corresponded to the remaining part of the textured surface to a larger trapezium (see Figure 4.9). No AOI was drawn for other parts of the stimuli, since the number of fixations toward them were very scarce.

2. I analyzed the difference among the three areas in terms of number of fixations and fixation duration. Since I was interested also in difference between lower and upper edges, given the asymmetry that some studies reported between contact with the ground and contact with the ceiling (cf. Bian, Braunstein & Andersen, 2002), I did not pool these data.

I ran two repeated-measures one-way ANOVA in order to detect differences between the conditions for the two dependent variables, number and duration, related to gaze. The first ANOVA was run considering the number of fixations as the dependent variable, and AOI type (upper edge, lower edge, body) as the factor. The test yielded a significant result for the effects of AOI type [$F_{(2,22)}$= 72.794, $p < .001$]. The mean number of fixations for each AOI is shown in Figure 4.10.

Comparing 95% confidence interval around the estimated means for each level, the number of fixations resulted to be larger for the level AOI = 3. There is no difference between AOI = 1 and AOI = 2. Given these results, it is possible to conclude that most of the fixations fell inside the target. One potential drawback of the use of fixations number as index of subject attention is that this index is highly dependent on the size of the AOI. As we may see from Figure 4.9, the area of the AOI body is larger than the other two.

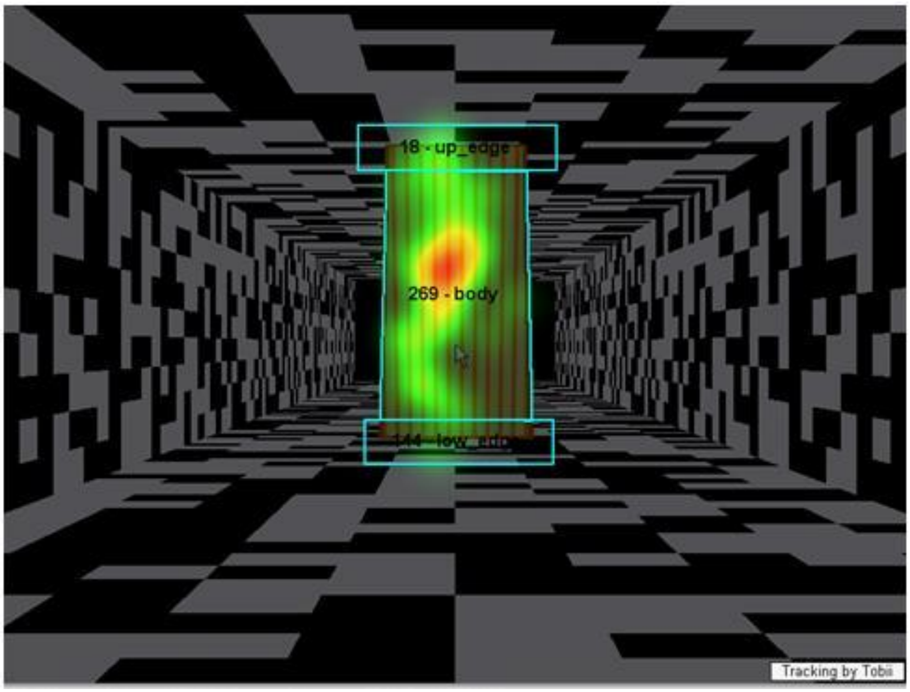


Figure 4.9. An example of the organization into Areas Of Interest of one stimulus. The red, yellow and green zones represent the number of fixation in a particular point for every subject: red color is assigned to the zones with the highest number of fixations, yellow color is assigned to the zones with an intermediate number of fixations, green color is assigned to the zones with a low number of fixation. Non-colored zone have not been fixated by any subject.

A further analysis regarded fixation time on each AOI. Nodine, Carmody & Kundel (1978) found that fixation time is often longer for areas which have higher informativeness. Furthermore, this measure has the advantage of being independent of the size of the AOI.

I ran a repeated-measure ANOVA, considering the duration of fixation as the the dependent variable, and AOI type (top edge, bottom edge, body) as the factor. The test yielded a significant result for the effects of AOI type [$F_{(2,22)} = 6.545$, $p = .006$]. The mean time of fixation for each AOI is shown in figure 4.11. Comparison of 95% confidence interval around estimated means shows a significant difference between the mean duration of fixation only between those in the AOI bottom edge and those in the AOI body.

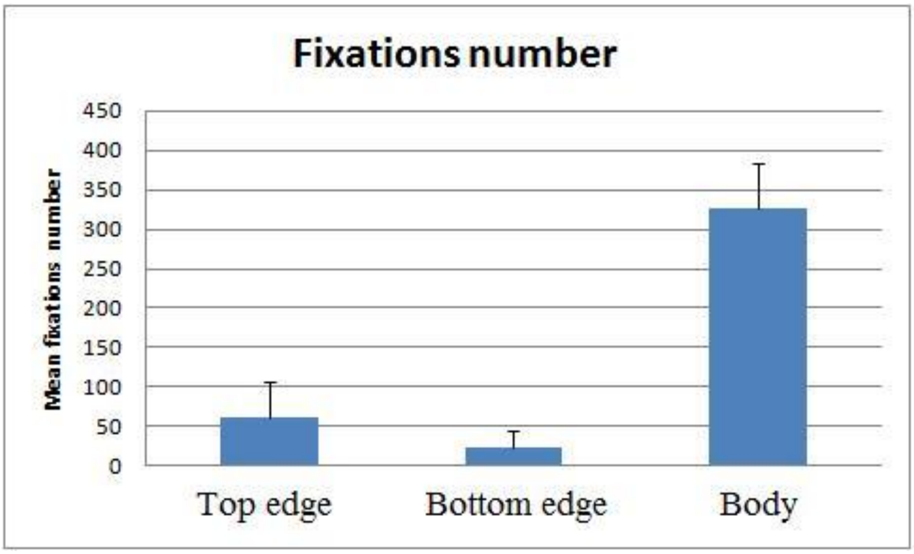


Figure 4.10. Plots of means fixations number to levels of AOI upper edge, lower edge, body). Length of vertical bars corresponds to the confidence interval at 95% confidence level.

Analysis reveals that:

a. subjects pay attention to the information conveyed by the two vertical extremities of the target, in which the optical contact is realized;

b. there is an asymmetry between top and bottom sides in their difference with the central section of the target. This result is quite interesting, since the reports from the “optical contact” literature mostly deals with ground contact, and a common result (cf. Bian, Braunstein & Andersen, 2005) is that contact with the ground is generally preferred as a source of information over the contact with the ceiling.

Of course, it is necessary to look at these results with caution. The upper and lower side of the target also convey information about linear perspective (when the target is seen as a regular shape, then if it is slanted downward, the upper side is larger than the lower side, and if the target is slanted upward, then the lower side is larger than the upper side). Another caution to take when considering this last result is that, even if subjects took into account information inherent in the upper and lower side, the response in the task showed that they did not use this information.

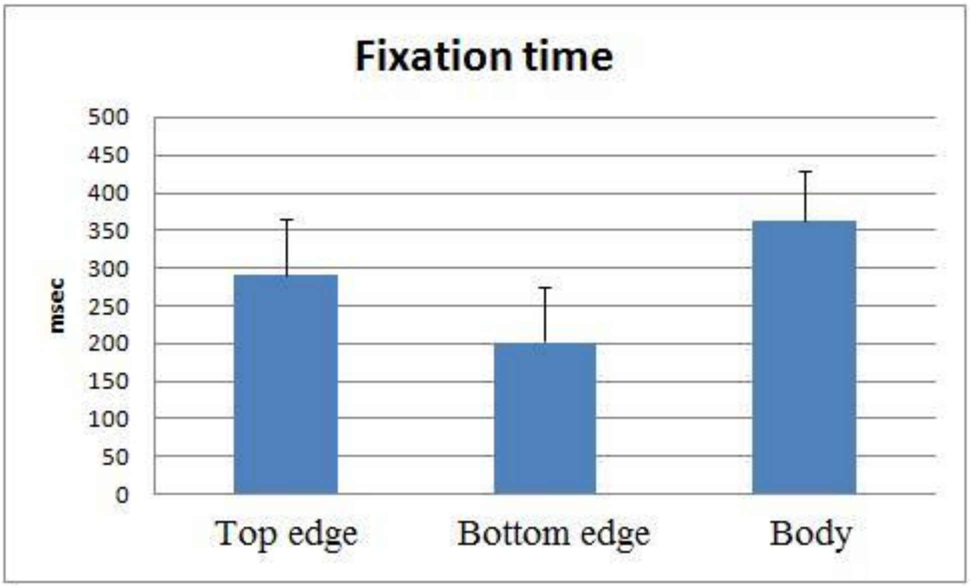


Figure 4.11. Plots of mean duration (expressed in milliseconds) of fixations for each fixation for the three levels of AOI upper edge, lower edge, body. Length of vertical bars corresponds to the confidence interval at 95% level of significance.

4.3.4 General discussion

The two experiments described in this section aimed at a better specification of the effect of the optical slant and vertical discrepancy among vertical extremities in determining apparent slant of two-dimensional figures. Results show that optical contact is critical in inducing perceptual slant if no other strong intrinsic factor is present on the target part of the stimulus. When this is not the case, and other powerful cues to depth are acting in the scene, then the role of optical contact strongly diminishes. The particular trend of the interaction effects found in the first experiment between shape, tilt direction and vertical discrepancy does not permit to draw any general conclusion about the way they interact. Further researches on this topic should consider these problems.

Chapter 5

Interaction between Foveal and Peripheral Information in Slant from Texture

5.1 Introduction

Textures are a fundamental source of information for determining depth-related characteristics of visual surfaces. There is ample evidence, starting with the studies of James Gibson (in which subjects were asked to perform a slant perception task), that the facility in discriminating differently slanted surfaces depends on specific properties of the textures on it. Gibson (1950a) found that it is easier to distinguish the slant of surfaces if the textures depicted on them are regular rather than irregular, i.e., regular textures are more informative than irregular textures.

Textures are generally composed of discrete local elements, usually referred to as *texels*[1]. The properties of texel-based textures conveying information about depth are called “texture gradients”, and represent the amount of variation of certain parameter as the surface recedes into depth[2]. Gradients can be divided in different modalities: one of the most frequently used categorization is the one distinguishing between *density gradient*, *scaling gradient* and *compression gradient* (Purdy, 1958; Gillam, 1968; Cutting & Millard, 1984; Blake, Bülthoff, & Sheinberg, 1993; Andersen, Braunstein and Saidpour, 1998).

