



Basic features in vision. An analysis of constraints on perceptually relevant sensory properties.

Dario Taraborelli

► To cite this version:

Dario Taraborelli. Basic features in vision. An analysis of constraints on perceptually relevant sensory properties.. Philosophy. Ecole des Hautes Etudes en Sciences Sociales (EHESS), 2005. English. NNT : . tel-00350276

HAL Id: tel-00350276

<https://theses.hal.science/tel-00350276>

Submitted on 6 Jan 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dario Taraborelli

BASIC FEATURES IN VISION

*An analysis of constraints on
perceptually relevant sensory properties*

Thesis submitted for the degree of Doctor of Philosophy
Specialty: COGNITIVE SCIENCE

Supervisor:

Roberto CASATI, Research Director, CNRS

Jury:

Jérôme DOKIC, Research Director, EHESS

Pierre JACOB, Research Director, CNRS

Pascal MAMASSIAN, Research Fellow, CNRS

Zenon PYLYSHYN, Professor, Rutgers University

ÉCOLE DES HAUTES ÉTUDES EN SCIENCES SOCIALES

October 6, 2005

BASIC FEATURES IN VISION
Dario Taraborelli

D. Taraborelli (2005) *Basic features in vision: An analysis of constraints on perceptually relevant sensory properties*, PhD dissertation, École des Hautes Études en Sciences Sociales, Paris, 2005

Last revision: October 15, 2005

An electronic reprint of this document is available at:

<http://dartar.free.fr/features>

Contents

PREFACE: IN QUEST OF BASIC FEATURES	15
 PART I INTERNAL CONSTRAINTS	 23
1 PRIMITIVENESS	29
1.1 Compositionality	31
1.2 Unanalyzability	38
1.3 Simplicity	45
1.4 Earliness	53
1.4.1 Attention, visual search and feature conjunction	54
1.4.2 Wolfe's test for feature basicness	59
1.4.3 Early selection vs. late selection theories	61
1.4.4 Conclusions	68
1.5 Low-levelness	69
1.5.1 Low-level vision	70
1.5.2 A short history of the notion of feature detector	71
1.5.3 Beyond feature detectors: multi-dimensional tuning and overlap- ping maps	75
1.5.4 Wiring and hierarchies	78
1.5.5 The perceptual role of neural activity in low-level vision	79
1.5.6 Conclusions	85
1.6 Conclusions: distinct criteria for primitiveness	86
 2 ASCRIBABILITY	 87
2.1 Locality	88
2.1.1 Intrinsic locality	91
2.1.2 Functional locality	93
2.1.3 Non-local properties	98
2.1.4 Feature ascription without location	99
2.2 Objecthood	100

2.2.1	Ascribable features and object-files	101
2.2.2	Features and Object-fixation	104
2.2.3	Features and object-persistence	107
2.2.4	Non-objectual features	111
2.2.5	Conclusions	114
3	ISSUES IN THE STUDY OF INTERNAL CONSTRAINTS	117
3.1	A case study: the Feature Binding Problem	117
3.1.1	Operational vs. non-operational formulations	118
3.1.2	Extensional characterizations of basic features for binding	122
3.1.3	A closer look at color as a basic feature for binding	125
3.2	Level conflation	127
3.3	Consilience strategies	129
	 PART II ADAPTIVE CONSTRAINTS	 135
4	ADAPTIVE CONSTRAINTS ON BASIC FEATURES	137
4.1	Steps towards a science of environments	137
4.1.1	From ecological to adaptive constraints	138
4.1.2	Adaptive constraints and perceptual processing	139
4.1.3	Adaptive constraints and the direct pickup of properties	140
4.1.4	Beyond internal constraints	141
4.2	The methodological landscape	142
4.2.1	Distributional cues and statistical learning	143
4.2.2	Ecological rationality and simple heuristics	147
4.2.3	Bayesian perception and perceptual inference	149
4.3	From environment structure to basic features	151
5	SENSORIMOTOR CONSTRAINTS	153
5.1	Sensitivity to sensorimotor couplings	154
5.2	Perceptual skills based on sensorimotor knowledge	156
5.3	Sensorimotor criteria for feature goodness	160
6	PERCEPTUAL SHUNTS	163
6.1	Robust environmental correlations	163
6.2	Perceptual shunts in a simplified environment	168
6.3	Perceptual shunts vs. perceptual inferences	174
6.4	Psychological evidence for shuntable properties	177
6.5	Sensorimotor constraints and perceptual shunts	180
6.6	New criteria for feature basicness	181

7	ISSUES IN THE STUDY OF ADAPTIVE CONSTRAINTS	183
7.1	Non-Inferentiality	184
7.1.1	Perceptual shunt: detection or direct pickup?	187
7.1.2	Perceptual shunt: what is represented?	189
7.2	Ecological Intelligence	190
7.2.1	Recruitment vs. Delegation	191
7.2.2	The virtues of dumb mechanisms	193
7.2.3	Perception as a heuristical process	194
7.3	Conclusions	196
	CONCLUSIONS	199
	BIBLIOGRAPHY	204
	LIST OF ABBREVIATIONS	225
	INDEX	225
	ACKNOWLEDGMENTS	232

List of Figures

1.1	Examples of geons and representative objects that can be constructed from geons Biederman (1990)	34
1.2	Examples of reductive determination of optimal features for 12 TE cells – from Tanaka (2003)	48
1.3	Single feature pop-out – from Wolfe and Cave (1999)	55
1.4	Feature conjunction search – from Wolfe and Cave (1999)	55
1.5	Evidence for High Level basis of Feature Search – from Hochstein and Ahissar (2001)	66
1.6	Stimulus effectiveness in single cell recordings of the macaque visual cortex – From Gross et al. (1972)	76
1.7	Overlapping feature maps in the cat’s visual cortex – from Hubener et al. (1997)	77
2.1	Frank Jackson’s <i>Many-Property</i> Problem	95
2.2	Object and their features. (A) A single blob: a <i>shapeless, preattentive object</i> ; (B) Featural variations on single blobs; (C) Emergent properties of two blobs – from Pomerantz et al. (2003)	113

List of Boxes

1.1	Marr's sketch-dependent primitives	33
1.2	Wolfe's list of basic features	60
1.3	Psychophysical Linking Hypothesis	73
2.1	Intrinsic locality constraint on basic features	91
2.2	Functional locality constraint on basic features	93
2.3	Ascribability constraint for basic features	103
2.4	Revised ascribability constraint for basic features	106
6.1	Perceptual shunt hypothesis	166
6.2	Perceptual shunt hypothesis for sensorimotor learning	180

To my parents

Preface: In quest of basic features

THE AIM OF THIS WORK is to explore a theoretical and methodological issue that has not attracted so far a large attention in perceptual science: the characterization of *basic properties for perception*. Basicness criteria are pervasive in the study of perceptual properties. Implicit basicness assumptions are conjured up each time in a functional explanation of perceptual skills a property is assumed to be *inferred* from another property; each time a property is considered *directly detected* by the sensory organs; each time a complex stimulus configuration is assumed to be analyzable by the perceptual system as a *conjunction* of simpler properties; each time a property or sensory pattern is taken as the *proper input for a perceptual module*, and so on. The interest of understanding what is a basic feature for a perceptual system is intimately related to what it means to provide a functional explanation of how perception works. In the general case, a theory of perception must address the question of how certain kinds of capability are possible in organisms endowed with specific sensory organs and functional architectures with given constraints. Observable (measurable) perceptual skills of an organism represent the classical *explanandum* of a theory of perception: the *explanans* is accordingly a description of:

- the kind of *properties* to which perceptual processing applies (i.e., the patterns in the stimulation that are parsed and encoded by sensory

organs and hence play a functional role as entry-level in the deployment of perceptual mechanisms);

- the nature of this *processing* (i.e., the operations the perceptual system performs on such properties to extract the information needed to control behavior);

Accounting for these two levels provides a characterization of the necessary conditions required for an organism to meet the demands of specific perceptual tasks. How to constrain the set of properties of the sensory pattern that are relevant for understanding perceptual capabilities, though, is a theoretically complicated issue. The problem has been synthetically formulated, in the case of vision, by Anne Treisman as follows: “The critical question is what counts as a feature for the visual system” (Treisman, 1986, p.1301). In one of her first syntheses on the *feature integration theory* she developed during the '80 she claims that this problem can be described as “the question of how to decide what is and what is not a functional feature in the language of visual coding” (Treisman, 1988, p.203). The focus of the present work is on understanding whether and how the notion of a basic perceptual feature can be constrained in such a way to become a robust theoretical notion.

The reasons why this operation is particularly challenging are manifold. Two reasons, in particular, are worth mentioning:

- A. In the last decades, the study of perception has become more and more *piecemeal*, and the idea of providing general theories of perception has become obsolete, favoring the study of specific classes of perceptual subroutines. The gain in analysis and descriptive accuracy has produced as an obvious consequence the fact that what counts as a relevant basic property is often reduced to a matter of terminology and adjusted as

a function of the specific explanatory requirements of the considered perceptual routines that are studied.

- B. As a consequence of this piecemeal approach and of the variety of experimental methodologies to study perceptual routines, there is hardly an agreement on a shared set of criteria for telling apart basic properties, i.e. properties that constitute the relevant input for perceptual processing, from patterns that are not relevant for perception;

Such reasons make the quest for criteria of basicness for perceptual properties an extremely delicate matter. More precisely, A. and B. both threaten the idea according to which some properties of sensory patterns might be easily singled out for the particular functional role they play in determining the proper input of perceptual mechanisms. For on the one hand, by reducing basicness to the mere issue of what variables have to be selected to provide a local explanation of some perceptual subroutines, different basicness criteria become hardly commensurable and can hardly be translated into general constraints on the functioning of perceptual systems. On the other hand, the lack of robust methodological strategies to compare and integrate these local explanations is likely to produce a number of theoretical artefacts, whenever distinct criteria just *happen* to converge on the same set of basic properties and this convergence is taken as a corroboration of the validity of such criteria. The methodological strategy that the present analysis adopts in order to disentangle different criteria for feature basicness consists in:

- reviewing and systematizing the use in the literature of different notions of “basic feature” to characterize properties of the visual stimulation that constitute the entry-level of perceptual processing;
- pointing out cases of possible conflation between distinct notion of

“basic feature”;

- delving into the (often implicit) assumptions underlying the choice of specific basicness criteria over others.

Different ways of regarding perception, its nature and its alleged goal result – I will argue – in different sets of criteria for what counts as perceptually relevant properties of the visual stimulation.

One might wonder whether clarifying the use of the concept of basic feature (so as to avoid conflation and warn against invalid explanatory strategies) is of theoretical interest. The very issue of characterizing basic features – one might argue – sounds genuinely empirical. Different organisms have evolved in such a way that they have zeroed in on specific sets of properties as functionally relevant patterns upon which to build perceptual capabilities. The study of constraints on their sensory organs – one might conclude – is sufficient to provide a thorough characterization of what are functionally relevant properties for understanding their perceptual skills. The problem – though – is much more complicated than it may seem. Assuming that the issue of what counts as basic features for perception can be solved by simply looking at those patterns of stimulation that are *compatible with the structure of sensory organs* is – as [Fodor and Pylyshyn \(1981\)](#) pointed out – a way to shift the burden of characterizing feature basicness to the question of what counts as a *sensory organ*. In this respect, understanding what properties constitute a relevant input for perceptual systems requires defining the nature, the scope and the functional boundaries of sensory organs. Individuating sensory organs is hence an issue that is tightly related to the definition of basic features. Witness of this relation, the lively debate between defendants of ecological approaches to perception – often dubbed as *direct perception* – (see [Gibson, 1966, 1979](#); [Michaels and Carello, 1981](#)) and defendants of *indirect*

perception (see [Fodor and Pylyshyn, 1981](#); [Rock, 1977, 1997b](#)), that has focused on the question of understanding what kind of properties perceptual systems can be said to *directly pick up* from the sensory stimulation. It is worth reporting a quote that make this point with particular strength:

Recent versions of the Establishment theory have sought to constrain the notion of direct detection by identifying the properties that are available without inferential mediation with those to which transducer mechanisms are sensitive. This transfers the problem of constraining “directly detectible property” to the problem of constraining “mechanism of transduction” and, contrary to the assumptions that appear to be widely made, specifying what is allowed to count as a transducer for the purposes of cognitive theory is a non trivial problem. For example, transducers are technically defined as mechanisms which convert information from one physical form to another. But this definition is entirely compatible with there being transducers for *any* pattern of stimulation to which the organism can respond selectively since *whole organisms* are, in that sense, transducers for any category to which they can reliably assign things; e.g. for sentences, or shoes, or, in Berenson’s case, for Da Vincis. This is precisely Gibson’s problem as it arises in the context of Establishment theories, and to fail to grasp its seriousness is to fail to understand the challenge that Gibson poses to the Establishment – [Fodor and Pylyshyn, cit.](#), p.157.

In order to address the question of what are good criteria for feature basicness (and hence what properties of the sensory stimulation are relevant for perceptual matters), I present two alternative options.

The first option – analyzed in **Part I** (Chapters 1, 2, 3) – consists in assuming that basicness can be defined by looking exclusively at the **internal constraints** on the architecture of perceptual systems. The idea that there

are properties that are more basic than others depends – according to this view – on a number of facts about the internal makeup of perceptual systems. Hence the necessity of studying internal constraints. Internal constraints determine the kind of properties perceptual systems *can process*: different kinds of internal constraints yield different classes of basic properties and different kinds of perceptual processing. I argue that internal constraints, albeit a necessary condition for the study of feature basicness, are not *per se* adequate to describe the totality of perceptual phenomena.

In **Part II** of the present work I suggest an alternative option. I submit that in order to single out properties that are functionally relevant for perceptual processing one cannot just take into account internal or architectural constraints on the structure of the perceptual system. Criteria for feature basicness – I argue – depend on the match of *internal* (or architectural) constraints with *external* constraints on the structure of the environment in which an organism is embedded. This match determines what I call *adaptive constraints*. The goal of this part is to focus on a number of potential candidates for basic features that have been systematically disregarded by mainstream perceptual research because of strongly internalist bias.

The conclusion of this work can be resumed in the idea that a considerable part of current perceptual science has failed to take into account the role external constraints play on the definition of properties that represent a relevant entry-level for perception. The spirit of this work is strongly Gibsonian, in that it stresses the importance of environment structure in the shaping of perceptual capabilities. Some of the proposals made in this work, though, will possibly go beyond the scope of ecological theories and contribute to individuate potential perceptual capabilities that have deserved so far only a minor attention in perceptual science.

GENERAL OUTLINE OF THIS WORK

PART I focuses on two classes of **internal constraints** on the definition of basic perceptual features. The first class of constraints (**Chapter 1**) is what I call **primitiveness constraints**, i.e. internal constraints on perceptual systems that allow to describe certain features as *primitives* with respect to other properties. I propose a number of independent criteria for primitiveness that – I argue – are often conflated in the literature and play a prominent role in the characterization of basic features. The second class of internal constraints that I present (**Chapter 2**) is what I call **ascribability constraints**, i.e. constraints related to mechanisms through which perceptual systems process specific properties of the sensory stimulation as *attributes* of individual entities (*spatial locations* or *objects*). I argue that the pairing of primitiveness and ascribability constraints is at the origin of the privileged role given in vision science to a specific class of properties over other properties of the sensory stimulation. In **Chapter 3** I focus on a case study – that of the *feature binding problem* – in order to show how the lack of clear-cut distinction of criteria for feature basicness based on internalist constraints produces a number of major methodological issues.

PART II extends the analysis of feature basicness criteria to adaptive constraints, where by adaptive constraints I mean the match - that I will articulate - of internal constraints of perceptual systems and external constraints on the environment structure. After an introduction to what I call adaptive constraints (**Chapter 4**), I present a paradigmatic case of perceptual mechanisms – *perceptual shunt mechanisms* – that rely on adaptively-defined features (**Chapter 5-6**). I finally analyze the consequences of taking into account environmental regularities in the definition of perceptually relevant basic properties and review a number of methodological issues in the study of

adaptive constraints, suggesting potential research directions (**Chapter 7**).

PART I

INTERNAL CONSTRAINTS

Internal constraints on basic features

WHAT DOES IT MEAN to study internal constraints on the notion of basic feature? The goal of part I of this work is to disentangle a number of independent criteria that are often conflated in the literature and that concur in the definition of what counts as relevant input for perceptual processing. The pervasiveness of these criteria - I will argue - depends on some general assumptions on the goal and functioning of perceptual systems that are largely endorsed in current vision science. These assumptions have focalized the attention of researchers on the study of the *internal resources* that allow perceptual systems to deliver to the organism reliable information on the basis of the sensory stimulation *in any condition*. Understanding and modeling perception is typically seen as a matter of describing how a *correct representation* of the visual scene can be built *in any condition* on the basis of information made available through the senses. The problem has been clearly stated in artificial vision: whereas it is (relatively) easy to construct a device able to parse and recognize objects of a given kind in an oversimplified environment, a crucial problem arises as soon as we want to make the perceptual skills of an artificial perceptual system *scalable*, i.e. make it able to cope with the variability of the visual world. The typical answer is that skilled perceptual systems are those systems that can flexibly make use of a rich set of inferential processes in order to *correctly interpret* sensory data.

This strongly internalist stance on the requirements of reliable perception is, I argue, at the origin of a particular way of regarding the functional role of some basic properties of the visual stimulation as input of perceptual processing. My goal in this chapter is to unpack these implicit assumptions and to shed light on potential problems that arise from the use of merely internal criteria to define feature basicness.

Perception has long since been considered as an interface level between sensation and cognition functioning as a *general-purpose device* with respect to its input. Whatever further capacity perceptual devices are designed to serve (like controlling action, reasoning or conceptualizing), their goal is to deliver reliable information on *any kind* of entity or property in the environment an organism can interact with through the senses.

I will start my analysis in the following chapters by tackling a widespread assumption according to which:

1. there is a class of properties of the visual stimulation that constitute the entry level of perceptual processing;
2. these properties (or *basic features*) are the constituents of the very first representation of the visual scene delivered by sensory organs;
3. any kind of further perceptual processing must start from such basic features in order to build a reliable representation of the distal sources of the stimulation.

This assumption gives to basic features a precise functional status with respect to perceptual processing: it specifies what counts as a functional input of perception. Given such constraints on the input of perceptual devices, we are now in condition to understand that a paradigmatic problem for perceptual processing consists in how to correctly interpret what the senses signal in

terms of basic features in order to deliver reliable representations. This idea is synthetically captured by [Treisman and Kanwisher \(1998\)](#):

The goal of perception is to account for systematic patterning of the retinal image, attributing features to their real world sources in objects and in the current viewing conditions. In order to achieve these representations, multiple sources of information are used, such as color, luminance, texture, relative size, dynamic cues from motion and transformations, and stereo depth; however, the most important is typically shape.

The kind of processing perceptual systems are designed to perform on the basis of sensory information consists - according to a longstanding tradition - in a *reconstruction* of the correct distal causes of sensory stimulation. Such reconstruction, it is assumed, is what allows the organism to acquire reliable information of its environment. Whether the process of correctly interpreting sensory information should be qualified as a process of actual *reconstruction* has been a largely debated issue that goes beyond the scope of the present analysis (see for instance [Edelman, 1994](#); [Tarr and Black, 1994a](#)). What is interesting, though, whatever position we take in such debate, is that typical problems in the study of perceptual processing arise from the fact that basic properties do not bear per se *sufficiently reliable information* to meet the perceptual needs of the organism. The preliminary representation of the visual scene delivered by the senses in terms of basic features is insufficient to provide cognitively reliable information: in order to provide the organism with reliable information to control behavior, sensory information needs - as a general rule - to be parsed, processed and matched with internal representations. Analyzing internal constraints on feature basicness means, then, articulating the distinct criteria that have been proposed in the literature to characterize

the entry-level of perceptual processing, i.e. the properties of the sensory stimulation from which perceptual systems must build a reliable representation of its distal sources. I maintain that internal constraints on feature basicness can be divided into two main families: primitiveness criteria and ascribability criteria. I will dedicate **Chapter 1** and **Chapter 2** of this work to the respective analysis of these two families of criteria.

Chapter 1

Primitiveness

BASIC FEATURES AS PRIMITIVES

ONE of the most common criteria for selecting relevant visual features consists in restricting them to *primitive visual properties*. Any perceptual mechanism able to “make sense” of sensory signals (i.e. extract from the sensory stimulation reliable information that can be used for different perceptual routines, like for instance object recognition) must be able to parse some specific patterns as its *entry level*. The minimal hypothesis I am assuming here is that any perceptual task requires that some patterns or properties in the sensory stimulation must be *processed* (parsed, filtered, and extracted) in order to provide valuable information to further perceptual mechanisms¹.

This first processing stage is what yields, according to mainstream perceptual research, a set of *primitive visual components*. Primitives can then be *prima facie* defined as the entry-level properties required by any kind of further perceptual processing. What is meant for *entry-level* is then the main problems

¹The only challenge to this almost trivial assumption might come from a radical defendant of a direct-perception paradigm, according to which no kind of processing is required for extracting reliable information from the senses. I address this issue in more details in Part II.

that must be addressed.

In this chapter, I survey a number of distinct characterizations of *primitiveness* and *entry-level* criteria that occur in the literature. Features can be defined as primitive for various reasons:

1. they enter in the *composition* of any derived visual properties;
2. they are functionally (*unanalyzable*) descriptors of the visual scene;
3. they represent the *simplest* perceptual systems are able to encode;
4. they are accessible *earlier* than other visual properties
5. they are encoded by the most peripheral areas of the sensory cortex.

In each of these cases, a number of implicit assumptions are put to work for characterizing respectively the ideas of *compositionality*, *unanalyzability*, *simplicity*, *earliness*, and *low-levelness*. The main purpose of this chapter is to review the assumptions behind each of these criteria and to warn against possible conflation between distinct notions of primitiveness. Compositional primitiveness is usually taken to be inseparable from unanalyzability which, in turn, is often assumed to entail simplicity; earliness is frequently associated to low-levelness and to unanalyzability: my aim is to show that each of these notions of primitiveness is independent from the others and, as a consequence, that linking them through stronger relations than required can generate potential theoretical artifacts.

Every notion of primitiveness described in this chapter is such in virtue of internal (or architectural) constraints on the considered perceptual systems. More precisely, a property can be said to be primitive with respect to the specific way in which the perceptual system *handles it* as opposed to other properties. Most of the frequent unargued overlaps between primitiveness

criteria depend on the fact that it is implicitly assumed that certain kinds of perceptual processing are inseparable from others, for example: that by extracting patterns eliciting selective activity in single neurons of the primary visual cortex, the perceptual system is *at the same time* building a set of compositional descriptors for the visual scene. There is no *a priori* reason, though, to assume that this must be the case. My focus in what follows is precisely on the fact that different classes of primitives result from *independent* (i.e. not necessarily functionally correlated) ways of handling properties of the sensory stimulation.

1.1 COMPOSITIONALITY

Visual primitives can be characterized as the basic components into which any complex property of the visual scene can be decomposed. I will call primitives that comply with this definition *compositional primitives* or *C-primitives*. According to this idea, visual primitives constitute a fixed set of descriptors (a “lexicon”) that allows, together with appropriate compositional rules (a “syntax”), the formal description of any element of an image.

This notion of a fixed repertoire of basic visual components represents the most familiar characterization of primitives in the framework of artificial vision, but is common to several approaches to the study of perceptual systems. A very similar idea can be found, for example, in the field of pattern recognition, where a structured set of measurable properties (generically called “features”), playing a special descriptive role for a pattern family, is defined a class of “primitive features”. The difference between generic features as measurements and primitive features lies in the fact that the latter can be restricted to members of the smallest set of features that, for any given pattern, allow to effectively match the pattern with a specific pattern family.

The idea that the first stages of perceptual processing can be described as mechanisms using a finite set of C-primitives and compositional rules has been implemented by a large number of models. The benefits of this approach basically consist in its economy (only a small set of descriptors and compositional rules is required) and the possibility of accounting for a large number of relations between image elements (identity, symmetry, similarity, inclusion) in terms of underlying relations between their components. An argument in this sense is given by [Treisman \(1986\)](#):

Although the identification of objects and events is likely to be the primary goal of perception, there are compelling reasons to believe that it is achieved through analysis or descriptive decomposition of the physical stimuli. The alternative would be a direct, unique, and unitary labeling response for each distinguishable occurrence of each possible object, event or state of affairs. Since these must be infinite in number, it seems unlikely that the economy of a finite brain could encompass the variety of perceptual experiences

Classical examples of models of visual perception using C-primitives as formal descriptors of visual representations can be found in [Marr \(1982\)](#), [Biederman \(1987\)](#) and [Koenderink \(1993\)](#). Let us take a closer look at each of these models.

In [Marr](#)'s model, each level of analysis (or "sketch") of the visual scene is characterized by a specific set of primitives that exhaust the formal description of all the relevant elements of this level: the *primal sketch*, for example, encodes local intensity changes through a specific (finite) set of primitives (EDGES, BARS, ENDS, BLOBS). The 2 1/2D *sketch* and the 3D *sketch* provide in turn a formal description of the image based on their own sets of primitives,

depending on the kind of operations each of them is supposed to enable. For each level of analysis (which Marr assumes to be hierarchically organized) a specific set of operations are performed on the visual input in order to deliver a specific representation of the visual scene based on a specific lexicon of properties.

Marr's sketch-dependent primitives

- **IMAGE**
Operation: represents intensity
Primitives: Intensity value at each point in the image.
- **PRIMAL SKETCH**
Operation: makes explicit information about the two-dimensional image, primarily the intensity changes there and their geometrical distribution and organisation.
Primitives: Zero crossings, Blobs, Terminations and discontinuities, Boundaries.
- **2 1/2 SKETCH**
Operation: makes explicit the orientation and rough depth of the visible surfaces, and contours of discontinuities in these quantities in a viewer-centred co-ordinate frame of reference.
Operation: Local surface orientation, Distance from viewer.
- **3D MODEL REPRESENTATION**
Operation: Describes shapes and their organisation in an object centred co-ordinate frame, using a modular hierarchical representation which includes volumetric primitives (i.e. represents volume of space that a shape occupies) as well as surface properties.
Primitives: 3D items arranged hierarchically, each one based on a spatial configuration of a few sticks or axes, to which volumetric or surface shape primitives are attached.

[Biederman's Recognition By Components Theory \(RCT\)](#)- a primitive-based model of object recognition - defines a fixed set of 36 geometric elements (called GEONS) whose spatial combination is used to account for the recognition of

virtually any kind of 3D object. Although this model has been proposed to account for some high-level perceptual processes like object recognition, it is a paradigmatic case of the how a finite lexicon of bulding blocks can be adopted to parse and describe in an economical way virtually any 3-dimensional object present in the visual scene.

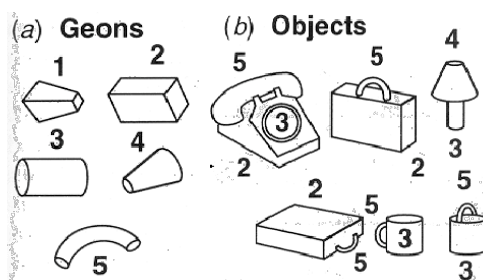


Figure 1.1: Examples of geons and representative objects that can be constructed from geons [Biederman \(1990\)](#).

