Constraints in vision
Outline of a set-theoretic approach

Luigi Burigana*

Università degli Studi di Padova, Dipartimento di Psicologia Generale, Via Venezia, 8, I-35131 Padova, Italy

Accepted 1 November 1998

Abstract

The idea is advanced that the highly complex relationship between stimuli and percepts in vision may be formally described as an extensive system of relational constraints, involving both stimulus and perceptual variables; some guidelines are also suggested for working out this idea and applying it to salient general questions of vision research. The paper starts by defining the basic notions of located variable and located constraint (Sections 2 and 3) with respect to a presumed homogeneous set of observational situations. Then the most important notions of a system of constraints and of pairing between a system of constraints and an assignment of values to stimulus variables are introduced (Sections 4 and 5). A set-theoretic way of representing the combined predictive power of such pairing is suggested, and some elementary properties of the concept are described. Lastly, a few comments are made on the possible use of the constraint paradigm in discussing some critical aspects of contemporary research on vision and some reasons are given for the suggested set-theoretic approach (Sections 6 and 7). © 1999 Elsevier Science B.V. All rights reserved.

1. Introduction: the multiple constraint paradigm

An observer has a visual experience when he/she looks at some portion of the surrounding physical reality. In a very general sense, the task of the science of vision is to study how the data in visual experience depends on optically important aspects of the observed reality or, more properly, how the contents in the resulting phenomenal scene depend on the given stimulus conditions. There are different ways of conceiving and pursuing such a general task. For example, one may study vision by trying to discover or

E-mail address: burigana@psico.unipd.it (L. Burigana)
reconstruct “perceptual processes”, i.e., examining the complex dynamics (transformations, causations, integrations, etc.) which presumably take place in the visual system, starting from the entering stimuli and ending with the perceptual response. Alternatively, vision may be studied by recording, analysing and combining regular relations between stimulus and perceptual data, as they are directly established through experimentation, without entering into the features of the underlying processes. We may qualify such alternative perspectives as the “genetic” and the “descriptive” approaches to the study of vision, thus applying a distinction used long ago by Franz Brentano for psychology in general (cf. Mulligan and Smith, 1988). The former is qualified as a genetic approach because it is directed to the reconstruction of the “genesis” or formation processes of perceptual results; the latter is qualified as a descriptive approach because all terms it refers to are observable (through direct perception or by optico-physical measurement) and it aims at an ordered and parsimonious representation of the network of relations connecting such terms.

The ideas we have in mind to expound in this paper are specifically related to the latter perspective, i.e., the descriptive approach in perceptual science. Questions concerning processes in the visual system are virtually extraneous to our analysis (except for one isolated mention in the last but one section). Even so, the set of concepts we find necessary in applying the descriptive approach proves to be quite complex, as different requirements and rules of variation come into play. (i) First of all, a distinction must be drawn between some levels of data, each level being qualified by a certain position in the organization of an observational case (perceptual episode) and by a certain way of empirical assessment. Specifically, we will refer to an S-level (the level of external surfaces, or distant stimuli), which is the set of optically relevant properties of the observed reality, as they may be measured through optico-physical procedures; an I-level (the level of retinal images, or proximal stimuli), which is the set of effects or conditions on the eye, as they may be determined by means of optical measurement or computation; and a P-level (the level of perceptual data), the set of contents in the perceptual scene, as it is directly experienced by the observer. (ii) At each level, distinction must be made between several types of data: at the P-level, for example, one may distinguish between colours, lightnesses, lengths, distances, orientations, etc., which represent separate kinds of possible contents in perceptual scenes. Under certain conditions, in relation to a presumed multiplicity of observational cases, each type of data defines a type of variable (a non-located variable) which may be referred to such multiplicity. Distinction may be made between S-, I- and P-variables, depending on whether the values they take on are data belonging to the S-, I- or P-levels. (iii) Further, we must consider a variety of constraints, which are here conceived as dependence relations acting on definite groups of variables (their fields of action). Constraints may differ from each other because of the number of variables they involve, the types of such variables, the nature of the relations they express, and so on. In particular, a constraint will be referred to as an S-constraint if all variables in its field of action are S-variables, an SP-constraint if it involves both S- and P-variables (but no I-variable), and so on. (iv) Lastly, for any given perceptual problem (applied to some universe of observational cases) a certain collection of constraints may be determined, which is the system of constraints relevant to that problem. This is a most important concept in our theoretical construction, since it
represents a large part of the power of stimulus conditions in specifying a selected perceptual aspect. As a rule, different perceptual problems lead us to consider different systems of constraints presumed to be relevant for them.

Let us illustrate these various concepts with the schematic example in Fig. 1, which refers to the problem of visual recovery of depth, i.e., the problem of perception of distances and relative positions in the third dimension. In the following picture, the observer’s position corresponds to the image plane, which represents the eye (more properly, the retina). Images projected on this plane by objects in the external world, as well as attributes and relations inherent in the images, are elements of the I-level of data. In this example, the observed portion of physical reality (the so-called layout of physical surfaces) consists of three planar objects lying perpendicular to the line of sight: a rectangle on the left, a disc in the centre, and another rectangle on the right. The three physical objects and all their optically relevant properties (physical shape, size, reflectance, distance from the observer and from one another, etc.) are part of the S-level of data. By looking at such a physical scene, the observer will experience a certain perceptual scene, supposedly comprised of three separate units, just like its physical counterpart. The three perceptual units and all their perceptual properties (apparent shape, size, lightness, phenomenal egocentric and exocentric distances, etc.) are elements of the P-level of data. Three simple comments are in order here. First, for the sake of simplicity, the observer is represented in Fig. 1 by a single image plane, but in discussing binocular vision two separate image planes should be considered, representing the left and right ocular retinas. Second, according to the descriptive approach, data at the S- and P-levels refer to one and the same base (the observed piece of reality), the difference being due to the way in which information is obtained: S-data are gained through physical measurement and analysis and P-data by direct visual observation.
Third, again for simplicity, we presume here that there is an exact correspondence between units in the physical and the perceptual scene, but this is not a general rule at all: there are cases in which perceptual organization thoroughly modifies the articulation of the scene (see, for example, the case of phenomenal transparency, discussed in Section 3).

Fig. 1 represents one single observational case, but a perceptual problem makes sense, as a scientific problem, only if it is referred explicitly or implicitly to some homogeneous class of possible observational cases (cf. the notion of a "domain of intended applications of a theory-element" in Balzer et al., 1987, Ch.II). For the example at hand, we might presume that the reference class is a variety of observational contexts similar to the case depicted in Fig. 1, in that each context presents one circular object in the centre and two rectangular objects on both sides, the three objects being perpendicular to the line of sight. The several possible cases in the class may differ, among other things, in the size of each recurrent object and their distances from the observer and from one another. If referred to such a universe of observational cases, each kind of content, at any level of data, takes on the shape of a variable. For example, the physical size of the disc may be different in different observational cases, so that it constitutes a variable on the presumed universe of cases: specifically, an S-variable, since the values it takes on are items at the S-level of data. Similarly, the perceived size of the disc, its apparent distance from the observer, apparent separation in depth from either rectangles, etc., may be qualified as P-variables on the same universe of observational cases: their possible values are elements at the P-level of data. Furthermore, the size of the retinal image of each external object, binocular disparity between retinal images of two external objects, etc., may be referred to as I-variables, since their values are items at the I-level of data.