Texel based representations of textures are not able to account for most of the surfaces composing our visual world: generally, natural textures are not composed by well-defined texels. A different approach, suitable to describe natural textures, consists in describing textures using spatial frequency-based analysis (Bajcsi & Lieberman, 1976; Malik & Rosenholtz, 1997; Sakai & Finkel, 1997; Rosas,

---

[1] Todd & Akerstrom (1987) defines “texels” as “the elements of surface/optical texture can [which] be thought of as bounded regions of one reflectance/luminance surrounded by a background of some other reflectance/luminance […] typically in one-to-one correspondence” (p. 287).

[2] Cutting & Millard (1984) defines “gradients “as “the sources of information that grade, or change, with visual angle as one looks from one’s feet upward to the horizon” (p. 198).

Wichmann & Wagemans, 2004; Velisavljević & Elder, 2006). The idea that spatial vision is accomplished through a spectral decomposition of the visual stimuli can be dated back to Campbell & Robson (1968), who found that mechanisms for determining the contrast of gratings could be understood in terms of decomposition into Fourier components of their waveform. An open question is whether these mechanisms can be useful to understand high-level tasks such as depth perception, as it has already been found for low-level vision tasks.

Sakai & Finkel (1995, 1997) proposed a model of spatial frequency decomposition based on the idea of *average peak frequency* (*APF*). According to their model, spatial frequency analysis is computed for each orientation of the image by means of differently oriented spatial filters. When the spectrum contains high peaks, the visual system is presumed to track single peaks frequencies. On the contrary, when strong and isolated peaks are missing, the visual system is presumed to track the mean frequency of peaks. Sakai and Finkel proposed that the larger the variation of peaks in the first case, or APF in the second case, the stronger the perception of slant. Different textures provide a difference in peaks frequency between the top and the bottom of images: the larger this difference, the better the observers' performance. Using different types of textures, characterized by definite spatial frequency spectra, Rosas, Wichmann & Wagemans (2004) produced a rank ordering of them along the dimension of facility in discriminating between different slants. This ordering partially reflects the informativeness of textures based on the APF model of Sakai & Finkel (1997), although results were not completely consistent for every subject they tested. Rosas et al. concluded that "spatial frequency characterization is part of the underlying mechanism yielding the observed rank-order", but "other mechanism are likely to be involved in this task" (Rosas, et al., 2004, p. 1533). In subsequent studies using eye tracking systems, Rosas (personal communication) tested if, during a slant discrimination task, human gaze tracks the location with the largest difference in peak frequency. He did not find consistent or recurrent gaze patterns for types of textures. Furthermore, he found that subjects had the tendency to fixate few locations, that fixations are concentrated in small portions of the image, often coinciding with areas with large changes in the APF. He confirmed that the strategy used by human observers cannot be fully explained by the APF model. The different performances of subjects may be explained by the different attentional strategies used by them: some subjects attended certain parts of the spectrum for longer periods than other subjects did.

Indeed, eye movement tracking can provide insightful suggestions on how humans observe and which kind of information they use when judging depth from texture. My current study uses this instrument, but it is not focused on tracking the specific location observed in order to estimate slant. The aim of my study is to understand if and how the information presented in different regions of the visual field interacts and jointly determines slant estimation. Several eye movement studies showed that the visual field can be divided into three main regions: foveal, parafoveal and peripheral (Rayner, 1998). Fovea extends out to 2° around fixation point, parafovea extends out to 5° around fixation point, and periphery contains the remaining portion of the visual field. This distinction has been introduced to take into account acuity limitation of the visual system: during visual exploration of a scene, we move our eyes so that the part of the stimuli we want to see clearly is inside the foveal region, which is the one with the best resolution. The study presented in this chapter does not directly address the problem of the acuity limitation: it has been designed to analyze how information presented within vs. without the fovea distinctly contributes to the perception of depth of textured surfaces.

In order to evaluate the role of peripheral and foveal information in Slant-from-Texture, I used the Gaze-Continent Window (GCW) Technique. In experiments using this technique, the stimulus display is continuously updated according to the subject’s current gaze position. Usually, a rectangular, circular or elliptical window is centred at the participant’s gaze position. As the participant moves his/her gaze, the window is positioned wherever the participant looks. In the most general form of GCW experiments, the stimulus information within the window is visible, while the stimulus outside the window is generally masked or blurred. At first, this technique had been proposed in reading studies (McKonkie & Rayner, 1975), but it has been immediately extended to further fields of research, like scene perception, face recognition, ergonomics or visual search studies (Saida & Ikeda, 1979; Rayner, 1998; Bertera & Rayner, 2000; Pomplun, Reingold & Shen, 2001). As far as I know, there are no previous studies that specifically addressed textured surface perception with these instruments, although some studies directly addressed the role of the field of view size or position.

Specifically, Blake, Bülthoff & Sheinberg (1993) systematically measured how texture informativeness changed as a function of the field of view size or location (respectively, at the bottom, at the middle or at the top of the image), and found that change in the field of view position affects performance. Knill (1998) systematically observed the effect of horizontal or vertical variations of the field of view size: he asserted that “changes in discrimination performance with changes in horizontal field of view size should primarily reflect limits on the range of spatial integration of texture information (as opposed to strategic focusing of attention on selected stimulus regions)” (p. 1692), while changes in horizontal field of view size should affect texture cue reliability[3]. In the experiments described below, the field of view shape remained fixed, but the information in it changed according to the gaze position, as my primary interest was to analyze how spatial integration of texture information occurs.

5.2 Experiment 1

The first experiment was aimed at obtaining a performance baseline for each subject, to be compared with data of the second experiment, in which discrepant information among different regions of the images was introduced. This experiment consisted in a repetition of experiment 1 of Rosas, Wichmann & Wagemans (2004), with a restricted set of stimuli.

---

[3] According to Knill (1998), “three independent factors determine the effects of vertical field of view size on texture cue reliability: the relative density of texture elements in different parts of an image, the extent spatial gradients contained within the image […], and differences in the relative contributions of individual texels to texture cue informativeness as a function of the position in the image” (pp. 1695-1696).

5.2.1 Experimental setup

Subjects. Five subjects (4 Ph.D. students and one visiting researcher at the University of Leuven) participated in the experiment; each subject was completely naïve as to the purpose of the experiment and all of them had normal or corrected-to-normal vision.

Apparatus. Stimuli were displayed on an Iiyama HM204DT Vision Master Pro 514 monitor. Screen resolution was set at 800x600 pixels. Head movements were constrained as much as possible using a head and chin rest, positioned at 60 cm from the screen. The monitor was completely covered using a black cardboard except for a circular aperture of 23 cm diameter, through which subjects were able to see the screen. The aperture subtended about 21.7° of visual angle at the subjects' eyes; its centre was at the centre of the screen. This setting was designed in order to reduce cue conflict due to the physical flatness of the screen.

Subjects' eye movements were recorded using SR Research Eyelink II ® oculography machine, at a sampling rate of 500 Hz. Before any recording session, a nine-point calibration of eye position was run. Calibration was typically accomplished in less than one minute, and gaze position error was below or equal to 1 degrees of visual angle. The temporal resolution of the system was 2ms.

Stimuli. The textures used were a subset of the ones used in Rosas, Wichmann & Wagemans (2004); namely Leopard (also called "Diffusion Map"), 1/f Noise (also called "Natural Noise" or "Pink Noise") and Perlin Noise (also called "Coherent Noise", cf. Perlin, 1985)[4]. For some example of these textures, see figure 5.1a, 5.1b, 5.1c.

---

[4] Leopard or Diffusion Map textures were generated using an algorithm proposed by Turk (1991), which generate leopard-skin-like textures; 1/F Noise textures were generated filling a plane with white noise, then "coloured" in the Fourier domain by scaling its amplitude spectra with a 1/f shaped cone and finally by computing the inverse Fourier Transform; Perlin Noise textures were generated using an algorithm proposed by Perlin (1985), with a proper setting of parameters so to realise cloudy-looking textures. Textures were mapped onto slanted planes using an algorithm proposed by Heckbert (1989). The procedures used to produce textures are fully explained in Rosas et al., 2004.

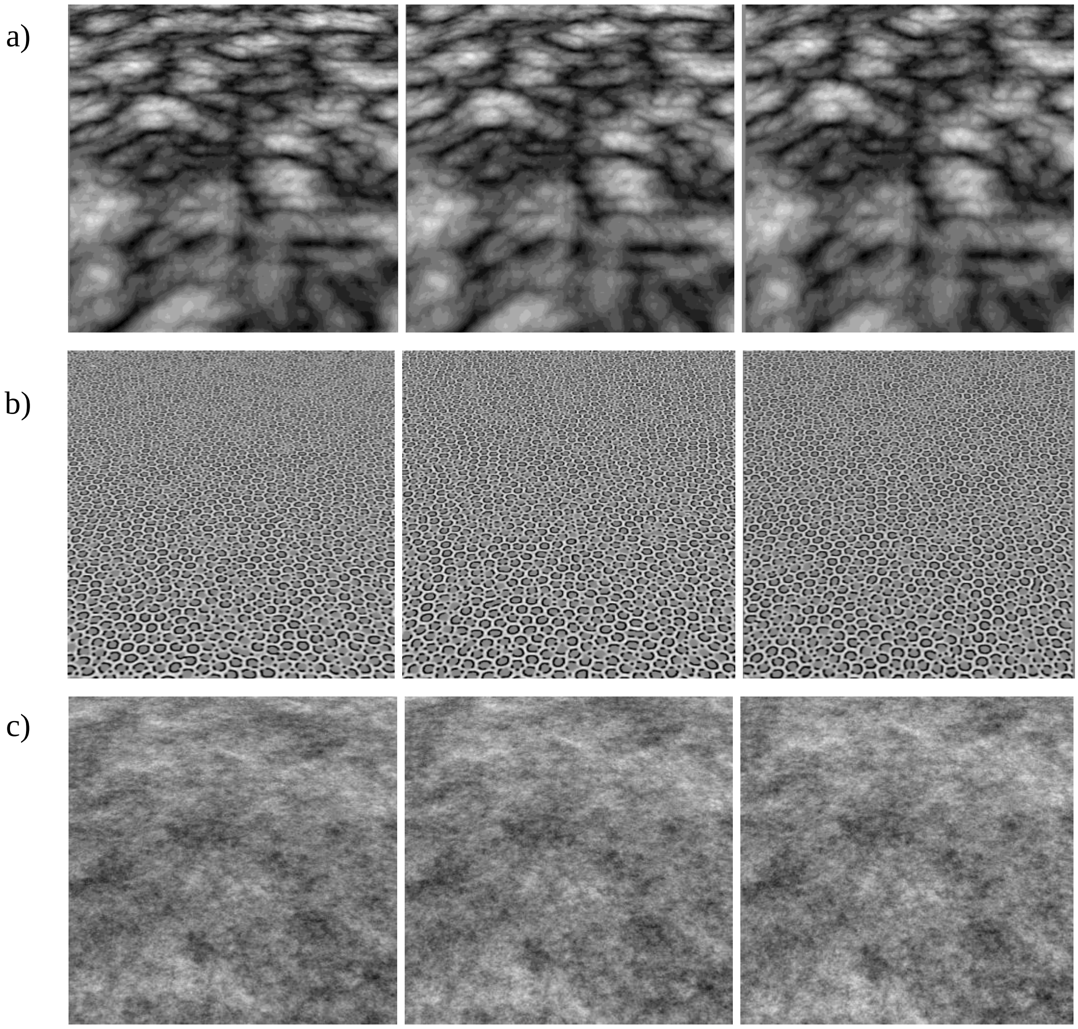


Figure 5.1. Examples of the different textures used in the two experiments. The textures of the surfaces in the first row correspond to Perlin Noise pattern, the textures in the second row correspond to Leopard pattern, and the textures in the third row correspond to 1/F Noise pattern. The slant depicted in the first column corresponds to 33°, the slant depicted in the second column to 37° (standard), and the slant depicted in the third column to 41°.