[Koenderink](#)'s model is meant to provide a formal description of the finite repertoire of properties that pertain to the structured activity of local operators in what he calls the “front-end visual system”. He assumes that this level of visual processing works in a purely “bottom-up, syntactical and pre-categorical fashion” and represents a bottleneck for all further perceptual processing, in the sense that only those properties that are encoded at this level are made available for further levels. Basic primitives at this level, that can be used to provide a compact description of a 2-dimensional image, are four different types *icons* that conver all possible variability in geometrical structure of patterns on the retinal image: UNIFORM PATTERN, BLOB, EDGE, and BAR.

Other paradigmatic examples of models of perceptual capacities based on lexica of C-primitives can be found in specific domains where percepts typically display a patterned structure².

²Paradigmatic cases are those of phonetic or graphemic parsing (see [Saffran et al., 2001](#))

These examples raise a number of general issues concerning the relation between abstractly defined compositional primitives and the possible internal constraints of perceptual systems.

C-primitives need not be perceptually salient properties

An interesting feature of Biederman's model is that it postulates the existence of primitives that seem to totally lack perceptual saliency. Indeed, the fact that perceptual systems might use a lexicon of basic components for internal reasons (due to architectural or processing constraints) does not entail that these components necessarily show up as perceptually accessible properties. It is perfectly legitimate to conceive of a system parsing the sensory stimulation according to an internal repertoire of compositional properties that do not show up as such in any measurable perceptual capability.

C-primitives can be hierarchically organized

Compositional primitives can be specific to particular stages of visual processing, and different stages can rely on distinct repertoires of primitives. There is no compelling reason to conceive C-primitives as forming a universal repertoire of basic properties. Granularity and format specifications for specific stages of perceptual processing can affect the way in which possible primitives are picked up. Moreover, compositional primitives for a given level of representation can be thoroughly re-describable in terms of more fine-grained properties, without losing their functional role of primitives for their specific level of application. The fact that, for instance, in Marr's model a given set of primitives can exhaustively describe the structure of the sensory

or visual parsing of uniformly connected blobs ([Palmer and Rock, 1994](#)). For a survey of fixed-primitive models of perception, see [Schyns et al. \(1998\)](#)

stimulation for the specific computational function relevant for its level of analysis is not incompatible with the fact that there might be vertical relations between different that allow to describe primitives at a given level in terms of primitives at a subordinate level. What is crucial, though, is that at each level primitives are the smallest descriptors performing the relevant operations for that level.

C-primitives are unable to provide flexible representations

The main benefit for perceptual systems to work as syntactic devices that parse complex patterns in terms of combinations of a small number of atomic properties and compositional rules is, as suggested above, the parsimony of their architecture. Selecting compositional primitives that can encode recurrent regularities of the visual stimulation can drastically reduce the complexity of computations performed by perceptual systems. The cost of this solution, though, is representational rigidity, i.e. the fact that such descriptions are functionally “blind” to properties that are neither part of the lexicon of primitives nor the result of composition of such primitives. The need for flexible representations can be accommodated with the notion of C-primitives by weakening the idea that *any* kind of perceptual processing must be based on the very same set of basic descriptors and by assuming, for instance, that different kinds of routines select a subset of compositional primitives and compositional rules that need not apply to other kinds of routine. The strong task-dependence of many visual routines can be seen as a challenge to the idea that there should be a set of basic descriptors of the visual stimulation that can feed any kind of further perceptual processing (see [Schyns et al., 1998](#)).

Static vs. dynamic C-primitives

A longstanding trend in the study of perceptual mechanisms has privileged static sensory configurations (like the retinal image) over dynamic sensory patterns. Accordingly, compositional primitives have systematically been restricted to static image properties. Yet there is no *a priori* reason why a formal description of visual basic components should not take into account dynamic (spatio-temporally extended) descriptors of the sensory array. Accounting for invariant properties of complex motion patterns is thoroughly compatible with the idea of a repertoire of elementary dynamic components.³ It has also been suggested that by considering the optical flow more primitive than the static retinal image, it is still possible to account for static properties as boundary properties of dynamic configurations.

Psychological reality of C-primitives

Perhaps the strongest objection to the descriptive adequacy of models of perceptual processing based on C-primitives is their validity to describe the internal structure of real-world, biologically constrained perceptual systems. Formal models of image description based on fixed lexica of primitives can perfectly account for optimal processing strategies implemented in artificial visual systems with not biological constraints. But from the fact that a lexicon of C-primitives (plus some compositional rules) is sufficiently rich to derive all relevant properties of an image, it hardly follows that actual perceptual systems use such compositional primitives for parsing the visual scene. Mainstream psychological models of visual perception assume that the first stages of visual processing are indeed devoted to extracting some

³This is actually an essential requirements of sensorimotor theories of perception, that I will review in Chapter 5

quasi-compositional primitive features (e.g. surfaces and edges) from the retinal image and using them to construct a first rough description of the stimulus structure. But this parsing can hardly be characterized as a strictly generative derivation of complex structures from a finite set of primitives plus compositional rules. Primitive features, as it will become clearer in the remainder of this chapter, is a complex notion that often results from the implicit lumping together of distinct criteria. One of the common abuses of language consists in taking incomplete and extensional characterizations of alleged basic components of the visual scene as C-primitives although they lack any genuine (formal) compositional role.⁴

To sum up, compositional primitives have deserved a large attention in formal theories of perceptual processing, in which fixed-repertoire of properties have been shown to provide a compact and powerful strategy to describe stimulus structure. Among the drawbacks of processing strategies based on such kind of primitives, though, I have mentioned the limited representational flexibility provided by fixed lexica and the problematic issue of understanding whether finite repertoires of primitives working as generative devices represent a psychologically plausible way of looking at real-world perceptual systems

1.2 UNANALYZABILITY

The *compositionality criterion* adopts a bottom-up strategy for defining good primitives. Good compositional primitives, as I have characterized them in the previous section, are those basic visual properties that – together with appropriate compositional rules – yield a description of any complex structures in a given visual scene. Now, taking the problem from the opposite

⁴This argument is further developed in my criticism of the *Feature Binding Problem* in Chapter 3, in which I point out that feature conjunction is often implicitly understood as a quasi-compositional process based on a fixed repertoire of properties.

perspective, one might ask what are the lowest units in the parsing of a visual scene that cannot be analyzed into simpler elements. It might be tempting to expect that this top-down strategy will lead exactly to our lexicon of compositional primitives, but it is not necessarily the case.

The problems of upward compositionality and downward analyzability need not be mutually dependent. There can be a good (minimal and sufficiently rich) lexicon of C-primitives for deriving any complex property P of the visual scene and yet the downward analysis of P into its constituents may not necessarily yield elements of the original lexicon. To put it in other words, the fact that there exists an effective function of upward derivation of a complex property P from a set of compositional primitives does not entail that the inverse procedure should be equally effective. In particular the downward analysis of a complex property P could stop, because of internal constraints, at some basic level where perceptual units are no more analyzable into smaller components that keep a functional relevance for perceptual processing.

Atomic features represent then a distinct set of visual primitives that do not need to overlap with compositional primitives. I will call such *unanalyzable* properties *atomic* primitives or *A-primitives*.

It should be made clear what is meant here by “atomic”. I am not arguing that there should be *absolutely* unanalyzable properties at the bottom of the hierarchy of features perceptual systems are sensitive to. I am rather saying that there can be basic properties that are unanalyzable for a perceptual system (or for a specific kind of perceptual processing), even if these properties are not per se *absolutely* unanalyzable (i.e., they could be further analyzable by other kinds of perceptual processors). As a consequence, the question whether A-primitives should be necessarily internally unstructured must be spelled out in terms of their relation to the specific kind of processing to which

they are relevant. A deflationary strategy for rejecting the idea of internally structured visual primitives is to adopt a radical *pointillistic* view. If one takes sufficiently fine-grained primitives – one may argue – like *pixel-level* properties, together with powerful compositional rules, any kind of relevant visual properties can be reduced to molecular aggregates of these properties.

The argument that every property can be redescribed in terms of pixel-level visual properties does not hold, unless pixel-level properties can be functionally accessible to the perceptual system. Atomic primitives must be identified by their functional role: it may well be the case that a specific kind of processing be only able to treat properties of a given format, without being able to access their internal structure or articulation.

To put it another way, A-primitives need not be internally unstructured, provided that they are not further analyzable by the specific kind of processing they are referred to. Let us consider for instance Biederman's GEONS: it is interesting to remark that these primitives are not internally unstructured, since each of them can be redescribed in terms of a 5-uple of invariant properties of edges (CURVATURE, PARALLELISM, CO-TERMINATION, SYMMETRY, CO-LINEARITY). But these n-uples of invariant edge properties cannot be considered per se as "primitives", since assumedly they are not the properties the system is functionally relying upon when building representations of the visual scene (they are *opaque* to the system and their further decomposition is only available in a theory that describes them). Downward analysis of visual structures for this specific kind of perceptual routine (3D shape recognition) stops at the level of geons and goes no further: this does not entail that geons are absolute geometrical atoms. Unanalyzability is not only relative to a level of description, but also to the functional role primitives play in perceptual processing. What is at stake in these cases can be described as a problem of

relevant *granularity* or, more generally, of relevant *representational format*: properties on which we focus in this section are primitive – I assume – with regard to processing constraints of perceptual devices, not necessarily for the theory that describes them.

The psychological literature is rich of examples of visual tasks that, although virtually possible through the recruitment of more fine-grained properties, actually rely on more complex, albeit (for the task under question) unanalyzable or atomic features.

Consider the case of the extraction of triangular configurations from more basic visual properties of the visual scene (Pomerantz, 1978). On the one hand, SLOPED LINES might be seen as good candidates of A-primitives to which complex properties as triangles can be reduced. It is however evident that the detection of differently oriented lines is not sufficient to encode the presence of a triangle. Segments must meet to create angles, but must not pass through one another as they would create intersections. Thus it would appear that the visual system needs to be equipped with VERTEX detectors too, and so on. Pomerantz suggests that the human visual system – even if it had the possibility to extract triangles from simpler features as oriented lines plus some other conditions – is actually privileging some *more complex* properties (*more complex* insofar as they are less local and they seem to be theoretically analyzable in terms of simpler properties) that allow reliable extraction of the target features (in the above example, the diagnostic features are ANGLES).⁵ Such complex, yet functionally unanalyzable properties are defined by Pomerantz as “emergent features” since their functional role cannot be explained in terms of the functional role of their components, which are *opaque* to perceptual processing.

⁵See also Ullman et al. (2002) for the role played by *moderately complex* features as functionally unanalyzable properties for specific kinds of perceptual processing.

Another way of characterizing this notion of relative unanalyzability is to say that the visual system is unsensitive to lower decompositions of A-primitives for specific kinds of routines. There are many possible characterizations of this idea of lack of sensitivity to properties at subordinate levels: the most current of these characterization is probably that of *format-specificity*. It is interesting to mention Pomerantz's own solution to this seeming puzzle, a solution that appeals to a sort of "reverse encapsulation" property. According to his hypothesis (that he calls "sealed channels hypothesis"),

[i]t is conceivable that higher-order features are derived from lower order ones in perceptual process, but that the subject is not able to make responses on the basis of lower-order feature detectors.⁶
(p.222)

Classical examples of encapsulated mechanisms are cases in which modules are not able to integrate "higher-level" information to solve specific kinds of routines. The idea of "sealed channels", in a sense, reverses the direction of processing opacity, by redirecting it from the periphery to the center rather than the other way round. This idea of opacity to "lower-level" features, as we will see, is crucial for explaining a number of counter-intuitive phenomena in alternative definitions of primitives. Moreover, being attuned to particular complex and unanalyzable patterns can give a perceptual system specific advantages from an adaptive viewpoint.⁷ The idea of A-primitives or emergent properties as the lowest functional units for a given kind of perceptual processing can be exemplified by a number of cases, that I will shortly review in what follows.

⁶I will return on this hypothesis in Chapter 6.

⁷See Part II for a further analysis of this issue.

Good gestalts

It is well known since the Gestalt school that some principles of visual organization play a major role in segmenting the scene into units that are preferentially selected over others in visual grouping tasks. Some of these principles can be seen as constraining the set of basic properties that cannot be further analyzed without losing their functional role: for example, the idea of UNIFORM CONNECTED REGIONS, as introduced by [Palmer and Rock \(1994\)](#), can be considered as a paradigmatic case of A-primitive for visual grouping: any proper part of a uniform connected region is insufficient to account for preferences in visual grouping. Hence, uniform connected regions are the smallest property that preserves its functional role. This kind of functional superiority of features of intermediate complexity with respect to their components or proper parts is analogous to known configurational effects in many kinds of visual object recognition tasks, in which the functionally relevant role of an item is lost as soon as we analyze it into its components.

Spelke Objects

The literature on concept acquisition often refers to properties or set of properties that allow children to bootstrap particular kinds of perceptual skills. An interesting example is offered by the notion of *Spelke objects* ([Spelke, 1990, 1993](#)), i.e. specific kinds of stimulus configurations that are assumed to play a crucial role during early developmental stages in bootstrapping the acquisition of the concept of a material object. Children tend to systematically privilege ([Casati, 2003](#)) over other configurations, in a number of experimental conditions, stimuli that are represented by CONNECTED, BOUNDED and COHERENTLY MOVING items. As soon as any of these properties is dropped (e.g., as soon as these items are kept bounded and coherently moving, but lose

connectedness), the predicted privilege disappears. In this sense, *Spelke objects* (that should be more neutrally described as *Spelke bundles of properties*, if not directly as *Spelke features*, (see [Fodor, 2001](#))) are an interesting example of A-primitives for the acquisition of objectual representations.⁸

Pylyshyn's FINGs

Another case of functionally unanalyzable features is that of stimulus properties upon visual tracking depends. Through a large number of contrast experiments, Pylyshyn and his collaborators have managed to show that a certain set of visual features are required for attracting what he calls visual indexes or FINSTs. These index-grabbing features (that he baptized FINGs, since they are interdefined with visual indexes) represent the basic, unanalyzable units that allow a visual item to be tracked across movement and distractors. However defined ([Clark \(2004\)](#) provides for instance an alternative characterization of the same set of properties), FINGs can be considered as atoms for multiple tracking routines, since any further decomposition of these configurations into simpler features results in a tracking failure.⁹

These examples are meant to illustrate paradigmatic cases of properties of the visual stimulation that cannot be further analyzed without losing the functional role for which they are recruited by specific kinds of perceptual processes. I have extensively discussed in which sense the notion of unanalyzability should be distinguished from that of compositionality. In the next section I will argue that primitiveness criteria based on functional unanalyzability should also be distinguished by primitiveness criteria based on representational *simplicity*.

⁸See also section 2.2 in which this notion is discussed in the context of objecthood criteria.

⁹A further analysis of requirements for index fixation can be found in Section 2.2.2.

1.3 SIMPLICITY

We have seen in the previous sections that the compositionality constraint and the unanalyzability constraint allow to tell apart two independent notions of “visual primitives”. Such primitives are commonly described as *simple properties* as opposed, respectively, to *composed properties* and *analyzable properties*. It is now time to investigate more closely this notion of visual “simplicity” and see if its meaning is completely exhausted by the former two characterizations of primitiveness. Key questions will be:

- What are criteria that can make a visual property *simple*?
- What criteria allow to define a visual property as *simpler* than another property?
- Are there criteria to characterize a set of *simplest* visual properties?

In this section, my aim is to show that not only “simplicity” is a concept that suffers from fatal theoretical ambiguities, but also that, in the case of visual primitives, there is room for a third independent meaning of “simple” on top of, and not reducible to, C-primitiveness and A-primitiveness. Disentangling the different meanings of “simplicity” - I argue - should help operationalize what is meant by “simple primitive”.

A good starting point for gauging the complexity of the notion of “simple” visual features is a work by Tanaka and collaborators, in which cortical sensitivity is studied for what he calls “moderately complex features” ([Tanaka, 2003](#)).

Looking for the most effective stimuli for a specific neuronal population (area TE in the inferotemporal cortex), Tanaka and collaborators discovered that single cells of this area respond selectively to what they dub visual properties of “intermediate complexity”. Faced with the overwhelming variety of

stimulus properties that might elicit selective activities in this area¹⁰, the team developed a strategy to isolate the allegedly “simplest” relevant visual properties. In order to determine “the minimal feature required for the maximal activation”, Tanaka adopts an empirical image reduction method: starting from an initial set of 3D objects images, the image eliciting maximal activation was progressively “simplified step by step to determine which feature or combination of features contained in the image was essential for maximal activation”.

Figure 1.2 displays the results of this reductive strategy: images to the left of the arrows represent the original images of the most effective object stimulus and those to the right of the arrows, the critical features determined by the reduction.

Let us try to make explicit some hidden assumptions in this apparently straightforward simplification strategy:

- (a) The initial set of items should be sufficiently rich not to bias the reduction procedure and produce experimental artefacts.
- (b) If images can be simplified “step by step”, there must be something like a *metrics* or *quantitative measurable criterion* to establish whether a stimulus configuration is “simpler” than another one.
- (c) The reference to “features or combinations of features” suggests that certain results of this reduction strategy are better described as *bundles* of features rather than as features *per se*. Then a criterion should be provided for distinguishing features from bundles of features that would

¹⁰Significantly, Tanaka states that “the variety of object features existing in the world is too great to test its entire range for a single cell while activity of the cell is being recorded” [p.90].

not count themselves as features.

- (d) Finding features that are the most effective or “essential” for maximal activation implies that no other set of visual properties different from these essential features should elicit equal or stronger activation of the same cells.

None of these caveats is explicitly taken into account by Tanaka and collaborators and the way in which the direction of the “reduction” procedure is established is – as the author himself acknowledges – “to a certain extent arbitrary”.

- (A) First, the starting class of stimuli among which the most effective one is selected is arbitrary (it is based on an intuitive selection of stimuli with presumable ecological validity).
- (B) Second, the criterion adopted for the “reduction” procedure is based on an intuitive geometrical notion of “simplicity” of the structure of the stimuli, that can hardly be expressed in algorithmic or measurable terms.
- (C) Third, the fact that the result of the image reduction strategy can be alternatively defined as a *feature* or as a *bundle of features* does not seem to matter to the experimenter. This is somehow surprising, since understanding whether a critical bundle of features can be further reduced into subordinate components that preserve an optimal response is exactly the goal of this experiment.
- (D) Fourth, and finally, the resulting critical features can hardly be considered “essential” to elicit maximal activation, since they are derived from an arbitrary set of initial images and reduced according to an arbitrary simplicity criterion. Another initial set of stimuli and a different reduction

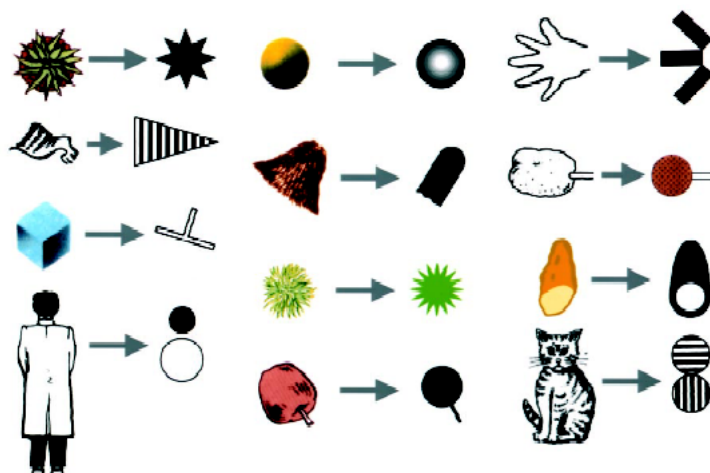


Figure 1.2: Examples of reductive determination of optimal features for 12 TE cells – from [Tanaka \(2003\)](#).

strategy might plausibly yield a different set of features that produce equal or greater activation of the same cell. Critical features resulting from the experiment are certainly *sufficient* to produce the observed patterns of neural response, but – given the above considerations – can hardly be considered as *necessary* or *essential*.

Given the lack of further constraints on the strategy of “image reduction” adopted by Tanaka, the only stable result of such experiment is that the transformation indicated by the arrow, in figure 1.2, preserves the amount of neural activation in the considered cell. Now, it is legitimate to ask for which reason stimulus configurations on the right of the arrow should be considered “simpler” than stimulus configuration on the left. It seems *prima facie* unlikely that there exist an effective algorithm for obtaining the different transformations of each kind of stimulus configuration in figure 1.2. It follows, then, that the reductive criterion adopted by the experiment can hardly provide a good (quantifiable) criterion for assessing visual simplicity.

Tanaka's use of an underconstrained notion of "simplicity" is not an isolated case. Empirical works in which simplicity criteria (either absolute or relative) are defined by drawing on the intuition of the experimenter are quite common in the literature. [Chen et al. \(2003\)](#) suggest that the finding that honeybees can be sensitive to topological properties of the stimuli is "surprising", because, among other reasons, these properties are generally regarded as complex and more difficult to derive than other properties. Claims of this kind not only show that the common-sense notion and the scientific notion of "simple" can often be in conflict, but also that a clear-cut scientific notion of "simplicity" is needed if one wants to make a grounded use in scientific explanations of concepts like that of "simple properties". Elliott Sober's seminal work on simplicity ([Sober, 1975](#)) has shown how pervasive intuitive notions of simplicity are in science and how urgent the need of disentangle them. In particular, he has showed that simplicity is a multi-dimensional criterion that can hardly be reduced to single comparisons on isolated variables.

Generally speaking, if we want to operationalize the notion of simplicity, what is needed is an explicit criterion that enables us to establish, given two distinct stimulus configurations, which is the simpler and which the more complex.

Arguably, the notions of compositionality and unanalyzability that I have introduced earlier in this chapter provide, on the one hand, two explicit conditions to establish the simplicity of a given visual property. On the other hand, there are other simplicity criteria that are not exhausted by these two conditions, and that I will analyze in what follows.

A viable strategy to operationalize the notion of simplicity of visual properties consists in assuming that perceptual systems are built to *find the simplest perceptual description consistent with the sensory input*. The idea of a *simplest perceptual organization* (and, accordingly, of *simple primitives* or *S-primitives*)

is closely related to notions such as parsimony of interpretation and economy of encoding that date back at least to Ernst Mach and have been largely drawn upon by the Gestalt school (see for instance [Koffka, 1935](#))

According to [Pomerantz and Kubovy \(1986\)](#), the simplest perceptual organization for a given sensory configuration can be either described as:

- (a) the perceptual organization providing the most concise and economical description of the sensory configuration;
- (b) the perceptual organization that provides the most economical explanation of sensory data, with respect to their distal causes.

These two distinct characterizations of simplicity as parsimony of sensory encoding and parsimony of perceptual interpretation have been considered for a long time two core principles of perceptual organization, respectively: the *Prägnanz* and the *Likelihood* principles.

The notion of *Prägnanz* is related to what today is better known as *economical coding* of sensory properties.

The existence of some internal redundancy or regularity in sensory patterns can be exploited by perceptual systems to encode more briefly (i.e., using shorter descriptions) the structure of the stimuli. If the stimulus displays some structural regularity, like for instance a redundant pattern, this can be used to provide a concise description of the stimulus configuration: COMMON FATE, in the case of Gestalt principles, allows to encode globally the identical behavior of several components (much as objects moving in the same direction) instead of encoding separately the behavior of each item. In this sense, parsing the scene according to a common fate principle will be systematically preferred by the perceptual system because simpler to encode.

The notion of *Likelihood* does not refer to the internal structure of the stimuli, but rather to the relation between proximal stimuli and their distal causes. Perceptual interpretations of sensory data have a high degree of Likelihood if they provide the most economic description of the environmental causes that produced them. A paradigmatic example of perceptual organization driven by a Likelihood principle is that of RIGID MOTION. Among the countless possible interpretations of the predictable way in which the sensory flow unfolds when an observer moves, postulating the existence of rigid objects may provide the most economic interpretation of the distal causes producing the motion pattern on the retina.

Since early on, *Prägnanz* and Likelihood have been considered as two opposing principles of economy in perceptual organization. [Koffka \(1935\)](#) already distinguished between two kinds of organizing forces in perception, the external and the internal. The external forces were presumed to be retinal in origin and acted to make the neural representation veridical to the distal stimulus. The internal forces were those acting within the dynamic field of the brain, often in opposition to the external forces.

It is an empirical, yet theoretically stimulating question to study under which conditions *Prägnanz* and Likelihood override each other, in particular to see under which conditions the organizational (*Prägnanz*-driven) process allows regularity to be imposed upon percepts at the expense of interpretation economy (likelihood-driven): there are cases in which a good solution from the point of view of *Prägnanz* is discarded because of its low likelihood, and viceversa.

A possible compromise between these two principles has been suggested within the framework of information theory.

Consider Attneave's concept of *economical coding*:

Suppose that what the system *likes* is short descriptions and that the image is progressively changed, within the constraints of the input, until its description is minimized. This way of looking at the matter, which is considerably different from the classical Gestalt point of view, has the advantage of taking into account not only intrinsic stimulus properties – that is, redundancy, uniformity, or homogeneity of the stimulus itself – but also schemata corresponding to familiar objects. If an input can be brought into conformity with a well-formed schema that is frequently used and to which a short symbol has been assigned, it might be described quite as economical as if it were intrinsically simple – [Attneave \(1954\)](#)

More recently, [Chater \(1996\)](#) has shown by relying on Kolmogorov’s theory of complexity that likelihood and simplicity (as MDL – Minimum Description Length) criteria are not in contrast and can be reconciled.

Since Likelihood requires taking into account constraints that are not only internal or architectural, I will return on this principle in the context of a discussion on adaptive constraints on visual features. So far, what I have suggested is that the idea of economical coding of sensory information (the direct heir of the Gestalt notion of *Prägnanz*) can provide a rigorous criterion (alternative to those of compositionality and unanalyzability) to operationalize the notion of simplicity for visual properties. Visual properties can be considered primitive according to this simplicity criterion if they allow the shortest description of the stimulus configuration. As such, they should be kept distinct from properties that can be qualified as primitive because *not further functionally analyzable* or because *compositionally simple*.

1.4 EARLINESS

A major experimental tradition in perceptual psychology has focused on the definition of those properties that the human visual system can extract “earlier” or “rapidly, automatically and efficiently”.

In particular, several behavioral investigations have convincingly shown that certain classes of stimulus properties can be considered as entry-level primitives from the point of view of behavioral access: architectural constraints on visual processing make specific classes of properties *earlier retrievable* than other classes of properties that require more elaborate processing or the allocation of supplementary resources. “Earliness” represents then another *prima facie* candidate that can be invoked for characterizing visual primitives. Yet, as soon as we try to articulate this notion of “earliness”, we realize that things are much more complicated than one might think.

Much as in the case of “simplicity”, earliness is a theory-laden notion that needs to be constrained in order to provide scientifically tenable (operational) criteria. In this section, I will consider some of the background assumptions that are drawn upon to define earliness of processing from the point of view of experimental psychology: this analysis will touch a number of neighboring issues that play a major role in the definition of “earliness” such as the idea of *computational load*, or the distinction between *automatic vs. controlled behavior*, *preattentive vs. attentional processing*, and *parallel vs. serial* processing.

A largely shared assumption in vision science is that the visual system, as any other cognitive device, has specific processing limits. These limits can be accurately measured by evaluating the average time needed to perform specific kinds of visual tasks or by testing the number of joint routines the visual system can perform at the same time. When the reaction times become

significantly longer than those required for control tasks or when the system is not able to accurately perform multiple tasks beyond a certain number, we say, with an often abused, albeit convenient, computer metaphor, that the system is *working under high computational load*, or, less metaphorically, that it is reaching its processing limits. *Attention* is the main device used by the visual system to cope with its processing limits, since it allows to distribute more resources to those visual processes that have higher priority, in cases in which multiple processes are involved.