Now let us consider the physical size of the disc, its physical distance from the observer, and the size of the image it projects on the retina. An elementary optico-geometrical argument shows that these quantities (two S-variables and one I-variable) are linked by a rule: for discs of the same physical size, the diameter of the retinal image is inversely proportional to the physical distance from the observer. This is a constraint, acting on three variables: specifically, it is an SI-constraint, as it involves S- and I-variables, or an optical constraint, being determined through optico-geometrical analysis. Symmetrically, we may consider the retinal size of the disc and its perceived external size and distance, and conjecture the following rule: for discs of the same retinal size, the apparent external size (expressed by the diameter) is directly proportional to the apparent distance from the observer. This too is an example of constraint, which may be named an IP-constraint, as it acts on I- and P-variables, or a psychophysical constraint, since it relates data in the proximal stimulus to data in the perceptual scene. Note that, if analysis is restricted to some subset of observational cases wherein the disc has a constant retinal size, then this rule comes down to a direct relationship between apparent size and apparent distance of the disc, i.e., a percep--percept or P-constraint. Also note that one could directly relate physical size and perceptual size, or physical distance and perceptual distance of the changing disc, thus obtaining SP-constraints. Similar constraints may also be defined for the other recurrent units in the scene (the rectangles). In the same vein, let us consider the physical distances of the rectangles from each other and from the observer, and the binocular disparity of their retinal images. Optical
analysis shows that a certain relationship is at work between these variables (binocular disparity can be computed from distances), which represents a new example of SI-constraint. We may also consider the transposed relationship, which connects binocular disparity with the apparent distances of the rectangles, as they are seen by the observer: this is an example of IP-constraint. Similar constraints may also be specified in comparing either rectangle with the disc. Now, in addressing the problem of perceptual recovery of depth positions, a variety of such relational constraints must be considered, together with other kinds of constraints, which are related to other stimulus cues for distance (convergence, perspective, relative size, etc.). These various constraints, acting on specific subsets of an exhaustive set of variables, make up a system of located constraints.

The example just described is relatively simple when compared with observational situations in real life, and yet it suffices to illustrate the following notable circumstance: when the descriptive approach is applied to perceptual problems endowed with some articulation (say, not problems on isolated stimulus–percept correspondences, as are treated in elementary psychophysics), the components in the observational context that are presumably relevant for a given problem make up a heterogeneous set of relations, which may be represented as a system of constraints acting on a definite set of variables. Besides stimulus–percept or psychophysical relations (SP- and IP-constraints), the system may comprise percept–percept or intra-perceptual relations (P-constraints); besides binary relations (acting on selected pairs of variables), there may be constraints of a higher order (ternary, quaternary, etc.); further, some constraints in the system may be of different types, some others may be of the same type but differ from one another in location (i.e., the subsets of variables on which they act). In the following, we will call this way of representing and addressing perceptual problems the *multiple constraint paradigm* (MCP). Our view is that this paradigm is the natural evolution of the descriptive approach, when it is faced with perceptual problems of some structural complexity.

The proposed name is our own, but the essential features of MCP are distinctly seen in some significant branches of modern perceptual science. For example, the paradigm is implicit in studies on integration of visual modules, concerning how several elementary processes which are relevant for a certain perceptual aspect (e.g., various cues to distance, or various factors for solid shapes) may interact with each other, and combine their effects in an unequivocal and coherent result (cf. Aloimonos and Shulman, 1989; Bulthoff, 1991; Pankanti and Jain, 1995; Uttal et al., 1996). The paradigm is also at the basis of studies on multiple recovery in vision, such as the recovery of “surface shape and orientation from texture”, “surface shape, albedo, and illuminant direction from shading” and “object shape and rotation sense from kinetic depth effect”, in which distinct aspects of the perceptual scene are seen to derive from one set of stimulus data (Witkin, 1981; Todd and Mingolla, 1983; Ramachandran, 1988; Zheng and Chellappa, 1991; Nawrot and Blake, 1993).

MCP is therefore provably and profitably at work in contemporary research on vision, but it lacks, in our view, a general formal expression, i.e., an explicit definition of the concepts which are essential and common to the applications of the paradigm. This impression matches a judgment expressed by Aloimonos and Shulman, who in the
Epilogue to their book on “integration of visual modules” state that “we need a formal theory for combining information from different sources” (Aloimonos and Shulman, 1989, p. 267). The aim of the present paper is to suggest some simple ideas on this issue, i.e., to draw a minimal conceptual frame for MCP, by defining and relating a few concepts which appear to be implicit in all its applications. Identifying and making explicit the basic concepts in a rigorous way is an initial and necessary stage for undertaking formal analysis of the paradigm, its conditions and implications.

The language we adopt for this formalization is elementary set theory, in order to ensure wide generality to the fundamental concepts, and to obtain a system that is flexible and faithfully applicable to various specifications of MCP (more detailed arguments for the set-theoretic approach will be presented in the concluding section). The main concepts are introduced progressively, starting from variables (Section 2), moving to constraints (Section 3), systems of constraints (Section 4), and pairings between systems of constraints and valuations of stimulus variables (Section 5). Besides definitions of these various concepts, some of their immediate facts and relationships will be noted, as the result of an initial inquiry into their formal properties. The paper makes use of some simple examples from perceptual science in order to illustrate the defined concepts, but it offers no fully extended demonstration of the adequacy of the suggested apparatus in dealing with specific research problems on vision. Nevertheless, in Section 6 we will mention and comment on some key concepts of current approaches to the study of vision which may be expressed, linked and clarified by using terms of the suggested conceptual system.

2. Variables and valuations

Intuitively, the concept of a variable refers to some kind of results that are (partially) different or unstable along a series of observations. Equally, the concept of a variable demands some stability along that series: results must be homogeneous to some extent, and must refer to a common and permanent base, so that they may be properly qualified as different values of one and the same variable.

These obvious requirements lead to the first components of our theoretical construction. First, a set \( O \) of possible observational cases is presumed, each case \( o \in O \) being a closed perceptual episode, which comprises data at separate empirical levels (primarily, the S-, I- and P-levels defined in the Introduction). Second, we assume a well-defined set \( U \) of recurrent units (i.e., specific objects or events, or selected parts of objects or events); each unit \( u \in U \) is presumed to occur (at a fixed level of data) in all cases in \( O \), with stable identity but with possibly and partly changing properties. Set \( U \) of recurrent units will serve as the required stable base for variable identification.

In relation to the given set \( O \) of observational cases, a located variable \( v \) is determined by three aspects: type \( X_v \), as a complete set of admissible values for the variable, state-function \( s_v \), as a function from \( O \) to \( X_v \), and location \( l_v \), which is a subset of set \( U \) of recurrent units. Type \( X_v \) is here conceived as an abstract and homogeneous inventory of possible properties (e.g., continuum of possible reflectances, variety of possible lengths, complete set of admissible distances, etc.). For each case \( o \in O \), the
coincidence of the same pairs in the values or, formally:

\[ y \]

the following, and is defined as follows: any two valuations

\[ V \]

and the exhaustive set

\[ (V \times W) \]

\( s \)

can be described as the set of pairs \( W(o) = \{(1, s_1(o)), \ldots, (m, s_m(o))\} \) of the entire set of variables \( V \) is a complete valuation, and the exhaustive set \( \{V(o): o \in O\} \) of such joint states is a subset of \( X \).

Compatibility between valuations, denoted by symbol \( \leftrightarrow \), plays an important role in the following, and is defined as follows: any two valuations \( y, y' \in Y \) are compatible with each other if coincidence of some of their pairs in the coordinates implies coincidence of the same pairs in the values or, formally:

\[ y \leftrightarrow y' \text{ if for all } (i_h, y_h) \in y, (i'_h, y'_h) \in y', i_h = i'_h \text{ implies } y_h = y'_h. \]
These facts are immediate: \( y \subseteq y' \) implies \( y \leftrightarrow y' \) (inclusion between valuations, as sets of pairs, ensures compatibility), if \( M|y \subseteq M|y' \) then \( y \leftrightarrow y' \) implies \( y \subseteq y' \), if \( M|y \cap M|y' = \emptyset \) (the two valuations are separate, i.e., they act on disjunct sets of coordinates) then \( y \leftrightarrow y' \). It is also directly seen that \( y \leftrightarrow y' \) iff \( y \cup y' \subset Y \), i.e., compatibility of two valuations means that their union is itself a regular valuation.