Procedure. A standard temporal Two-Alternative-Forced-Choice (2AFC) procedure was used: participants had to choose which of two textured planes rendered at physically different levels of slant and in temporal succession appeared steeper. Subjects were instructed that the more a surface seemed to be near the horizontal plane, the less steep it had to be considered. Each texture was tested separately: this means that in every single block of trials, every image to be judged had always the same type of texture. The order of presentation of each texture was randomised across subjects. For each texture, seven different instances were created. For each texture and each slant level, one of these seven instance were displayed randomly, to avoid possible effects of learning specific features of the images (unrelated to slant) in doing the task.

The experiment was preceded by a training phase 50 trials long to familiarise the subjects to the task; during the training phase, another type of texture (“random circles”) was used (Figure 5.2). Prior to every block of 50 trials, a nine-point calibration of eye position was performed again: calibration was accepted only if a subsequent validation procedure showed an average tracking error smaller than 1°.

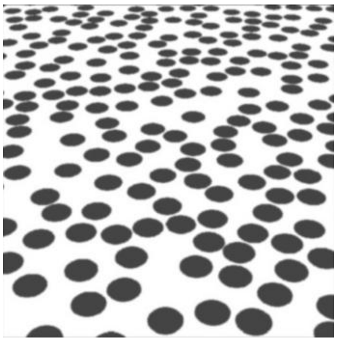


Figure 5.2. An example of the stimuli used during the training phase.

At the beginning of each trial, a “drift correction” was performed using a single white circle with a black dot in its centre, which appeared centred on the screen[5]. Subjects were requested to fixate the black dot and press the space bar: if their eyes were properly fixating the dot according to the eye tracker, the trial began. The screen became blank for 250 msec, and the first image was displayed for 800 msec; after that, a blank screen appeared for 250 msec and the second image was displayed for 800 msec[6]. After the second image disappeared, the screen again became blank and a short tone marked the beginning of the response phase. Subjects had 1500

---

[5] This procedure is necessary to prevent errors due to small displacement of the cameras as subjects moved their head.

[6] The blank screens were all set to a mean luminance.

msec to provide an answer using the keyboard, although they were requested to respond as fast as possible. Standard stimuli value was fixed at the value of 37° degrees of slant (slant is measured from the horizontal). Data were collected using a constant stimuli procedure in two sessions: an initial estimation of the psychometric function was obtained using test values suggested by results from Rosas et al. (2004). Six different comparisons between standard and different levels of test stimuli were tested for 50 trials each, for a total of 300 trials. Each session lasted about 20-25 minutes; subjects were allowed to take a break every 50 trials. The order of comparison of standard and test stimuli was completely randomised.

For every subject I estimated a psychometric function around 37° degrees of slant (slant is measured from the horizontal). Data were fitted using the Psignifit Toolbox, which implements the constrained maximum-likelihood method proposed by Wichmann & Hill (2001a, 2001b). Fits were done using a logistic function as underlying shape.

Using this preliminary estimate, some critical values were selected in order to obtain a more reliable estimation of the psychometric function, by pooling these data with those previously obtained. Each psychometric function was obtained using data recorded in a number of trials varying from 400 to 600 in number.

5.2.2 Results

A different psychometric function was obtained for each texture and for each subject. Each of the functions represents the probability of the functions that a target surface of a variable simulated slant could be judged as more slanted then a standard surface of fixed simulated slant (37°). Psychometric functions are depicted in Table 5.1. Confidence intervals were calculated using the parametric bootstrap procedure proposed by Wichmann & Hill (2001b). The 1/F Noise psychometric function for one of the participants (VB) is not displayed, since his answers were completely random and obtained from an insufficient number of trials.

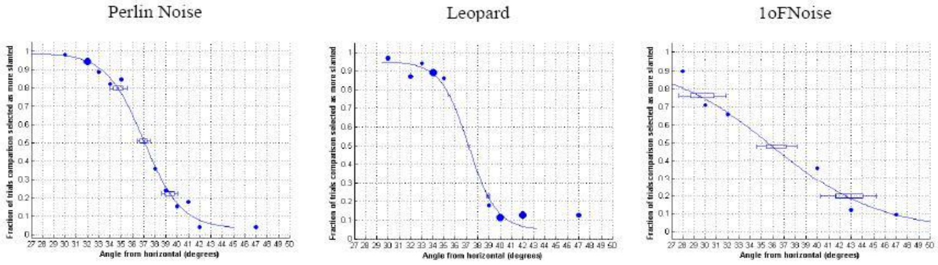
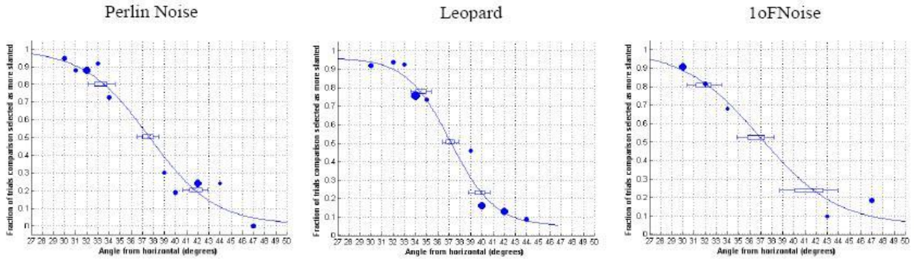
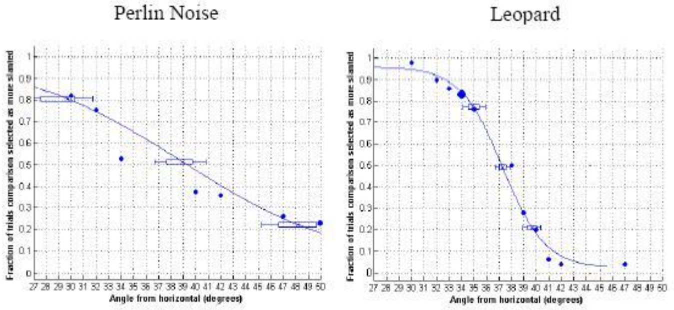
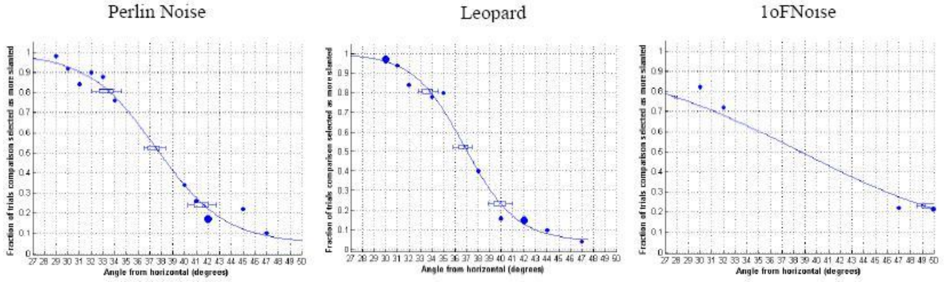
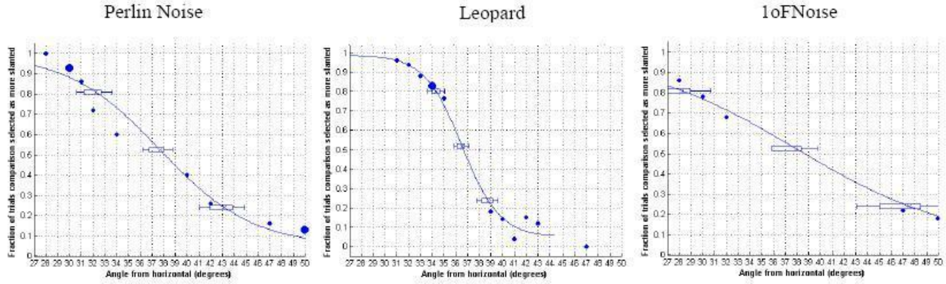
JU	
BO	
VB	
WS	
SP	

Table 5.1. The psychometric functions estimated for the textures observed by participants in experiment 1. Logistic functions (curves) were obtained by fitting them to the proportions of responses "test is more slanted/target is less slanted" against the standard stimulus. Each row corresponds to the data collected from one of five participants. Circle

size is proportional to the number of trials for each target measure. 68% and 95% confidence interval are displayed for percentage of 20%, 50% and 80% fraction of trials perceived as more slanted than the standard (test stimuli values are reported on the abscissa, which ranges from 27° to 50°).

If any texture were good in informativeness, then it would be easier to discriminate between different levels of slant expressed by it. Thus, difficulties in discriminating among different textures would be reflected by the shape (steepness) of corresponding psychometric functions.

I ran a one-way repeated-measures ANOVA in order to detect differences between the three levels of factor “texture”, considering the value of the slope of the psychometric functions at 50% of responses as the dependent variable (thus, a set of 3×5-1 = 14 measures). The test yielded a significant result for the effect of Type of texture [$F_{(2,6)}$= 12.615, p = .007, η^2_p= .808] (cf. Figure 5.3).