Mainstream psychological research assumes that a number of visual routines are performed “rapidly”, “effortlessly” and “automatically” in the sense of “pre-attentively”, i.e. without the need of deploying any kind of attentional mechanism (Neisser, 1967). Among these routines, visual search has become a paradigmatic case-study for the investigation of preattentive vision.

What does it mean to be processed preattentively? According to (Wolfe, 1998b) *any* kind of visual processing of an item prior to the act of attentional selection can be defined “preattentive”. This provides a viable criterion to describe a class of features as *processing primitives* (or *P-primitives*). Properties that are processed preattentively are those properties that can be used to subsequently drive the deployment of attention.

1.4.1 ATTENTION, VISUAL SEARCH AND FEATURE CONJUNCTION

At the beginning of the 1980’s, Anne Treisman and collaborators (Treisman and Gelade, 1980) introduced a distinction between two modes of visual search. The first mode is what they called parallel search for single features, a visual search style characterized by pop-up effect of the target prior to attentional selection. Consider a basic example of visual pop-out:

The RED bar on the left side of figure 1.3 or the HORIZONTAL bar on the right



Figure 1.3: Single feature pop-out – from [Wolfe and Cave \(1999\)](#).

can be said to pop-out since it is not necessary to search them by separately analyzing each item.

Such examples of visual processing can be taken as a paradigmatic case of *automatic, parallel* and *preattentive* routine. Pop-out effects have been used to contrast preattentive and parallel processing with a different style of visual search which is *non-automatic, serial* and *attentional*. Treisman and Gelade claimed that whereas the first mode of visual search affects the individuation of single features, the latter is put to work as soon as a subject is required to detect conjunctions of visual features.

Consider the following example of what they call “feature conjunction” (Fig.1.4):¹¹

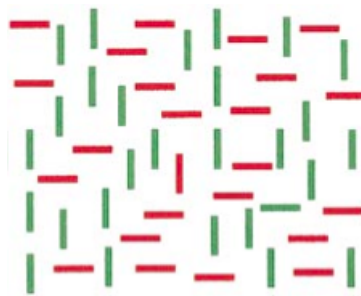


Figure 1.4: Feature conjunction search – from [Wolfe and Cave \(1999\)](#).

¹¹I stress the use of “conjunction” in this context since I argue that the characterization of this condition as an actual *feature conjunction* can be challenged: see the discussion below.

The [RED & VERTICAL] bar in this example does not popup among distractors (that are either [GREEN & VERTICAL] or [RED & HORIZONTAL]). Finding an item instantiating the “conjunction” of two simple features like RED and VERTICAL among other distractor-conjunctions is a task that requires a serial inspection, enabled by attentional selection of single item after single item.

This evidence was used to support the claim that a number of simple visual features, that the visual system allegedly encodes at a very low-level (in distinct feature maps of the primary visual cortex), can be processed pre-attentively, whereas their conjunction (that requires a matching between different feature maps) depends on the intervention of selective attention. Attention has since been considered as the necessary condition for “gluing” or “binding” single features “together”.

Now, this claim and the corresponding theory (*Feature Integration Theory* - FIT) have been challenged along many different lines and have undergone a number of refinements since their first formulation. In particular, several contrast experiments have contributed to reject the somewhat harsh dichotomy suggested by Treisman’s seminal work, showing that:

- (A) *Similarity* between the target item and distractors, regardless of the presence of unique preattentive features, *and the number of distractors both affect response time in conjunction search* (Quinlan and Humphreys, 1987; Treisman, 1988).
- (B) *Many classes of target-distractor configurations that produce pop-out effects are asymmetric*, i.e. search for the presence of a feature is more efficient than search for its absence (Treisman and Gormican, 1988; Treisman and Souther, 1985; Wolfe, 2001).
- (C) *A number of feature conjunctions can be efficiently detected preatten-*

tively. There is large evidence in the literature that a number of feature conjunctions can be detected effortlessly and without the need of serial selection of visual items (See [Green \(1991\)](#) for an early review of such cases).

For Feature Integration Theories, the most uncomfortable of these counter-examples is certainly the latter (c). The fact that certain feature conjunctions can be detected preattentively sheds light on some implicit assumptions in the traditional framework. In particular, (c) can entail two distinct things:

1. *The notion of a single basic feature has to be revised, so as to include stimulus configurations that would be intuitively considered as feature conjunctions.*

Let us recall the two examples above. It is interesting to remark that there is no clear criterion according to which the first task (Figure 1.3) should be characterized as a case of *single-feature detection* while the second a case of *feature-conjunction detection*. What enables us to say that what we label as [RED & VERTICAL] is a conjunction of features, while RED is a simple one? Most researchers would answer that this characterization is consistent with local single-cell sensitivity in the primary visual cortex. But if we leave physiological considerations apart, and remain with the conceptual limits of the notion of feature as operationalized in the preattentive vision paradigm, there is no *a priori* reason to reject the idea that [RED & VERTICAL] could be considered as a simple feature in its own right that cannot be further decomposed for specific kinds of task).¹² Moreover, textural interpretations of preattentive processing ([Julesz, 1981](#)) have seriously threatened the

¹²See the similar point discusses in paragraph 1.2.2 on the case of emergent features.

idea that features that allow efficient preattentive search should not include structured configurations or a subclass of those properties that traditional approaches would consider feature *conjunctions*.

2. *A strict distinction between preattentive and attentional processing must be abandoned.*

Further empirical evidence has led to a progressive refinement of the visual search paradigm and given rise to “Guided Search” theories, in which the *serial vs. parallel* and *preattentive vs. attentional* distinctions are progressively abandoned in favor of a more neutral *efficient vs. inefficient* distinction (Treisman and Sato, 1990; Wolfe et al., 1989). (Wolfe et al., 1989)’s Guided Search (GS) model accounts for the efficiency of many conjunction-search tasks by proposing that preattentive feature processes “guide” the deployment of attention in conjunction search. Attention is drawn to a given item as the result of both bottom-up processing (encoding distribution of featural differences in the stimuli) and top-down processing (expectations driving the subject’s behavior on the target properties).

Generally speaking, the number of variables and possible biases of single experimental protocols on visual search tasks are so complex that it has become virtually impossible to give a unitary interpretation of the whole set of data and draw a sharp line between preattentive and attentional processing, parallel and serial search, single feature vs. feature combination tasks, respectively: to date, there is no single visual search theory that can account for the plethora of data described in the literature (Wolfe, 1998a). Nonetheless, the quest for basic features in visual processing is still attractive and is far from being abandoned. Former criteria based on pop-out effects

have been dismissed as too simplistic to define what count as P-primitive.

1.4.2 WOLFE'S TEST FOR FEATURE BASICNESS

Lacking a clear-cut criterion, (Wolfe, 1998b) has proposed a sort of informal test that should allow one to tell “safely” if a visual property belongs to the set of basic features.

Significantly, he rejects the interpretation (1) of (c), namely the fact that the notion of basic visual features should be weakened so as to allow structured stimulus configurations, like moderately complex textural patterns. “One could propose – he observes – “that these conjunctions have featural status but this seems unparsimonious. It is one thing to propose that there are parallel processors for a set of basic features like color, orientation, size, and so forth. It is something else again to argue for parallel representations of all the pairwise (and, perhaps, 3-way) combinations of that initial list. This rapidly leads to combinatorial trouble”. This statement clearly shows that it is implicitly assumed that a good list of P-primitives should respect some kind of compositionality and unanalyzability constraint (see above, paragraphs 1.2.1 and 1.2.2): basic features should be restricted to a *small* lexicon of moderately unanalyzable properties.

Wolfe surveys in turn a number of plausible candidates for basic features, as well as a number of “dubious” cases. Although he provides no explicit criterion for assessing a visual property’s goodness as a candidate for the set of basic features, his examples suggest that a test of basic feature goodness might consist of the following conditions.

Wolfe's list of basic features

- **Probable featural dimensions:** color; luminance onset; luminance polarity; orientation; aspect ratio; size; curvature; Vernier offset; motion; stereoscopic depth and tilt; pictorial depth cues; shape; line termination; closure; glossiness; number.
- **Doubtful cases:** intersection; lighting direction (shading); novelty; letter identity; alphanumeric category.
- **Probably non-features:** faces; optic flow; color change; 3d volumes; "your name".

As summarized by (Pomerantz et al., 2003), "good" candidates for basic features according to Wolfe are characterized by:

- efficient search (near-flat search slopes, or pop out);
- effortless texture segregation;
- search asymmetries;
- tolerance for distractor heterogeneity.

At face value, the idea of a "basic feature" or P-primitive that Wolfe endorses appears much more restrictive than the notion emerging from this test: in particular, the paradigmatic examples chosen by Wolfe seem to presuppose a first selection of properties as *local*, *moderately complex* and plausibly *low-level* features of the visual stimulus. *Locality* excludes from the set of good properties configurational properties that might span on larger areas or on the whole retinal image and not just on small retinal regions. *Lack of complexity* derives from Wolfe's refusal of textural patterns as good candidates (see above). *Low-levelness* is implied by the fact that basic features should correspond to properties encoded by feature maps in the primary visual areas. It should be noted that Wolfe's list is based on the single paradigm of visual

search, so properties mentioned in this list should be actually regarded as good candidates for visual-search primitives. In no way they are meant to exhaust the set of possible properties of the stimulus that can be processed “early” or “preattentively”. Among “bad candidates”, we find a number of more *global*, moderately *complex* and allegedly *high-level* visual properties (like for instance FACES) for which it has been shown that subjects display *early sensitivity* and efficient preattentive detection (see for instance [Enns and Rensink, 1990](#)).

We have then two quite conflicting, if not opposite views about how to characterize P-primitives.

On the one hand we have a “spurious” set of constraints consistent with most phenomena studied within the visual search paradigm. The impurity of these constraints comes from the fact that:

1. they do not allow for a clear-cut distinction between properties that would count as good basic features and properties that would not;
2. they implicitly endorse a number of independent assumptions (low-levelness, compositionality, complexity) that are not necessarily implied by the notion of earliness.

1.4.3 EARLY SELECTION VS. LATE SELECTION THEORIES

On the other hand, we have an alternative set of constraints that can (at least as legitimately) be used to characterize P-primitives. Models of the kind that I presented so far have been labeled by some authors as “early selection” theories ([Chen, 2001](#)). In what follows, I will try to characterize the alternative way in which “late selection” theories account for P-primitives.

A very common assumption of early selection theories is that basic features should be restricted to properties that count as low-level from the point of view of the functional specialization of the visual system: this implicit assumption results in an unjustified conflation of *early processing* and *low-level processing*. Even if structural constraints like those emerging from the study of the functional specialization of visual areas can restrict the number of possible early visual routines, it is misleading to think that these constraints determine *per se* what processes (and consequently what kind of visual properties) the visual brain performs earlier. As Palmer puts it:

It is tempting to try to translate the terms "early" and "late" into simple brain locations. The problem is that massive backward connections from higher levels to lower levels throughout the visual system make such translation difficult, if not impossible. Processing that goes on in a given area might be functionally either early or late, depending on whether it happens without or with the benefit of feedback from higher levels and depending on the higher levels from which it might receive feedback – [Palmer \(2002\)](#)

Low-levelness as a structural property of perceptual systems provides an independent criterion for assessing primitiveness that should be kept distinct from earliness, unless there are grounded arguments to merge the two criteria (see below, paragraph 1.2.5).

Since low-level and early processed visual features do not overlap, we might try to articulate the relation between low-levelness and earliness, by either maintaining that:

T1 Low-level properties are a subset of early accessible properties

T2 Early accessible properties are a subset of low-level properties

T3 Low-level properties and early accessible properties are two (possibly partially overlapping) sets of features none of which includes the other.

My conclusions, as it will become clearer at the end of this paragraph, argue in favor of (T3). In particular, I will suggest that (T1) should be rejected on the basis of empirical evidence showing that a number of low-level properties are not accessible *at all* or only at *later* stages of visual processing. (T2) should in turn be rejected on empirical grounds, since a number of early accessible properties do not qualify as low-level. Hence, the only viable possibility is (T3).

It is important to stress that even if the set of properties captured by these two criteria were exactly the same, it would still be legitimate to hold a conceptual distinction between the two: a relation of inclusion or substantial overlap could be just a case of extensional coincidence that – lacking any further characterization – can hardly be considered as explanatorily significant as such.

As we will see, the first general conclusion that can be drawn by the following discussion is that low-levelness provides neither a sufficient nor a necessary condition for early accessibility.

Not all low-level properties are early accessible

The first issue (T1) – whether all low-level properties need also be early accessible – is *prima facie* contradicted by the fact that a number of paradigmatic low-level properties (like BINOCULAR DISPARITY, one of the features eliciting selective activities in single cells of the primary visual cortex) do not seem accessible *at all at a behavioral level* (and *a fortiori* not *early accessible*) to a

subject during discrimination or visual-search tasks

This should come as no surprise given that, as I mentioned above, the way in which information encoded by low-level areas is made available to further processing is far from being understood.

Not only there are low-level properties that are simply unavailable to behavioral response and consequently extremely hard, if not impossible, to extract, but there is also a number of preemption cases in which low-level features are harder to extract than not-so-low-level features (i.e., properties that are usually taken as more “expensive” from the point of view of visual processing).

The problem of the relative lack of accessibility of low-level features has been recently addressed by several investigations that showed difficulties met by subjects in performing tasks based on properties that are known to be extracted by the most peripheral areas of the primary visual system. [Rensink and Enns \(1995\)](#) have shown for instance that a number of properties traditionally considered as visual primitives from the point of view of cortical specialization are actually extremely difficult to access:

Features are sometimes thought to include the set of “visual primitives” (i.e. the properties directly obtained from the spatiotemporal filters at the earliest [in the sense of “lowest”] levels of processing). Strictly speaking; however, visual search data provide evidence only of structures that can be rapidly accessed to make a conscious report of target presence. These rapidly accessed structures need not be visual primitives; indeed, some are complex scene-based properties obtained through rapid-interpretation processes. Conversely, there is also no a priori reason for visual primitives to be rapidly accessible.

A strategy that can be adopted to refute (T1) consist then in showing that the universal claim according to which *rapidly constructed visual primitives are always rapidly accessed* does not hold:

Primitives must always be accessible; they can never be preempted by the more complex structures they form. If such preemption exists, it would indicate that search cannot rapidly “reach down” to the lowest level of visual processing. This in turn would imply that the features of visual search correspond not to visual primitives but to structures formed at some higher level of processing –
(p.103)

The fact that there are cases of low-level grouping that preempt simple image measurements like segments (i.e., the grouped items) provides a direct empirical refutation of the universal claim mentioned above: some low-level properties are indeed harder to extract than derived or allegedly less-low-level stimulus configurations.

Not all early accessible properties are low-level properties

The second kind of evidence that can be used to assess the relation between earliness and low-levelness comes from some relatively recently discovered effects of pop-out for high-level properties. This shows that, conversely to what we showed above, nothing prevents high-level properties from belonging to the set of features that can be efficiently processed preattentively (T2).

Consider the following examples:

3D-DEPTH CUES (B) and DEPTH-FROM-SHADING CUES (D) – as respectively studied by [Enns \(1990\)](#) and [Ramachandran \(1988\)](#) – are high-level features that pop-up during visual search tasks whereas their equivalent 2D isoluminant arrangements (A,C) do not (see figure 1.5). High-level features for which

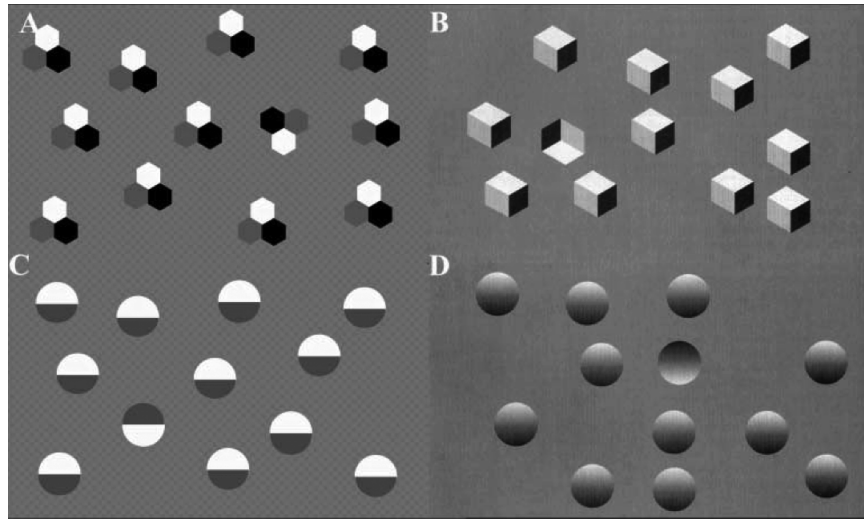


Figure 1.5: Evidence for High Level basis of Feature Search – from [Hochstein and Ahissar \(2001\)](#).

pop-out effects or effective visual search have been demonstrated include LIGHTING DIRECTION, SURFACE SLANT, FACIAL EXPRESSION, and even ONE’S OWN FACE.

What makes these properties high level is not just their intuitive complexity (if compared to the traditional basic feature lists of early selection theories) but the fact that none of them can apparently be described in terms of local image measurements.

Already [Wolfe et al. \(1992\)](#) realized that features that pop-out are better described in terms of categories than as measurable spatial characteristics of the stimuli that are extracted by low-level filters. [Hochstein and Ahissar \(2001\)](#) suggest that what marks them as high-level features is the fact that they are *encoded in high-level areas of the visual cortex*. [Enns and Rensink \(1990\)](#) suggest they are *scene-based properties*, that can only be obtained through image interpretations. In any case, these studies show that the traditional restriction of processing primitives to basic Wolfian features must be abandoned.

Another series of works do not rely on the visual search protocol but rather on rapid visual categorization tasks. They have recently investigated the ability of the human visual system to access in extremely limited time frames global and complex properties of the stimuli. Several studies have shown that whenever human subjects are actively searching for a complex category of objects whose appearance is highly variable, it appears that global and “semantic” (i.e., scene-based) properties can be accessed rapidly and outside the focus of attention (Li et al., 2002). In particular, it has been argued that the “gist” of a visual scene could be available preattentively (Rensink, 2000). Gist-oriented studies have demonstrated that the general meaning of the visual scene (in the sense of a broad categorization of “*what the scene is about*”) can be more easily and efficiently reported than its precise and local details: the visual system, under conditions of high computational load, so to say “guesses” about what is present in the scene by exploiting a small set of visual cues. This “vision at a glance” (Hochstein and Ahissar, 2002) occurs earlier than a “vision with scrutiny” and is assumed to rely on the use of high-level representations instead of low-level image measurements. Attention, in these “late selection theories”, cannot be considered any more as the glue that enables explicit object perception from the conjunction of basic features: on the contrary, attention is required for allowing the visual system to “reach down” to such basic features. What counts as a P-primitive according to early-selection theories is hence incompatible with what late-selection theories claim, namely that detailed features of single items are *not* early primitives, since they are only accessible at later stages of processing.¹³

¹³It is worth noting that some philosophical arguments that largely draw on *early selection theories* in order to empirically corroborate *a priori* hypotheses on perceptual reference (see for instance Campbell, 1997) endorse obsolete assumptions that have been strongly challenged in the psychological literature.

1.4.4 CONCLUSIONS

If we now return to the problem of finding a criterion for selecting good primitives from the point of view of processing stages, we realize that things are far more complicated than a single straightforward test can decide. The first conclusion that can be drawn from the analysis of this section is the necessity to acknowledge the irreducible plurality of visual routines that involve feature selection and extraction. Defining earliness criteria cannot be restricted to a single paradigm as the one provided by visual search, since considering different routines will yield different (and sometimes conflicting) characterizations of earliness: a prerequisite for the study of earliness is then to define a taxonomy of the kind of routines that the visual system can perform and that can affect dramatically the way in which the same kinds of visual properties are processed

More specifically, in this paragraph I have tried to challenge a number of implicit assumptions about processing primitives:

Earliness does not imply low-levelness

Assuming that early accessible features must somehow coincide with visual primitives as characterized by the study of the functional specialization in the primary areas of visual cortex can significantly bias the study of rapid visual routines. Low-levelness is neither a sufficient nor a necessary conditions for picking out early accessible visual properties.

Earliness does not imply simplicity

The idea that only simple visual properties are rapidly available to behavioral response is a hypothesis that should be discarded on the evidence that the visual system is able to rapidly and efficiently extract extremely complex (in

the sense of “semantic”) and highly variable scene properties.

Earliness does not imply locality

As a corollary of (A), it is often assumed that preattentive vision only has access to properties that correspond to local image measurements. Gist sensitivity clearly shows that early accessible properties need not be restricted to local neighborhood features that match the receptive field sensitivity of single-cells.

Earliness does not imply pre-categorical processing

Several cases of preemption effects of structured and category-related stimulus configurations on their simpler components suggest that early accessible properties should not be restricted to properties that can be considered pre-categorical.

In the following section I will address in more details the characterization of low-level primitives that has been partially anticipated in this section for the sake of the exposition.

1.5 LOW-LEVELNESS

One of the most important criteria to determine feature primitiveness consists in identifying visual primitives with those properties that are encoded at the output of sensory transducers: primitive visual features, in this sense, correspond to those properties of the visual stimulation that are encoded in the *early* visual system, i.e. elicit selective patterns of neural activity in the *most peripheral areas* of the visual cortex.

The notion of “earliness”, as we saw in section 1.4, is not completely immune from ambiguity. What counts as early in the visual system from the point

of view of functional anatomy should not be confused with what counts as early from the point of view of behavioral accessibility. However traditionally, the processing of visual stimulation that takes place in the most peripheral areas of the visual system is defined as “low-level vision” (as opposed to mid-level or high-level vision): I hence adopt this notion of *low-levelness* — and accordingly I talk of *low-level primitives* or L-primitives — to introduce the class of constraints on basic features that I intend to articulate in this section.

Although established by an honored tradition of empirical research, low-levelness criteria for visual primitives deserve a careful analysis, not only for the prominent position they occupy in the family of constraints on primitive features, but also for the strong implicit assumptions upon which they impinge: understanding what properties is encoded by a certain signaling process, what is the perceptual role (if any) of such a signal and to what extent such processing is stimulus-driven are some of the delicate issues that I will tackle in this section.

1.5.1 LOW-LEVEL VISION

It might seem arbitrary to assume that there are *low-level* processing stages in the functioning of perceptual systems that substantially differ from others with respect to the definition of *basicness*. Yet, a longstanding tradition (whose most representative formulation is probably to be found in [Marr \(1982\)](#)’s classical work) has defended the view according to which the initial stage of perception can be demarcated as a level characterized by purely bottom-up and stimulus-driven processing.

I will not address in this chapter the longstanding debate about the distinction

between sensory properties and perceptual properties.¹⁴ I will restrict my analysis to a modern reformulation of the idea that part of the perceptual phenomena are characterized by purely *bottom-up*, *stimulus-driven* processes. In the case of vision, this idea is grounded in two hypotheses, defining the goal and nature of low-level visual processing:

- (A) The goal of low-level vision is to perform a set of *measurements* of specific dimensions of the visual stimulation, and hence to signal the “presence” of such features to higher perceptual stages.¹⁵
- (B) Such measurements are purely bottom-up, stimulus-driven and not modulated by the internal “knowledge” of the system.

In what follows, my aim is to flesh out these two hypotheses and their consequences on the main line of my analysis. I will start by framing the crucial notion of *feature detection* mechanisms that is implied by hypothesis (A).

1.5.2 A SHORT HISTORY OF THE NOTION OF FEATURE DETECTOR

Thinking of low-level vision as a system devoted to the measurement of specific stimulus dimensions leads us to discuss the fundamental notion of *feature detection*.¹⁶ Low-level vision can be said to measure the quantity of *P*-ness present in specific areas of the visual field, where *P* is one of the possible dimensions along which the stimulus in that area of the visual field can vary. To quote a classical work by [Lettvin et al. \(1959\)](#),

¹⁴See the introduction to Part I on internal constraints.

¹⁵I use in this context the term “dimension” as a synonym of “feature” to refer both to properties that can vary on a continuum of values and properties that can only assume one value among a discrete set of possible values. In this sense, I do not follow [Treisman \(1986\)](#) who restricts the notion of “dimension” only to the first kind of properties.

¹⁶For a historical survey on the origins of the notion of *feature detection*, see [Martin \(2000, 1994\)](#).

[w]hat, then, does a particular fiber in the optic nerve measure?

We have considered it to be how much there is in a stimulus of that quality which excites the fiber maximally, namely that quality.

A mechanism selectively triggered by the presence of X in the visual stimulation is usually described as a *detector* of property X or a X -*detector*.

Whether the activity of a feature detector bears any relevance to the understanding of perceptual phenomena and, more specifically, whether there is any significant relation between properties detected by neurons (or neural populations) in low-level areas of the visual system, on the one hand, and perceptually salient properties, on the other hand, is an issue that has dramatically oriented modern research in perceptual science. An answer to such question, implicit in most modern neurophysiological studies of perception, is provided by what Barlow (1972) called a *neuron doctrine for perceptual psychology*. Such a doctrine has become one of the main theoretical tenets of the first scientific attempts to understand the relation of neural activity in the cerebral cortex to perceptual capabilities. The doctrine can be summarized in two independent statements that establish a methodological link between the study of the neurophysiology of the visual cortex and the psychology of vision.

From perceptual discrimination to neural activity

The first assumption of the *neuron doctrine* establishes a *reducibility relation*, i.e. that any perceptual discrimination must be grounded in differences of neural activity at the level of single cells. The original formulation of this hypothesis is due to Horace Barlow who called it the *Psychophysical Linking Hypothesis*:

Psychophysical Linking Hypothesis

Whenever two stimuli can be distinguished, in normal life or in a psychophysical experiment, then proper analysis of the impulses occurring in *a single neuron* would enable them to be distinguished with equal or greater reliability – Barlow (1985a, p.134)

This hypothesis states that behavioral discrimination of visual properties must be reducible to differences in patterns of neural activity at a certain level of analysis. Barlow was certainly not the first one to assume that such a link between perceptual capabilities and underlying neural activity in the visual cortex should exist. The first discoveries on the relation between sensation and patterns of neural activity in cortical cells date back to the late 20's and to the investigations of Barlow's predecessor E. Adrian ([Adrian 1928](#), cit. in [Martin 2000](#)), to whom we owe one of the most fundamental principle of modern neurophysiology, i.e. that the intensity of sensation is proportional to the frequency of sensory nerve impulses.

From neural activity to perceptual content

The second assumption at the basis of the *neuron doctrine* establishes the converse *explanatory relation*, i.e. the significance of patterns of activity at the level of single cells for explaining perceptual capabilities. In the remainder of this section, I will focus on this explanatory relation that lies at the core of Barlow's neuron doctrine and that can be summarized as follows:

- (a) patterns of activity in single neurons in the primary visual cortex signal the presence of visual properties in their receptive field;
- (b) such property-selectivity determines the perceptual role of the neuron in the architecture of the visual system;

Since the early 1930's, empirical investigations on the functional anatomy of the visual cortex have been directed to understanding the meaning of patterns of activity in single neurons.