Compatibility is first used in defining some basic operations on sets of valuations. Let \( T, Z \) be subsets of \( Y \). The *restriction* of \( T \) with respect to \( Z \) and the *disjunction* of the two sets are specified by the following formulae:

\[
\begin{align*}
  r(T, Z) &= \{ t \in T : t \leftrightarrow z \text{ for some } z \in Z \} \\
  d(T, Z) &= \{ y \in Y : y = t \cup z \text{ for some } t \in T, z \in Z \} \nonumber
\end{align*}
\]

In other words, restriction \( r(T, Z) \) is the set of valuations in \( T \) that are compatible with at least one valuation in \( Z \), and disjunction \( d(T, Z) \) is the set of unions between compatible valuations in the given sets. If \( Z = \{ y \} \) (conditioning set \( Z \) is a singleton), restriction is denoted simply by \( r(T, y) \), so that:

\[
r(T, y) = \{ t \in T : t \leftrightarrow y \};
\]

accordingly, \( r(T, Z) = \bigcup_{y \in Z} r(T, y) \) for any \( Z \subset Y \). These are two of various immediate properties of the given concepts:

\[
\begin{align*}
  &\text{if } T \subseteq T' \text{ and } y \supseteq y', \text{ then } r(T, y) \subseteq r(T', y') \quad (1) \\
  &\text{if } T \subseteq T' \text{ and } Z \subseteq Z', \text{ then } r(T, Z) \subseteq r(T', Z') \quad (2)
\end{align*}
\]

In addition, for any \( K \subseteq M \) and \( Z \subseteq Y \) such that \( K \subseteq M|Z \) for all \( z \in Z \), the *projection* of set of valuations \( Z \) on set of coordinates \( K \) can be considered, which is denoted by \( p(Z, K) \) and is defined as the set of valuations in \( Y|K \) which are compatible with some valuations in \( Z \), i.e.:

\[
p(Z, K) = \{ y \in Y : M|y = K, y \leftrightarrow z \text{ for some } z \in Z \}. \quad (3)
\]

It is seen that projection, so conceived, is a special case of restriction, in that \( p(Z, K) = r(Y|K, Z) \).

The last definition in this section regards a correspondence between single partial valuations (or sets of partial valuations) and sets of complete valuations. For any \( y \in Y \), the corresponding *extension* \( E(y) \) is the set of complete valuations which are compatible with \( y \) and more generally, for any \( Z \subset Y \), the corresponding extension \( E(Z) \) is the set of complete valuations which are compatible with at least one valuation in \( Z \), so that \( E(Z) = \bigcup_{y \in Z} E(y) \). Alternatively, the two concepts may be expressed by using the restriction operation, specifically:

\[
\begin{align*}
  E(y) &= r(X, y) \\
  E(Z) &= r(X, Z). \quad (4)
\end{align*}
\]

Properties (1) and (2) directly imply that:
if \( y \subseteq y' \), then \( E(y) \subseteq E(y') \)  

(5)

if \( Z \subseteq Z' \), then \( E(Z) \subseteq E(Z') \).  

(6)

It is also easily seen that:

if \( y \rightarrow y' \), then \( E(y \cup y') = E(y) \cap E(y') \)  

(7)

\[ E(Z \cup Z') = E(Z) \cup E(Z') \]  

(8)

\[ E(d(Z, Z')) = E(Z) \cap E(Z') \]  

(9)

Note that, because of the last two equations, the range of mapping \( E \) (as a subset of the power set of \( X \)) is a lattice with respect to operations of union and intersection.

3. Constraints on variables

The formal definition of constraints is given in two consecutive steps: first, we introduce non-located (i.e., types of) constraints, and then describe located constraints (i.e., instantiations of types of constraints). This is because data in observational cases may be subjected to distinct regularities that are of the same abstract type but apply to different sets of variables: such regularities qualify as distinct instantiations of the same non-located constraint.

A non-located constraint is here conceived as a kind of local and conditional regularity: local, in that each of its applications acts on a limited subset of the whole set of variables; and conditional, because its predictions are conditioned by some assumptions concerning the relevant subset of variables. More precisely, a non-located constraint \( \gamma \) refers to an unspecified set of located variables \( W = \{v_i, \ldots, v_j\} \) and is composed of four distinct conditions: \( \mu \gamma \), which specifies the number \( k \) of variables in \( W \); \( \tau \gamma \), which qualifies types \( X_{i_1}, \ldots, X_{i_k} \) of those variables, \( \lambda \gamma \), which concerns their locations \( l_{i_1}, \ldots, l_{i_k} \), and \( \sigma \gamma \), which applies to the values of state-functions \( s_{i_1}, \ldots, s_{i_k} \).

Constraint \( \gamma \) as a conditional law, states that, for any set of variables \( W \subseteq V \), if it complies with conditions \( \mu \gamma, \tau \gamma \) and \( \lambda \gamma \), then the joint state of variables will satisfy condition \( \sigma \gamma \). Thus, constraint \( \gamma \) may be formally expressed as:

\[
(\forall W \subseteq V)(\mu_\gamma(W) \land \tau_\gamma(W) \land \lambda_\gamma(W) \Rightarrow (\forall o \in O)(\sigma_\gamma(s_{i_1}(o), \ldots, s_{i_k}(o))))
\]  

(10)

where \( \{i_1, \ldots, i_k\} = M|W \) is the set of coordinates of variables in \( W \). If \( k \) is the number determined by \( \mu \gamma \), and \( T_{i_1}, \ldots, T_{i_k} \) are the types (as sets of values) presumed by \( \tau \gamma \), then condition \( \sigma \gamma \) specifies the following subset of the product-set of the types:

\[
S_{\gamma} = \{t_{i_1}, \ldots, t_{i_k}\} \in T_{i_1} \times \cdots \times T_{i_k} : \sigma_\gamma(t_{i_1}, \ldots, t_{i_k})
\]  

(11)

so that the last segment of the logical formula may be rewritten as:

\[
(\forall o \in O)\{(s_{i_1}(o), \ldots, s_{i_k}(o)) \in S_{\gamma} \},
\]

which expresses the prediction implied by constraint \( \gamma \).
Located constraints, we said, are instantiations of non-located constraints, related to specific sets of variables. Accordingly, any located constraint \( c \) will be described as a pair \((\gamma, W)\), in which \( \gamma \) is the type of constraint \( c \), and term \( W \) is its location (the field-of-action of \( c \)). It has been pointed out that, for a given universe of observational cases, there may be several instantiations of the same non-located constraint; in the suggested symbolism, those distinct instantiations are described as pairs \((\gamma, W), (\gamma, W')\), ..., with a common first component \( \gamma \) but different second components \( W, W' \), ... representing different subsets of \( V \). A most important concept is given, for any located constraint \( c = (\gamma, W) \), by its extension \( E(c) \), which is defined as the set of all complete valuations which satisfy the predictive condition \( \sigma_{\gamma} \) of type \( \gamma \). In formal terms:

\[
E(c) = \{ x \in X : (x_{i_1}, \ldots, x_{i_k}) \in S_x \},
\]

where \( \{i_1, \ldots, i_k\} = M/W \) and \( S_x \) has been described in Eq. (11). Conditions \( \mu_{\gamma} \), \( \tau_{\gamma} \) and \( \lambda_{\gamma} \) are presumed to be true on \( W \) (they are critical for “locating” or “instantiating” a general constraint), so that if constraint \( \gamma \) expressed by Eq. (10) is a valid empirical law, then we can predict that \( V(o) \in E(c) \) for all \( o \in O \), i.e., \( V(O) \subseteq E(c) \): this is the set-theoretic representation of the predictive power of located constraint \( c \).

Let us present some examples of the use of such concepts.

(i) The following figure illustrates a simple rule of geometric optics: if \( u \) (a unit in the observed reality) is an oblongate object lying perpendicular to the line of sight, then its length \( s \) and distance \( d \) from the nodal point of the eye and the length \( s' \) of its projection \( u' \) on the image plane (representing the retina) are related by the following equation:

\[
\frac{s}{d} = \frac{s'}{d'}
\]