I have also analysed single subjects’ performance by comparing 95% confidence the extension of 95% confidence intervals around the value of the slope at 50% of each psychometric function (see Figure 5.4). Generally, performance with Leopard textures was the best for every subject (although confidence intervals at 95% between Leopard and Perlin Noise partially overlap for subject JU, BO and SP), and performance with Leopard was better than 1/F Noise (an overlap between confidence intervals is found only for subject BO). These results confirm the pattern of results found by Rosas et al. (2004). Leopard texture is generally easier to discriminate than Perlin Noise, which is generally easier to discriminate than 1/F Noise texture.

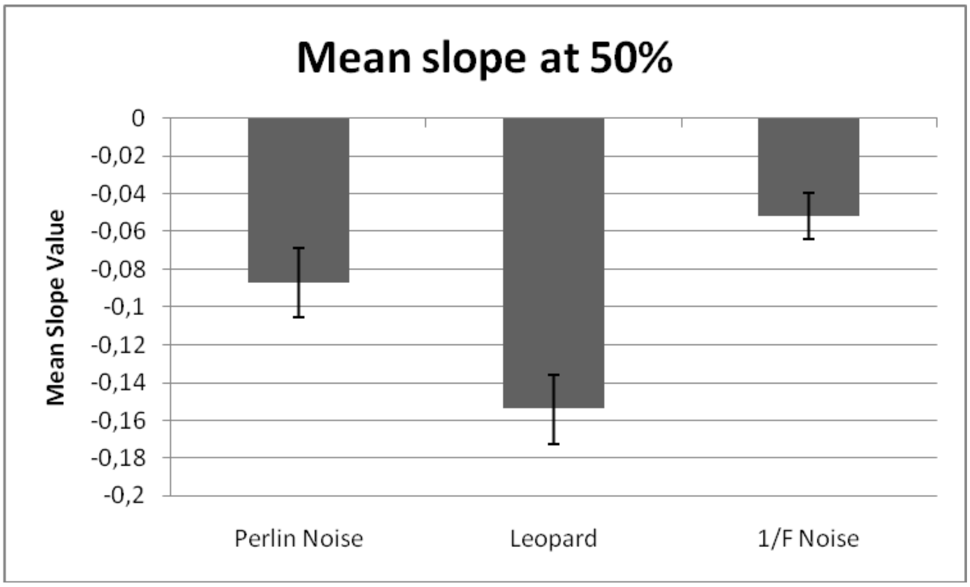


Figure 5.3. Mean slope values at 50% of the psychometric functions obtained in experiment 1. Texture type is reported on the abscissa, slope of the psychometric function at 50% is reported on the ordinate. Length of vertical bars corresponds to twice the standard error of corresponding means.

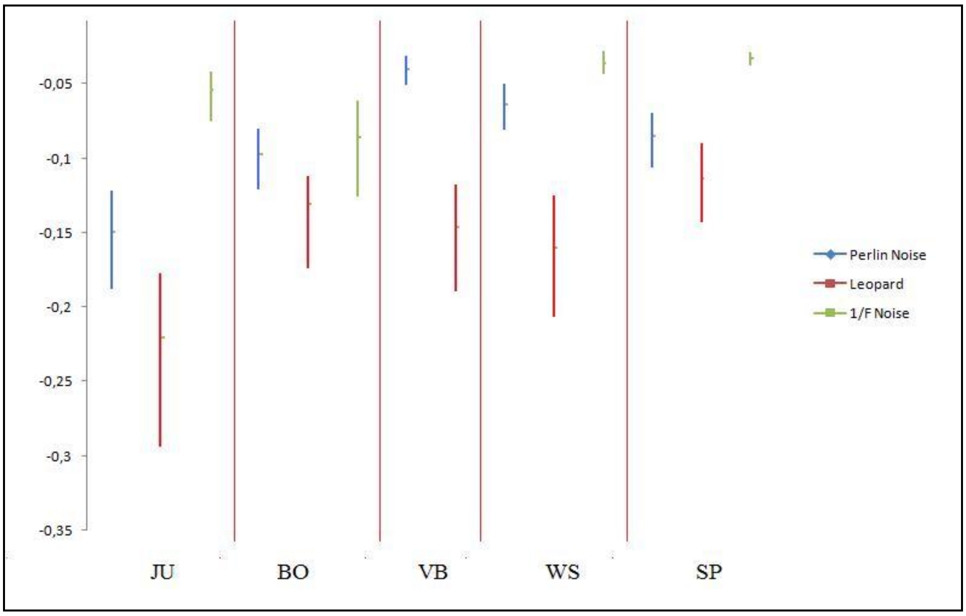


Figure 5.4. Lower and upper bound for 95% confidence intervals around the value of slope at 50% of each psychometric function, considering each participant and each texture.

In the aforementioned paper, these differences were explained in terms of changes of *APF* in each texture pattern, as suggested by Sakai & Finkel (1998): Leopard spans bigger changes than Perlin Noise does, and Perlin Noise spans bigger changes then 1/f noise (cf. Rosas et al., 2004, p. 1531). According to the model, a larger change of the APF within an image should improve the perception of slant for it, and this is confirmed by the results of the experiment.

5.3 Experiment 2

The second experiment constitutes the core of the study presented in this chapter: it specifically addresses the integration of the information at different positions in the visual field for perceptually determining the slant of a textured surface. Baseline values obtained from the first experiment are compared with those obtained when discrepant information between periphery and fovea is presented, in order to explore how they interact, under the hypotheses that if subjects rely only on local properties of the texture, participants should not draw their attention toward peripheral portion of the surface, so that their performance should not be affected by the presence or the size of a discrepant information outside GCW.

5.3.1 Experimental setup

The main characteristic differentiating the current experiment from the previous one is the introduction of a Gaze Contingent Window (GCW) so that the portion of the stimulus visible inside it could depict either the test term or the standard term of the constant stimuli method. Setting, task and participants were the same as in the previous experiment. Standard stimuli slant value was fixed at 37° (slant was measured from horizontal).

In this experiment, the portion outside the GCW displayed always a surface at 37 degrees slant (that is, the standard stimulus): only the image inside the GCW was actually depicting a change of slant between the two intervals in a trial (see Figure 5.5). The GCW was also used for the standard stimulus.

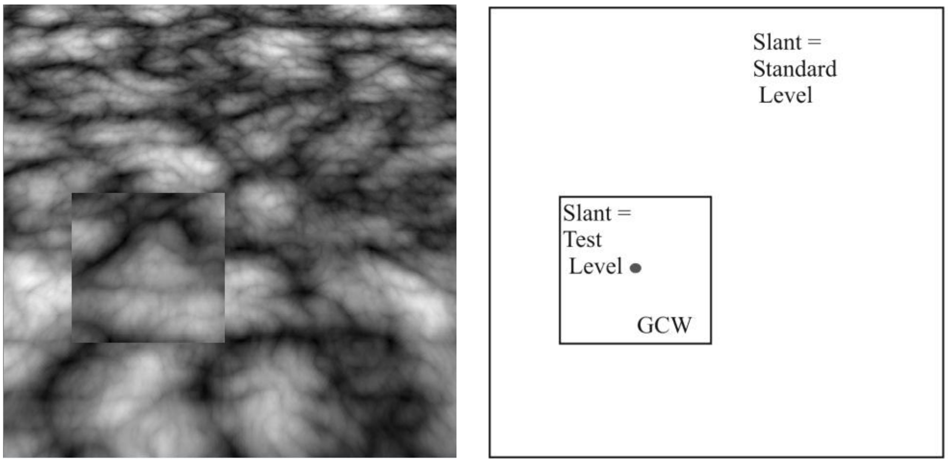


Figure 5.5. The picture on the left shows an example of how the stimuli changed after the introduction of the GCW. The picture on the right shows corresponding slant value for each portion of the stimulus. The small circle inside the square corresponds to current eye position for that configuration of the stimulus.

GCW had a square shape, and its size varied in three different levels, which are 3.6°, 5.4° and 7.2° of visual angle.

In this experiment I started by selecting a test slant such that it produced more than 90% of correct discrimination in Experiment 1 for each subject. Using this first estimate, some critical values were selected in order to obtain a more reliable estimation of the psychometric function, by pooling these data with those previously obtained. Each psychometric function was obtained using data computed on data from 200 to 400 in number.

Logistic functions were fitted to the data using the constrained maximum-likelihood method proposed by Wichmann & Hill (2001a, 2001b) implemented in the Psignifit Toolbox.

5.3.2 Results

The gathering of data for the second experiment had some difficulties. One of the risks of the experimental setting I used was that subject could realise the presence of the GCW, given the discontinuity along the border between the surface depicted inside and the surface depicted outside the GCW. Generally, no one of them reported its presence, except for subject SP during one of his last sessions. Because of this, data of SP collected after that moment were unreliable, and the subject reported that he found some difficulties to accomplish the task (he continued to focus his attention on the borders). Moreover, only two subjects, BO and JU, completed the set of trials, while subject WS and subject VB did not. Because of all these difficulties, I decided to split the analysis for each texture. Given the low numerosity of the samples and the presence of missing data across conditions, I ran three separate ANOVA for repeated measures, considering GCW size as the within factor and slope at 50% of the psychometric functions as the dependant variable. When data were missing, no data for that specific subject and for that specific texture were considered in the analysis (see Table 5.1 for a brief summary of data analysed in the second experiment)[7]. The general trend of results is displayed in Figure 5.6.