The first formulation of a modern notion of “feature detection” goes back again to the seminal work of H. Barlow in the 1950's and to the contemporary research of several groups of neurophysiologists. When Hubel and Wiesel (1959) reported the sensitivity of their cells to the orientation of edges it seemed natural to dub these cells “edge detectors”. In the same year, the notion of “feature detection” was brought to prominence thanks to the influential paper by Lettvin, Maturana, McCulloch, and Pitts (1959), which put forward the thesis according to which patterns of activation in low-level visual neurons signal the detection of specific visual properties. The conclusion reached by Lettvin and collaborators after studying the specific response profile of single cells in the frog's retina, was that specific cells could be considered *bug detectors* because of their selective detection of *bug-related features*. It is worth reporting the conclusion of the article, for its theoretical interest:

The operations [performed by the frog's retina], thus, have much more the flavor of perception than of sensation if that distinction has any meaning now. That is to say that the language in which they are best described is the language of complex abstractions from the visual image. We have been tempted, for example, to call the convexity detectors “bug perceivers”. Such a fiber responds best when a dark object, smaller than a receptive field, enters that field, stops, and moves about intermittently thereafter. The response is not affected if the lighting changes or if the background(say a picture of grass and flowers) is moving, and is not there if only the background, moving or still, is in the field.

Could one better describe a system for detecting an accessible bug? – [Lettvin et al.](#), cit., p.253-254

Two crucial facts emerge from this conclusion:

- the first fact is that patterns of neural activity can legitimately be described in the *language of perception*, i.e. the language of “complex abstractions from the visual image”;
- the second fact is that the perceptual role of neurons is determined by specific constraints on the environment of the organism.¹⁷

These two considerations, together with the discoveries from which they were drawn, cleared the path to a major trend of neurophysiological studies aiming at understanding what properties in the visual stimuli *maximize* activity in single cells. If the profile of activity of neurons in the visual cortex is determined by the visual properties of the stimulation to which they preferentially respond, it becomes empirically possible to study the specific visual features single neurons are tuned to.

1.5.3 BEYOND FEATURE DETECTORS: MULTI-DIMENSIONAL TUNING AND OVERLAPPING MAPS

The feature detection paradigm enjoyed a long fortune and was essential to foster the development of the first empirical investigation in the microarchitecture of the brain. Current neurophysiology has got rid of the somewhat simplistic assumptions of Barlow’s feature detection model since a long time. One of the first impasses encountered by this old methodological paradigm, when applied to studying the functional specialization of the visual cortex,

¹⁷This second aspect, i.e. the *environmental closure* of perceptual systems, will be extensively discussed in Part II

has been how to identify the *most effective* properties that maximize activity in single cells. It became evident pretty early that neurons in the visual cortex respond differently to different stimulus dimensions and there is no straightforward strategy for determining which among the infinite number of possible visual properties are likely to maximize patterns of activity in individual cells. Moreover, single neurons can be tuned to stimulus dimensions in different ways: they can signal the presence of a certain property in a portion of the receptive field (like differences in intensity between the center and the periphery of their receptive field), signal analogically the value of a certain property (like *edge orientation*), or even the degree of similarity of a visual pattern to a certain property.

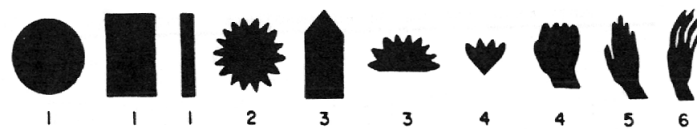


Figure 1.6: Stimulus effectiveness in single cell recordings of the macaque visual cortex – From [Gross et al. \(1972\)](#)

In figure 1.6, an example is reported by [Gross et al. \(1972\)](#) (cit. in [Barlow 1985a](#)), in which the number under each object shows the authors' subjective assessment of the strength of the response the object evoked in single cells of the macaque's cortex. This response increases from left to right, as -allegedly- similarity to a macaque hand increases.

Neurophysiological investigations of the visual cortex have provided so far massive data about the specialization of specific neural populations.¹⁸ Perhaps the most significant discovery in the study of neural specialization consists in the discovery that neurons tuned to code for the same visual feature are organized in the mammal visual cortex in *retinotopic maps* (i.e., maps that

¹⁸For a recent review of studies of functional anatomy of the visual system see [Lennie \(1998\)](#)

preserve the topological structure of the retina), thus suggesting that each map may provide an independent description of the visual scene. An interesting aspect of feature maps is the fact they *overlap* in the primary visual areas of the mammal visual cortex: the same neurons can be differentially tuned to specific values of *multiple visual dimensions*. This has been taken as evidence supporting the hypothesis that the goal of low-level vision is to perform a *multi-dimensional analysis* of retinal stimulation. Figure 1.7 illustrates the idea of multi-dimensional analysis by showing the overlap of different features to which single neurons of the cat's visual cortex respond. (A) displays the sensitivity to OCULAR DOMINANCE (the gray and white blobs) compared to sensitivity to EDGE ORIENTATION (the colored lines), while (B) displays sensitivity to OCULAR DOMINANCE compared to sensitivity to low SPATIAL FREQUENCY (the dark gray blobs) in the same neural population. Each neuron of the considered population, hence, codes for a certain value of SPATIAL FREQUENCY, OCULAR DOMINANCE and EDGE ORIENTATION.

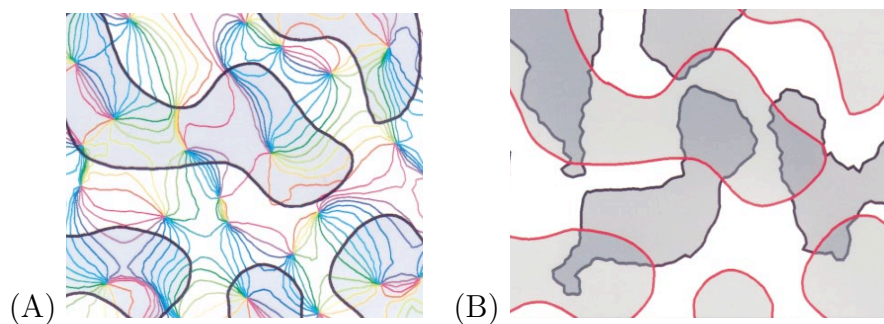


Figure 1.7: Overlapping feature maps in the cat's visual cortex – from [Hubener et al. \(1997\)](#)

Although many explanations have been given for the existence of overlapping feature maps and neurons performing multi-dimensional analysis,¹⁹ this phenomenon raises some theoretical issues that are relevant for the present inquiry.

¹⁹The most important of which has suggested that cortical organization in overlapping maps is a matter of optimization of spatial coverage, (see [Swindale, 2000](#))

The *psychophysical linking hypothesis* has been discarded as an inappropriate characterization of the functional role of single cells. A weaker and more prudent version of this hypothesis, though, is still implicitly at work behind the studies of cortical specialization of specific neural circuits, insofar as they require the understanding of what aspect of the stimuli is responsible of the measured patterns of neural activity. What count as independent stimulus dimensions need to be established in advance at a behavioral level in order to test what distinct stimulus dimensions a specific neural populations effectively responds to. Moreover, the fact that the same cells or cell populations might respond selectively to two independent stimulus dimensions might suggest a revision of the behavioral criteria used to determine relevant stimulus dimensions. Consider, for example, a single neuron responding selectively and with the same intensity to the presence in its receptive field of either property P or property Q . Shall we characterize the effective feature for this neuron as the $P \vee Q$ disjunction, or rather decide to revise our previous taxonomy of properties and introduce in our set of candidates for basic visual features a new property R replacing the above disjunction? This oversimplified example is meant to show that even the most elementary hypothesis about the existence of detectors in the primary visual cortex for a given dimension of the sensory stimulation is *theory-laden*, in that it depends on a previous characterization of what counts as independent dimension: establishing independent stimulus dimensions, hence, requires the adoption of criteria for feature goodness.²⁰

1.5.4 WIRING AND HIERARCHIES

Accurate descriptions of the featural preferences of different visual areas have been recently integrated with the study of cortical *connectivity*, which

²⁰The risk of circularity in the definition of “feature” (as “whatever property can be encoded by a feature detector”) was first pointed out by [Koenderink \(1993\)](#)

has provided evidence for the hierarchical relations between distinct areas and suggested the existence of distinct stages and autonomous processing streams (“visual pathways”) within the primary visual cortex (see [Felleman and Van Essen, 1991](#); [Zeki and Shipp, 1988](#)).

It is tempting to interpret the existence of cortical areas with selective featural preference and particular connectivity relations to other areas as straightforward *visual modules* (see [Burr, 1999](#); [Swindale, 1998](#)), i.e. anatomically isolable systems subserving specific operations in the functional architecture of vision. In past years, such evidence has been used to provide alleged “direct demonstrations” concerning the localization of perceptual processors in the visual cortex and to ascribe to anatomically isolable populations a dedicated functional role.²¹

Yet some care is needed in inferring from this kind of evidence hypotheses about the functional organization of the visual system. The general picture presented so far, according to which the visual system is organized as a *hierarchical processing system* for the extraction of properties of increasing complexity, on the basis of a first level of basic feature detection, has been challenged in a number of ways. In the next section I address such criticisms, in order to clarify whether the notion of feature detection and of L-primitives as characterized so far is tenable.

1.5.5 THE PERCEPTUAL ROLE OF NEURAL ACTIVITY IN LOW-LEVEL VISION

The idea outlined in the previous section can be summarized as follows:

- (a) the visual system can be globally seen as a hierarchically organized processing system whose aim is to analyse the visual scene at different

²¹The most paradigmatic case is the claim that the area responsible for color perception had been identified in the visual cortex (see [Lueck et al., 1989](#); [Zeki et al., 1991](#))

levels of complexity;

- (b) the lowest level of the visual systems consists of families of cells whose goal is to provide a description of the distribution of basic features (like ORIENTED EDGES) in the visual scene;
- (c) higher order cells depend on low-level cells for their input and are dedicated to integrate information from low-level cells to form more abstract and location-independent neural representation of the visual scene (like SHAPE);
- (d) at the apex of the visual hierarchy, populations of cells fire to signal the presence of specific visual configurations before the eye, thus mediating recognition of visual objects or extraction of complex elements of the visual scene.

Although still defended by a number of authors (see, for instance, [Shadlen and Movshon, 1999](#)), such view has been recently challenged by many, on both empirical and theoretical grounds. It is not my aim to provide here an accurate survey of the different kinds of criticisms that have been raised against this view. Instead, I will select some of the criticisms that have had a particular deflationary role towards the hypothesis that low-level vision is devoted to the description of the visual scene along a number of distinct basic featural dimensions. Challenging the fact that neurons code for specific features of the visual stimulation has major consequences on the claim that basic features are those described by patterns of cortical sensitivity of primary visual areas.

Perceptually significant activity occurs not only at the level of single cells

One of the major criticisms against what has been called the “hegemony of the single neuron” (Douglas and Martin, 1991) is the claim that neural populations, more than single cells, are the basic functional units of visual processing. Theories of *population coding* have emphasized the importance of operations performed by families of neurons and stressed how the attempts to synthesize from single-neuron measurements confront one with the dilemma that much of the recorded activity is irrelevant to the behavior studied. Redefining basic features as those properties that are encoded by neural populations — instead of single neurons of the primary visual system — does not really threaten the hypothesis of the existence of L-primitives. Yet, the consequences of this approach on the traditional assessment of the perceptual significance of feature maps for the understanding of what counts as “basic” should not be overlooked: if neurons are not any more the units of perceptually-relevant activity, and they are massively modulated by the activity of neighboring cells, it becomes much harder to establish the precise architectural role played by a cortical area within the visual system.

Response patterns of a neuron have no direct link to its perceptual role

In even stronger terms, MacKay (1985) has argued against the idea that studying response profiles of visual neurons in the primary visual cortex might provide an explanation of their *perceptual role*. Instead of thinking that the significance of “feature sensitivity” in the visual system consists in a primitive description of the visual scene in terms of the firing rates of distinct classes of feature detectors, he suggested that alleged “feature sensitivity” might emerge from neurons tuned to the *co-variation of sensory stimuli* across

ocular movements and other kinds of motion schemes.²² This means that low-level visual neurons, far from providing a “symbolic description of the visual scene”, should be seen as devices that are meant to guarantee the stability of the visual scene over transformations, by signaling *what co-varies with what* or the way in which parallel sensory signals *co-vary* with each other and with ongoing motor activity.

Perception is represented by the activity of updating the conditional state of the internal organization to match current sensory data. (...) The function of visual information-processing cells is not to ‘name’ the stimuli (...) The task of symbolizing the perceived world could well be a more central process, which requires from the sensory system not pictures or descriptions but an array of selective clues to help it “home in” on the appropriate conditional readiness to reckon with that world. If so, we must be prepared to look for quite different kinds of link between striate cortical activity and pattern recognition — [MacKay](#), cit. p.50-51.

The deflationary impact of MacKay’s position on the principle according to which neural activity of single cells can be interpreted if not in terms of feature detection, at least in connection with some aspect perception, is hardly negligible. If patterns in the activity of single neurons are related to the detection of covariant sensory events, then the hypothesis that their response profile can be linked to a definite perceptual role is seriously undermined.

²²For a detailed discussion on sensorimotor constraints on feature selection, see Chapter 6

The idea of receptive-field content as phenomenally relevant is flawed

The assumption according to which featural preference in patterns of neural activity of single cells is relevant for understanding their perceptual role (and might hence contribute to the definition of basic visual features) has also been recently criticized on a more theoretical level by some authors, like [Noë and Thompson \(2003\)](#).

Noë's argument starts by addressing the claim made by [Teller and Pugh \(1983\)](#) that any explanation of perceptual phenomena based on neurophysiological data should consist of *linking propositions* establishing a normative relation between particular patterns of neural activity and particular kinds of perceptual content. [Teller and Pugh \(1983\)](#) write:

Most visual scientists probably believe that there exists a set of neurons with visual system input, whose activities form the immediate substrate of visual perception. We single out this one particular neural stage, with a name: *the bridge locus*. The occurrence of a particular activity pattern in these bridge locus neurons is necessary for the occurrence of a particular perceptual state; neural activity elsewhere in the visual system is not necessary. The physical location of these neurons in the brain is of course unknown. However, we feel that most visual scientists would agree that they are certainly not in the retina. For if one could set up conditions for properly stimulating them in the absence of the retina, the correlated perceptual state would presumably occur — [Teller and Pugh](#), cit., p. 581

The assumption of neural-perceptual bridge laws, that applies as a general premise to most empirical investigations on the neural correlates of conscious

perception, lies also at the core of those studies which aim to identify the *featural preference* of specific classes of neurons as the condition for explaining the *perceptual experience of that feature*. Noë and Thompson discuss a number of reasons why *receptive-field content* and *perceptual content* are incommensurable and why a simple methodological constraint (as the necessity of studying isomorphic relations between some aspects of neural processing and some aspects of perceptual experience) has often been turned into an explanatory strategy (what they call *Matching-Content Doctrine*). Noë and Thompson attack in particular alleged neural explanations of perceptual experience as the following:

“Subject *A* is visually conscious of a red patch in the upper-left quadrant of her visual field *because* a neuron in her visual cortex is firing whose receptive-field overlaps with the position of this red patch”

The conclusion drawn from the review of a case studies of alleged neural explanation of perceptual experience (*binocular rivalry*) is that the idea of the existence of a *neural content* (or, more specifically, of a *receptive-field content*) has often been misused to formulate pseudo-explanations of what constitutes the perceptual experience of a given property. Much as Mac Kay’s argument, this hypothesis on the significance of neural activity could threaten the validity of most functional explanations of perceptual skills based on the idea of that pattern of neural activity can be directly matched with specific kinds of perceptual content.²³

²³I will articulate and develop a very similar argument in my criticism of the *Feature binding problem* in Chapter 3

1.5.6 CONCLUSIONS

In this section, my aim was to frame a notion of primitives based on constraints of neural sensitivity at the lowest stages of visual neural processing — which I called *L-primitives*. I introduced here a class of constraints (*low-levelness*) that is commonly evoked to provide an alternative characterization of what counts as a basic feature and I have outlined the general theoretical framework (whose historical origins can be traced back to Barlow’s “neuron doctrine for perceptual science”) within which such constraints can be applied. I have suggested that although endorsed by an established tradition of empirical research, such theoretical framework is threatened by a number of controversial assumptions on the alleged *perceptual role* of pattern of activities in single neurons and neural populations. The very assessment of what counts as a primitive feature from the point of view of neural sensitivity seems hardly separable from a number of assumptions on what perceptual role (if any) can be attributed to patterns of neural activity. The outcome of this analysis is the claim that if anything like an L-primitive exists, its relevance as a theoretical notion for the explanation of perceptual processes depends on the endorsement of some strong methodological assumptions. The traditional notions of *receptive field content*, *feature detection* and *feature maps* are embedded in a specific theoretical framework that has been criticized in recent years under many respects, both empirical and theoretical. This is not meant to underestimate the results that have been possible thanks to this methodological background, but to point out a number of possible shortcomings of explanations of perceptual phenomena that take for granted the idea that features to which neurons in the primary visual cortex seem to be tuned should be considered as the set of basic properties upon which a description of the visual scene is built. I will present in Chapter 3 a

paradigmatic case of explanatory strategy that assume that basic features should be restricted to those properties that are encoded in low-level vision's feature maps and I show how some explanatory weaknesses are likely to emerge from a poor or inconsistent characterization of what counts as a visual primitive property or from an uncritical endorsement of the assumptions implied by the notion of L-primitive.

1.6 CONCLUSIONS: DISTINCT CRITERIA FOR PRIMITIVENESS

I have reviewed in this chapter the formulation of five distinct criteria available in the literature that concur to the characterization of the notion of *primitive features*. The outcome of this analysis has been an articulated clarification of the distinct and irreducible dimensions of what constitutes *perceptual primitiveness*. The goal of the following Chapter is to address a second family of internal constraints on basic feature: a family of criteria that assume that perceptually relevant properties of the sensory stimulation are those susceptible of being ascribed to single visual items.

Chapter 2

Ascribability

VISUAL FEATURES AS ASCRIBABLE PROPERTIES

A WIDESPREAD TREND in perceptual science has focused on the study of visual features as *attributes* of individual entities. It seems natural to assume that as soon as a property is processed by the visual system, it must be encoded as a property *of some entity*. Being able to refer a property to an entity seems a major requirement for the functioning of perceptual systems: object recognition and identification, perceptual judgments regarding part-whole relations, as well as the ability of building incremental representations from serially scanning different parts of the visual scene are admittedly capabilities that draw on referential mechanisms to pick up individual entities and ascribe properties to such entities.

Instantiability or *ascribability conditions*, I assume, constitute together with primitiveness conditions the second major class of internal constraints on what count as a basic feature for perceptual systems. The aim of this chapter is to explore the two main sets of ascribability constraints that occur in the literature, the first related to features as attributes of **spatial locations**, the second to feature as attributes of **visual objects**. The outcome of this

analysis will be an assessment of the notion of localizable-features and object-related features, their use in the literature and their mutual relation. In Chapter 3 I will draw on this characterization, as well on the characterization of basic features as *primitives*, to address some issues raised by the use of “basic feature” as a theoretical notion.

2.1 LOCALITY

The first class of ascribability-related constraints that I address in the present chapter concerns *spatial location*. The fact that perceptual features can be ascribed to spatial locations and hence dealt with as *attributes of spatial locations* is an idea with a longstanding and honored tradition, both in psychology and in philosophy of perception. In the philosophical tradition, *feature placing* – an organism’s ability to ascribe perceptual attributes to spatial locations – has been considered as one of the major requirements for any system able to build accurate representations of the world (Clark, 2000; Strawson, 1959). We can say that a system lacking the ability to represent the spatial location of specific perceptual attributes (as it is the case in specific neuropsychological syndromes) would be a system incapable of object identification and recognition, two of the core functions of perceptual systems. The ability to single out spatial locations and to ascribe to them perceptual properties seems a necessary requirement to qualify a system as a *perceptual* system.

Spatial location and spatially-localizable visual attributes have played a prominent role not only in philosophy of perception but also in scientific explanations of perceptual abilities. The centrality of feature placing as one of the most fundamental mechanisms in perception could not be expressed better than by quoting David Marr’s famous statement according to which

the main goal of vision is to specify *what is where*:

What does it mean *to see*? The plain man's answer (and Aristotle's, too) would be, to know *what is where* by looking. In other words, vision is the process of discovering from images *what* is present in the world, and *where* it is – Marr (1982, p.3).

Assuming that perceptual systems are designed to inform an organism about the instantiation of specific properties at specific locations, one of the major goals of modern perceptual science has been the investigation of two orders of mechanisms: those that allow an organism to succeed in *perceptual selection of spatial locations* and those that mediate *feature placing tasks*. Clark (2001) has suggested that the existence of these two kinds of mechanism is motivated by their distinct logical roles: on the one hand – he argues – we need *proto-referential mechanisms*, i.e. mechanisms enabling an organism to single out individual entities (like a single spatial location of the visual field); on the other hand, we need *proto-predicative mechanisms* that allow the ascription of properties to the selected entities. Neither mechanism alone – he argues – is sufficient to account for the complexity of feature placing behavior. Puzzling as it may seem, the dominant paradigm in the last decades has acknowledged the existence of these two orders of issues, but has assumed that a single mechanism is sufficient to address both problems: *selective attention*. Space-based attention (Posner, 1980) has become the key to understand both how organisms manage to refer to spatial locations in the perceptual field *and* to retrieve properties instantiated at such locations.

In order to understand the nature of proto-referential and proto-predicative capabilities mediated by selective attention, I will briefly recall the framework that is traditionally adopted to characterize pre-attentive as opposed to

attentive vision.¹ Even if selection of spatial location is considered as the result of the deployment of attention, pre-attentive vision cannot be considered as a processing stage that totally lacks spatial features. Rather, pre-attentive vision qualifies as a sort of raw measurement of the spatial distribution of multiple dimensions of the stimuli. Local discontinuities in these measurements are the reason that explains pop-up effects and peculiar efficiency in visual search routines (see figure 1.2 of the previous chapter). A pop-up effect, in this sense, trivially allows a subject to refer to a specific location of the perceptual field and to ascribe a specific property (the property triggering the pop-up) to that location. We have seen, though, that this ability decreases as soon as the distribution of features does not allow easy segmentation and the task demands get more complex: the individuation of items characterized by *conjunctions* of specific features requires a serial “scan” of locations, that is assumed to be mediated by selective attention. In the general case, in order to access the location of a visual target and to perform proto-predicative operations, attentional mechanisms must be activated. Selective attention is then the condition that makes spatial locations accessible to the perceptual system and prone to proto-predication. The relation of selective attention to spatial location, though, is far from straightforward. On the one hand, spatial locations (e.g., local featural discontinuities in the visual scene) *pre-exist* to the deployment of attention, since they are precisely what attention *selects*, and can be accessed, in some particular cases, independently from attention (as in pop-up conditions). On the other hand, once attention has parsed specific spatial locations, further operations on these locations (like ascription of feature conjunction) become possible. How are we then to characterize the difference between the (implicit) spatial character of preattentive vision

¹This analysis has already been introduced in sections 1.4 and 1.5

and the (explicit) spatial character of attentive vision? I will tackle this issue – that is crucial to understand the specific kinds of constraints imposed to perceptual features – by introducing two distinct notions of locality for features: their *intrinsic locality* and their *functional locality*.

2.1.1 INTRINSIC LOCALITY

Intrinsic locality pertains to the fact that features are extracted by the visual system as properties of specific and delimited portions of the visual field. This amounts to say that the first draft of the visual scene built by the visual system consists of *local* measurements of Q_{ness} , where Q is any of the separable dimensions that can be detected by neurons in the primary visual system and *local* is defined in relation to a specific neighborhood (in the case of vision the size of receptive fields of single neurons). We have seen in section 1.5 that low-levelness criteria add specific constraints on the nature of properties that can be detected by neurons in the primary visual system. Intrinsic locality adds a further constraint on basic features:

Intrinsic locality constraint on basic features

Basic features are features that can be extracted as properties of local neighborhoods by feature detectors in the primary visual cortex.

This constraint restricts basic features to those properties that elicit selective patterns of activity of single neurons of the primary visual cortex and that can be extracted as *local properties* of receptive fields. It is in virtue of this constraint that neurophysiologists have been able to describe basic features as those stimulus dimensions that are recorded in feature maps in the primary visual cortex. Without an intrinsic locality constraint, it would make little sense to define maps of features that encode the spatial distribution in

the visual field of specific stimulus dimensions. Only (intrinsic) local properties can by definition be recorded by feature maps.

Since intrinsic locality means that the presence of a specific local feature is encoded at a given position of the visual field, it does not mean yet that the system can access that location and make information about that location available to further processing. As [Briand and Klein \(1989\)](#) have pointed out,

[t]here is a big difference between the *registration* of features by peripheral visual channels and the *detection* of those features, both of which can be considered encoding processes.

Simply *being recorded at a given location* does not grant the system the ability to refer to that location (except from cases of major discontinuities in feature distribution that cause pop-up effects). A feature map contains large quantities of information sustaining the spatial discriminability of the features it registers, but not yet the relation of features to spatial locations. Intrinsic locality is functional to particular forms of spatial coding that have been selected by evolution to solve specific problems raised by visual processing. Retinotopy – the fact that feature maps preserve the topology of the retina so that adjacent receptive fields in the retina are encoded in adjacent positions in the feature map and non-adjacent fields by non-adjacent positions – is possibly one of the smartest examples of strategies that exploit locality as a key to optimize encoding ([Swindale, 2000, 1998](#)) or to optimize computation of relations between featural information relevant to the same location ([Green, 1991](#); [Van der Heijden, 1995](#)). But again, *spatial coding* (or coding by spatial position) need not have any particular relation to coding *of* spatial relations ([Wolff, 2004](#)). It is controversial whether intrinsic locality plays any role in mediating perceptual judgments about a feature being located at a specific region of the visual field: such judgments require a second notion of locality,

irreducible to intrinsic locality.

2.1.2 FUNCTIONAL LOCALITY

I call this second notion of spatial location *functional locality*, since it deals with the capacity of a system to process a feature *as an attribute* of a specific location. Functional locality encompasses intrinsic locality in that it makes explicit the fact that a specific property can be *ascribed* to a specific spatial location. Whereas properties that are intrinsically local are simply local due to an accident of the encoding system, properties that are functionally local are properties that the system represents as attributes of particular entities. More specifically, functional locality allows to formulate a locality constraint on feature selection that is logically independent from the intrinsic locality constraint:

Functional locality constraint on basic features

Basic features are features that can be ascribed by the observer to delimited spatial regions of the visual field.

The logical independence between intrinsic and functional locality is evident if we consider that intrinsic locality is neither sufficient nor necessary for functional locality:

1. Features that are parsed by the primary visual system as properties of local neighborhoods defined by single-neuron receptive fields may not become available to the observer as attributes of specific locations of his visual field, and hence not be functionally local.
2. Properties that do not match the intrinsic locality constraint can still be considered as local features according to the functional locality con-

straint, provided they can be handled as attributes of spatial locations of the visual scene for the observer.

More particularly, not all functionally local properties can be encoded in feature maps: feature maps encodability is related to intrinsic locality, i.e. the arbitrary structure of encoding devices in the primary visual system. But functionally local properties include properties that the observer can ascribe (at different scales) to portions of the visual field, not just to locations defined by receptive fields of neurons of the primary visual cortex.