where \( d' \) is the separation of the nodal point from the image plane (see Fig. 2). This optical rule may be described as a constraint \( \gamma \), acting on variables at the S- and I-levels of data (it qualifies as an SI-constraint). Condition \( \mu_{\gamma} \) states that three variables \( s, d, s' \) are involved; condition \( \tau_{\gamma} \) specifies that \( s \) and \( s' \) are lengths and \( d \) is a distance; condition \( \lambda_{\gamma} \) requires that \( u' \) and \( u \) are corresponding units (at the two different levels) and that \( u \) lies perpendicular to the line of sight; predictive condition \( \sigma_{\gamma} \) is expressed by Eq. (12). Furthermore, if term \( u \) is replaced by \( u'' \), denoting the given object as it appears to the observer, and accordingly terms \( s \) and \( d \) are replaced by \( s'' \) and \( d'' \), representing the perceived length and egocentric distance of \( u'' \), then another constraint may be considered: it is a psychophysical constraint \( \delta \), which involves I-variable \( s' \) and P-variables \( s'', d'' \) (it qualifies as an IP-constraint). Whereas premises \( \mu_{\delta} \), \( \tau_{\delta} \) and \( \lambda_{\delta} \) are directly adapted from the corresponding conditions of constraint \( \gamma \), there is dispute in psychophysics concerning the proper expression of predictive condition \( \sigma_{\delta} \). The simplest option is to presume that \( \sigma_{\delta} \) is transcribed from Eq. (12) by replacing S-variables with corresponding P-variables: this defines the basic version of the “size–distance invariance hypothesis”, as a conjectural psychophysical law (cf. Kilpatrick and Ittelson, 1953; Epstein et al., 1961).
(ii) The second example is offered by stereo vision, as it applies to very simple observational contexts, of the kind illustrated in Fig. 3. The terms to be considered are: two linear vertical objects $u_1$ and $u_2$ in the physical scene, their projections $u'_1$, $u'_2$ on the left and $u'_1$, $u'_2$ on the right image planes, egocentric distance $d$ of the nearer object ($u_1$ in this example), separation in depth $e$ between the objects, binocular disparity $b'$ of their retinal images ($b'$ may roughly be identified with the difference between separation of images on the left retina and separation of images on the right retina). Optical analysis shows that, for a good approximation, the following equation holds true:

$$b' = ie/(d^2 + de),$$

where $i$ is interpupillary distance. This equation may be thought of as the $\sigma$-condition of an optical constraint, acting on $S$-variables $d$, $e$ and $I$-variable $b'$; the $\mu$-condition of the constraint specifies that three variables are involved; the $\tau$-condition qualifies variable $d$.
as an egocentric distance, variable $e$ as an exocentric distance, variable $b'$ as a measure of binocular disparity; the $\Lambda$-condition defines correspondences between the units to which such quantities refer. Here too, besides the optical constraint, we might consider the psychophysical constraint which relates binocular disparity $b'$ to perceived egocentric distance $d''$ of the nearer object and perceived exocentric distance $e''$ between the objects, and assume the optical formula as a tentative model for the $\sigma$-condition of the psychophysical constraint. In this way, from Eq. (13), through substitution of terms and inversion, the following formula is obtained:

$$e'' = b'd''^2/(i - b'd'').$$

This formula is mentioned in the literature as a possible rule for perceptual estimation of depth separation between two objects, on the basis of perceived egocentric distance and binocular disparity (cf. Gogel, 1960; Ogle, 1962).

(iii) The third example refers to a rather different perceptual problem, i.e., the emergence of phenomenal achromatic transparency using an episcotister, as shown in Fig. 4. This figure represents an observational situation wherein the physical scene is comprised of a rectangular surface with an incomplete disc in front of it, both lying perpendicular to the line of sight. The rectangular surface is opaque and gray in colour, and divides into two parts $u_1$ and $u_2$ of different gray shades and of correspondingly different luminances, $l_1$ and $l_2$, in the given illumination conditions. The incomplete disc, denoted by $u_3$, is itself opaque and gray — $l_3$ denotes its luminance and $t$ is the measure of its incompleteness, i.e., the areal proportion of the open sector — and is presumed to rotate at fusion speed during observation (it is used as an episcotister). In such stimulus conditions, the retinal image will consist of four parts $u_1'$, $u_2'$, $u_3'$, $u_4'$, whose illuminances $l_1'$, $l_2'$, $l_3'$, $l_4'$ are related to the properties of the observed reality by the following equations, deriving from Talbot-Plateau’s law in optics (cf. Pirenne, 1962):

$$l_1' = l_1, \quad l_2' = tl_1 + (1 - t)l_3,$$
$$l_3' = tl_2 + (1 - t)l_3, \quad l_4' = l_2.$$

Fig. 4. Basic observational context for the episcotister model on phenomenal transparency.
By applying the suggested categories, this system of equations may be conceived as the \( \sigma \)-condition of an optical constraint; the \( \mu \)-condition for the constraint specifies that eight variables are involved, the \( \tau \)-condition qualifies \( l_1, l_2, l_3 \) as luminances, \( l'_1, l'_2, l'_3, l'_4 \) as illuminances and \( t \) as a proportion, the \( \lambda \)-condition characterizes the general architecture of the observational context, such that the relevant variables are properly located. In the presumed situation, the prevailing perceptual result is a scene with a rectangular opaque surface comprised of two parts, \( u'_1 \) and \( u''_1 \), with different values \( l''_1 \) and \( l''_2 \) of phenomenal luminance (brightness); and a stationary, complete, transparent disc \( u''_2 \), which appears to be in front of the rectangular surface, and is qualified by a certain degree \( l''_3 \) of phenomenal luminance and a certain degree \( t'' \) of phenomenal transparency. By substituting these four \( P \)-variables to the corresponding \( S \)-variables in the given optical equations, and solving for unknowns \( t'' \) and \( l''_3 \), the following formulae are obtained:

\[
\begin{align*}
t'' &= \frac{(l'_2 - l'_3)}{(l'_1 - l'_3)} \\
l''_3 &= \frac{(l'_4 - l'_2)}{(l'_1 - l'_4)}
\end{align*}
\]

which correspond to Metelli’s equations on phenomenal transparency (Metelli, 1970, 1974). This pair of formulae may be thought of as the \( \sigma \)-condition of a psychophysical constraint (an \( IP \)-constraint), which is conjectured and abstractly determined on the basis of an optical principle (the \( \sigma \)-condition of an SI-constraint).

The examples just described have been introduced to illustrate the following basic aspects of relational constraints: each located constraint refers to a definite set of variables (its field of action); constraints may be classified, at a primary level, by referring to the gnosiological kind of variables they involve (SI-constraints, \( IP \)-constraints, etc.); each constraint is qualified by a certain predictive condition, which amounts to a relation between possible values of the variables in the field of action; this predicted relation is conditional upon some specific premises concerning the number, type and location of the involved variables, so that a non-located constraint has the general shape of a formal implication. These are the aspects most important for the following analyses, but they certainly do not exhaust the complexity of the concept in question. Among other aspects and problems, let us mention the qualification of constraints according to the formal kind of variables they involve (dichotomous or categorical, possibly expressed by scales at some measurement levels, etc.); algebraic expressions of predictive conditions of constraints (in terms of equations, inequalities, logical formulae, etc.); requirements and procedures for empirical testing of relational constraints, as hypothetical empirical laws; and, last but not least, ways of arriving at conjectural constraints which are potentially relevant for vision. Let us briefly comment on this last issue, specifically referred to the production of conjectural psychophysical constraints (hypothetical \( IP \)-constraints).

In this respect, the given examples have the following in common: they each comprise an optical and a psychophysical constraint, the latter being derived from the former through a sort of transposition (S-variables in the optical constraint are replaced by corresponding \( P \)-variables, the new formula being possibly transformed or “inverted” into an algebraically equivalent expression). This kind of example has been adopted here
basically for the sake of convenience: our purpose was to illustrate concisely both SI- and IP-constraints, using examples in which predictive conditions are expressed by simple equations. But now some qualifications are in order, on this specific point.

First: The method of transposition, as a direct way of producing conjectural psychophysical constraints based on known optical constraints, is not uncommon in visual science (various significant examples, taken from both the psychology of visual perception and computer vision, are analytically discussed in Burigana, 1999). This method may be viewed as part of a more general heuristic attitude in the study of vision, sometimes called “vision is inverse optics” (cf. Poggio et al., 1985). Second: The method of transposition is simply one of a variety of procedures used by researchers for determining conjectural dependence relations involving perceptual variables: psychophysical constraints may also be conjectured or verified by combining or modifying constraints previously derived through transposition, or by applying procedures which are totally independent of optical analogies (see next paragraph). As regards the model proposed here, distinctions relating to the origin of various constraints are not of primary importance: a system of constraints, as defined in the next section, may be heterogeneous in this respect, in that it comprises psychophysical constraints that are conceived or verified by means of different procedures. Third: The method of transposition, which appears so simple in execution – physical variables in an optical formula are substituted by their perceptual counterparts – is in fact a cumbersome and questionable operation, in the scientific view, since it results in transferring a regularity from one scientific level to another scientific level. This transfer operation is not illegitimate in an absolute sense but, in order to be accepted as plausible, it generally requires strong assumptions concerning structural similarities and correspondences of data between the two scientific levels (an analysis of assumptions implicit in some applications of the method is made in Burigana, 1999, Ch.3).