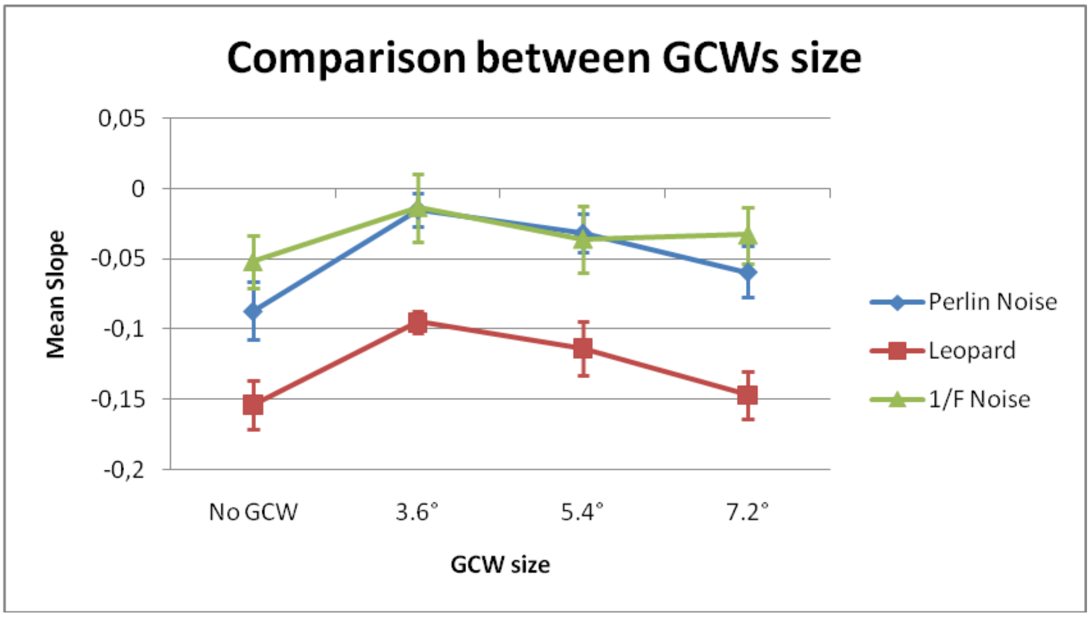


Figure 5.6. Plots of means of 50% slope to levels of GCW size for each texture. Length of vertical bars is twice the standard error of corresponding means.

---

[7] Given the low number of observations used for these analyses, I previously checked that every necessary assumption for ANOVA was satisfied.

The three-separate ANOVA, yielded significant results for the main effects of factor GCW size [$F_{(3,\ 9)}$ = 4.451, p = .035, η^2_p= .52] in Perlin Noise condition, significant results for the main effects of factor GCW size [$F_{(3,\ 12)}$ = 3.968, p = .035, η^2_p= .498] in Leopard condition, and significant results for the main effects of factor GCW size [$F_{(3,\ 12)}$ = 3.968, p = .035, η^2_p= .498] in 1/F Noise condition, for an alpha = .05.

Subject	Perlin Noise	Leopard	1/F Noise
JU	X	X	X
BO	X	X	X
VB	X	X	
WS	X	X	
SP		X	

Table 5.1. Data used in the Analysis reported in Section 5.3.2

These results show a general effect of the GCW size on each of the texture types I analysed: it seems that subjects take peripheral information into account when they are asked to judge the slant of a surface. In order to deepen the analysis on these results, I made the comparison between 95% confidence intervals around the value of the slope at 50% of the psychometric function for each subject, texture and GCW size, by checking if there were overlaps among confidence intervals.

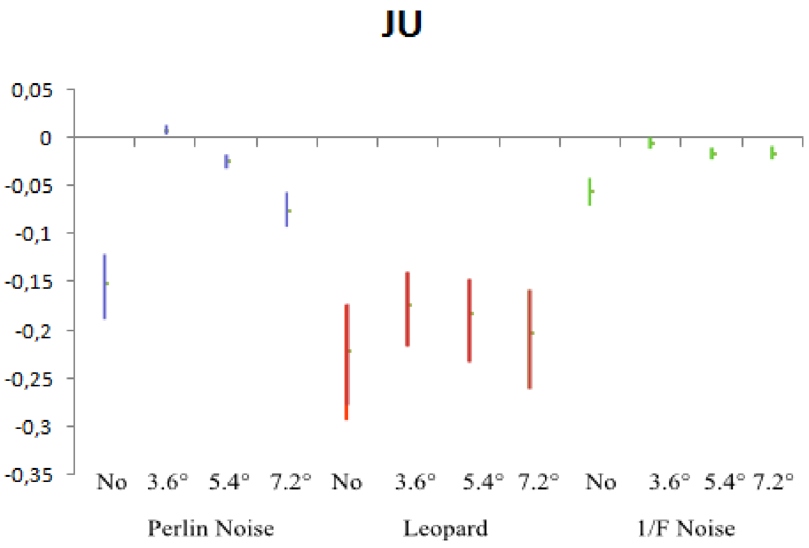
The effect of the application of a discrepancy between foveal and peripheral information on Leopard textured surfaces was effective only on subject WS and VB, and only when the size of the GCW was 3.6° or 5.4° (so, no effect was found when the size was 7.2°).

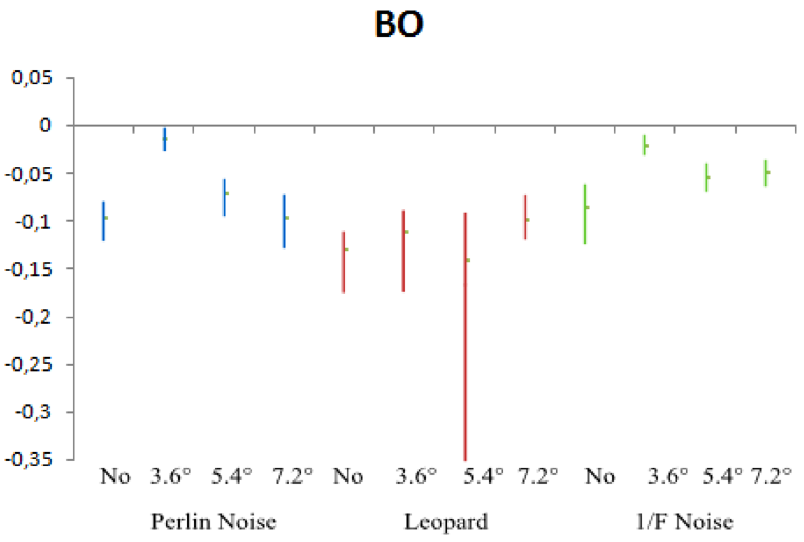
The effect of the application of a discrepancy between foveal and peripheral information on Perlin Noise textured surfaces was effective for all four subjects w performed this condition. Two cases deserve special comments. One comment refers to the fact that subject WS had a better performance when GCW size was 3.6° than when GCW size was 5.4°. Moreover, the performance in the first case is very similar to the one without the application of the discrepancy, while the performance in the second case is worse than the one measured during the first experiment. I interviewed the subject at the end of the experiment, and he reported that he suddenly realised that using a different strategy when solving the task could improve his performance. In particular, during all the previous sessions he fixed his gaze around the central position of the screen, while in these last ones he started to fixate on the upper part of the screen (an asymmetry in informativeness between upper portion and lower portion of a textured surface was also reported by Knill, 1998). Session with 5.4° size preceded the one with 3.6° size. I may surmise that, if subject WS would had not changed his strategies during the experiment, maybe the same trend observed for the other subjects would have been found.

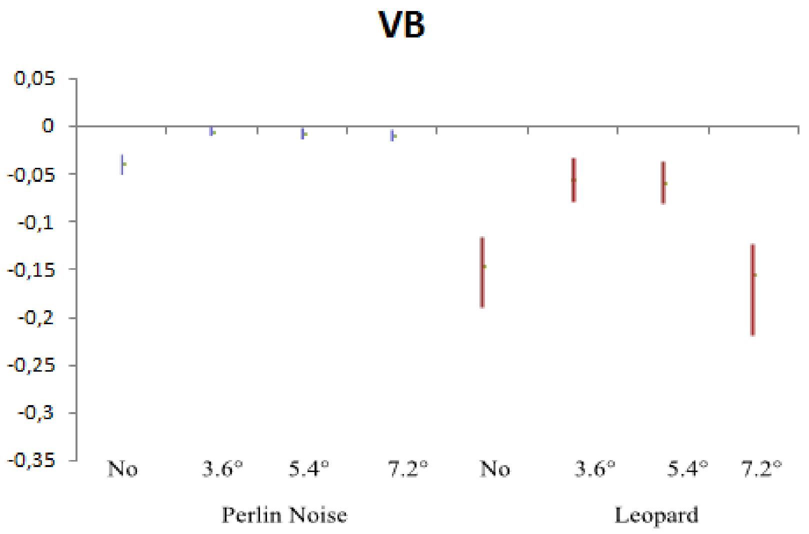
The second comment regards subject BO. Differently from JU and VB, he performed worse than he did in the baseline condition (i.e., without any discrepancy between different portion of the stimulus) only when GCW had the smallest size (see graphs in Figure 5.7).

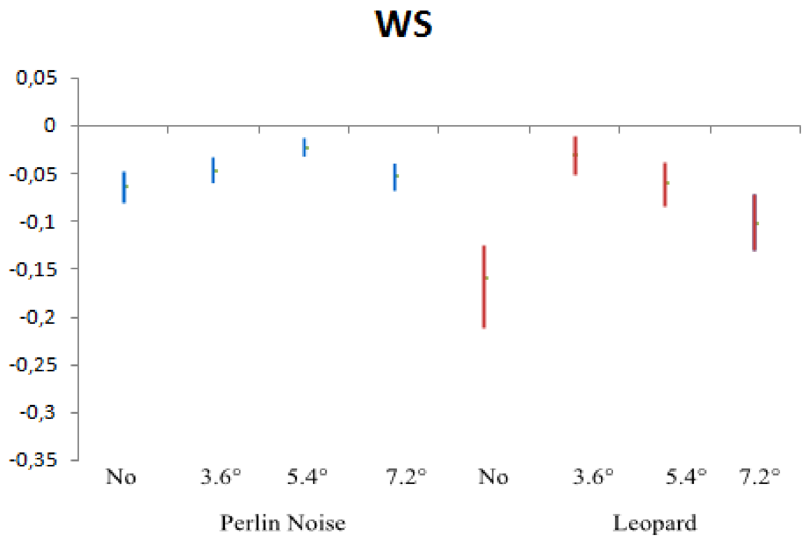
Given this trend of results, it seems legitimate to conclude that the effect of discrepancy is generally stronger in case of Perlin Noise texture than Leopard texture.

The effect of the application of a discrepancy between foveal and peripheral information on 1/F Noise textured surfaces is effective for both subjects who performed this condition. The results reflect for both of them what have already been found in the case of Perlin Noise condition: concerning subject JU, there is a difference between all the conditions in which GCW were introduced and the baseline one. Concerning subject BO, a difference in performance was found only between the GCW smallest size condition and the baseline condition.