Although intrinsic and functional locality should be regarded as two logically independent issues, it is somewhat surprising that they have often been considered as two complementary aspects of the very same problem in the study of early vision. The idea, put forward by *Feature Integration Theories* (FIT) and largely shared in mainstream perceptual psychology in the last decades, can be resumed as follows: preattentive vision feeds a number of feature maps with measurements of the local distributions of specific stimulus dimensions in the visual field. At this level, the visual system only possesses a number of rough drafts of the distribution of properties in the visual field, and it has no access to the specific featural content of a given location. In order to mediate perceptual judgments of the form *spatial location X has property P*, two conditions must be met: 1) a specific spatial region must be selected and 2) the relevant features need to be “tagged” as attributes of *that spatial region*. Why these operations are problematic can be illustrated through a famous example, first introduced by Frank Jackson ([1977](#)), and known as the *Many-Property Problem*.

The Many-Property Problem

Let us imagine an organism endowed with some basic sensory discrimination skills but unable to establish a link between sensory qualities and spatial

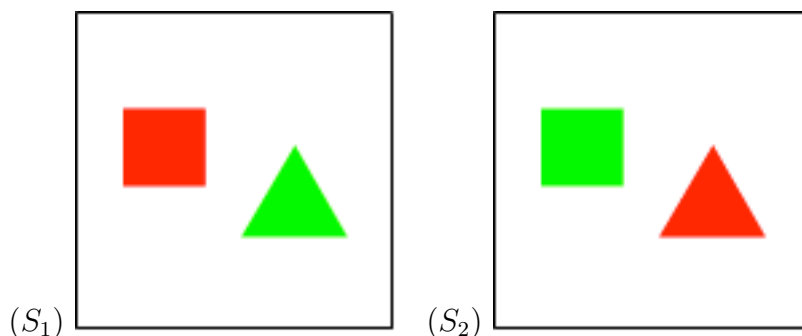


Figure 2.1: Frank Jackson's *Many-Property* Problem

locations. Now suppose this organism is displayed a scene – let's call it S_1 – containing a red square (R,S) and a green triangle (G,T). The organism in question will certainly be able to detect the presence of something red *and* something square *and* something green *and* something triangular. But in no way it will be able to detect that there is a particular entity that is both red and square and that is distinct from a second particular entity that is both green and triangular. To put it differently, this organism will not be able to detect any difference between scene S_1 as described above and a second scene – let's call it S_2 in which a red triangle (R,T) and a green square (G,S) are presented. In both cases, our organism will be able to detect the presence of something red *and* something square *and* something green *and* something triangular. But lacking the capability to tag some of these sensations as belonging to a particular entity and some other sensations as belonging to another particular entity, our organism will not make any distinction between S_1 and S_2 .

Now, it seems that any creature capable not only of elementary sensory discrimination but basic perceptual skills (like perceptual identification and recognition of particulars) needs to be able to solve the Many-Property problem. If, say, predators are identified as conjunctions of REDNESS and HAIRINESS (but not through REDNESS or HAIRINESS alone), avoiding predators requires the ability to parse individual entities with *both these qualities*, i.e. the ability to perceive both features as attributes of the same entity (the same

visual object or the same portion of the perceptual field). We can even imagine that two entities (a predator and a non-predator) are not discriminable on the basis of their observable features, but only by discriminating the spatial distribution of their features. Structural descriptions of the relations between different parts of an object require the ability to ascribe specific attributes to specific locations, lacking which an organism will not discriminate between innocuous animals with red paws and hairy nails and dangerous predators with hairy paws and red nails.

The problem that FIT have tried to address is, in other words, how to get from *features* encoded by feature maps to *individuals*, in particular to entities that can be identified as conjunctions of multiple features. FIT proposes that the perceptual system solves the problem of the correct ascription of visual features to individuals (a version of the *Feature Binding Problem*, hereafter FBP) by driving selective attention on spatial locations. As Attneave (1974, p.109) put it, in one of the first modern formulations of the feature binding problem, this is tantamount to asking “*where* is the glue that holds quite different *what* properties together”.

Considered from a different perspective, the *Feature Binding Problem* can appear like a processing limitation problem with interesting functional consequences. Due to processing limits, the visual system can only process feature conjunctions (that require visual attention) by “packing” them into units through individual attentional scans: it cannot process multiple feature conjunctions at once. The consequence, according to defendants of FIT is that attention as the mechanisms mediating object awareness can be spread over a limited number of items.² Selective attention, in this sense, subserves

²Evidence that feature conjunction under high attentional load is very poorly performing and gives rise to *illusory conjunction* has been used as a rationale for *feature integration theories*, see Treisman and Schmidt (1982)

at the same time two distinct functions: a proto-referential function (it allows locations to be selected) and a proto-predicative function (it allows conjunction of features to be ascribed to the appropriate item). The latter point is captured by [Clark \(2000\)](#), who observes that:

while features are general terms, open to multiple instantiation, binding requires singular terms, the picking out of places. The work of binding is the work of identifying the subject matters of the various feature maps. It is not conjunction, but rather joint predication (p.16).

Given this characterization of the role binding plays in allowing feature integration or spatial anchoring of features, we can now move to some of the major criticisms raised against this model. FIT have been the object of a large debate between the 1980's and the 1990's. The debate has addressed many levels:

1. the *rationale* of the problem: *why is there a feature binding problem to be solved?*
2. the putative *solution* to the problem: *what are the mechanisms that allow the brain to solve the binding problem?*

Both the rationale and the specific solutions to the feature binding problem have been challenged on different grounds by several authors (see for instance [Garson, 2001](#); [O'Regan and Noë, 2001](#); [Rensink, 2000](#); [Taraborelli, 2002](#); [Van der Heijden, 1995](#); [Wolff, 2004](#)). In what follows, I will review some of the main arguments that have direct consequences on the issue of basic features and their ascription to spatial locations.³

³For a more comprehensive discussion of the FBP, see Chapter 3.

2.1.3 NON-LOCAL PROPERTIES

I have already mentioned (section 1.5) that the idea according to which basic features should be restricted to properties detected by single cells in the primary visual areas has been criticized on many grounds. Similar criticisms apply to the idea, put forward by FIT, that basic features should be restricted to the local measurements of stimulus dimensions encoded in feature maps. There are many examples of features that elicit selective activity in the primary visual areas and that are not reducible to properties of local portions of the visual field.

In a comparison between feature placing theories and visual indexicality theories (see section 2.2 for a further discussion) [Clark \(2004\)](#) observes that certain properties can indeed be ascribed to individuals but not to spatial locations: *direction of motion*, *motion trajectory*, “*glistening*” or “*shimmering*” just to mention some exemplar cases, are all candidate of basic features that cannot be taken into account by ordinary mechanisms of feature placing, since they refer to temporally extended spatio-temporal regions, entities that are not compatible with the narrow constraints of FIT.

If we put dynamic properties apart, even within static properties there are several cases of features not matching the narrow constraints on locality set by FIT. [Koenderink \(1993\)](#), for example, provides a formal taxonomy of types of properties that can be processed by the *front-end visual system* and that extend beyond local properties:

A property is said to be: **punctual** if it is defined with respect to a single point [...]; **local** if it is defined in terms of spatial derivatives at a point [...]; **multilocal** if it is defined in terms of local properties taken at distinct points [...]; **global** if it is defined with respect to the substrate as a whole; (p. 62)

If visual features are no more to be restricted to local features (neither insofar as they are recorded at discrete portions of the visual field nor insofar as they can be considered attributes), it is clear that the arguments adopted by FIT to justify the binding problem and the need of a local tag for visually encoded features lose much of their interest.

2.1.4 FEATURE ASCRIPTION WITHOUT LOCATION

For decades, the binding problem, i.e. how to correctly ascribe featural content belonging to a specific location has puzzled neuroscientists and psychologists alike. The main rationale behind it was evidence of segregated processing for distinct feature maps on the one hand, and the need of joint predication of properties belonging to the same visual item to mediate correct object perception on the other. Spatial location and selective attention (as the only mechanism able to anchor vision to specific spatial location) have been considered as the necessary requirements for the solution of this problem, until a number of new empirical studies and theoretical positions appeared, with a strongly deflationary view on the role played by location in mediating proto-referential capabilities.

Defendants of FIT have argued that to show the reality of feature binding as a real psychological problem one just has to consider that under high attentional load features systematically fail to be conjuncted. The reply has been that humans are actually very poorly performant in feature placing in general and that mislocalizations are more the rule than the exception ([Prinzmetal and Keysar 1989](#) cit. in [Green 1991](#), p.396). On similar grounds, [Rensink \(2000\)](#) has observed that the binding problem might be ill-posed since what is relevant for different perceptual routines might be simply a virtual representation of objects and their features depending on the specific

task requirements. As a general rule, he observes:

only a few objects need to have a coherent representation at any instant; detailed information about any object must be made available whenever requested (p.1475).

Pylyshyn has provided what appears to be the best demonstration (on both empirical and logical grounds) that FIT cannot solve the problem of proto-reference they have been trying to address by appealing to space-based attention. Not only in many tasks spatial location is not accessible, whereas objects can be perfectly parsed and accessed in their featural content, but spatial location itself is often not sufficient to provide vision with the kind of objective connection between features and the “*bearers*” of such features that space-based selective attention was assumed to give. The relevant bearers of basic perceptual features, Pylyshyn argues, are not to be found in spatial regions, but in *visual objects* that can have extremely unreliable (if any) connections with the locations they occupy in the visual field.

2.2 OBJECTHOOD

The main class of constraints on feature ascribability that I will consider in this section is related to the notion of *objecthood*, a concept that has played an increasingly central role in the explanation of several perceptual capabilities. Although the debate is open about what should count as a perceptual object and what is the relation between this theoretical notion and its commonsense counterpart (see [Casati, 2004](#)), mainstream research trends in perceptual psychology, developmental psychology, neuropsychology and neuroscience often refer to perceptual structures with “objectual properties” as the main units of our perceptual experience.

The fundamental units of conscious perceptual experience are objects and events. Although to philosophers it may seem a matter of choice whether we take nonconcrete qualities or concrete spatially or temporally bounded particulars as the basic units from which to build a descriptive system, phenomenologically a strong case can be made for the primacy of objects – [Treisman \(1986\)](#)

The aim of the present section is an exploration of those aspects in the literature on the characterization of visual objecthood that suggest specific constraints on feature selection. In order to understand how objecthood constraints restrict good features to a limited set of perceptual properties, we need to introduce some basic notions and a number of theoretical distinctions.

2.2.1 ASCRIBABLE FEATURES AND OBJECT-FILES

Ascribing a feature to a visual item is considered by many as one of the most basic operations enabled by perceptual capabilities, the equivalent of *predication* in the domain of language and reasoning (see for instance [Clark, 2001](#)). Arguably, the centrality of features ascribability to objects is not only relevant at a behavioral or phenomenological level of description of perceptual phenomena, but also at the level of the functional analysis of the visual system.

According to some researchers in perceptual science – at least those who have challenged the primacy of space-based attentional mechanisms⁴ – the goal of visual attention in early visual processing is to parse the scene for individual entities (*visual objects*) and to build compact descriptions of such entities (descriptions of their *attributes* or *observable features*), in order to make these

⁴Consider, in contrast, theories as those reviewed in the previous section, according to which visual attention selects objects by selecting their locations

descriptions available to more central mechanisms. In this respect, we can say that the visual scene is processed and described by the human visual system as a set of individuals that *bear* certain features. The “language” of low-level vision – to quote a recent synthesis by Cavanagh (2003) – is committed to an ontology whose basic constituents are, on the one hand, *visual objects* (structures that “play a role similar to that of individuals referred to by *proper names* in ordinary language”), and, on the other hand, *visual features* that can be ascribed to them (much as linguistic *predicates* can be ascribed to individuals). The view according to which visual objects and object-ascribable features are the basic constituents of the first description built by the visual system has blazed the trail to a number of empirical works that have focused on mechanisms that mediate the perceptual ability to *build objectual representations*, to *maintain reference* to them and to *encode specific properties* in such representations.

Amongst the various attempts to operationalize the idea of feature-ascribability, the notion of an *object-file* deserves a special attention. The origins of the concept of object-file can be traced back to a paper by Ann Treisman (1977), in which she made the claim that the goal of attentional mechanisms in vision is to parse the visual scene for individual items and to enable the creation of temporary representations of such items that are required for performing further operations (like categorization, identification, motor interaction etc.). Visual objects are hence those visual items that attentional mechanisms select as the fundamental units of visual perception: selective attention picks out properties from the sensory array that belong to an individual item and *binds* them together into an individual representation of that item (to deliver a compact and reliable description of the item’s properties to enable reference to that object).

According to this idea, visual objects have to be conceived as *bundles of properties* that are bound to specific visual items through selective attention. The account has been later brought to prominence thanks to an influential series of papers by Kahneman, Treisman and collaborators ([Kahneman et al., 1983](#); [Treisman and Kahneman, 1983](#)), who articulated it in terms of the creation and mobilization of *object-files*. An object-file, they claim, is an *episodic visual representation* that allows incoming sensory data to be collected from the attended object and to be updated when changes are detected. Object-files are conceived as structures whose psychological reality is determined by the necessity of postulating temporary stable representations to *encode* specific properties as *attributes* of an item, to represent object's *persistence* over time and over featural change, and to solve possible ambiguities raised by multiple instantiations of the same properties in the visual field.⁵ The very definition of object-files leads us to a first characterization of an ascribability constraint on basic features:

Ascribability constraint for basic features

Basic features are those properties of the visual scene that are encodable as attributes of a visual object, i.e. properties that can be stored in the temporary representation of that object (an *object-file*).

It is important to note that an encoded property can be *stored*, *modified* and *removed* from an object-file, without disrupting the file itself. The removal of a property from an object-file does not entail the disruption of the object-file, which - instead - persists across featural change. The stability of object-files over change of their encoded features has a number of interesting consequences. Object-files are allegedly those structures that enable, at any given instant t ,

⁵See Chapter 3 for a detailed discussion on this problem, known as the *Many Property Problem*

perceptual judgments of the form:

Object P *has* _{t} property F

The object-dependence of ascribable features has a number of interesting consequences:

- basic features are restricted to those properties that can be referred to individual visual items;
- constraints on the nature or visual objects determine what features are eligible as attributes of that object (and, hence, encodable in object-files);
- perceptual access to objects has a fundamental primacy over perceptual access to their features.⁶

2.2.2 FEATURES AND OBJECT-FIXATION

If an object needs to be picked out before features can be ascribed to it, then a mechanism for object individuation is needed. We have mentioned that *Feature Integration Theories* used to postulate selective attention on spatial locations as the mechanism enabling object-file representations to be built. Treisman's model is consistent with the fact that object-files can be constructed *independently* from the encoding of specific features, since the individuation of a visual object depends on its unique spatial location in the visual field.

In a series of papers, Pylyshyn and collaborators (2000; 2001) have challenged the idea that visual objects are picked up by referring to their location. They have shown that a large number of operations on visual objects (in particular *Multiple Object Tracking* - MOT) are indeed possible *without* postulating

⁶The primacy of objects on their ascribed features is consistent with what has been traditionally called object- or configurational superiority.

mechanisms for accessing their spatial location in the visual field. This challenge to Treisman’s theory has lead to a revision and partial amendment of the notion of an object-file.

On the one hand, in Pylyshyn’s model, object-files are conceived of as temporary representations that encode information about visual objects. Object features, accordingly, are still conceived as properties that can be ascribed to such individual items and stored in object-files as attributes of these items .

[This kind] of property determines the object’s appearance – what a particular object looks like – including its color, shape, lightness, and texture. We call these *featural properties* – [Scholl et al. \(submitted, p.2\)](#).

A file “content” is defined as a bundle of properties that the visual system *binds* to an individual object and that can be retrieved on demand at a given instant t , to mediate perceptual judgments, including categorization and identification.

On the other hand, Pylyshyn introduces a major conceptual change in the object-file notion. If so far the characterization of an object-file is consistent with Treisman’s original formulation, a considerable difference emerges as soon as we look at the interplay between the content of object-files and the conditions that are required for the *opening* of an object-file. As a consequence of Pylyshyn’s criticism against the role of location in objecthood fixation, visual objects as represented by object-files need no more be restricted to *localizable* visual items: they can include items susceptible of being tracked through *visual indexicality* mechanisms and *not* anchored to specific spatial locations. This new condition on objecthood fixation entails a partial but fundamental revision of the ascribability constraint:

Revised ascribability constraint for basic features

Basic features are those properties of the visual scene that are encodable as attributes of a visually *trackable* object, i.e. properties that can be stored in the temporary representation of an item (an object-file) insofar as it is selected and tracked through visual indexicality mechanisms.

Pylyshyn's slight revision of the notion of object-file is important not only for its deflationary impact on the role of *spatial location* (which implies a radical revision of *objecthood* conditions), but also for its implicit consequences on the selection of object-related *features*. In the framework of Pylyshyn's model, it is possible to introduce a distinction between two classes of visual features that are related to objecthood.

1. Object fixation properties

The first class of properties relevant for visual objecthood is defined by those properties that allow a visual object to be picked out or individuated and, hence, an object-file to be opened. Such properties *need not* be part of the properties that are *encoded* in the object-file's content and may not necessarily be retrieved and used in perceptual judgments. They should not be conflated with object-ascribable features.

2. Object encoded properties

The second class of properties relevant for visual objecthood is defined by those properties that can be properly described as "attributes" of a visual item, i.e. features that – to quote Pylyshyn – "determine an object's appearance", and that as such can be stored, encoded and removed from an object-file and retrieved in perceptual judgments. This class of properties is independent from the class of *object fixation properties* that enable object individuation and should not be conflated with the former.

It should be noted that Pylyshyn’s distinction concerns two distinct *functional roles* of object-related features, not necessarily the existence of two mutually exclusive sets of properties: it can well be the case that a single property be (extensionally) part of both classes. The functional role played by the same property in the two cases, though, should be kept distinct.

2.2.3 FEATURES AND OBJECT-PERSISTENCE

If we endorse the view on visual objecthood presented so far, we must acknowledge that *fixating* and *maintaining* reference to a visual item seems a fundamental ability required for the explanation of object-related perceptual skills. We have seen that object-files can be considered as the temporary stable representational structures that allow particular bundles of properties to be ascribed to an item and we have distinguished two different classes of properties that play a functional role in relation to visual objects. It is legitimate to ask at this point under which conditions items encoded by object-files can *persist* over time and – in particular – if the *preservation* of an object-file is somehow mediated by the detection and representation of particular visual properties.⁷ In particular, it is plausible to assume that properties that enable an object representation to ‘stay alive’ over time might constitute a separate class from object fixation properties and encodable features. Pylyshyn (2004) introduces accordingly a third class of properties related to visual objecthood.

3 Object preservation properties

The third class of properties relevant for visual objecthood is defined by those properties that allow a visual object to *persist* and the corre-

⁷I owe most of the ideas in the present section to discussions with Benjamin Sylvand and Ángeles Eraña.

sponding object-file to be *preserved* over time. Such properties should in principle be distinguished from the other two classes of object-related properties.

Although the postulation of three distinct classes is conceptually justified, especially in cases in which one has to establish whether, say, certain object-file fixation properties are at the same time relevant for object-file preservation, it is arguable whether such theoretical distinctions can be easily translated into methodological criteria for the study of object-related capabilities. *Under which conditions* can one exclude, for example, that a property allowing an object to persist over time be considered as an *encoded feature*?

Pylyshyn has repeatedly insisted on the fact that a number of operations on tracked visual objects are possible in cases in which *no featural information* about that object can be accessed. Properties that guarantee the persistence of an object over time – he argues – are by definition *not* part of an object-file's content. But, if this is the case, on which basis can we establish that a property like COHERENT MOTION (one of the most important object-file preservation properties) does not belong to the set of an object's ascribable features, that typically include its TEXTURE, SHAPE or COLOR? Can we say that the perceptual system is not *accessing* or *tracking* by picking out visual items with COHERENT MOTION across the visual scene? An implicit constraint – based on an intuitive idea of what it means to *ascribe a property to an object* – seems in this case to be tacitly adopted to restrict the set of properties that can be stored in and accessed from an object-file. While the SHAPE of an object is usually taken as a paradigmatic case of object attribute that can be encoded in an object-file, a property like COHERENT MOTION seems *prima facie* hard to fit with a cluster of properties specifying an object's appearance. But if we set intuitions aside, there are no reasons for

excluding that the perceptual system can encode and maintain in a temporary representation of an individual visual item properties that are non sensory, or less salient than those accessible at a phenomenological or behavioral level.

This example shows that the generic characterization of encodable features as properties that specify “how an object looks like” is too weakly constrained to be used in order to make interesting predictions on the emergence, persistence or disruption of visual objects. Further constraints on what is meant by *ascribability* (or *encodability* of a feature into an object-file) are needed if claims like the one mentioned above (according to which object-file fixation properties are not accessible as features belonging to that item) are to be made.

The difficulty inherent in the translation of Pylyshyn’s criteria into general methodological constraints on the study of object-related visual properties becomes even more evident as soon as we consider their application to alternative characterizations of visual objecthood that have been proposed in the literature.

Consider for example the notion of objecthood that emerges in the study of early object-related perceptual abilities in children, as studied by developmental psychologists such as Liz Spelke. [Spelke \(1990, 1994\)](#) has proposed a list of properties that determine which visual items are systematically preferred over other others by children in perceptual tasks (such as parsing a visual scene) at an early stage of cognitive development, hence framing the nature of children’s perceptual representations of objects. A summary list of these properties includes: COHESION, SOLIDITY, MOTION CONTINUITY, CONTACT. The conjunction of these properties determines a notion of visual object that is commonly referred to as *Spelke object*. A theoretical link between this notion of visual object and the kind of object representations that seem to be

at work in the case of adults' perceptual tasks requiring divided attention or multiple tracking, has been proposed by [Carey and Xu \(2001\)](#), who suggested that both in adults and infants the same kind of temporary representations (precisely the above mentioned *object-files*) are used by the perceptual system to pick out and maintain reference to individual objects.⁸

Now, if this unification hypothesis holds, we are entitled to test Pylyshyn's distinction between three classes of properties with distinct functional roles on the list of properties invoked by Spelke. It is quite clear, on the one hand, that properties defining *Spelke objects* can all be seen as belonging to the class of *object fixation properties*: if any of this properties is lacking or systematically disrupted, we must expect that no Spelke-objectual representation can be formed (i.e., *no object-file can be opened*) and hence that the infant cannot parse the considered entity as an individual item. On the other hand it is debatable whether each of these properties is required to *maintain* perceptual reference to an object, once this is individuated. For example, we might ask (and empirically test) whether and in which cases a certain object fixation property like SOLIDITY can be dropped once an object is individuated without disrupting the object's persistence. Such a distinction has crucial consequences for accounting for the difference between cases in which a single object (perceptually) *ceases to exist* and cases in which an object undergoes major changes *without ceasing to exist* for the perceiver.

Furthermore, it is interesting to consider a possible extension of Pylyshyn threefold distinction to include a fourth class of object-related properties, that we might dub *object disruption properties*: the fact that, on the one hand, certain properties must be constantly accessed, or systematically reiterated to the perceiver, in order for an object to persist does not exclude that,

⁸For a critical assessment of this proposal of theoretical unification see [Casati \(2004\)](#).

on the other hand, the instantiation of a specific *disruption property* in the visual field determines the object's disruption, thus overriding preservation properties. A further class of properties, not reducible to the absence of object preservation properties, can then lay a legitimate claim to existence. Pylyshyn's criteria can be taken as *prima facie* candidates for the *general* theoretical notions required for studying the different classes of properties that are relevant to understand how perceptual systems handle objectual entities. Lacking explicit methodological conditions to establish whether a specific property belongs to any of these classes (and hence to ground hypothesis such as the "non-encodability" of certain sets of features), this classification might though be far from feasible.

2.2.4 NON-OBJECTUAL FEATURES

The definition of different classes of features related to objecthood has been fundamental for providing a more robust theoretical framework for the study of some important visual skills in humans. A large part of contemporary research in perceptual science actually assumes objectual features as *basic features*. Assuming that the most basic description of the environment delivered by perceptual systems consists of representations whose content includes individual entities and ascribable-features has, though, a number of consequences that should not be underestimated.

By focusing on different notions of what counts as *basic perceptual features*, ecological and sensorimotor theories of perception (see Chapters 5-6) have challenged (more or less explicitly) the idea according to which objects and object-ascribable features should be considered as core notions for the understanding of the functioning of perceptual systems. Even if our environment is populated by material entities that (intuitively) seem to be central in our

perceptual experience (and in conceptual access to such experience), the validity of theories taking objects - however defined - as basic ingredients of perceptual processing can be challenged on different grounds.

Much as in the case of the criticism of the study of feature integration – which, as we have seen, gives spatial location too central a role in the explanation of perceptual capabilities – mainstream research in perceptual science (including Pylyshyn’s work) admittedly endorses a strong notion of objecthood and feature ascribability directly drawn from phenomenology. It is arguable, though, whether any guarantee can be provided that feature ascribability is not an explanatory artefact resulting from our pre-reflective way of thinking of the world and its intuitive ontology.

In particular, it can be shown that in contemporary research not only when analyzing the featural content of a particular stimulus systematic preference is given to object-ascribable features, but – even when the study explicitly focuses on object-centered mechanisms – features that do not match the above constraints are seldom taken into account.

Consider the following case. In his effort towards a principled approach to the study of *basic perceptual features*, James Pomerantz has recently suggested we should try to adopt a bottom-up approach and study feature emergence starting from the *simplest* stimulus configuration ([Pomerantz et al., 2003](#)), i.e. starting from the simplest stimulus configuration (the “Ganzfeld” of the Gestalt tradition) and progressively increasing complexity. As an example he takes the case of a *single blob*, a “shapeless, preattentive object”.

Pomerantz observes that whereas this particular configuration can only vary along three dimensions (POSITION, SIZE, COLOR), as soon as we add a second object of the same kind, two new features emerge (PROXIMITY, ORIENTATION), neither of which can be described as a feature belonging “to a single element

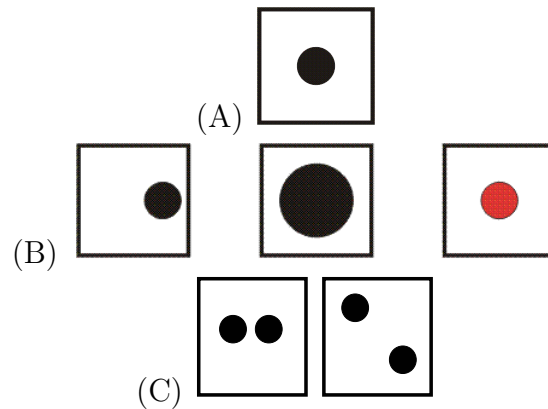


Figure 2.2: Object and their features. (A) A single blob: a *shapeless, preattentive object*; (B) Featural variations on single blobs; (C) Emergent properties of two blobs – from [Pomerantz et al. \(2003\)](#)

alone”. The second configuration, accordingly, can be described as containing three basic, object-centered features and two emergent, non-ascribable features. Now, one might ask why a property like *being aligned with another single blob on a -45° -sloped line* should not be considered a basic, object-ascribable feature. Examples of this kind of ascribable yet configurational properties (like, *being surrounded by at least two identical blobs within a visual angle of n°*) are not difficult to make up. Such properties are plausible examples of perceptually relevant properties on ecological grounds, in cases of pattern parsing and recognition. Yet, they seem to be systematically excluded from the set of good properties that can be ascribed to objects.