We conclude this discussion by referring to one alternative way of establishing constraints involving perceptual variables, which is quite different from the transposition method, and is well illustrated by Lukas (1987). The problem in this application regards perceptual constancy of size, and refers to observational contexts of the kind we considered in example (i) concerning the size–distance invariance hypothesis. Specifically, it is presumed that a set \( O = \{o_1, \ldots, o_n\} \) of observational cases similar to that in Fig. 2 is given, and that each case \( o_j \) (for \( j = 1, \ldots, n \)) may be described as a pair \((d_j, s'_j)\), \(d_j\) being the egocentric physical distance of the oblongate object in the scene and \(s'_j\) the visual angle corresponding to that object (the visual angle is closely related to the length of the retinal projection). Further, it is presumed that a binary relation \( R \) is determined on set \( O \) through psychological experimentation, where \( o_j R o_l \) (for \( 1 \leq j, l \leq n \)) means that the object in case \( o_j \) perceptually appears to the observer at least as long as the object in case \( o_l \). Then, for empirical relational system \((O, R)\), a possible “additive model of size constancy” is considered, which formally consists of three positive real-valued functions \( f, g, h \), so that: \( o_j R o_l \) iff \( h(o_j) \geq h(o_l) \), \( h(o_j) = g(d_j)f(s'_j) \) for all \( 1 \leq j, l \leq n \). Reference is made to the theory of “additive conjoint measurement”, which defines the conditions which empirical relation \( R \) must satisfy so that the “additive model of size constancy” may be applied, i.e., empirical system \((O, R)\) may be numerically represented by measurement scales \( f, g, h \) in the stated way.
There are considerable differences between this approach, which is inspired by the representational measurement theory, and the traditional way of discussing the relation between perceptual size and perceptual distance, which we applied in example (i). These differences regard the empirical base (given by relation \( R \) of perceptual comparison, in this alternative approach), numerical expression of data (according to measurement scales \( f, g, h \), in the new approach), conditions for validity of the model (i.e., assumptions for the legitimacy of a transposition, or formal requirements of relation \( R \)), possibilities and procedures for empirical testing of the model, and so on. We point out, in particular, that if functions \( h \) and \( g \) are interpreted as measures of the perceptual size and perceptual distance of objects in the observational cases, then equation \( h(o_j) = g(d_j)f(s_j) \) qualifying the “additive model of size constancy” is structurally similar to equation \( s'' = d''s'd' \) expressing the size–distance invariance hypothesis. This is a constraint involving perceptual variables which, in the way used by Lukas (1987), is independently obtained as the solution of a representational measurement problem, not as the result of a copying operation on an optical regularity, i.e., not as the result of a transposition procedure.

4. Systems of constraints

In the Introduction, the following circumstance was emphasized: in addressing perceptual problems on articulated observational contexts, several dependence relations must be considered, as a rule, which are potentially relevant for their solutions. This notion is now formalized into the concept of a system of constraints, defined as a set of located constraints and abstractly described by the following equation:

\[
C = \{c_1, \ldots, c_n\} = \{(\gamma_1, W_1), \ldots, (\gamma_n, W_n)\}.
\]

In other words, a system of constraints \( C \) is a set of instantiations of non-located constraints \( \gamma_1, \ldots, \gamma_n \) on specific locations \( W_1, \ldots, W_n \) within the presumed exhaustive set \( V \) of located variables. We emphasize again that system \( C \) may comprise constraints of various kinds (SI-, IP-, P-constraints, etc.), as well as separate instantiations of one type on distinct locations in \( V \) (this means that there may be repetitions along the list \( \gamma_1, \ldots, \gamma_n \) of the types of constraints constituting \( C \)). System \( C \) may be depicted as a hypergraph on \( m \) nodes with \( n \) hyperedges (Berge, 1973): the nodes correspond to the variables in \( V \) and the hyperedges represent the constraints in \( C \), as relations on definite subsets of variables. This representation resembles that used in AI to describe a “network of constraints” (cf. Montanari and Rossi, 1991; Gyssens et al., 1994). In fact, our definition of a system of constraints, specifically referred to perceptual problems, is abstractly similar to that of a network of constraints, as it is applied to a variety of computational problems in AI (belief maintenance, scheduling, logic programming, temporal reasoning, etc.; cf. Dechter, 1992; Kumar, 1992; Tsang, 1993).

The concept will be illustrated by referring to the example given in the Introduction (Fig. 1), concerning perceptual reconstruction of the third dimension. We proceed in an orderly way by identifying units, variables and constraints, in three consecutive steps.
**Units**: Let $u_1$, $u_2$, $u_3$ denote the rectangle on the left, the circle in the centre and the rectangle on the right, as items at the S-level of data; similarly, let $u_1'$, $u_2'$, $u_3'$ and $u_1''$, $u_2''$, $u_3''$ denote the corresponding units at the I- and P-levels of data. We have already noted that, in the special kind of observational cases of the example, exact correspondence is presumed between the units at the three levels of data, i.e., there is no distortion concerning the number or identity of the units. **Variables**: At the S-level, let $s_1$, $s_2$, $s_3$ be the geometrical sizes of objects $u_1$, $u_2$, $u_3$, and $d_1$, $d_2$, $d_3$ their physical distances from the observer. At the I-level, for $1 \leq i \neq j \leq 3$, let $s_i'$ be the size of projected unit $u_i'$ and $d_{ij}'$ the binocular disparity associated to projected units $u_i'$ and $u_j'$. At the P-level, let $s_i''$, $s_j''$, $d_{ij}''$ be the phenomenal sizes and $d_{ij}''$ the phenomenal egocentric distances of units $u_i''$, $u_j''$, $u_3''$. **Constraints**: For the presumed kind of observational cases, both a set of SI-constraints and a set of IP-constraints may be considered. The former, for $1 \leq i \neq j \leq 3$, includes an optical constraint $\alpha(s_i', d_i', s_j')$ which relates the size of I-unit $u_i'$ to the size and distance of the corresponding S-unit $u_i$, and an optical constraint $\beta(d_i', d_j', b_{ij}')$ which derives the binocular disparity between I-units $u_i'$ and $u_j'$ from the distances of S-units $u_i$, $u_j$. The latter, for $1 \leq i \neq j \leq 3$, includes a psychophysical constraint $\epsilon(s_i', s_j'', d_{ij}''')$ (a version of the size–distance invariance hypothesis) which connects the phenomenal size and phenomenal egocentric distance of P-unit $u_i''$ to the size of the associated I-unit $u_i'$, and a psychophysical constraint $\zeta(b_{ij}', d_{ij}'''', d_{ij}''')$ which connects the phenomenal egocentric distances of P-units $u_i''$, $u_j''$ to the binocular disparity of the corresponding I-units $u_i'$, $u_j'$. In addition, a P-constraint $\eta(s_i'', s_j''')$ may be conjectured: it is promoted by the equality in shape between units $u_i''$, $u_j''$, and suggests that they should also be equal in size (this corresponds to the “relative size cue” for perceptual distance; cf. Gogel and Sturm, 1971). Overall, by representing the variables as nodes and the constraints as hyperedges, the hypergraph in Fig. 5 is obtained. Note that there is some inconsistency in this system of constraints, as

![Fig. 5. Hypergraph of a system of constraints.](image)
constraint $g(b_1, d_n^1, d_n^2)$ supports $d_n^1 > d_n^2$, while constraint $h(s_n, s'_n)$, combined with $e(s_2, s'_2, d_4^n)$ and $e(s_3, s'_3, d_5^n)$, favours $d_4^n < d_5^n$.