JU
0,05
0
-0,05
-0,1
-0,15
-0,2
-0,25
-0,3
-0,35
No
3.6°
5.4°
7.2°
No
3.6°
5.4°
7.2°
No
3.6°
5.4°
7.2°
Perlin Noise
Leopard
1/F Noise


BO
0,05
0
-0,05
-0,1
-0,15
-0,2
-0,25
-0,3
-0,35
No
3.6°
5.4°
7.2°
No
3.6°
5.4°
7.2°
No
3.6°
5.4°
7.2°
Perlin Noise
Leopard
1/F Noise


VB
0,05
0
-0,05
-0,1
-0,15
-0,2
-0,25
-0,3
-0,35
No
3.6°
5.4°
7.2°
No
3.6°
5.4°
7.2°
Perlin Noise
Leopard


WS
0,05
0
-0,05
-0,1
-0,15
-0,2
-0,25
-0,3
-0,35
No
3.6°
5.4°
7.2°
No
3.6°
5.4°
7.2°
Perlin Noise
Leopard

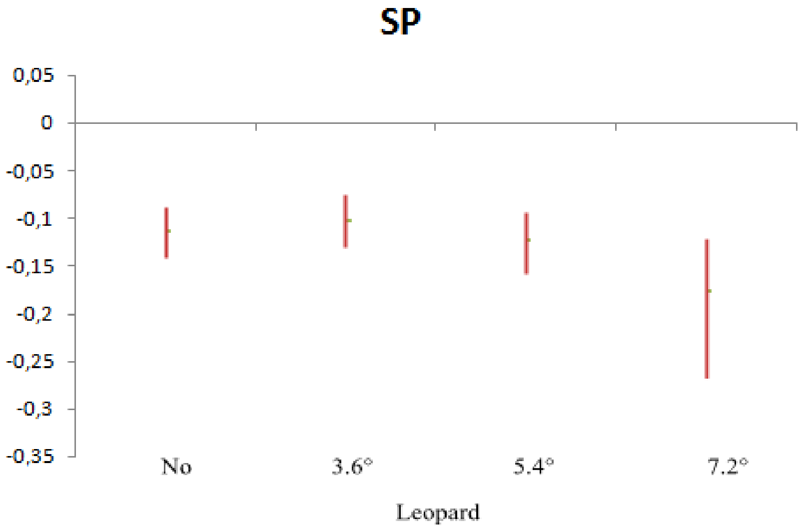


Figure 5.7. Lower and upper bound for 95% confidence intervals around the mean for slope at 50% for each participant and each texture. Bars referring to the same textures has the same colour (Perlin Noise ones are blue, Leopard ones are red, 1/F Noise ones are green). Each graph corresponds to the data obtained for each subject. Subjects VB, WS and SB did not perform the 1/F Noise condition in the second experiment. Moreover, as already explained in the text, subject SP had problems during Perlin Noise condition in the second experiment.

These results, and the estimation of the effect size (see page 115), suggest that 1/F Noise condition was affected more severely than Leopard condition by the introduction of discrepant information among distinct portions of the stimuli.

5.4 General discussion

The study described in this chapter specifically addresses the theme of spatial integration of texture information. I hypothesized that, if subjects focus their attention on selected and limited stimulus regions when judging slant from texture, the introduction of a discrepancy between local and peripheral information should not modify their performance. A local strategy can be effective only when strong local information is available. In case the changes in peaks are less reliable, given the introduction of discrepant information, and the system needs to rely on textures with small changes in APF, as happens in the two noise- based textures condition, a discrepancy between local and peripheral information affects the subjects' performance more dramatically. The results described above partially confirm these hypotheses, since the presentation of discrepant information using a GCW really made the task more difficult for the participants in the experiment. Given this general trend, a detailed analysis of the performance of each subject showed strong differences between subjects. Leopard textured surface should be the one less affected by discrepant information between local and peripheral portion of the stimuli, but two subjects showed an effect when incongruent information was presented. It is interesting to notice that these two subjects were also those who had the worst performances in general for each texture, and that after one of them (WS) changed his strategy and started to fixate a different part of the image, he succeeded in improving his performance. These outcomes suggest the importance of the role of attentional factors (i.e., the part of the stimuli to which subjects attend) in the task here described, and how different strategies make a different use of local and peripheral information.

Chapter 6

General Conclusions

As I stated in the introductory section, what mainly distinguishes my research work from previous ones on similar topics is the type of perspective adopted when considering stimulus information about depth. Whereas in most of the studies about depth perception, depth cues are classified according to “qualitative-categorical” criteria, I used a different criterion, based on distinguishing the various sources of information according to the “topological/spatial” relations they have in the visual scene. I use the notion of topological/spatial relations with two different meanings, and the experiments described in Chapters 3, 4 and 5 have been planned in order to cover both of them.

One of the two meaning corresponds to the topological/spatial relations among the different parts composing the spatial layout. In particular, the experiments described in Chapters 3 and 4 were specifically aimed at testing the interaction of one of these factors, the “optical contact”, with other intrinsic and extrinsic factors.

The other meaning refers to the relations between different portions of the visual field. In particular, the experiments described in Chapter 5 were specifically aimed at understanding how information presented in the fovea interacts with information presented in the peripheral portion of the eye in determining the perceptual slant of a textured surface.

Indeed, I paid more attention in studying the first type of meaning, but I found it necessary, for the sake of completeness, also to present an experimental research covering the second one, in order to describe in detail this important distinction.

Experiments described in Chapters 3 and 4 analysed the role of what I referred to as “extrinsic factors”: results from the experiments showed that these sources of information are generally effective in determining the slant of a target object. Nevertheless, if some “intrinsic factors” residing on the object are strong enough, these factors can reduce the effectiveness of sources of information of different types. In general, extrinsic factors, like “optical contact“, play important roles in determining depth-related properties in more natural and richer contexts (e.g., real environments), as Ni, Braunstein & Andersen (2005) found when analysing the interaction between shading and optical contact. As a matter of fact, these authors showed that whereas in simple configurations the stimulus shading appears to overcome the effect of optical contact, in more complex situations, where more than one shadow appear in the environment, optical contact is used to disambiguate among possible and plausible interpretations of the same visual scene.

The types of information manipulated in the experiments reported in this thesis illustrate the subtleness and intricateness of problems of vision science, by focusing on some factors that are generally neglected in other researches. I tried to highlight this complexity, at an abstract level, in the first two chapters, where I discussed some premises to formal analysis of the theory of depth perception using a set-theoretic approach. In particular, the second chapter provides an extension of the premises and comments introduced in the first chapter: the chapter is specifically aimed at extending the idea of integration among depth-cues in order to take into account how information residing on different parts of the visual environment may interact in supporting depth vision. Furthermore, the distinction between extrinsic and intrinsic properties, based on set-theoretic terms introduced in Chapter 1, is proposed. This idea is not really new in the vision science literature: among others, Gogel (1978) distinguished between absolute cues, i.e., “the factors that determine the perceived characteristics of an object independently of other objects”, and relative cues, i.e., “the factors that change the perception when other objects are introduced” (p. 126). In my theorization, I tried to give a wider background to this concept, and specify its relation with other sources of information.

One of the notable concepts within the discussion presented in the first two chapters is the characterisation of the ‘constraining effect’ of depth-cues as the central component of a ‘psychophysical inference rule’. I drew the distinction between a ‘punctual’ constraining effect (as a function from the range of a stimulus variable to the range of a perceptual variable) and a ‘limiting’ constraining effect (as a function from the range of a stimulus variable to the power set of the range of a perceptual variable). In other words, “constraints” and “constraining effects” can be conceived as set of rules or relations concerning some variables in their dynamic aspects, i.e., in the changeable values they take on during a certain series of observational cases. They express the dependence that on some occasions reduces the indeterminacy of the values taken on by some variables. Formal terms adopted in the first two chapters have been helpful in fixing such ideas and discussing them.

The characterisation in terms of constraining effects is a step towards discussing certain elementary aspects of depth-cue theory in purely set-theoretic terms. Suggestions in this direction are given in Section 1.4.1, in commenting on the general problem of ‘integration of depth information among cues residing on the same aggregate’, and in Section 2.1.1, in commenting on the general problem of “integration of depth information among cues residing on different aggregates”. The general idea of “relational constraint”, and the references I made to the “Constraint Network” theory (see pages 17-18), were a guide for the ideas I presented in this thesis and my research work, since the early stages of it.

This orientation came to my attention by considering one of the problems with contemporary investigation of interaction of depth-factors, both in human vision and machine vision: several theories about cue integration resort to highly complex models with selective requirements (e.g., regularization theory, theory of Markov fields, and so on), and lack of simplicity. On the contrary, the proposed theoretical paradigm requires a minimum of primitive terms and assumptions. I suggested that the paradigm known as “Constraint Satisfaction Problem” (Montanari, 1974; Dechter, 2003) appears to be suitable as a general frame for discussing constraints in the psychophysical context. Some of the examples I provided in the text (see, for instance, Sections 1.2.4, 1.3, 2.1) constitute elementary applications of these ideas[1].

The concept of constraint has been also taken into consideration, in my thesis, to account for results of the experiments described in Chapter 3. In particular, I used this idea to characterize the interaction of “process by optical contact” and “process

---

[1] In those examples, I did not use the terminology of “Network of Constraint Theory”, but the approach I used in finding a solution constitutes an elementary example of their application, in particular for the case discussed in Section 1.3.

by inverse optics” on determining the slant of the straight linear figure used in the stimuli, as explained in Chapter 3. In particular, the concept referred to as “process by inverse optics” is a typical exemplification of how a conjectural psychophysical constraint may be produced: considering some generally valid optical law (i.e., a law that allows predicting properties of a retinal image from properties of the physical world), a conjectural psychophysical constraint may be derived by changing the order of variables (i.e., producing a rule that allows predicting properties of the physical world from properties of the retinal image, cf. p. 59).

Further exemplification of this type of approach can be found in the explanation of the combined effects of variables in the fourth chapter, in which I hypothesized an “accumulation effect” of texture and linear perspective information in commenting results of Experiment 2 (the description of accumulation procedure in constraints terms have been described in Chapter 1, p. 29)

A psychophysical paradigm based on “Constraint Networks” may be classified in the category of “strong-fusion models” (cf. Section 1.4 about depth cues integration modalities), since it is based on interactive or holistic principles, and does not need a modular organisation of information. As Johnston, Landy, Maloney and Young (1991) pointed out, one of the main problems of “strong fusion models” is their huge complexity and difficulty in modelling. An approach like the one I considered is based on sets of rules (or relational constraints) that can be derived, e.g., by inverse optics or using a limited set of ecologically-plausible prior constraints. As I expressed in Chapter 1, variables playing a part in a psychophysical analysis may differ widely from one another as the richness or density in information of their possible values is concerned. One advantage of an approach based on “relational constraints” is its great flexibility, i.e., it is suitable to combine heterogeneous variables and rules, constituting a possible alternative to those models which require, for instance, mechanisms such as “cue promotion” or “ancillary measures” (cf. Johnston, Landy, Maloney & Young, 1995). Furthermore, the characterization offers not only the possibility of integrating information from heterogeneous sources, but also, as I tried to illustrate in Chapter 2, to take into account and describe integration of information among cues carried by different units.