Even if one assumes that such intuitively contrived properties are not properly speaking *objectual* or *object-ascribable*, there are cases of configurational properties that seem to be extremely relevant to object-centered perceptual abilities. Consider as an example the role of configurational properties like the relative PROXIMITY between three distinct visual items. Although this property does not represent a paradigmatic case of object related feature (it is typically assumed to be a non-ascribable feature), it seems that many

object-oriented perceptual skills might benefit from sensitivity to such a property. In particular, one might redescribe the traditional MOT experiments by looking at the mutual relations between objects and investigate if the same ability could not be explained by referring to configurational and dynamic properties of the ensemble of tracked items. This does not necessarily mean going back to a location-based model of object individuation, at least not in the traditional sense of FIT. A configurational, non-objectual property that might be invoked as a case of property mediating object-persistence might be the property of maintaining *proximity* between items below a given threshold. Perceptual judgments on a property P being an attribute of an individual visual item (i.e. the paradigmatic case of judgments relying on feature-ascribability) might well be mediated by mechanisms that do not rely on encoding of properties ascribed to individual items, but rather on relational or configurational properties between objects that are parsed from the visual scene. This opens the question of the existence of *non-objectual features* that might play a functional role in the constitution of visual objecthood (see ?), a question that has been largely disregarded in contemporary perceptual literature.

2.2.5 CONCLUSIONS

Although feature ascribability to perceptual objects has been a central concept in contemporary perceptual science, it seems that the study of the relation between perceptual objecthood and feature ascribability still lacks a robust methodological foundation. Questions like the following have been only partially (if ever) addressed in the literature.

- What criteria are necessary to yield perceptual objecthood in a given modality?

- What principles allow to distinguish between *encodable* and *non-encodable* features?
- Provided there is a principled way to tell apart encodable and non-encodable features, is there evidence that some non-encodable features might play an explanatory role in the understanding of how perceptual systems make feature-ascribability possible?

I have suggested that – although our phenomenology, our language and our conceptual structure are strongly biased towards a notion of object-ascribability – properties relevant to the understanding of object-ascribability reach beyond the traditional notion of encodability (see above, the *revised ascribability constraint*). Properties mediating the perceptual equivalent of *predication* should therefore not be restricted to phenomenologically characterized attributes of perceptual items. I have argued that *object-fixation properties*, *object-preservation properties*, *object-disruption properties* and (at least a class of) *configurational properties* might contribute to feature-ascribability as much as (traditionally defined) object-encodable properties.

Phenomenology (and, more generally, our naive way of parsing the world into relevant entities) might prove a misleading source of evidence for developing the appropriate theoretical notions to be adopted in the study of perceptual systems and of properties they are supposedly sensitive to. I have argued that implicit factors of this kind, though, systematically orient perceptual research on objecthood towards the selection of features that are encodable as attributes of static, bounded and connected perceptual items. Properties that might not seem *prima facie* encodable into object-files (in particular global scene properties, relational properties involving multiple items or properties of events) are nonetheless potential candidates to the class of features relevant to the understanding of ascribability skills. Such skills not only deserve

further empirical investigation: they also require a more constrained way of establishing the meaning of *object-ascribable property*.

Chapter 3

Issues in the study of internal constraints

THE GOAL of this chapter is to put to work the analysis of different feature basicness criteria, by tackling a paradigmatic case of processing problems - the *Feature Binding Problem* - involving a notion of basic feature based on internalist constraints. My goal will be to show that the rationale for this problem is ill-posed, partly because of the use of an inappropriate notion of basic feature.

3.1 A CASE STUDY: THE FEATURE BINDING PROBLEM

The *Feature Binding Problem* (FBP) can be considered, in its most common formulation, as the problem of correctly reconstructing the properties belonging to a single perceptual entity on the basis of sensory information available from the retina. As such, it stands out as a paradigmatic case of processing problem arising from the background assumption presented at the beginning of part I: how can correct perceptual output be obtained from informationally poor sensory information?

We already met the FBP in the context of the analysis of earliness constraints, where it was introduced as the subject matter of *Feature Integration Theories*. According to these theories, correctly representing the visual scene requires

selecting and binding together the features of single visual items as they are encoded in the primary visual system. I will not recall here the different lines of criticism against *Feature Integration Theories* that I presented earlier: I will focus instead on the *rationale* for this problem. I will actually adopt the FBP as a testbed for assessing the validity of the thesis that I have defended so far: namely, that there are several, distinct criteria of feature basicness, whose theoretical independence is often overlooked in the literature. The conclusion of this analysis will be that certain formulations of the FBP are a theoretical artifact resulting from the conflation of basicness criteria and the corresponding levels of descriptions.

3.1.1 OPERATIONAL VS. NON-OPERATIONAL FORMULATIONS

In order to clarify the main point of my argument, I will start by reviewing some common formulations of the FBP from the literature:

A1. If, as we have reason to believe, color and form are processed in separate parts of the nervous system, why does one not simply perceive circle, triangle, blue, green without knowing which form has which color? – [Attneave \(1974\)](#), cit. in [Green \(1991\)](#)

A2. Most of the objects, people, and scenes we perceive produce complex, multidimensional, changing patterns of stimulation on the retina ? At least some of their attributes appear to be registered by independent neural channels, specializing in different aspects, such as orientation, color, spatial frequency, brightness... But this immediately raises the question of how the component properties are resynthesized into the correct compounds, so that we correctly see the shirt as blue and the trousers as gray, for example, rather than the reverse – [Treisman \(1977\)](#)

A3. It is now well established that in the early visual system of primates there are at least ten distinct visual areas, arranged in a branching hierarchy. Different cortical areas specialise, to some extent, in different features, one responding mainly to motion, another to colour, etc. As one proceeds to areas higher in the hierarchy the mapping of the visual field onto the cortical surface tends to become more diffuse. This is not however how we see the world. Our inner visual picture of the external world has unity. How then does the brain put together all these different activities to produce a unified picture so that, for example, for any object the right colour is associated with the right shape? – [Crick \(1984\)](#), cit. in [Van der Heijden \(1995\)](#)

A4. The various visual properties of objects in the field of view start out unified. The shape, the color, size, and direction of movement all come from the same object. They come from the same place in space and they co-occur in time. When one feature disappears typically they all disappear. But when the object is processed by the nervous system, at least some dimensions get parsed and are analyzed in different brain regions or in different cells within the same brain region. Single cells recordings and pattern of deficit following neurological disorder all suggest that different regions in the occipital, temporal, and parietal cortex process different features emanating from the same object. This distributed processing raises a problem. Suppose that two or more objects are present in the field of view, each having a different color, different shape, different location and the like. If one part of the brain codes color, for example, and another codes form, then how is it that later in the processing it is determined which color goes with which form ? What is the mechanism of reassembly ? – [Keele et al. \(1988\)](#)

A5. The physiological evidence for the binding problem comes from studies of neurons in extrastriate visual cortex of primates. One key observation is that different features of an object are processed to a certain extent by different neurons within the visual system. Logically, in order to identify the shape, color and motion of a stimulus, the visual system must somehow integrate the activity of these different shape-selective, color-selective, and motion-selective neurons. When only one stimulus is present in the visual field, this is not a difficult problem because these features can only be assigned to one possible stimulus. However, when multiple stimuli appear together in the visual field, which is the typical situation in « real-world » scenes, the visual system must assign the correct color, shape and motion signals to each object. – [Reynolds and Desimone \(1999\)](#)

Although such formulations present a number of surface commonalities, a first important distinction must be drawn between what I call *operational* and *non-operational* characterizations.

The common assumption shared by all of these formulations is that distributed sensory coding of basic features raises a problem for perceptual processing: the fact that basic features are encoded in a segregated way by the visual system and that correct perception requires such information to be bound together in order to be referred to one and the same visual item, is an issue that must be addressed by perceptual processing. A difference arises, though, in the way different formulations characterize the rationale for the required integration of features. Some of them (for instance **A1.-A3.**) assume that

a binding mechanism is required in order to explain the *phenomenological unity* of the percept, i.e. to justify the fact that we do not perceive objects as separate bundles of features but as *wholes* specified by joint features. I will call these formulations *non-operational* in that they do not assume that binding is justified on the basis of further processing requirements, but is needed in order to explain phenomenological unity of perception. A second class of formulations (such as **A4.** and **A5.**), consider the FBP as a merely *operational* problem, i.e. a problem arising from the way specific patterns of the sensory stimulation are encoded and from the necessity of signaling joint information to further processing mechanisms.

We can grant, on the one hand, to operational formulations the status of genuine empirical questions. I argue, on the other hand, that non-operational formulations of the FBP are flawed because of the reference they make to a theoretically weak notion as that of phenomenological unity.¹ It is hard to see how non-operational formulations could provide an account of incoherent phenomenology. What would correspond to the visual experience of *unbound* features? Or to perceive visual objects *incoherently*? None of the phenomenological formulations has actually provided a contrast class to frame the notion of an alleged unity of the perceptual experience of visual objects.² I agree on this point with [Clark \(2001\)](#) where he observes:

It is also risky to define “unity” or “coherence” in terms of what is experienced to be unified or coherent. This maneuver shifts the burden onto some definition of what it means to be experienced as coherent; and any such definition is likely to be even more elastic than an account of what it is to perceive “one” object (p.3).

¹This argument is further developed in [Taraborelli \(2002\)](#).

²[Treisman and Gelade \(1980, p.100\)](#) actually note that, unless fixed by focal attention, the features of objects “may be free floating spatially, as well as unrelated to one another”.

The strategy of making reference to feature conjunction errors as an example of incoherent phenomenal experience (see Briand and Klein, 1989; Donk, 1999; Treisman and Schmidt, 1982; Tsal, 1989) actually turns the problem into a performance issue and consequently into an operational one. I consequently assume that genuine formulations of the FBP should be restricted to operational ones and I will focus, accordingly, in the remainder of my analysis on formulations that assume the rationale of this problem as a consequence of processing constraints.

The general form taken by operational formulations of the FBP can be described as an instance of the *many-property problem* first introduced by Jackson (1977), which I already discussed in the *Earliness* section (1.4). How can the visual system succeed in signaling the joint belonging of featural information to a single visual item? As stated in A5., the problem is to understand how multiple featural information processed by the primary visual system can be *tagged* as referring to the same visual item.

An answer to an operationally formulated FBP should then address the following three questions:

1. In which cases is a feature binding process needed?
2. What are candidate computational mechanisms for the accomplishment of feature binding?
3. What class of features do binding mechanisms apply to?

I will focus in particular on the third question, which directly impinges upon the characterization of the notion of a basic feature. I argue that in characterizing the class of features to which feature binding applies, a large number of operational formulations found in the literature run into level-conflation problems. In what follows, I first review some of the current

characterizations of the class of features that must be bound according to defendants of the existence of a genuine FBP. In particular, I analyze more closely the claim according to which one of the basic features involved in feature binding is color. I conclude by pointing at some major methodological issues that undermine the validity of a FBP as it is commonly formulated in the literature.

3.1.2 EXTENSIONAL CHARACTERIZATIONS OF BASIC FEATURES FOR BINDING

Characterizing which properties belong to the class of basic features to which binding mechanisms apply is a highly problematic issue which has not been explicitly addressed in the literature: no clearcut criteria establishing what properties of the sensory stimulation are susceptible of being bound has been provided so far and the issue has hardly deserved attention.

Proposers of alleged computational mechanisms for the solution of a FBP, though, have provided plenty of examples of basic features, thus offering a sort of *extensional characterization* of properties to which binding mechanisms allegedly apply. They have failed to provide, though, a specification of the necessary requirements basic features for binding have to meet. Let us see some examples (emphasis is mine):

Some of those attributes are pictorial features like LINE ORIENTATION, TEXTURE, COLOR, SIMULTANEITY OF APPEARANCE, and COMMON MOTION, but others require more complex information about such things as 3D SHAPE, LIGHTING, and OBJECT SURFACE PROPERTIES. (...) Simple features such as COLLINEARITY, COLOR, TEXTURE, and COMMON MOTION (...) might easily be extracted from representations in primary visual cortex – [Shadlen and Movshon \(1999\)](#).

In order to identify the SHAPE, COLOR AND MOTION of a stimulus, the visual system must somehow integrate the activity of these different shape-selective, color-selective, and motion-selective neurons – [Reynolds and Desimone \(1999\)](#)

The difficulty in solving the [feature binding] problem lies in the fact that a number of different subcues need to be integrated with each other to yield reliable segmentation. In

the visual modality, for instance, relevant subcues are MOTION, COLOR, TEXTURE, STEREO DEPTH, COHERENT EDGES, and simple or known SHAPES. – [Von der Malsburg \(1999\)](#)

At face value, such feature lists suggest that basic features for binding grossly correspond to stimulus dimensions that are encoded in feature maps in the primary visual cortex (like LINE ORIENTATION). But other features that often occur in such lists (like 3D SHAPE or SURFACE SLANT) actually do not refer to properties that are encoded in the primary visual cortex, but to properties that can be considered basic insofar as they produce effortless textural segregation and preattentive access on the basis of psychophysical tests (see above, section 1.4). There are also some interesting asymmetries: on the one hand, COLOR occurs in virtually *any* list of basic features for binding, as well as SHAPE; on the other hand BINOCULAR DISPARITY, although it is an extensively studied property of single-cell response profile in primary visual cortex, is mentioned with much lesser frequency.

Beside these extensional characterization, one might wonder whether there is a principled way to frame the class of properties to which feature binding applies. How many features need to be bound in order to yield “correct object perception”, as [Treisman and Schmidt \(1982\)](#) suggest? Is any of these feature an indispensable attribute? Are there features that can be considered as basic but to which feature binding mechanisms do not apply? An explicit answer to these questions cannot be found in the literature. The lack of explicit criteria for framing the notion of basic features for binding and the proliferation of extensional, often elliptic and only partially overlapping characterizations suggest two possible interpretations.

Under a first, stronger reading one might argue that the lack of a precise characterization of the class of properties to which binding mechanisms apply reflects the assumption that there is *no such a set of features*: the actual

kind of properties to be bound depends on the specific requirements of each perceptual task. This implicit assumption, albeit endorsed by many opponents of traditional *Feature Integration Theories* (see for instance [Rensink, 2000](#)) does not imply, though, that *any* property of the sensory stimulation can be a candidate for feature binding: this is a strongly deflationary reading of the FBP and it is unlikely to correspond to what most formulations found in the literature assume.

Under a weaker and more plausible reading, the lack of a precise characterization does not depend on the fact that there is no such a class of basic features for binding, but rather on the assumption that the precise definition of what properties belong to this class can be settled on empirical grounds. But this is precisely *the* problem that I have been addressing in the first chapters of this work: there is no unique empirical criterion that can be adopted to frame the set of basic properties that constitute the functional input of perceptual processing. Each criterion is theory-laden and different basicness criteria yield different sets of properties. Hence, shifting the burden of the definition to a set of criteria upon which there is no general agreement does not represent a valid strategy either.

This impasse in defining the precise class of features susceptible of being selected by feature binding mechanisms is a first piece of evidence towards the main thesis I aim to defend, namely that most formulations of the FBP rest on an inappropriate characterization of the notion of feature. Before directly addressing this question, though, let us take a closer look at an example of the way in which a property like COLOR is dealt with in the feature binding literature.

3.1.3 A CLOSER LOOK AT COLOR AS A BASIC FEATURE FOR BINDING

As I suggested in the previous paragraph, among the stimulus dimensions that are allegedly encoded in feature maps and that are susceptible of being applied a binding mechanism, COLOR has a prominent place. What property corresponding to COLOR is encoded in feature maps, though, is an issue that has not been completely settled. The proponents of *Feature Integration Theory* themselves have switched between two different positions during the elaboration of their hypothesis, first assuming that a single feature map is responsible for encoding color distribution in the visual scene and, later, suggesting that there are at least different feature maps for different hue values (see [Treisman, 1988](#)). Color is systematically referred to in this tradition as one of the features extracted by the primary visual cortex and encoded in retinotopic maps, although there is empirical evidence for different areas responsible for coding distinct aspects of chromatic stimuli *beyond the primary cortex*. Moreover, there is large evidence supporting the hypothesis that the primary cortex is actually *not* involved in encoding the perceptual property that we generally refer to when talking about COLOR, but merely local wavelength differences. As [Zeki and Marini \(1998\)](#) suggest, whereas V1 is responsible for wavelength discriminations, the actual processing underlying color perception must take place at a higher level (V4), where the first large-scale wavelength comparison are performed. [Van Essen et al. \(1992\)](#) already pointed out that the relation between receptive field properties in the primary cortex and their role in perception is far from settled and wavelength selective neurons may have many alternative functions that are unrelated to color perception (see for instance [De Yoe and Van Essen, 1988](#)). Hence assuming that properties encoded in Treisman's feature maps correspond to properties encoded in the primary visual cortex requires at least some

prudence. Now, when referring to COLOR as a basic feature to which feature binding mechanisms apply, what notion are we actually considering? At least three potential candidates can be identified:

- COLOR₁ as a property encoded in (some) retinotopic map of the visual system.
- COLOR₂ as a property eliciting effortless texture segmentation and effective visual search in preattentive vision.
- COLOR₃ as a phenomenologically accessible attribute of a visual item.

Surprisingly, all these properties are taken in *Feature Integration Theories* as descriptions of the very same property (COLOR *tout court*) at different levels of analysis. We have shown, though, that this cannot be the general case for *any* feature, since earliness, lowlevelness, and phenomenal saliency (as specified in the objecthood section) are independent criteria that yield different and only partially overlapping sets of properties. What these theories implicitly assume, on the contrary, is that:

1. there is a class C₁ of properties that are encoded in dedicated feature maps in the visual cortex;
2. this class of properties matches a class C₂ of preattentive features that can be individuated through behavioral tests (like visual search);
3. this class matches in turn a class C₃ of properties that have phenomenological salience as attributes of visual items;

Integration of distinct featural information at any of these levels, it is accordingly assumed, must *show up* at each of the other levels: Feature binding (at least in the original formulation of *Feature Integration Theories*) can then be

described as the mechanism that, by joining the signal relative to featural dimensions encoded in cortical maps (C_1), mediates behavioral access to conjunctions of features (C_2), which in turn causes our phenomenal experience of visual items as unitary conjunctions of visual attributes (C_3). This picture, although intuitively sensible and adopted by several philosophers as a plausible model for understanding mechanisms of perceptual reference (see [Campbell, 2002, 1997](#); [Clark, 2000, 2001](#)), runs into some major methodological issues that - I argue - affect the very justification of the existence of a FBP.

3.2 LEVEL CONFLATIONS

The analysis of the case of color allows us to draw some considerations on the notion of basic feature for binding. Most operational formulations of the FBP rely on the assumption that segregation of features at the level of their neural encoding must *show up* at the level of performance and that this segregated encoding is actually the cause of failures in feature conjunction. This hypothesis has recently been countered by several lines of criticism which pointed out how such formulations suffer from a methodological conflation between distinct levels of description (see [Garson, 2001](#); [Taraborelli, 2002](#); [Van der Heijden, 1995](#); [Wolff, 2004](#)). The existence of psychophysical evidence for a FBP - the criticisms run - may be independent from the fact that certain stimulus properties are encoded by the visual cortex in separate maps. Neural segregation does not necessarily imply behavioral segregation; conversely, behavioral integration of visual properties does not necessarily require integration of the underlying neural vehicles. By *not necessary*, I mean to suggest that although it *can* be the case that some properties of psychological segregation are directly caused by properties of underlying neural representations, the *latter does not necessarily entail* the former, as

many programmatic statements seem to tacitly assume. [Millikan \(1993b\)](#) has drawn attention on the capillarity of level conflation in perceptual research. She points out how tempting it is in scientific explanations of perceptual phenomena to apply a strategy of *content internalization*, i.e. to project a set of properties of perceptual content onto properties of the alleged vehicles of this content (in this case neural activity) and to claim that this isomorphism of properties represents a genuine explanation of what causes perceptual content.

The error to be eradicated, then, certainly is not that of positing intermediaries. Postulation of intermediaries of some kind is essential to understanding perception and thought. The error is that of projecting, without argument, chosen properties of what is visaged or conceived onto these intermediaries and vice versa. The error is equally that of taking this sharing of properties to constitute an explanation of mental representing. – [Millikan](#), *cit.*

I maintain that two (symmetrical) fallacious moves of this kind can be identified in the formulation of the FBP:

- A. **Internalization of unity:** conjunction of featural representation in perceptual performance is projected onto unity of neural vehicles (whereas success in feature conjunction can be completely independent of actual integration of underlying signal in neural processing).
- B. **Externalization of segregation:** segregation of feature processing at a neural level is assumed to show up at the level of perceptual performance (whereas segregation of neural encoding may be completely independent of failures in features conjunction).

Not only we can challenge along these lines the legitimacy of the claim that neural segregation raises a problem for behavioral performance. On the basis of the analysis done in the previous chapters, we can challenge the very assumption according to which there is a genuine explanatory value in cross-level accounts of the activation of the “same” property, e.g. when we say that the firing of certain cells in the visual cortex *explains* the perception of a given property as localized at a given position of the visual field. [Hurley \(1998\)](#) observes to this regard:

Firing-pattern synchrony is sameness in the type of firing pattern; the suggestion is that it codes for sameness of object in content. It is of course an empirical question whether this hypothesis is true. However, we should not suppose that sameness of object represented must be encoded by sameness of firing pattern. There might be a specific systematic variation in firing pattern, without loss of information, between cell populations in different areas. Systematic difference rather than sameness of firing pattern might in principle code for the sameness of object, though it would again be an empirical question whether it ever does. (p.43)

There are hence conceptual reasons to assume that the explanatory link connecting patterns of neural activity in the visual system, perceptual performance and phenomenological saliency might actually be much more complicated than advocates of cross-level explanations have argued.³

3.3 CONSILIENCE STRATEGIES

Let us try to sum up some conclusions of this analysis. [Clark \(2004\)](#) observes that the concept of a “feature” refers to at least three distinct kinds of uses:

³See [Teller \(1984\)](#); [Teller and Pugh \(1983\)](#). For a criticism of these positions see [Noë \(2002\)](#). A more detailed analysis of this debate can be found at the end of section 1.4.

1. properties of the sensory stimulation that selectively trigger activity of cells in early visual areas;
2. independent dimensions of sensory discriminability, indentifiable through behavioral tests;
3. general terms used in feature-placing sentences (such as ‘here is green’, or ‘there is brighter than here’).

I have argued in the previous sections that one of the major sources of explanatory fallacies consists in assuming that there might be *more than a simple homonymy* – as the one suggested by Clark – between different notions of basic features. The explanatory strategy against which I am arguing is the one that takes for granted that there is a subset of properties of the visual stimulation upon which multiple criteria converge (like *earliness*, *low-levelness*, *simplicity*, *unanalyzability*, *compositionality*, *localizability*, *object-ascribability*) and consequently takes this simple convergence as an explanation of the mutual relations between each of these criteria. Instead of being an explanatory virtue, I argue, the fact that the “same” property matches different kinds of criteria at a time should warn one against embracing the tempting conclusion that for this very reason there should be an immediate explanatory connection between any of such criteria. Examples of similar explanatory strategies are common:

- “It is because a cell fires in MT signaling MOTION that I am perceptually aware of visual MOTION in this specific region of the visual field”.
- “It is because attention binds together the location of COLOR, SHAPE and TEXTURE from distinct feature maps that I can consciously perceive an object at that location with that specific COLOR, SHAPE, and TEXTURE”.

- “It is because visual stimuli are parsed in terms of COLOR, SHAPE and TEXTURE that any percept results from compositional rules applied to COLOR, SHAPE and TEXTURE”.

Taking the convergence of several criteria as an explanatory virtue is what, in philosophy of science, has been often called a *consilience strategy*. A consilience strategy literally consists in “a *jumping together* of knowledge by the linking of facts and fact-based theory across disciplines to create a common groundwork for explanation” (see [Wilson, 1998](#)). Considered by some as a virtue of cross-theoretical explanations and by others as a risky explanatory strategy, consilience-based explanations - I assume - are what created a consensus on the existence of a set of properties that (a) are given a privileged status over other patterns of the visual stimulation and (b) constitute the functional input of perceptual processing.

I have showed that there are several basicness criteria that can be distinguished on conceptual grounds and argued that there is no *a priori* reason why such criteria should converge on the same set of properties: as a matter of fact, I have reviewed empirical evidence indicating that each of these criteria selects distinct sets of properties that do not necessarily overlap: a visual feature can be retrieved preattentively *without* necessarily being encoded in the primary sensory cortex; it can be functionally unanalyzable for the visual system *without* being compositionally relevant; it can be encoded as an attribute of a visual object *without* being represented at a precise location in the visual field etc.

Through the analysis of the case of the *feature binding problem* I have tried to show that the assumption that neurophysiologically-defined features should match behaviorally-defined features which should in turn match phenomenally-defined features relies on a major conflation between distinct levels and criteria

for feature basicness. The goal of this criticism is *not* to deny the existence of signal integration mechanisms or the legitimacy of postulating binding processes in general as valid computational strategies: binding mechanisms as solutions developed by the brain to encode specific spatio-temporal co-occurrences of two or more neural events have an obvious computational interest (Von der Malsburg, 1995, 1999). What I argued, instead, is the fact that typical formulations of the FBP rely on some fallacious assumptions, like the fact that segregated coding of properties from the visual stimulation should require some kind of integration mechanism in order to yield correct perception and avoid false conjunctions in perceptual performance. Assuming that the visual system has to solve a problem of feature binding *in any case* in order to yield a correct representation of visual entities means accepting the fallacious argument according to which:

1. there is a set of basic features that are required in order to correctly parse and perceive any visual entity;
2. these features are encoded in segregated maps by the visual system;
3. correct perception can only occur after such features have been bounded together and represented as belonging to the same visual item.

The ultimate interest of this analysis can be seen in the fact that consilience strategies have contributed to enforcing a *standard view* about the architecture of perception and the relation between basic features and perceptual processing which has masked a number of interesting perceptual phenomena. In the second part of this work I will develop an alternative hypothesis on what might constitute basic features, not relying on merely internalist constraints but taking into account the match between adaptive needs of an organism and the contingent structure of its environment. I will argue that such an

alternative account of what might constitute the proper functional input of perceptual processing can reveal some genuine perceptual phenomena that are not captured by traditional internalist criteria for the selection of feature basicness.

PART II

ADAPTIVE CONSTRAINTS

Chapter 4

Adaptive constraints on basic features

4.1 STEPS TOWARDS A SCIENCE OF ENVIRONMENTS

“**T**HE REALISTIC SCIENCE of organisms, biology, needs as its counterpart a realistic science of environments”.

With this programmatic statement, [Smith \(1999\)](#) concludes a survey of the main theoretical paradigms which have taken the *organism-environment* integrated system as the proper level of analysis for the understanding of perceptual phenomena. Environments, he argues, considered as the specific contexts into which organisms are embedded, need to be studied as the partitions of the physical world that are cognitively relevant for such organisms. Environments so construed need to become the subject matter of a scientific investigation if we want to identify what aspects of reality are relevant for perception. Studying different kinds of environments for different classes of organisms means understanding how perception is *tuned* to the specific properties of the environment into which the organism fits. A science of environments should aim at finding the appropriate level of description of environmental regularities relevant for the perceptual goals of the organism in question.

Smith's programmatic statement can be taken as a premise for my investigation into the issue of what role environments play in the determination of basic features for perceptual systems. This chapter aims to answer two main questions:

- What does it mean to study the relevance of specific environmental settings for an organism's perceptual system?
- How can a specific environment structure determine what counts as relevant perceptual input for this organism?