Coming back to general aspects, we note that the overall predictive power of a system of constraints $C = \{c_1, \ldots, c_n\}$ may be adequately represented in extension $E(C)$, which is the set of complete valuations that satisfy the predictions of all constraints in $C$, i.e.:

$$E(C) = \{x \in X : x \in E(c_1), \ldots, x \in E(c_n)\}$$

$$= E(c_1) \cap \cdots \cap E(c_n).$$

This is exactly the analogue of the set of “admissible solutions” for a “constraint satisfaction problem” (Mackworth, 1992), and of the set of “consistent labelings” for a “relaxation labeling problem” (Hummel and Zucker, 1983). Since $C$ is defined as a set of located constraints, one presumes that, for each $c_j = (\gamma_j, W_j) \in C$, premises $\mu_j$, $\tau_j$ and $\lambda_j$ of type $\gamma_j$ are satisfied by $W_j$ on each observational case $o \in O$; if (as expected) the same is true for predictive conditions $\sigma_j$ $(j = 1, \ldots, n)$, then we should have $V(o) \in E(C)$ for all $o \in O$, i.e.:

$$V(O) \subseteq E(C).$$

(14)

In other words, the set of joint values taken on by variables in $V$ on cases in $O$ should be included in the extension of the system of constraints.

This is the formal expectation of the model, which in practice may prove not to be exactly true. In fact, constraints are here conceived as kinds of scientific hypotheses: they are conjectural predictive regularities on perceptual processes, to be tested and refined through experimentation. What is more, constraints are regularities which apply locally to restricted subsets of the general set of variables, and to selected portions of the spatio-temporal extent of the observational cases. Thus, it may happen that each constraint in $C$, in isolation, is a plausible rule, but system $C$ as a whole is not consistent (in this respect, constraints are like the local principles of geometric coherence in impossible figures; cf. Sugihara, 1982; Kulpa, 1987). For these reasons, the prediction of Eq. (14) may not be confirmed. It is possible that $E(C) \neq \emptyset$ but $V(o) \not\subseteq E(C)$, in which case system $C$ would prove to be formally consistent (there is some complete valuation which complies with all constraints in $C$), but some observational cases disagree with some predictions of constraints in $C$. More strongly, it is possible that $E(C) = \emptyset$ (which obviously implies $V(o) \not\subseteq E(C)$), in which case system $C$ would be formally inconsistent, i.e., at some points some predictive conditions of constraints in $C$ conflict with one other, so that no complete valuation can obey all the constraints in the system.

The latter possibility is significant for theoretical analysis, since it characterizes the “cue conflict paradigm” frequently used in perceptual research. This issue will be examined more closely in Section 6. In cases of inconsistency of system $C$, consistency may be restored by ignoring some of its constraints; this means that only selected subsystems of $C$ are considered. The predictive power of any subsystem $D = \{c_{j_1}, \ldots, c_{j_p}\} \subseteq C$ is expressed by its extension:

$$E(D) = E(c_{j_1}) \cap \ldots \cap E(c_{j_p}),$$

(15)
which is the set of complete valuations complying with all constraints in \( D \). It is immediately seen that, for all \( D, D' \subseteq C \):

\[
\text{if } D \subseteq D', \text{ then } E(D) \supseteq E(D') \tag{16}
\]

\[
E(D \cup D') = E(D) \cap E(D'). \tag{17}
\]

Consistent subsystems are subsets of \( C \) with non-empty extensions; Eq. (16) shows that the set of consistent subsystems is an order ideal in the power-set of system \( C \).

5. Pairing of constraints and valuations

The specific phenomenal scene an observer perceives in a given observational situation may be interpreted as the result of two complementary classes of factors: a network of relational rules, which instantiate some general principles of conditioning and interaction, and a set of input data, which are offered by and strictly contingent upon that particular situation. In our theoretical frame, the former class of factors are represented as an extensive system \( C \) of located constraints, while the latter class may be described as a (partial) valuation \( y \) which specifies the values taken on by some set of S- and/or I-variables (stimulus variables) in the presumed observational situation. On a similar basis, Clark and Yuille (1990) distinguish between “a priori (scene independent) constraints” and “scene dependent sources of information” (p.158). We note, however, that in our view located constraints are themselves “scene dependent” to some extent, in that each constraint \( c_j = (\gamma_j, W_j) \) acts on a definite set \( W_j \subseteq V \) of located variables, which is supposed to satisfy premises \( \mu_j, \tau_j \) and \( \lambda_j \) of type \( \gamma_j \).

Thus, we are led to consider any pairing \((C, y)\) between the presumed system of constraints \( C \) and some valuation \( y \) on a set of S- and/or I-variables. More generally, for the needs of formal analysis, we may refer to any pairing \((D, Z)\) between some subsystem of constraints \( D \subseteq C \) and some subset of valuations \( Z \subseteq Y \). To each pairing \((D, y)\), with \( D \subseteq C \) and \( y \in Y \), an extension \( E(D, y) \) is associated, which is the set of complete valuations that are compatible with valuation \( y \) and comply with all constraints in \( D \). In other words, the extension of a pairing is the intersection of the extensions of its components (as they are defined by (4) and (15)), i.e.:

\[
E(D, y) = E(D) \cap E(y). \tag{18}
\]

More generally, for each \( D \subseteq C \) and \( Z \subseteq Y \), we define:

\[
E(D, Z) = E(D) \cap E(Z), \tag{19}
\]

i.e., the extension of pairing \((D,Z)\) is the set of complete valuations which are compatible with some valuations in \( Z \) and comply with all constraints in \( D \); equivalently:

\[
E(D, Z) = \bigcup_{y \in Z} E(D, y).
\]

On an abstract level, Eq. (18) describes mapping of \( 2^C \times Y \) in \( 2^X \); similarly (19) specifies
mapping of \( 2^C \times 2^Y \) in \( 2^X \) \( (2^C, 2^Y \) and \( 2^X \) are the power-sets of sets \( C, Y \) and \( X \)). The following formulae show some elementary properties of these mappings:

\[
\text{if } D \subseteq D' \text{ and } y \subseteq y', \text{ then } E(D, y) \supseteq E(D', y') \tag{20}
\]

\[
\text{if } y \leftrightarrow y', \text{ then } E(D \cup D', y \cup y') = E(D, y) \cap E(D', y') \tag{21}
\]

in view of (5), (16), (7) and (17). In addition:

\[
\text{if } D \subseteq D' \text{ and } Z \supseteq Z', \text{ then } E(D, Z) \supseteq E(D', Z')
\]

\[
E(D \cup D', Z) = E(D, Z) \cap E(D', Z) = E(D', E(D, Z)) \tag{22}
\]

\[
E(D, Z \cup Z') = E(D, Z) \cup E(D, Z')
\]

\[
E(D, d(Z, Z')) = E(D, Z) \cap E(D, Z') \tag{23}
\]

because of (6), (16), (17), (8) and (9). Eqs. (21), (22) and (23) show that the ranges of mappings (18) and (19) within power-set \( 2^X \) are closed to intersection (i.e., they are closure systems), provided that \( E(C) \) is a proper subset of \( X \).

The consistency of a system of constraints may now be redefined in a relative form with respect to a partial valuation. That is, any subsystem of constraints \( D \subseteq C \) is said to be \textit{consistent} with a given valuation \( y \in Y \) if \( E(D, y) \neq \emptyset \), i.e., when there is at least one complete valuation that includes \( y \) and complies with all constraints in \( D \). In view of remarks in the previous paragraphs about the joint action of a system of constraints and a set of input data in the construction of a perceptual scene, it is the relative concept of consistency which really matters in formal analysis. Any subsystem of constraints is absolutely inconsistent (in the sense defined in Section 4) if and only if it is inconsistent with respect to all valuations in \( Y \). Moreover, it may be directly inferred from Eq. (20) that, for all \( D, D' \subseteq C \) and \( y, y' \in Y \):

\[
\text{if } D \supseteq D', \text{ } y \supseteq y', \text{ and } D \text{ is consistent with } y, \text{ then } D' \text{ is consistent with } y'.
\]

This property means that, for any fixed valuation \( y \), the set of all subsystems of constraints consistent with \( y \) is an order ideal in power-set \( 2^C \), the order ideal being completely determined by its maximal members, which are the \textit{maximal consistent subsystems} with respect to valuation \( y \). If the complete system \( C \) is inconsistent with a given set of input data \( y \), then each maximal consistent subsystem represents the result of a minimal reduction of \( C \), such that consistency with \( y \) is achieved. Such subsystems have the following properties, which result directly from Eq. (21):

\[
\text{if } D \text{ is a maximal consistent subsystem with respect to } y \text{ and } c \in C - D, \text{ then } E(c) \cap E(D, y) = \emptyset;
\]

\[
\text{if } D \neq D' \text{ are maximal consistent subsystems with respect to } y, \text{ then } E(D, y) \cap E(D', y) = \emptyset. \tag{24}
\]