An approach based on the combination of constraints is not really new in the vision science literature. This idea is clearly present, for instance, in Sedgwick (2001). On commenting the most traditional models of depth cue integration, the author stated that: "the conditional probabilities of Bayesian statistics are conceptually related to [...] environmental constraints [...]. These constraints, however, being based on geometry, optics, and the persisting physical qualities of the environment, are often *determinant*, or non-probabilistic” (p. 154). Then, he concluded: “another non-modular way of modelling the combination of multiple sources of information is as the interaction of a *large number of conditional inference rules*, such as form the basis of expert systems“ (p. 154, italic mine). These *conditional inference rules*, as proposed, for instance, by Sedgwick (1987), resemble the representation in terms of “network of constraint”.

In conclusion, I intend to remark some of the limitations of the present research. The premises to formal analysis outlined in Chapter 1 and 2 are suitable for extending psychophysical models to the general idea of constraints, providing a rigorous and well-constructed framework to this effect. Indeed, the “Network of Constraints” theory has its major strength in proposing several algorithms and computational procedure for solving the general category of problems referred to as “Constraint Satisfaction Problems” (see Chapter 1, page 17, note 8). The class of algorithms proposed inside this framework is quite rich, and contains several sophisticated solutions (cf. Rossi, Beek, Walsh, 2006). In my research work, I decided to focus on more basic, preliminary questions, and I did not consider the specific procedures for solving problems: the examples I provided are comparatively simple in their conceptual profile and do not require any sophisticated algorithms to be solved. Neither had I made any special comments on this problem on my experiments, since the set of rules used to explain the results - I am referring in particular to experiment 3 - are very simple, as simple is the type of stimulus configuration I used. In my opinion, the real strength of the models I proposed can be fully evaluated in more complex and realistic situations, in which the set of rules being in action is rich and the interaction is complex. I hope that my contribution could constitute an initial step in this direction of research on depth vision.

References

Agresti, A. (2002). *Categorical data analysis.* New York: Wiley.

Aloimonos J., & Shulman, D. (1989). *Integration of Visual Modules: An Extension of the Marr Paradigm*. Boston: Academic Press.

Andersen, G.J., Braunstein, M.L., & Saidpour, A. (1998). The perception of depth and slant from texture in three-dimensional scenes. *Perception, 27*, 1087-1106.

Bajcsy R., & Lieberman, L. (1976). Texture gradient as a depth cue. *Computer Graphics and Image Processing, 5*, 52-67.

Bauermeister, M. (1978). Differences between right versus left lateral body tilt in its effect on the visual and tactual perception of verticality. *Psychological Research, 40*(2), 183-187.

Berkeley, G. (1709). An essay towards a new theory of vision. Reprinted in *The works of George Berkeley*, vol 1., (2005, pp. 229-293). Adamant Media Corporation: Boston (MA).

Bertera, J.H., & Rayner, K. (2000). Eye movements and the span of effective stimulus in visual search. *Perception & Psychophysics, 62*, 576-585.

Bian, Z., Braunstein M.L., & Andersen, G.J. (2005). The ground dominance effect in the perception of 3-D layout. *Perception & Psychophysics*, *67,* 802-815.

Bian, Z., Braunstein, M.L., & Andersen, G.J. (2006). The ground dominance effect in the perception of relative distance in 3-D scenes is mainly due to characteristics of the ground surface. *Perception & Psychophysics, 68*, 1297-1309.

Blake, A., Bülthoff, H.H., & Sheinberg, D. (1993). Shape from texture: Ideal observers and human psychophysics. *Vision Research, 33*, 1723-1737.

Boring, E.G. (1942). *Sensation and perception in the history of experimental psychology*. New York: Appleton-Century-Crofts.

Braunstein, M.L., & Payne, J.W. (1968). Perspective and the rotating trapezoid. *Journal of the Optical Society of America, 58*, 399-403.

Bruno, N., & Cutting, J.E. (1988). Minimodularity and the perception of layout. *Journal of. Experimental Psychology: General, 117*, 161-170.

Bruce, V., Green, P.R., & Georgeson, M.A. (1996) *Visual Perception: Physiology, Psychology and Ecology*. (3rd Ed.) Hove: Psychology Press.

Brunswik, E. (1966). Reasoning as a universal behavior model and a functional differentiation between "perception" and "thinking". In K.R. Hammond (Ed.), *The psychology of Egon Brunswik* (pp. 487-494). New York: Holt, Rinehart and Winston.

Bülthoff, H.H. (1991). Shape from X: Psychophysics and Computation. In M. Landy and A. Movshon (Eds.), *Computational Models of Visual Processing*, 305-330, Cambridge (MA): MIT Press. .

Bülthoff, H.H., & Mallot, H.A. (1990). Integration of Stereo, Shading and Texture. In A. Blake and T. Troscianko (Eds.) *AI and the Eye*, (pp. 119-146), John Wiley Sons: Chichester.

Burigana, L. (1999). Constraints in vision. Outline of a set-theoretic approach. *Mathematical Social Sciences, 38*, 315-341.

Burigana, L., & Martino, F. (2009). Constraining effects by single and combined cues to depth. Premises to formal analysis. Submitted to *Teorie e modelli*.

Burton, G., & Turvey, M.T. (1990). Perceiving lengths of rods that are held but not wielded. *Ecological Psychology, 2*, 295-324.

Burton, H.E. (1945). The optics of Euclid. *Journal of the Optical Society of America, 35*, 357-372.

Campbell, F.W. and Robson, J.G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology, 197*, 551-566.

Carr, H.A. (1935). *An Introduction to Space Perception*. New York: Longmans, Green.

Clark, J.J., & Yuille, A.L. (1990). *Data fusion for sensory information processing systems.* Norwell: Kluwer Academic Publisher.

Creem-Regehr, S.H., Gooch, A.A., Sahm, C.S., & Thompson, W.B. (2004). Perceiving virtual geographical slant: Action influences perception. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 811-821.

Cutting, J.E. (1997). How the eye measures reality and virtual reality. *Behavior Research Methods, Instruments & Computers, 29*, 27-36.

Cutting, J.E., & Millard, R.T. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General, 113,* 198–216.

Cutting, J.E., & Vishton, P.M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein and S. Rogers (Eds.), *Handbook of perception and cognition. Vol. 5: Perception of space and motion* (pp. 69-117). San Diego: Academic Press.

Da Silva, J.A. (1985). Scales for perceived egocentric distance in a large open field: comparison of three psychophysical methods. *The American Journal of Psychology, 98*, 119-144.

Dechter, R. (2003). *Constraint processing*. San Francisco: Morgan Kaufmann.

Dosher, B.A., Sperling, G., & Wurst, S.A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research, 26,* 973-990.

Ellis, B. (1991). *Scientific essentialism*. Paper presented at the 1991 conference of the Australian Association for the History and Philosophy of Science (reported in Weatherson, 2006).

Ernst, M.O. and Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429-433.

Fechner, G.T. (1860/1966). *Elements of psychophysics* (D.H. Howes, E.G. Boring, Trans.), Holt, Rinehart and Winston: New York.

Feria, C.S., Braunstein, M.L., & Andersen, G.J. (2003). Judging distance across texture discontinuities. *Perception, 32*, 1423–1440.

Freeman, R. B., Jr. (1966). Effect of size on visual slant. *Journal of Experimental Psychology, 71*, 170-176.

Freeman, W.T. (1994). The generic viewpoint assumption in a framework for visual perception. *Nature*, 368, 542-545.

Freeman, W.T. (1996). Exploiting the generic viewpoint assumption. *International Journal of Computer Vision, 20*, 243-261.

Gibson, J.J. (1946). Perception of distance and space in the open air. In J.J. Gibson (Ed.), *Motion Picture Testing and Research (AAF program Report #7)*. Reprinted in D.C. Beardslee & M. Wertheimer (Eds.), *Readings in Perception.* Princeton: D. Van Nostrand.

Gibson, J.J. (1950a). *The perception of the visual world.* Boston: Houghton-Mufflin.

Gibson, J.J. (1950b). The perception of visual surfaces. *American Journal of Psychology, 63*, 367-384.

Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Gillam, B.J. (1968). Perception of slant when perspective and stereopsis conflict: Experiments with aniseikonic lenses. *Journal of Experimental Psychology, 78,* 299-305.

Gogel, W.C. (1965). Equidistance tendency and is consequences. *Psychological Bulletin, 64*, 153-163.

Gogel, W.C. (1972) Depth adjacency and cue effectiveness. *Journal of Experimental Psychology, 92*, 176–181.

Gogel, W.C. (1976). Adjacency and attention as determiners of perceived motion. *Vision Research, 16*, 839-845.

Gogel, W.C. (1978). The adjacency principle in visual perception. *Scientific American, 238*, 126-139.

Gordon I.E. (1989). *Theories of visual perception*. New York: Wiley.

Gregory, R.L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London B, 290*, 181-97.

Grove, P.M., Ono, H., & Kaneko, H. (2003). T-junctions and perceived slant of partially occluded surfaces. *Perception*, *32,* 1451-1464.

Hadamard, J. (1923). *Lectures on the Cauchy Problem in Linear Partial Differential Equations*. London: Yale University Press.

Hatfield, G.C., & Epstein, W. (1985). The status of the minimum principle in the theoretical analysis of visual perception. *Psychological Bulletin, 97*, 155-186.

He, Z.J., Wu, B., Ooi, T.L., Yarbrough, G., & Wu, J. (2004). Judging egocentric distance on the ground: Occlusion and surface integration. *Perception*, *33*, 789-806.

Heckbert, P.S. (1989). *Fundamentals of texture mapping and image warping* (Technical Report CSD-89-516). Berkeley: University of California, Berkeley.

Helmholtz, H. von (1866) *Treatise on physiological optics*, *vol.* 3.. J. P. C. Southall trans., 3rd ed., 1925, Menasha, WI: The Optical Society of America.

Hillis, J.M., Watt, S.J., Landy, M.S., & Banks, M.S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, *4*, 967-992.

Howard, I.P., & Rogers, B.J. (2002). *Seeing in depth. Volume 2: Depth perception.* Toronto: Porteous.

Jameson, D., & Hurvich, L.M. (1959) Perceived color and its dependence on focal, surrounding, and preceding stimulus variables, *Journal of the Optical Society of America*, *49*, 890.