4.1.1 FROM ECOLOGICAL TO ADAPTIVE CONSTRAINTS

Behind Smith's proposal lies Gibson's definition of an *ecological niche*. According to [Gibson \(1979\)](#)'s view, each type of organism is *tuned* in its perception and action to targets that belong to a specific level of description of physical reality. These targets – what Gibson calls “affordances” – are the environmental correlates of some adapted traits of the organism which – as a whole – form what Gibson calls the organism's *ecological niche*. An ecological niche is, then, the ensemble of patterns, entities and properties that fit the behavioral needs of an organism. The insistence with which Gibson has stressed the importance of studying environments at the level of an organism's niche and denied the relevance of the study of physical properties of sensory stimuli has often discredited ecological approaches as incompatible with a naturalistic explanation of perception. Gibson's notion of affordance, although theoretically stimulating, has been attacked as too weakly constrained in order to function as a theoretical notion in the study of perception. Basically, it has been argued ([Fodor and Pylyshyn, 1981](#)) that lacking a clear-cut characterization, the notion of affordance cannot have a genuine explanatory power, since it can

be adapted at will to account for any kind of alleged sensitivity an organism may display towards properties of its environment.

The recent revival of Gibsonian theories has stressed the role of ecological invariants to challenge the traditional understanding of what are the relevant properties that constitute the functional input of perceptual processing. The approach outlined in this chapter can be seen as both an *extension* and a *specification* of the answer that ecological theories have given to the question: *what properties of the sensory stimulation are relevant for perceptual systems.*

4.1.2 ADAPTIVE CONSTRAINTS AND PERCEPTUAL PROCESSING

The *extension*, on the one hand, consists in considering environmental constraints from an *adaptive* perspective. Although compatible with a loose ecological reading, the thesis that I will defend in Part II of the present work is that the principle according to which the environmental niche has a direct role in the determination of properties relevant for perceptual processing must be grounded in an adaptive hypothesis. Such adaptive hypothesis should make explicit the actual increase in fitness for the organism resulting from the adoption of particular kind of informational regularities in its environment. Reformulating the idea of direct perception into an *adaptive* framework allows us to understand *why* the contingent structure of the environment in which the organism is embedded should be taken into account for explaining the emergence and functioning of a number of perceptual skills. In order to provide more than a mere description of abstract informative relations, a *perceptual* hypothesis on the role of adaptive constraints must be formulated, together with some hypotheses on computationally plausible mechanisms that exploit such constraints. The main limits of the ecological approach - I argue - can be found in the fact that it has assumed as an object of investigation

the laws of ecological optics without actually addressing the question of how perceptual mechanisms might be designed to take advantage of such laws. Hence, studying *adaptive* and not merely *ecological* constraints means asking to what extent the kind of rich informational regularities studied by ecological theories (and many others) can be integrated into a robust perceptual hypothesis: the goal, then, is to specify the nature of candidate processing strategies that are likely to result in an increased cognitive benefit for the organism compared to other strategies.

4.1.3 ADAPTIVE CONSTRAINTS AND THE DIRECT PICKUP OF PROPERTIES

The *specification* of the ecological view offered by the present proposal consists in describing a class of mechanisms that might be used to provide a more constrained notion of *affordance*. The class of mechanisms that I will address in the present section – the exploitation of distributional regularities in the organism’s environment for a number of smart perceptual skills – can be seen as an attempt to define in a rigorous in which sense certain highly informational properties can be considered as *directly picked up* by perceptual systems. The aim of my analysis in part II, in other words, is to outline an alternative account of basicness criteria, to explain how certain properties of the stimulus might acquire their perceptual relevance from the fact of encoding certain environmental regularities with a high adaptive potential. To do this, I will adopt the following strategy:

1. I will try to characterize the sensory environment of an organism as a *patterned domain* in which specific features are characterized by robust *distributional properties*.
2. I will argue that the perceptual relevance of such patterns is determined

by two factors: by their coinstantiation in the environment with cognitively valuable properties that are advantageous for the organism and by the benefit resulting to the organism from relying on such patterns.

3. Such benefit, finally, is what makes perceptual strategies based on those sensory patterns adaptive over other perceptual strategies.

The expected conclusion of the following chapters will be a new formulation of what constitutes a basic feature with an adaptive value for environmentally bounded organisms.

4.1.4 BEYOND INTERNAL CONSTRAINTS

Extending the study of the constraints on the selection of relevant perceptual variables to adaptive constraints – i.e. to environmental constraints with an adaptive value for the organism – as we will see in Chapter 7, has a number of crucial consequences. I have argued in part I that current vision science has privileged a specific kind of issues as prototypical cases of perceptual problems deserving explanation. These problems can be qualified as internalist issues, in that they focus on the necessary internal requirements perceptual systems have to possess in order to be able to deliver reliable information on *any* kind of property and entity of the environment. Such problems, as I argued in Chapter 3, start from the assumption that basic features do not provide *per se* sufficiently reliable information on the distal sources of the stimulation. Consequently, since correct perception means reconstructing the correct distal sources on the basis of poorly informative sensory, successful perceptual mechanisms are those that possess adequate internal resources (like rich inferential mechanisms) to correctly interpret sensory data. I have stressed that such problems do not take into account the structure of the organism's

environment, precisely because they are meant to be common problems that face the organism in *whatever environmental conditions* it happens to be. This strongly internalist stance has resulted in the restriction of relevant perceptual inputs to properties defined on the basis of processing constraints. The role played by *environmental constraints* in shaping perceptual systems has been substantially neglected by mainstream research.

In which sense taking into account adaptive constraints can make a substantial difference? I will argue that the study of adaptive constraints is likely to shed light on a different class of properties of the visual stimulation, that - contrary to internalistically defined basic features - have a highly informative value for the organism. There are good reasons to assume that certain properties of the stimuli, within sufficiently narrow environmental context, maximize the organism's cognitive utility over other strategies. This might in principle provide a plausible explanation for a number of empirical data hardly fit the standard view.

I will dedicate the remainder of this Chapter to a short analysis of the existing methodological paradigms that have stressed the functional importance of environmental regularities in explaining adaptive perceptual and cognitive capabilities. In the following two Chapters (5-6), I will introduce a perceptual hypothesis (the "perceptual shunt hypothesis") drawing on the exploitation of environmental regularities. In Chapter 7, finally, I will address some major consequences related to the study of adaptive constraints and their potential impact on standard theories of perception.

4.2 THE METHODOLOGICAL LANDSCAPE

The role of an organism's environment in the explanation of its perceptual abilities has been addressed by a large literature spanning from developmental

psychology, to cognitive ethology, computational vision, and experimental psychology. We should distinguish here two distinct orders of problems that are relevant for the present analysis.

- The first, more abstract order of questions concerns the *epistemological* issue of what it is meant by *internalization of environmental regularities*. What it means for a specific perceptual capability to mirror environmental regularities that might have been selected during phylogenetic or ontogenetic development is a general question that I will discuss in Chapter 7.
- The second, more methodological level concerns some specific *methodological* proposals that have developed interesting theoretical notions for understanding what we mean by “environment structure”. In particular, such paradigms have tried to characterize what counts as an *environmental regularity* and how to describe the *structure* of an organism’s environment in a way that might be relevant for explaining some environmentally-tailored perceptual abilities.

In what follows, I will focus on the second point, by reviewing the main theoretical contributions developed within three distinct methodological paradigms that have dealt with the issue of studying ecological regularities.

4.2.1 DISTRIBUTIONAL CUES AND STATISTICAL LEARNING

The first research programme that deserves consideration is that part of developmental psychology that has recently focused on the study of distributional cues and their role in the acquisition of perceptual and linguistic abilities in children.

Considerable attention has been recently given to the study of mechanisms allowing infants to acquire some complex structures that are required for the emergence of full-fledged linguistic competence. Several researchers have embraced a moderately empiricist view according to which the first stages of language acquisition in human beings can at least partially be explained by looking at the statistical regularities of the linguistic environment to which infants are exposed. Although dominating theories of language acquisition have emphasized the role played by innate and experience-independent mechanisms (see [Chomsky, 2000](#), in particular the arguments that innate mechanisms are essential to account for the precocity and robustness of linguistic acquisition, even when appropriate and rich stimuli are absent), it is hardly deniable that a number of experience-dependent factors are crucial for bootstrapping linguistic development. Recent research works have demonstrated that infants possess powerful mechanisms of statistical learning that allow the extraction of salient regularities from their linguistic environment, thereby vindicating the idea that experience may play a more important role in the acquisition of language than existing theories have suggested so far.

[Saffran et al. \(1996a,b\)](#) have shown that segmentation of words from fluent speech can be accomplished by 8-month-old infants based solely on the statistical relationships between neighboring speech sounds, what they called the *transitional probability* between syllable pairs. Infants' precocious ability to extract an alleged complex property like word boundaries might then be explained, they argued, by looking at infants' sensitivity to the distribution of low-level properties of the speech stream that *happen* to be coinstantiated with word boundaries. This contingent but robust correlation of low-level features and high-level properties in the linguistic environment might be then the explanation of why children manage to cope with the seemingly

overwhelming complexity of word boundaries extraction. Computational studies ([Redington and Chater, 1997](#); [Redington et al., 1998](#)) and analyses of linguistic corpora ([Durieux and Gillis, 2000](#)) have confirmed that distributional properties of the linguistic environment might explain the early ability of extracting morphological, syntactic and semantic structures based on their cooccurrence with low-level features of language input.

[Kuhl \(2000\)](#) on the basis of recent results in developmental psychology focusing on the study of distributional cues in language acquisition, has proposed a number of general principles that might clear a path towards a new account of the interplay between innate factors and environmental constraints on language acquisition. Among the tenets of this new view of language acquisition, three are particularly relevant for our analysis:

- (i) infants initially parse the basic units of speech allowing them to acquire higher-order units created by their combinations;
- (ii) the developmental process is not a selectionist one in which innately specified options are selected on the basis of experience;
- (iii) rather, a perceptual learning process (...) commences with exposure to language, during which infants detect patterns, exploit statistical properties, and are perceptually altered by that experience;

Taken together, these principles suggest that infants are: “neither the tabula rasa that Skinner described nor the innate grammarians that Chomsky envisioned – [Kuhl](#), cit. p.11856.

Let us consider a little closer the implications of these tenets.

The first interesting aspect concerns the new moderate notion of *innateness* emerging from this paradigm. What is innate regarding language, it seems, is not much a universal grammar, containing already all the possible structures that are to be selected through experience, but rather a set of *biases* that place constraints on perception and learning during exposure to ambient language. These constraints are recruited thanks to their adaptive role in the

specific linguistic environment in which infants develop. Organisms endowed with such biases are in fact better performing in abstracting those higher-order properties that are required for linguistic competence. Second, the interesting aspect is the idea that being exposed to an environment with strong distributional regularities and having the capability to detect some statistically salient patterns in the environment actually *warps* perception in service of the acquisition of specific skills. Language experience not only produces a change in infants' discriminative abilities, it results in a "mapping" that literally alters perception, giving more relevance to certain classes of patterns over others.

It is an open empirical question to clarify the scope of such statistical learning abilities. It is still unclear whether this kind of statistical learning is *language-specific* or it can be regarded as an instance of a more general learning mechanism applicable to a broad range of distributional analyses of environmental input.

Still, we can assume that the general requirement for this kind of learning might be compatible with a number of different domains: the idea that the perceptual environment must be shaped as a *structured domain*, i.e. a domain in which some patterns – compatible with the subject's perceptual devices – occur with statistic regularity. It is in virtue of

1. this patterned structure;
2. the robust distributional correlations between certain classes of sensory patterns and higher-order, cognitively relevant properties in the linguistic environment;
3. the existence of specific perceptual biases tailored to these patterns;

that relatively "dumb" but highly adapted mechanisms can give rise to smart

perceptual and cognitive capabilities such as the ones displayed in linguistics statistical learning.

4.2.2 ECOLOGICAL RATIONALITY AND SIMPLE HEURISTICS

Heuristics-based theories of perception have emphasized the benefits deriving from the environmental closure of perceivers and the fact that structured environments can work as *external cognitive enhancers* for organisms with limited computational capabilities. The existence and use of simple heuristics based on environmental closure has been addressed by a number of works which have defined the paradigm known as *Ecological Rationality* (Gigerenzer and Selten, 1999; Gigerenzer and Todd, 1999; Todd, 1999b).

The basic assumption of the ecological approach to rationality is that the structure of an organism's environment is the main factor determining the success of its cognitive capabilities. Traditional internalist criteria – that have considered as a benchmark for cognitive systems the achievement of “general-purpose, optimal performance in any situation, no matter how rare; for any price, no matter how costly; and for any reward, no matter how meager” (Bullock and Todd, 1999) – must be replaced, following the defendants of the ecological rationality approach, by an *externalist performance metric*:

The extent to which an organism fits its niche, or a mechanism matches the problem it faces, is the extent to which it meets the demands of its environment. [...] The assessment of candidate cognitive mechanisms must be sensitive to facts concerning environment structure –Bullock and Todd, *cit.*

Stressing the role of environmental structure (as opposed to internalist criteria) for assessing the performance of a cognitive system and adopting bounded rationality (as opposed to general-purpose, unbounded rationality) as the

appropriate framework for studying intelligent behavior is the necessary requirement for understanding the selection of particular cognitive strategies in *real world, biological decision systems*. Real decision systems (like most biological cognitive systems) must employ *limited information* to make choices in *specific situations* and under a *limited amount of time*. Strategies that enhance an organism's performance under these constraints are hence good candidates as *adaptive mechanisms* for real world organisms.

I will not address here the various issues raised by the study of cognitive capabilities based on so called "fast and frugal heuristics". One aspect that I would like to retain, though, of this paradigm, for the sake of the present analysis, is that adopting an ecological rationality perspective in the study of perception requires the development of a theory of environmental structure and an account of the way in which this structure can be *measured*.

If different environment structures favor different cognitive mechanisms, what is needed is an account of how a given environment structure determines what are the successful cognitive mechanisms. Studying the *frequency structure* of the environment (the distribution of items of possible choice within the decision domain) or its *significance structure* (the manner in which items of possible decision differ in terms of their consequences for the organism's goal), represent two paradigmatic ways of *measuring* an environment's structure and assessing what cognitive strategies will be favored. In the specific case of perceptual mechanisms, the study of environment structure will require an understanding of the distributional properties of specific properties and their reliability as cues for accessing higher-order properties. Assessing what are good properties for environmentally bounded perceptual systems, hence, requires studying the contingent structure of the environment in which they fit.

4.2.3 BAYESIAN PERCEPTION AND PERCEPTUAL INFERENCE

The problem of understanding what are good perceptual properties for organisms that are embedded in specific environments has become one of the major trends of investigation in *Bayesian approaches* to perception. Although complementary to studies based on the hypothesis that biological organisms are sensitive to frequency distributions, bayesian models of perception ([Knill and Richards, 1996](#)) have addressed the same basic issues studied by researchers in the field of ecological rationality, i.e. the problem of understanding how external regularities make perceptual decisions based on the selection of specific properties *reliable* within specific environmental settings.

By adopting a Bayesian framework, such studies have managed to describe the conditions that must be ideally respected for a property to be a “key feature”, a property “unlocking reliable inferences about the world” ([Jepson and Richards, 1992](#)). Although valid for ideal situations in which many factors are abstracted, the definition of formal criteria to decide if a property may count as a key feature with regard to a specific environment has been particularly crucial because it has provided a rigorous characterization for the intuition that perceptual systems might be tuned to particular properties because of their “suspicious” or “non-accidental” character ([Barlow, 1985b](#)).

[C]onsider configurations of features that exhibit very special relations to one another, such as two line segments which intersect to form a *T* or a *V*, or two line segments that are collinear. As noted by many, intuitively, such coincidences imply very special “suspicious” and informative events. Surprisingly, however, in an unrestricted context, such as a world in which sticks are positioned arbitrarily, the observation of a “non-accidental” feature typically does *not* imply the intended world property. [...] Context plays

a crucial role. To correct this situation, the corresponding world event must express a generic regularity in that context – [Jepson and Richards](#), cit., p.84

This amounts to acknowledge the fact that some specific properties of sensory stimulation, that are in the general case not informative, can acquire a high degree of informativeness within sufficiently narrow contexts.

Past studies in perceptual psychology have investigated the role of “non-accidental” properties like STRAIGHTNESS, COTERMINATION, PARALLELISM, RIGIDITY, COLINEARITY OR SKEW SIMMETRY in narrowing the scope of perceptual inferences. The bayesian approach has developed a unitary framework for explaining the special status such properties play within *given* environmental settings and can hence be considered one of the most powerful tools for describing how the contingent structure of a specific environmental context can modulate the informativeness of specific features.

It should be noted, though, that Bayesian models as such do not constitute full-fledged explanations of the emergence of specific perceptual skills: they do not address the question of the *origins* and *nature* of the probabilistic knowledge internalized by an organism, nor the compatibility of the model with the actual adaptive constraints met by real world biological organisms. Nonetheless, they are the best available abstract model to define a property’s *informational goodness* with reference to a specific environment structure. We will see in the following Chapter how some basic Bayesian concepts can be adopted to illustrate the informational value of certain classes of features within specific environments.

4.3 FROM ENVIRONMENT STRUCTURE TO BASIC FEATURES

The paradigms reviewed in this Chapter represent three major methodological proposals that have tried to articulate the notion of environment structure and its relevance for the understanding of perceptual phenomena. In different ways, they offer a conceptual framework for defining what counts as a *structured or patterned environment* and what predictions can be done on the performance of perceptual systems that are tailored to such an environment. In the following chapters I will propose a perceptual hypothesis (the *perceptual shunt hypothesis*) that owes much of its theoretical background to these methodological paradigms. I will introduce this hypothesis by tackling a class of theories – *sensorimotor theories* – that have been recently proposed alternative accounts of the explanation of some perceptual capabilities (**Chapter 5**). My goal will be to show that such theories can be seen as an instance of a more general class of perceptual mechanism, that I will analyze in depth in **Chapter 6**.

Chapter 5

Sensorimotor constraints

A CLASS OF interesting adaptive constraints in the selection of basic features comes from recent *sensorimotor theories of perception*. (O'Regan, 2004; O'Regan and Noë, 2001; Philipona, 2004) have proposed a research programme based on a fundamental hypothesis according to which representing systematic correlations of sensory and motor patterns can provide an organism with the requirements for a number of perceptual skills. The interest of considering these theories for the present analysis derives from the particular constraints that such theories have postulated on relevant input of perceptual processing. The particular reason why sensorimotor theories of perception represent a good source of *adaptive constraints* for the selection of basic features, is that they take into account how the organism is embedded in an environment that displays some highly informative regularities. The class of perceptually relevant patterns of stimulation of an organism, according to theories, are constituted by the class of co-occurring *sensory and motor patterns* constrained by the structure of the physical world and by the bodily structure of the perceiver. The goal of this chapter is to briefly outline the main hypotheses of sensorimotor theories of perception in order to characterize some cases of sensorimotor explanations of perceptual abilities that are

particularly relevant for the present analysis.

5.1 SENSITIVITY TO SENSORIMOTOR COUPLINGS

One of the tenets of sensorimotor theories of perception is the idea that subjects are equipped with a capability to monitor and represent regular properties of co-occurring *sensory* and *motor* patterns and that such invariant properties of *sensorimotor couplings* are among the basic patterns of stimulation to which perceptual processing applies.

Learning such regular correlations between sensory and motor patterns – which O'Regan and Noë (2001) call *sensorimotor contingencies* – is then a matter of encoding statistical regularities of the patterns the perceiver is systematically exposed to during his motor explorations of the environment. Let us consider, for example, two classes of motor schemes a subject may perform while visually fixating an object: a subject can perform a *head rotation* or a *lateral translation* while maintaining his eyes fixed on an object. These two kinds of motor scheme are regularly associated in our environment with two different types of dynamic sensory patterns, in the case of vision two distinct kinds of optical flows. Sensorimotor theories hold that there are some invariant properties in the co-occurrence of the optical flow associated with each of these specific motor schemes that a subject can extract and process for performing some perceptual tasks.

The ability to extract and represent such invariant properties from their regular instantiation in the sensorimotor environment is what we may call a *genetic hypothesis* for sensorimotor learning.¹ Such hypothesis holds that in

¹The analysis on sensorimotor constraints on feature selection presented in this chapter draws on materials of a work in progress by Taraborelli and Mossio (2005). In particular I rely on this work for the formulation of the basic hypotheses behind sensorimotor skills as a prominent case of representational abilities.

order to acquire sensorimotor knowledge,² a system must be able to *extract* and *represent* some invariant properties of sensorimotor couplings, namely:

- (a) Perceptual systems must be sensitive to systematic correlations between motor patterns and sensory patterns
- (b) If such correlations are sufficiently *robust* in the environment, such sensitivity can bootstrap a learning process
- (c) This learning process results in the *representation* of sensorimotor invariants.

For this hypothesis to hold, in other words, three distinct conditions must be met: first, the system must possess sensory devices tuned to detect the coinstantiation of specific patterns (a); second, the sensorimotor environment must be sufficiently stable, i.e. must respect a minimal regularity allowing a subject's internalized sensorimotor invariants to *reliably* represent actual sensorimotor couplings (b); finally, the system must be able to store this knowledge in a format that might be retrieved whenever actual sensorimotor couplings are experienced (c).

A crucial condition for sensorimotor theories - that is relevant for the present analysis - is the availability of robust distributional regularities in the organism's environment. Sensorimotor environments can be qualified as patterned domains in that they present a number of robust invariances that allegedly bear a high informativeness for the organism within sufficiently narrow contexts. The existence of such distributional regularities and their coinstantiation with cognitively valuable information for the organism is what give to sensorimotor constraints, I maintain, the status of adaptive constraints.

²I use here the notion of "sensorimotor knowledge" to refer to any internal state a perceptual system has acquired from its past exposure to the environment that the perceiver may use to parse and categorize ongoing sensorimotor correlations.

5.2 PERCEPTUAL SKILLS BASED ON SENSORIMOTOR KNOWLEDGE

If we grant that perceptual systems have the capability to represent invariant properties of sensorimotor couplings, we can ask how such sensorimotor knowledge can be exploited to retrieve cognitively valuable information. Defendants of sensorimotor theories have argued that a number of perceptual abilities, traditionally accounted for by referring to specific kinds of neural processing of the sensory stimulation, should instead be explained by referring to the sensorimotor knowledge acquired by the perceiver. I will consider hereafter two cases of perceptual abilities based on alleged sensorimotor knowledge and argue that if such explanations are supported by empirical evidence, they can shed light on interesting adaptive constraints on perceptually relevant patterns of the stimulation.

COLOR PERCEPTION

Color perception is traditionally assumed as the result of a complex set of processing stages of the visual stimulation meant to determine the chromatic values of a given surface as a function of the wavelength of the stimulation and of global luminance conditions ([Zeki and Marini, 1998](#)). Although models of chromatic invariance have provided robust explanations for a number of perceptual judgments under specific luminance conditions, they have arguably not been able to provide a full account for the specific relations between colors in the perceptual space ([Philipona, 2004](#)). Sensorimotor theories of color perception (see [Broackes, 1992](#)) have recently challenged the idea that the structure of perceived color might be reduced to computational processing of specific properties of the retinal sensory stimulation *independently from* a contribution from motricity. Following this lines, O'Regan and collaborators have designed a number of empirical tests as well as formal models

(Philipona et al., 2005) to investigate the hypothesis according to which the perceptual structure of color may be explained in terms of a subject's *mastery of sensorimotor contingencies*.

Under the present view of what seeing is, the visual experience of a red color patch depends on the *structure of the changes* in sensory input that occur when you move your eyes around relative to the patch, or when you move the patch around relative to yourself [...] the sensation of “red” comes from the structure of changes that is caused by “red” – (O'Regan and Noë, 2001, p.951-2)

In particular, some interesting predictions have been formulated on the expected consequences on perceived color of manipulating a subject's sensorimotor knowledge. Assuming that the perceived chromatic quality of a visual item depends on the systematic changes it produces on the sensory patterns through motor interaction, Bompas et al. (2002) have tried to investigate to what extent a subject's judgments on the perceived color of an object could be altered by systematically modifying the sensory changes produced on the retina during eye movements. The experiments consisted in systematically manipulating the transformation of visual stimuli co-occurring with eye motion in order to force the perceiver to learn new sensorimotor correlations between his actions and the resulting sensory patterns.

The idea, partially confirmed by these experiments,³ is that after a period of sensorimotor re-training a subject will not report any significant chromatic difference between a red patch turning to green each time the subject performs a specific eye motor patterns: the sensorimotor training, it is argued, affects the way in which the subject associates REDNESS to specific sensorimotor

³The fact that the results were less significant than predicted was justified by the authors as due to the limited plasticity of an adult's visual system and to the short period of adaptation undergone by the subjects.

couplings; the new sensorimotor knowledge the perceiver acquires as a result of this training gives him the perceptual impression that the green patch in peripheral vision and the red patch in central vision are *the same color*. What is interesting for our analysis in this proposal is the fact that such studies seek to reduce some specific perceptual skills consisting in the detection of a perceptual property to the ability of monitoring some low-level regularities in sensory and motor patterns that are systematically associated to this property within a sufficiently constrained environment. We should stress that whereas sensorimotor laws connecting motion and sensation of the organism can be described in terms of nomic correlations, the relation between regularities in sensorimotor patterns and perceived color is - I assume - contingent, in that it depends on the specific environmental context in which the organism is embedded.

SPATIAL PERCEPTION

A second interesting case of perceptual skills based on the representation of sensorimotor regularities is the extraction of spatial properties. [Philipona et al. \(2003\)](#) suggested that a perceptual system can virtually extract a number of spatial properties of its environment by merely relying on a set of sensorimotor rules internalized during active exploration of the environment. This amounts to saying that perceiving an object's *distance* or *size*, for instance, is a matter of exploiting the appropriate sensorimotor rules acquired during past experience.

Let us consider the following example. Be $D(P)$ the *relative distance* of P from the observer and $D(Q)$ the relative distance of Q from the observer, where P and Q are objects of equal size. What properties are used by a perceptual system for estimating if $D(Q) < D(P)$?

Traditional approaches, on the one hand, take difference in size of the retinal projection of the objects as an example of relevant variable that perceptual systems must extract. This assumption derives from basic considerations of projective geometry: the size of the projections of two identical segments on the retina is inversely proportional to their relative distance. Since the relative distance is the (external) property that has to be estimated, then a difference in size of retinal projections is taken as the relevant sensory variable exploited for solving this specific spatial task. Relevant stimulus properties are then those that a geometrical mapping associates with external spatial properties.

Following a sensorimotor approach, on the other hand, relevant properties for estimating the distance of objects are properties of dynamic sensory patterns associated with a specific class of motor schemes. For example, a perceiver's lateral translations will be regularly associated with a kind of optical flows in which two different angular velocities are detectable: the closer an object is to the perceiver, the bigger the angular velocity of its retinal projection when the perceiver performs lateral translations. Relevant stimulus properties are such that, *insofar as they are coupled with specific classes of motor schemes*, they allow the system to discriminate between two different classes of sensorimotor correlations, namely to represent different sensorimotor invariants. What enables a perceiver to make a distinction between a close and a distant object is then the alleged ability to discriminate the sensory pattern each of them produce when the subject performs specific classes of motor schemes.

Again, the relevance of this kind of explanation for our analysis lies in the fact that the perception of spatial properties of the perceptual environment is reduced, according to sensorimotor theories, to the ability of extracting and representing the cooccurrence of sensory and motor patterns: couplings of

sensory and motor patterns are then given a crucial functional role as input for perceptual processing in such theories.