For any pairing \((D, y)\), if \( D \) is consistent with \( y \), we may consider the set of
coordinates whose values are uniquely determined by the joint action of $D$ and $y$. This set will be denoted as $F(D,y)$ and is formally specified by the following equation:

$$F(D, y) = \{ i \in M : |p(E(D, y), i) = 1 \},$$

where $p$ is the projection operation given in (3). The definition itself implies that, for each consistent pairing $(D, y)$:

$$F(D, y) \supseteq M[y].$$

Property (20) also ensures that, for all consistent pairings $(D, y), (D', y')$:

if $D \subseteq D'$ and $y \subseteq y'$, then $F(D, y) \subseteq F(D', y').$

Let us assume that valuation $y$ describes a set of input data (i.e., an assignment of values to some stimulus variables), which is the most natural presumption in our perspective. Each P-variable in the set of variables $V\backslash F(D, y)$ thus represents a perceptual aspect unambiguously determined or “specified” by input data in $y$ combined with constraints in $D$, while any P-variable in $V-V\backslash F(D, y)$ preserves some uncertainty, in that it admits two or more distinct values in spite of the joint constraining action of $y$ and $D$. Of course, a perfect model would consist of a system of constraints $C$ which is absolutely consistent and such that $F(C, y) = M$ for all valuations $y$ on a suitable set of independent stimulus variables.

6. Constraints in perceptual science

In the next paragraphs we will discuss in greater detail the plausibility and possible use of the multiple constraint paradigm (MCP) in the theory of visual perception. This section of the paper will thus resume and work out the suggestions given in the Introduction, in which MCP was qualified as a possible way of conceptualizing the complex state of affairs in the study of visual phenomena.

We started by noting that the psychophysical correspondence between stimuli and percepts most often proves to be a relationship of very high complexity. Research on vision has explored not only the dependence of a single perceptual content on several stimulus components (e.g., multiple “factors” for perceptual grouping, various “cues” for perceptual distance and depth, etc.), but also the joint dependence of distinct perceptual contents on the same stimulus component or set of stimulus components (e.g., visual reconstruction of “surface shape and orientation from texture”, “object shape and rotation sense from kinetic depth effect”, etc.). In general, the stimulus–percept relationship looks like a highly complex many-to-many correspondence. Now, MCP seems suitable to represent such a complex situation and to support its analysis. The stimulus–percept relationship may be modelled as a pairing $(C, W)$, where $C$ is a system of psychophysical constraints (SP- and/or IP-constraints) and $W$ is a selected set of stimulus variables (S- and/or I-variables): the many-to-many complexity of the relationship would be preserved, in that a single perceptual variable could be related through distinct constraints to various stimulus variables, and conversely for stimulus...
variables. For any observational case \( o \in O \), pairing \((C, W)\) is a consistent model if \( E(C, W(o)) \neq \emptyset \), a sufficient model if \( F(C, W(o)) = M \), and a faithful model if \( E(C, W(o)) = \{ V(o) \} \); in other words, a model is faithful for a certain observational case if it unambiguously predicts the values actually taken on by all variables (in particular, perceptual variables) in that specific case.

Reference to MCP also may help in explaining some terms of the multi-factor approach to vision, such as the “mutual disambiguation” of stimulus cues (Bülthoff and Mallot, 1990; Landy et al., 1995), “over-specification” and “under-specification” of a perceptual content (Massaro, 1988). For example, if variables \( v_i \) and \( v_r \) represent stimulus aspects which separately influence some \( P\)-variable \( v_r \), then the mutual disambiguation of \( v_i \) and \( v_r \) in a given observational case \( o \) may mean the following: projection \( p(E(C, \{ i', s_r(o) \}, \{ i'^n, s_r(o) \})), i \) (i.e., the set of values admissible for \( v_i \), that is defined by the assignment of values \( s_r(o) \), \( s_r(o) \) to variables \( v_j \), \( v_r \), and by constraints in \( C \)) is smaller than projections \( p(E(C, \{ i', s_r(o) \}), i) \) and \( p(E(C, \{ i'^n, s_r(o) \}), i) \), Further, if \( v_i \) is a specific perceptual variable and \( W = \{ v_{i1}, \ldots, v_{in} \} \) is a set of stimulus variables, then to state that \( W \) under-specifies \( v_i \) in some observational case \( o \) means that \( |p(E(C, W(o), i))| \geq 2 \) (i.e., input data in \( W(o) \) together with constraints in \( C \) are not strong enough to reduce the set of values admissible for \( v_i \) to a singleton, so that model \((C, W(o))\) is not sufficient relative to variable \( v_i \)). Similarly, to state that \( W \) over-specifies \( v_i \) in \( o \) means that there are proper subsets \( W' \neq W'' \) of \( W \) so that \( |p(E(C, W'(o)), i)| = 1 = |p(E(C, W''(o)), i)| \) (i.e., there are limited parts of \( W(o) \) that suffice to specify the value of \( v_i \), by means of constraints in \( C \)).

Theories on visual perception may include process-like implications, in which some steps are supposed to precede other steps, or some factors are presumed to act on situations prepared by other factors. This dynamic (or genetic) side of perceptual theorization clearly appears in the use of expressions such as “propagation of disambiguation” (Marr, 1982, p.141), “cue promotion” (in the so-called “modified weak fusion model”; cf. Landy et al., 1995), “perceptual causation and interaction” (Gogel, 1973; Epstein, 1982), and so on. MCP is generally compatible with such dynamic implications of perceptual science: in explaining the composition of a perceptual scene, some priorities may be hypothesized between multiple constraints, i.e., some constraints may be presumed to act before and prepare the action of others. Eq. (22) exemplifies a type of formal result which may be useful in this perspective. It implies, by iteration, that \( E(D, y) = E(c_{j_p}, E(c_{j_{p-1}}, \ldots, E(c_{j_1}, y) \ldots)) \) for any set of constraints \( D = \{ c_{j_1}, \ldots, c_{j_p} \} \subseteq C \) and any valuation \( y \in Y \), i.e., the overall predictive power of a set of constraints \( D \) may be expressed progressively, through successive exploitation of the single constraints making up the set.

We considered above the notion of over-specification, describing the possibility that two different subsets of stimulus data may be distinctly sufficient to determine a certain perceptual aspect, through the action of suitable constraints in the given system. For ordinary visual conditions, we may safely presume that the values supported for the same perceptual aspect by distinct stimulus data most times are in substantial agreement. But in experimental research on vision the following method is frequently used: by means of special devices, the experimenter sets up a conflictual observational context in which two distinct parts of the set of input data, both relevant for one selected perceptual
aspect, support two quite different values for that aspect. This is sometimes called the cue conflict paradigm (Blake et al., 1993; Buckley and Frisby, 1993). The method serves at least two important scientific purposes: first, it is used to estimate the relative strength of different stimulus factors for a fixed perceptual aspect; second, it allows us to explore the rules of combination or interaction of the effects of those different factors. The cue conflict paradigm has been widely used in recent research on perceptual recovery of solid shapes and distances from multiple cues to depth; in particular, it is the basis of the so-called “perturbation analysis” applied to the study of such perceptual phenomena (Maloney and Landy, 1989; Young et al., 1993; Landy et al., 1995). But this paradigm has a very long tradition in perceptual science: it is found, for example, in some of the figural cases discussed by Wertheimer (1923), in which the strength of different factors of perceptual grouping is comparatively estimated by orienting their effects in opposite directions.

Within MCP, a conflictual visual context may be described as an observational case $o$, such that the system of constraints $C$ acting in it is not consistent with valuation $W(o)$ of a selected set $W \subseteq V$ of stimulus variables. The inconsistency is due to the fact that distinct parts $W'(o), W''(o)$ of $W(o)$ specify different values for a fixed $P$-variable $v \in V$ by means of constraints in $C$, so that no complete valuation on $V$ can include $W(o)$ and at the same time comply with all constraints in $C$. By pursuing the set-theoretic approach, we may conjecture that, in a conflictual case $o$, the perceptual system will ignore some constraints in $C$ (and the data in $W(o)$ which are involved only in those constraints), and will operate on some subsystem $D \subseteq C$ which is maximal consistent with $W(o)$. Because of Eq. (24), experimentation should make it possible to identify the maximal consistent subsystem preferred by the visual system, and thus to evaluate the relative importance of the various constraints in $C$ and input data in $W(o)$.