Jeavons, P., Cohen, D., & Cooper, M.C. (1998). Constraints, consistency and closure. *Artificial Intelligence, 101*, 251-265.

Knill, D. C. (1998). Discrimination of planar surface slant from texture: human and ideal observers compared. *Vision Research*, *38*, 1683-1711.

Knill, D.C., & Saunders, J.A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*, 2539-2558.

Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace & World.

Koning, A., & Van Lier, R. (2003). Object-based connectedness facilitates matching. *Perception & Psychophysics*, *65*, 1094-1102 .

Kubovy, M. (1986). *The psychology of perspective and Renaissance Art*. Cambridge, England: Cambridge University Press.

Landy, M.S., Maloney, L.T., Johnston, E.B., & Young, M. (1995). Measurement and modelling of depth cue combination: In defence of weak fusion. *Vision Research*, *35*, 389-412.

Lawson, R., Bertamini, M., & Liu, D. (2007) Overestimation of the projected size of objects on the surface of mirrors and windows. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 1027-1044 .

Loftus, G., & Mackworth, N.H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Performance and Perception, 4*, 565-572.

Longuet-Higgins, H.C. (1986). Visual motion ambiguity. *Vision Research*, *26*, 181-183.

Loomis, J.M., & Knapp, J.M. (2003). Visual perception of egocentric distance in real and virtual environments. In L.J. Hettinger & M.W. Haas (Eds.), *Virtual and adaptive environments* (pp. 21-46). Mahwah: Lawrence Erlbaum Associates.

Madison, C., Thompson, W., Kersten, D., Shirley, P., & Smits, B. (2001). Use of interreflection and shadow for surface contact. *Perception and Psychophysics, 63*, 187-194.

Malik, J., & Rosenholtz, R. (1994). A computational model for shape from texture. In Bock, G.R., & Goode, J.A. (Eds.), *Higher-Order Processing in the Visual System* (pp. 272-286). Chichester: John Wiley & Sons.

Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition*, *81*, B1-B9.

Mamassian, P., & Landy, M. S. (2001). Interaction of visual prior constraints. *Vision Research*, *41*, 2653-2668.

Marr, D., & Nishihara, H.K. (1978). Representation and recognition of the spatial organisation of three-dimensional shapes. *Proceedings of the Royal Society, London. Series B*, *200*, 269-294.

McCarley, J.S., & He, Z.J. (2000). Asymmetry in 3-D perceptual organization: Ground-like surface superior to ceiling-like surface. *Perception & Psychophysics*, *62*, 540-549.

McConkie, G.W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*, 578-586.

Meese, T.S., & Holmes, D.J. (2004). Performance data indicate summation for pictorial depth-cues in slanted surfaces. *Spatial Vision*, *17*, 127-151.

Meng, J.C., & Sedgwick, H.A. (2001). Distance perception mediated through nested contact relations among surfaces. *Perception & Psychophysics, 64*, 1-14.

Meng, J.C., & Sedgwick, H.A. (2002). Distance perception mediated across spatial discontinuities. *Perception & Psychophysics, 63*, 1-15.

Montanari, U. (1974). Network of constraints: Fundamental properties and applications to picture processing. *Information Sciences, 7*, 95-132.

Metzger, W. (1971). *I fondamenti della psicologia della Gestalt.* Firenze: Giunti & Barbera. [Original edition: *Psyhologie. Die Entwicklung ihrer Grundannahmen seit der Einführung des Experiments.* Darmstadt: Steinkopff, 1941].

Nakayama, K. (1994). James J. Gibson – An appreciation. *Psychological Review, 101*, 329-335.

Ni, R., Braunstein, M.L., & Andersen, G.J. (2004). Perception of scene layout from optical contact, shadows and motion. *Perception, 33,* 1305-1318.

Ni, R., Braunstein, M.L., & Andersen, G.J. (2005). Distance perception from motion parallax and ground contact. *Visual Cognition, 12*, 1235-1254.

Ni, R., Braunstein, M.L., & Andersen, G.J. (2007). Scene layout from ground contact, occlusion, and motion parallax. *Visual Cognition, 15,* 46-68.

Nodine, C. F., Carmody, D. P., & Kundel, H. L. (1978). Searching for Nina. In J. W. Senders, D. F. Fisher, & R. A. Monty (Eds.), *Eye movements and the higher psychological functions* (pp. 241- 257). Hillsdale: Erlbaum.

Oruç, I., Maloney, L.T., & Landy, M.S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research, 43*, 2451-2468.

Perlin, K. (1985). An image synthesizer. *Computer Graphics, 19*, 287-296.

Poggio, T., Torre, V. and Koch, C. (1985). Computational vision and regularization theory. *Nature, 317*, 314-319.

Pomplun, M., Reingold, E.M., & Shen, J. (2001). Investigating the visual span in comparative search: The effects of task difficulty and divided attention. *Cognition, 81*, 57-67.

Purdy, W.C. (1958). The hypothesis of psychophysical correspondence in space perception. *Dissertation Abstracts International, 19*, 1454-1455 (University

Microfilms No.58-5594). Reported by Andersen, Braunstein & Saidpour (1998).

Ramachandran, V.S. (1988). Perception of shape from shading. *Nature, 331*, 163-166.

Rayner, K., (1998). Eye movements in reading and Information Processing: 20 years of research. *Psychological Bulletin, 124*, 372-422.

Reinhardt-Rutland, A.H. (1999). The framing effect with rectangular and trapezoidal surfaces: Actual and pictorial surface slant, frame orientation, and viewing condition. *Perception, 28,* 1361-1371.

Rock, I. (1983). *The logic of perception*. Cambridge(MA): MIT Press.

Rosas, P., Wichmann, F.A., & Wagemans, J. (2004). Some observations of the effects of slant and texture type on slant-from-texture. *Vision Research*, 44, 1511-1535.

Rosas, P., Wichmann, F.A., & Wagemans, J. (2007). Texture and object motion in slant discrimination: Failure of reliability-based weighting of cues may be evidence for strong fusion. *Journal of Vision*, 7, 1-21.

Rossi, F., Beek, P.V., & Walsh, T. (2006). Constraint Satisfaction: An Emerging Paradigm. *Handbook of Constraint Programming. Foundations of Artificial Intelligence*. Amsterdam: Elsevier

Saida, S., & Ikeda, M. (1979). Useful Visual Field Size for Pattern Perception. *Perception & Psychophysics, 25*, 119-125.

Sakai, K., & Finkel, L.H. (1995). Characterization of the Spatial Frequency Spectrum in Perception of Shape-from-Texture. *Journal of the Optical Society of America A, 12*, 1208-1224.

Sakai, K., & Finkel, L.H. (1997). Spatial-frequency analysis in the perception of perspective depth. *Network: Computation in Neural Systems, 8*, 335-352.

Sauer, C.W., Braunstein, M.L., Saidpour, A., & Andersen, G.J. (2002). Propagation of depth information from local regions in 3-D scenes. *Perception*, 31, 1047-1059.

Sauer, C.W., Saidpour, A., Braunstein, M.L., & Andersen, G.J. (2001). Perceived depth of 3-D objects in 3-D scenes. *Perception,* 30, 681-692.

Schlosberg, H. (1941). A scale for the judgment of facial expressions. *Journal of Experimental Psychology, 29,* 497-510.

Sedgwick, H.A. (1986). Space perception. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of perception and human performance. Volume 1: Sensory processes and perception* (pp. 21.1-21.57). New York: Wiley.

Sedgwick, H. (1987) Layout2: A production system modeling visual perspective information. *Proceedings of the 1st IEEE International Conference on Computer Vision*, 662-666.

Sedgwick, H.A. (2001). Visual space perception. In E. B. Goldstein (Ed.) *Blackwell Handbook of Perception* (pp. 128–167). Malden: Blackwell Publishers.

Sinai, M.J., Ooi, T.L., & He, Z.J. (1998). Terrain influences the accurate judgment of distance. *Nature*, *395*, 497-50.

Stevens, K.A. (1981). The information content of texture gradients. *Biological Cybernetics,* 42, 95-105.

Swoyer, C. (2000). Properties. In E. N. Zalta (Ed.), *Stanford Encyclopaedia of Philosophy*. URL: <http://plato.stanford.edu/entries/properties/>

Taylor, P. (1960). *The notebooks of Leonardo da Vinci*. New York: New American Library.

Todd, J.T. (1985). Perception of structure from motion: Is projective correspondence of moving elements a necessary condition? *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 689-710.

Todd, J. T. & Akerstrom, R. A. (1987). Perception of three-dimensional form from patterns of optical texture. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 242-255

Todd., J.T., & Mingolla, E. (1983). The perception of surface curvature and direction of illumination from pattern of shading. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 238-252.

Turk, G. (1991). Generating textures on arbitrary surfaces using reaction-diffusion. *Computer Graphics (SIGGRAPH '91 Proceedings),* 289-298.

Vallentyne, P. (1997). Intrinsic Properties Defined. *Philosophical Studies, 88,* 209-219.

Van Lier, R., van der Helm, P., & Leeuwenberg, E. (1994). Integrating global and local aspects of visual occlusion. *Perception, 23*, 883-903.

Velisavljevic, L., & Elder, J.H. (2006). Texture Properties Affecting the Accuracy of Surface Attitude Judgements, *Vision Research, 46*, 2166-2191.

Wade, N.J., & Finger, S. (2001) The eye as an optical instrument: from camera obscura to Helmholtz's perspective. *Perception, 30*, 1157–1177.

Watson, J.S., Banks, M.S., von Hofsten, C., & Royden, C.S. (1992) Gravity as a monocular cue for perception of absolute distance and/or absolute size. *Perception, 21*, 69–76.

Weatherson, B. (2006). Intrinsic vs. Extrinsic Properties. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. URL: < http://plato.stanford.edu/entries/intrinsic-extrinsic/>.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. II. *Psychologische Forschung, 4*, 301-350.

Wichmann, F.A., & Hill, N.J. (2001a): The psychometric function: I. Fitting, sampling and goodness-of-fit. *Perception and Psychophysics 63*, 1293-1313.

Wichmann, F.A., & Hill, N.J. (2001b): The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception and Psychophysics 63*, 1314-1329.

Wu, B, Ooi, T.L., & He, Z.J. (2004) Perceiving distance accurately by a directional process of integrating ground information. *Nature*, *428*, 73-77.

Wu, B., He, Z.J., & Ooi, T.L. (2007). Inaccurate representation of the ground surface beyond a texture boundary. *Perception*, *36*, 703-721.