This refinement of MCP is closely connected with the set-theoretic approach, in that the additional assumptions deal with operations of a selective type: the selection of a maximal consistent subsystem of constraints, and possibly also that of a complete valuation in its extension (if the chosen subsystem is not formally sufficient). Similar ideas may be found in some contemporary theoretical assumptions: for example, the “vetoing hypothesis” or the “winner-take-all principle” as applied to vision with conflictual stimulus data (Bülthoff and Mallot, 1988; Clark and Yuille, 1990, §3.3; Johnston et al., 1993). There are other plausible ways of theorizing about perceptual processing in conflictual contexts. One may conjecture, for example, some “combination rule” (in the form of a weighted sum, an average, a multiplication, etc.) which uniquely determines the value of a perceptual variable by combining different values supported for that variable by separate parts of the input data (cf. Bruno and Cutting, 1988; Maloney and Landy, 1989; Young et al., 1993). But these developments, which involve operations on numerical values, are not typical of a pure set-theoretic approach.

Lastly, we must point out the significance of MCP in analysing multiple visual illusions, i.e., perceptual phenomena in which various contents in the perceptual scene are altered with respect to the corresponding physical properties of the observed reality. Examples are the celebrated artifacts of Adelbert Ames (distorted room, rotating trapezoidal window, distorted chair, etc.; cf. Ittelson, 1952), the so-called “false-depth parallax anomalies” or “motion parallax illusions” (Rock, 1983; Gogel, 1982, 1990),
the joint variability of several perceptual features in a bistable figure (Necker’s cube, Rubin’s examples on figure–ground inversion, etc.; cf. Attneave, 1971; Rock et al., 1994). In a multiple illusion, distinct perceptual errors appear to be mutually dependent, according to principles of complementarity: for a fixed observed reality, errors concerning some perceptual properties must be compensated by errors on other perceptual properties, so that the perceptual scene as a whole is compatible with the given set of stimulus conditions. Now, such a situation may be represented as a network of P-constraints deriving from proper SP- or IP-constraints through sectioning operations: a psychophysical constraint which involves two or more perceptual variables gives rise to regular dependence at the perceptual level, if the stimulus data are fixed and variations are induced in some of the involved perceptual variables. Multiple illusions are commonly regarded as highly significant situations in the study of the rules of perceptual organization.

7. Epilogue: reasons for the set-theoretic approach

The framework presented in this paper is set-theoretic in nature, in that its formal components are solely grounded in concepts from set theory. What are the reasons for this choice? What are the expected benefits of the set-theoretic approach in treating problems in the science of vision? Comments in this last section are intended to provide some answers to these questions.

We conceive set theory as a basic language particularly suitable for defining plain, flexible scientific models, i.e., models with a minimum of requirements of the real contexts or problems to which they may be applied. Contemporary investigation on the interaction of factors in human and machine vision resorts to various formal and computational paradigms (fusion models, regularization theory, theory of Markov fields, etc.) which, in some applications, result in highly complex models with selective requirements. Instead, we need a theoretical framework with a minimum of primitive terms and assumptions. The first reason is that we intend to show that a variety of important concepts currently used by researchers on vision may in fact be made explicit with a relatively simple set of general prerequisites. This demonstration was done in the previous section. We do not mean, of course, that all formal developments in modern visual science may be grounded on a purely set-theoretic platform (we have just mentioned, for example, the elementary concept of a “combination rule” relating to the cue conflict problem, which cannot be directly located within the proposed framework). What is meant here is that a relatively simple network of assumptions is sufficient to support the rigorous definition of some crucial and representative concepts in modern visual science.

The second reason has to do with the possible heterogeneity of terms (factors, variables) and rules (relational constraints) which may be involved in the analysis of a certain perceptual result. Such a circumstance, not uncommon in the study of vision, is typically illustrated by multi-factor models of perceptual reconstruction of the depth dimension (visual recovery of distances in depth, solid shapes, etc.). A model of depth perception generally considers a variety of factors (cues to depth, such as ocular
accommodation, convergence, binocular disparity, occlusion, motion parallax, etc.), each factor being qualified by a certain mode of action in conditioning the perceptual result. But there are, in general, conspicuous qualitative differences between factors (some are optico-geometric quantities, others relational conditions, yet others dichotomous variables, etc.), between corresponding rules of action (some are expressed by equations, others may be represented as logical formulae) and between the kinds of information provided by the factors (concerning distance from the subject, distance between objects, or simply the order in the third dimension; cf. Cutting and Vishton, 1995). In dealing with a composite problem of this sort, it is in fact impossible to represent the whole structure in terms of a specialized calculus (e.g., as a system of numerical equations on variables which are all quantitative, or as a system of logical formulae on variables which are all dichotomous, etc.). So we are forced to look for a comprehensive basic calculus constituting a common platform for all the various terms in the problem: that calculus is here identified with set theory. Based on this choice, the contribution of each factor is represented as a subset of a common space (i.e., an extension in the universe of complete valuations; cf. Sections 2 and 3), and interactions between factors are described and analysed in terms of relations and operations (inclusion, intersection, restriction, etc.) between such subsets.

Somehow related to this argument is a property inherent in the concept of a pairing \((C, W(o))\) between a system of constraint \(C\) and valuation \(W(o)\) of a set of stimulus variables \(W\) in an observational case \(o\) (cf. Sections 5 and 6). In a sense, both \(C\) and \(W(o)\) are sources of information which take part in determining the perceptual result: the former is a normative source (a system of rules), the latter a source strictly contingent upon the presumed observational case. Within the suggested framework, the informational contributions of both sources are represented in a uniform way, as subsets \(E(C)\) and \(E(W(o))\) of universe \(X\) of complete valuations, and their combined effect is described as the intersection \(E(C, W(o)) = E(C) \cap E(W(o))\) of the two subsets.

One further perspective for an appreciation of the suggested set-theoretic approach concerns its possible analysis and development, as a mathematical construction. There are in the model some elements of structural substance which allow us to consider such a possibility. The model basically concerns systems of relations (constraints), acting locally on a large set of dimensions (variables); the relations are represented as subsets of the direct product of all dimensions, and may be subjected not only to ordinary Boolean operations (union, intersection, etc.) but also to the more specialized operations (restriction, disjunction, projection, etc.) defined in Section 2. Properties and implications of such a composite system of operations constitute the matter for a possible formal investigation. In particular, systems of relations are here represented as families of subsets of a common general space, and this fact suggests the opportunity of exploring the algebraic properties of such set-theoretic structures. Some very short steps in this direction are made in Sections 2, 4 and 5, where we noted that certain families of sets defined there qualify as “lattices”, “closure systems” or “order ideals”. These are small particles of an abstract algebraic investigation, which seems to us worthy of being carried out.

The last comment regards prospects for effective computational use of the suggested framework. In applying it to specific problems in visual science (reconstruction of depth
dimension, grouping in the scene, organization of light and colours, etc.), the ultimate goal is to determine a valuation of the set of perceptual variables which is compatible with both the given valuation of the set of stimulus variables and the system of constraints presumed to be at work in the observational context. Substantially, this is a goal of scientific prediction and resembles the “constraint satisfaction problem” considered in the field of artificial intelligence: given a “network of constraints” on a set of variables, find an assignment of values to the variables so that it complies with all constraints in the network. Referring to various specifications of such a general problem, researchers in information sciences and artificial intelligence have developed a variety of algorithms and computer programs for finding single solutions or sets of alternative solutions (cf. Tsang, 1993). In principle, these computational resources are also available for dealing with problems of integration of visual factors, provided that they may be formulated in the suggested set-theoretic terms.

Acknowledgements

The author wishes to express his sincere thanks to one anonymous referee, for constructive criticisms and useful suggestions on previous versions of the paper. He also thanks Ms Gabriel Walton, for revision of the English text.

References