5.3 SENSORIMOTOR CRITERIA FOR FEATURE GOODNESS

The most interesting aspect of sensorimotor approaches to perception reviewed in this Chapter is the way in which they orient the characterization of relevant variables for perceptual processing. The central explanatory role given by sensorimotor theories to dynamic properties of the sensory patterns is consistent with a large literature that has criticized the primacy given by mainstream perceptual science to static properties of sensory patterns as basic perceptual properties. Vision science has for a long time considered dynamic properties of retinal patterns as properties that are *derived*, *later extracted*, or *reducible to* static properties of the retinal image, appropriately integrated over space-time. Starting from ecological approaches to the study of perception (Gibson, 1979), though, dynamic properties of the sensory stimulation have been given a central explanatory role, often insisting on the fact that they should be considered in many respects more *primitive* than static properties. Along the same line, the sensorimotor paradigm suggests that perceptually relevant for perceptual processing are dynamic patterns of the stimulation. So far there is nothing radically new compared to mainstream vision science. The interesting constraints introduced by sensorimotor theories concern the way in which relevant properties of the sensory patterns are characterized. Since the basic functional units of sensorimotor learning are sensorimotor couplings, the following criteria on feature basicness can be introduced:

- (1) Basic properties of the sensory array should be identified with those dynamic properties that are susceptible of being systematically coupled with motor patterns: among the countless sets of dynamic patterns that

can be described on the retinal stimulation, for example, the subset of patterns that systematically co-occur with a specific class of motor schemes of the perceiver (like, for instance, head rotation) should be considered as relevant input for perceptual processing.⁴ This is coherent with some remarks by MacKay (1985) regarding some specific patterns of neural activity in the visual cortex that can hardly be described in relevant functional way with respect to perceptual content, but that acquire a potential functional justification as soon as they are interpreted as patterns signaling the co-occurrence of sensory transformations with underlying oculomotor patterns:

an alternative interpretation for the significance of “feature sensitivity” in visual cortical cells [is possible]. Instead of seeing it as leading simply to a primitive description of the visual scene in terms of the firing rate profile of “feature detectors”, it seems attractive to see it as helping to segregate sensory signals whose main information content has to be extracted by discovering *what co-varies with what* (and in what ways). (...) The categories to which they are sensitive must, of course, be those that are likely to co-vary in a functionally meaningful way as the projected retinae rove over the visual world during oculomotion or locomotion.

- (2) SENSORIMOTOR COUPLINGS themselves, i.e. regular coinstantiations of properties of sensory and motor patterns that bear contingent correlations with some cognitively valuable properties (like COLOR) in a given environmental setting, can be considered *as such* basic properties that

⁴It is interesting to observe that the introduction of constraints on relevant sensory patterns that are controlled by classes of motor patterns is formally analogous to the one introduced by motor theories of speech perception, according to which the relevant segmentation and parsing of sensory patterns into perceptually relevant units is controlled by articulatory skills (Lieberman and Mattingly, 1985; Lieberman et al., 1963).

constitute the entry-level of perceptual processing.

I maintain that these criteria provide two constraints on *feature basicness* that significantly challenge the traditional view based on internal constraints of perceptual systems. The second, in particular, counts as a prominent example of an *adaptive constraint* insofar as it derives from the specific match between stimulation patterns co-instantiated with cognitively valuable information in virtue of a contingent relation valid within a specific environmental setting. I will show in the next chapter how this kind of adaptive constraints on perceptually relevant patterns can be described as an instance of a more general mechanism that I call *perceptual shunt*.

Chapter 6

Perceptual shunts

IN THIS CHAPTER, I introduce a hypothesis for a class of perceptual mechanisms, drawing on some proposals formulated in developmental and perceptual psychology, that I call *perceptual shunts*. According to this hypothesis, a number of complex perceptual skills should be understood by looking at the subject's use of some patterns of the sensory stimulation that bear a contingent but informationally reliable relation with cognitively valuable properties within some sufficiently narrow environmental settings. By articulating this hypothesis, I will show how it introduces a class of adaptive criteria on the characterization of basic features that challenge the standard view according to which relevant properties of the sensory stimulation should be characterized by relying on merely internal constraints. I argue that if this hypothesis holds, then a relatively unexplored class of perceptual capabilities could be opened to empirical investigation.

6.1 ROBUST ENVIRONMENTAL CORRELATIONS

In a debate about the prospects of an empiricist approach to the study of cognitive development, [Keil \(2000\)](#) addresses the issue of understanding:

how far “up stream” domain specific specializations exist as we consider the flow of information from sensory transducers themselves “up” to the highest levels of cognition.

He reports a work by [Johnson and Morton \(1991\)](#) in which the authors – he argues – embrace an *enlightened empiricist view* for the explanation of how a specific class of perceptual skills are acquired by children¹. The hypothesis concerns an alleged mechanism for the development of computationally affordable and reliable mechanisms for face perception. Johnson and Morton grant the newborn something like a 3-blob inverted triangle detector that matches the stimuli corresponding to eyes and mouth of human individuals and that enables the infant to “lock onto” faces. A specialization to process face-like information, they argue, might then plausibly arise from the contingent fact that the perceptual environment of newborns is populated with 3-blob inverted triangles that happen to be coinstantiated with faces. Whether this correlation between simple triangular configurations and faces should be regarded as a necessary requirement for bootstrapping more fine-grained and flexible mechanisms of face recognition or as providing as such a reliable basis for efficient perceptual performance in the early stages of cognitive development is a debatable issue. The interest of this proposal for the present analysis, though, consists in the fact that it provides the rationale for a general hypothesis on the functioning of a class of perceptual capabilities based on contingent environmental correlations. Such studies suggest a prototypical case of highly specialized and dedicated (domain specific) perceptual capabilities for picking out cognitively valuable information for the organism, by relying on the extraction of patterns of the visual stimulation that are

¹It is an *enlightened* empiricism, he observes, in that it allows a domain specific processing system for allegedly complex configurations like faces to have a specific neural instantiation.

coinstantiated with such properties in a given environmental setting. To put it differently, for a number of complex perceptual skills that are usually taken to require a considerable cognitive investment, the possession of skilled inferential abilities and previously acquired world knowledge, it is plausible to assume, on the contrary, that some relatively rigid, cognitively affordable and noninferential mechanisms can be exploited in virtue of contingent environmental correlations. As long as the organism is embedded in an environmental niche in which these correlations hold, the retrieval of some sensory patterns in the sensory stimulation can provide a sufficiently reliable strategy to retrieve cognitively valuable information² from the environment.

Although Keil's examples are meant to provide evidence for the existence of cases of *cortical specialization* based on the recruitment of low-level features as cues for bootstrapping sensitivity to high-level properties, I borrow here his terminology to define a class of noninferential capabilities (that I will call "perceptual shunts") based on the existence of strong correlations in the perceptual environment of the organism that do not necessarily imply strong constraints on their cortical realization. I define the perceptual shunt hypothesis as follows:

²I will use hereafter the notion of *cognitively valuable* information to refer to any kind of information that is advantageous for the organism to increase its survival; I assume that such loose characterization includes properties that, although not strictly related to the organism's survival, are adaptive insofar they allow him to increase its utility with respect to its ordinary routines

Perceptual shunt hypothesis

Whenever a property P which is cognitively valuable for an organism S is robustly coinstantiated within a given environmental context with a sensory property Q and S is endowed with perceptual mechanisms to pick out Q , we will say that S is able to *shunt* P , i.e. to pick out and track occurrences of property P by picking out and tracking occurrences property Q .^a

^aIf shunt can apply to P in virtue of Q we can also say that Q can be considered, within the same environmental context, as a *perceptual proxy* for P

A way to rephrase this characterization consists in saying that within a given environmental niche in which the occurrence of property Q is robustly correlated with the occurrence of property P , we will not be able to estimate any significant difference at the level of performance, *ceteris paribus* between an organism endowed with perceptual devices for picking out Q and an organism able to pick out P .

A paradigmatic case of a perceptual mechanism that can be subsumed under the definition of perceptual shunt is [Lettvin et al. \(1959\)](#)'s case of bug detectors in the frog visual system. It is worth reporting an extensive conclusion of their paper:

The output from the retina of the frog is a set of four distributed operations of the visual image. These operations are independent of the level of general illumination and express the image in terms of 1) local sharp edges and contrast, 2) the curvature of edge of a dark object, 3) the movement of edges, and 4) the local dimmings produced by movement or rapid general darkening. (...) We have described each of the operations on the retinal image in terms of what common factors in a large variety of stimuli cause response and what common factors have no effect. What, then, does a particular fiber in the optic nerve measure? We have considered it to be how much there is in a stimulus of that quality which excites the fiber maximally, naming that quality. The operations thus have much more the flavor of perception than

of sensation, if that distinction has any meaning now. That is to say that the language in which they are best described is the language of complex abstractions from the visual image. We have been tempted, for example, to call the convexity detectors "bug perceivers." Such a fiber (operation 2) responds best when a dark object, smaller than a receptive field, enters that field, stops, and moves about intermittently thereafter. The response is not affected if the lighting changes or if the background (say a picture of grass and flowers) is moving, and is not there if only the background, moving or still, is in the field. Could one better describe a system for detecting an accessible bug?

The case described by [Lettvin et al.](#), independently of its actual neural implementation, can be qualified as an instance of a perceptual shunt mechanism insofar as the frog (F) possess a perceptual mechanism for picking up the sensory patter Q : "a dark object, smaller than a receptive field, [that] enters that field, stops, and moves about intermittently thereafter" and, within the F 's environment, the property Q is robustly coinstantiated with the property P : "being a bug". We will then say that the F is able to shunt P if it can pick up and track occurrences of P by picking out and tracking occurrences of Q .

Now imagine that a different species of frog (R) is endowed with perceptual devices for recognizing black bugs not only by detecting their color, shape and motion patterns but also by detecting the precise pitch of their buzz. Having a secondary mechanism for pitch discrimination is essential to R insofar as in a neighboring swamp where it uses to search for food there are dark bugs that buzz at a different pitch but that are inedible. Now, if the definition of shunt applies to F we can affirm that, *ceteris paribus*, in F 's environment we will not be able to estimate any significant difference in performance between F and R (although on its very first tour to another swamp F will be in serious

danger).

Now, apart from exemplifying cases of perceptual skills that can be considered as instances of perceptual shunt mechanisms, we can provide a simple formalization of the way in which contingent correlations result in reliable information within sufficiently narrow environmental contexts.

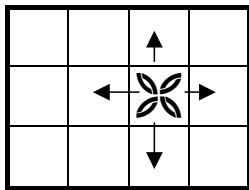
6.2 PERCEPTUAL SHUNTS IN A SIMPLIFIED ENVIRONMENT

We can characterize the functioning of “perceptual shunt” mechanisms for a cognitively valuable property X based on picking out of a property P in a given environment W by using a simple basic Bayesian framework.

Let a generic **world** W_n be defined as a bi-dimensional space with a discrete number of locations (which we identify as a (m, n) matrix).

	$m1$	$m2$	$m3$	$m4$
$n1$				
$n2$				
$n3$				

The inhabitants of such a world – I will call them **shunters** – are very simple creatures that can occupy one location at a given time and are able to move around the world by shifting from one location to one of the contiguous locations.



Shunters are endowed with very primitive sensory devices that allow them to perform basic sensory measurements that we will call **temperature**, which

is a function of the number of neighboring edges they can detect from a given location.

We will say that a shunter is *cold* or *feels temperature* P_0 when it senses no neighboring edge:



A shunter is *warm* or *feels temperature* P_1 when it senses one neighboring edge:



A shunter is *hot* or *feels temperature* P_2 when it senses two distinct edges:



A shunter is *hottest* or *feels temperature* P_3 when it senses three distinct edges:




Shunters *die* as soon as they feel temperature P_4 , i.e. when they sense four edges surrounding them:



We define a **sensory feature** any particular temperature measurement a shunter can perform without dying. Given this simple characterization of shunters and their environment, we can describe how shunters may display knowledge of the structure of their world by relying on regular correlations between world properties and sensory features like temperature.

WORLD1


W_1 is a world in which its inhabitants can experience a difference between two sensory states. It consists of a (3,2) grid of locations.

	<i>m1</i>	<i>m2</i>	<i>m3</i>
<i>n1</i>			
<i>n2</i>			

We define Q as the property *being in a corner* and B as the property “*being in contact with a world boundary*”. Given the structure of W_1 we can assume as a general axiom that $Q \supset B$. In W_1 there is a systematic correlation, on the one hand, between the property *being in a corner* (Q) and detecting the sensory feature P_2 , and, on the other hand, between *being in contact with a world boundary but not a corner* ($G := Q \wedge \neg B$) and detecting the sensory feature P_1 . We will say that in W_1 a shunter can shunt property Q *being in a corner* by simply detecting property P_2 . The reliability of this shunt, in this case, is granted by the fact that the posterior probability $p(Q|P_2, W_1)$ is equal to 1.

WORLD2

W_2 is a slightly larger world than W_1 consisting in a (3,3) grid of locations.

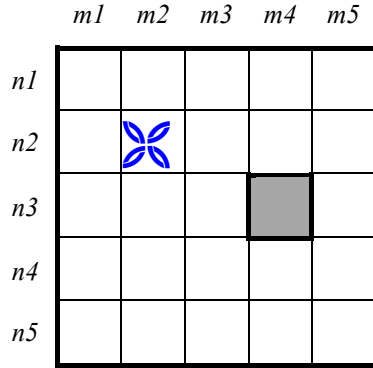
	<i>m1</i>	<i>m2</i>	<i>m3</i>
<i>n1</i>			
<i>n2</i>			
<i>n3</i>			

Properties Q and B are exemplified in W_2 . In W_2 we can define the further property F “*being free*” or “*being in contact with no world boundary, corner or obstacle*”. By our definition, if a shunter is free (F) then it is *cold* (i.e. it feels a temperature P_0). Trivial as it may seem, there is then a systematic correlation ($p=1$) between the detection of a property P_0 and the property of being free from any boundary. On the contrary, in W_2 a straightforward

correlation (with $p=1$) between property $\neg Q$ “*not being in a corner*” and a specific sensory feature is lost, since a shunter which is not in a corner can either feel a temperature P_0 or P_1 , for instance if it moves to (m_2, n_2) or in (m_1, n_2) . Depending on the amount of errors a shunter can suffer, I will assume that its performance based on shunting mechanisms will be more or less reliable to a degree of probability corresponding to the posterior probability p .

WORLD3

W_3 is a world in which a shunter cannot exploit straightforward correlations (i.e. with posterior probability $p = 1$) between world properties and temperature. W_3 consists of a (5,5) grid of locations, one of which is occupied by an impenetrable obstacle:



In W_3 we can hence introduce the further property O “*being in contact with an obstacle*”. We can also adopt a basic bayesian framework to evaluate the relative posterior probability for a world property to be shunted on the basis of a specific temperature, given the prior distribution of world properties in the environment.

$$p(P_0|W_3) = \frac{5}{24} - \text{‘feeling cold’}$$

$$p(P_1|W_3) = \frac{14}{24} - \text{'feeling warm'}$$

$$p(P_2|W_3) = \frac{5}{24} - \text{'feeling hot'}$$

$$p(Q|W_3) = \frac{4}{24} - \text{'being in a corner'}$$

$$p(B|W_3) = \frac{16}{24} - \text{'being in contact with a world boundary'}$$

$$p(G|W_3) = \frac{12}{24} - \text{'being in contact with a world boundary but not a corner'}$$

$$p(O|W_3) = \frac{4}{24} - \text{'being in contact with an obstacle'}$$

$$p(F|W_3) = \frac{5}{24} - \text{'being free'}$$

Thanks to Bayes theorem, knowing the prior probability and the distributions of properties in the world, we can express the posterior probability that a shunter will be able to shunt a particular world property given the temperature it detects:

$$p(X|P_n, W) = \frac{p(X|W) * p(P_n|X, W)}{p(P_n|W)}$$

Let us consider a few examples:

1. *What is the probability in W_3 that a shunter is in contact with an obstacle when it is feeling warm?*

$$p(O|P_1, W_3) = \frac{p(O|W_3) * p(P_1|O, W_3)}{p(P_1|W_3)} = 4/24 * 3/4 * 24/14 = 3/14$$

$$\Rightarrow p(O|P_1, W_3) = \mathbf{.21}$$

2. *What is the probability in W_3 that a shunter is in a corner when it is feeling hot?*

$$p(Q|P_2, W_3) = \frac{p(Q|W_3) * p(P_2|Q, W_3)}{p(P_2|W_3)} = 4/24 * 1 * 24/5 = 4/5$$

$$\Rightarrow p(Q|P_2, W_3) = \mathbf{.80}$$

3. *What is the probability in W_3 that a shunter is in contact with a world boundary but not a corner when it is feeling warm?*

$$p(G|P_1, W_3) = \frac{p(G|W_3) * p(P_1|G, W_3)}{p(P_1|W_3)} = 12/24 * 11/12 * 24/14 = 11/14$$

$$\Rightarrow p(G|P_1, W_3) = \mathbf{.78}$$

4. *What is the probability in W_3 that a shunter is free when it is feeling cold?*

$$p(F|P_0, W_3) = \frac{p(F|W_3) * p(P_0|F, W_3)}{p(P_0|W_3)} = 5/24 * 1 * 24/5 = 1$$

$$\Rightarrow p(F|P_0, W_3) = \mathbf{1}$$

With the only exception of the cold-freedom correlation (4), all other correlations are not true, i.e. they have different degrees of posterior probability in W_3 . Bayesian theories of perception (Knill and Richards, 1996) have developed a rigorous formalism to model the reliability of inferences in which posterior probabilities are inferior to 1 and to establish the conditions under which a property can be said to be a ‘good feature’ (Jepson and Richards, 1992) from the point of view of its diagnosticity about world properties.

Two general considerations can be drawn from this simple formalization of shunting mechanisms.

- Perceptual shunts should not be restricted to cases in which the posterior probability of shunting an environmental property on the basis of a sensory measurement is equal to 1. We can grant shunters the possibility of using mechanism that are reliable at probabilities inferior to 1 under the further assumption that the tradeoff between successful shunts and errors still results in a benefit for the shunter.
- There can be different characterizations for the sources of prior probabilities that are needed for justifying the reliability of shunting mechanisms, ranging from a *radically empiricist option* (according to which prior probabilities have been built through an exhaustive experience of the sensory consequences of exploring the world and learning its structure) to a *nativist one* (according to which shunters are equipped with a pre-specified knowledge of priors, in this case we would tend to describe this native, pre-built knowledge as a *natural constraint* implemented in their perceptual devices, possibly because of a successful adaptive history of the shunter's ancestors in the same environment).

6.3 PERCEPTUAL SHUNTS VS. PERCEPTUAL INFERENCES

There are other potential explanations for successful performance of an organism in situations in which, within specific environmental contexts, correlations of specific sensory patterns with world properties are reliable for the perceiver. There is a large literature on perceptual inferences and heuristics that are supposedly used by perceptual systems to reliably *infer* cognitively valuable properties from their robust co-occurrence with some observable features in specific environmental settings (see for instance [Hoffman, 1998](#)). [Feldman \(1999\)](#) suggests a general framework for understanding how a perceptual

system, embedded in a highly regular environment, might use basic heuristic rules to efficiently access hidden properties:

[C]onsider our caveman Ugg and his poisonous hemlock. Say that in the universe in general there is no relationship between color and edibility; but that inside Ugg's valley (the Boolean predicate *in_valley* set to true) blue fruits are poisonous and yellow fruits are edible, while outside the valley the reverse is true. Hence by hypothesis the universal theory T_0 contains the following set of sentences:

$$fruit(x) \wedge blue(x) \wedge in_valley(x) \Rightarrow \neg edible(x)$$

$$fruit(x) \wedge \neg blue(x) \wedge in_valley(x) \Rightarrow edible(x)$$

$$fruit(x) \wedge blue(x) \wedge \neg in_valley(x) \Rightarrow edible(x)$$

$$fruit(x) \wedge \neg blue(x) \wedge \neg in_valley(x) \Rightarrow \neg edible(x)$$

Now, Ugg lives in the valley so for him the predicate *in_valley* is true. Hence although the universal theory T_0 does *not* entail the rule “blue fruit are poisonous”,

$$T_0 \not\Rightarrow [blue(x) \wedge fruit(x)] \Rightarrow \neg edible(x),$$

Ugg's refined theory $T_{Ugg} = T_0 \cup in_valley(x)$ *does* entail this rule:

$$T_0 \cup in_valley(x) \Rightarrow [blue(x) \wedge fruit(x)] \Rightarrow \neg edible(x)$$

Feldman's notion of theory refinement states that if the refined theory of the world T_n held by a perceiver is true and this theory supports a rule R ,

then we can say that R is *justified*. The idea that in order for a rule R to be justified a perceiver should *hold* a specific *theory of the world* can be spelled out in different ways: one needs to specify what is meant by “holding” and by “theory of the world”³. Feldman’s position, though, seems to suggest that, no matter how such a “theory of the world” is internalized by the perceiver, using the rule means using an *inferential ability*.

There is a subtle but fundamental difference between the idea that a perceiver having a good theory of its environment might justifiably use some inferential rules within its environment to retrieve a complex property P and the idea that a perceiver might rely on shunt mechanisms for retrieving this property. In Chapter 7, I will defend the idea that shunt mechanisms *do not involve inferential abilities* of any kind. As we will see, tracking a cognitively valuable property P in virtue of its coinstantiation with a property Q in a given environment *does not mean* that the perceiver is actually *inferring* P from Q . What we can say, on the contrary, is that, if certain environmental conditions are met and the perceiver is equipped with devices for detecting property Q , then we can describe the perceiver as a system which is able to “shunt” P on the basis of Q : in the same environment being able to directly pick up P without relying on Q will not result in any significant improvement in the performance of the perceiver.

I will come back later to the idea that shunt mechanisms should be kept distinct from inferential mechanisms. The only point that has to be retained from this comparison with the literature on perceptual inferences is that the existence of shunt mechanisms that work on the basis of reliable correlations suggests that many seemingly complex perceptual skills can actually be redescribed as forms of highly adapted sensitivity to particular sensory patterns within

³See Chapter 7 for further discussion on issues related to statistical learning to environmental regularities

specific environments. It is arguable, though, whether a substantial part of the literature on implicit perceptual inferences might be or not reformulated in terms of shunt mechanisms.

6.4 PSYCHOLOGICAL EVIDENCE FOR SHUNTABLE PROPERTIES

Given our characterization of how shunt mechanisms work, we can now briefly survey the psychological literature to individuate some paradigmatic cases of complex or nonobservable properties that are likely to be ‘shunted’, given their strong correlation with simple sensory patterns in our environment. Most of such cases, as I suggested in the previous section, have been studied as examples of perceptual *inference*. I submit, though, that this sensitivity to complex, cognitively-valuable properties based on their co-occurrence with observable sensory features can be accounted for without referring to inferential capabilities: what the perceiver is doing while picking out and tracking a cognitively valuable property can be described in many cases as a mere sensitivity to its co-occurrent sensory patterns that – to rephrase Keil’s formulation – are environmentally ‘locked onto’ that property. I will consider in what follows some cases of properties that can be considered as paradigmatic examples of shutable properties.

(A) Perceptual agency and animacy

A number of early studies have demonstrated the existence in human beings of a robust ability to perceive entities as animate or endowed with agency ([Heider and Simmel 1944](#). See [Scholl and Tremoulet 2000](#) for a review of recent research directions). Certain simple visual patterns can give rise to percepts with properties that are typically related to high-level cognitive processing: this phenomenon has been demonstrated to be *perceptual* in nature and not involving any kind of conceptual knowledge,

being essentially stimulus-driven, automatic, encapsulated and cross-cultural. Although it has been shown that the properties that trigger the perception of animacy are basically related to the kinematics of the stimuli, it is still unclear what is the actual class of specific motion cues that can be exploited to shunt perceptual animacy. Yet, the relevance of these studies to our analysis lies in the potential reducibility of a case of alleged cognitive processing to the simple sensitivity and ability to track sensory pattern that are highly correlated in our perceptual environments with animacy:

After all, it is of no great surprise that one can *conceive* of some visual object as causing some action, as animate, or as anything you wish. But to the degree that such phenomena reflect perceptual processing, their existence is more interesting: they suggest that perceptual processes have more to do with domains previously considered to be purely cognitive – (Scholl and Tremoulet, 2000, p.305)

(B) Perceptual causality

The perception of simple motion displays as causal events has been largely studied since the work of Michotte (1946/1963). Michotte's model has been adapted and extended to explore the existence of a rich catalog of *functional relations* related to causality (see White and Milne 1997, 1999).

It is interesting to remark that most studies, including contemporary ones, have insisted on defining such phenomena as cases of perceptual *illusion* or *impression* of causality, thus assuming that the proper level for the understanding of causal phenomena is conceptual and not perceptual.

I maintain, on the contrary, that the perceptual shunt framework can account for the genuine perceptual nature of these phenomena: given the regularity of patterns that are associated with causal events, it is reasonable to assume that perceptual systems have adapted to automatically and reliably individuate causal patterns by detecting the occurrence of particular classes of motion patterns. Perceptual causality can be shunted on the basis of specific sensory patterns that have a particular saliency in our environment.

(C) Gist extraction

The *abstract meaning of a scene* or “gist” ([Rensink, 2000](#)) is allegedly a high-level property that has been considered for a long time as the result of prior extraction and identification of perceptual objects. Recent studies have demonstrated, on the contrary, that scene gist appears to be extracted rapidly ([Biederman, 1981](#)), without attention ([Li et al., 2002](#); [Oliva and Schyns, 1997](#)) and possibly on the basis of the statistics of low-level sensory features. The idea defended by many authors consists in assuming that the visual system is able to rapidly determine a scene gist (which can provide important constraints on the kind of objects to be expected) by relying on simple measurements like the distribution of line orientations, colors or coarse blobs in the image ([Oliva and Schyns, 2000](#)). Although many such phenomena can be explained by saying that the detection of some specific cues is sufficient to retrieve perceptual scene schemes stored in memory ([Arbib, 1990](#); [Intraub, 1997](#)), it is an open empirical question to understand whether some kinds of gist might be robustly correlated with environmental regularities independently of a subject’s past experience and, hence, to investigate whether being able to shunt gist on the basis of basic sensory patterns might be a plausible

perceptual strategy.

6.5 SENSORIMOTOR CONSTRAINTS AND PERCEPTUAL SHUNTS

As I suggested at the end of Chapter 5, some of the alleged perceptual abilities enabled by sensorimotor learning exemplify the very same rationale described for shunt mechanisms. We can now reformulate the claim that sensorimotor knowledge can explain certain types of perceptual capabilities as follows:

Perceptual shunt hypothesis for sensorimotor learning

Whenever a task involving the extraction of a cognitively valuable property P is given to the perceiver, if a robust correlation exists in the environment between some invariant properties $I(S, M)$ of sensorimotor couplings and this property P , then an organism will be able to pick out and track property P by extracting the cooccurring sensorimotor patterns $I(S, M)$.

In other words, if in the perceiver's environment the presence of a certain property P is regularly coinstantiated with low-level properties of sensorimotor patterns, then the latter might serve as a proxy for extracting the former.

The attempt to explain, say, *color perception* in terms of the detection of specific sensorimotor patterns is perhaps the best illustration of the fact that such sensorimotor-based skills can be formulated as specific cases of perceptual shunt mechanisms, i.e. mechanisms valid in the context of stable perceptual environments in which robust correlations hold between low-level patterns and complex properties.

Again, we will not say that the perceiver is *inferring*, be it explicitly or implicitly, the presence of a complex property P from the detection of co-occurring sensorimotor patterns. As we will see later, sensorimotor capabilities,

much as perceptual shunt mechanisms, constitute *noninferential capabilities* that allow a perceiver to directly access complex, hidden or nonobservable properties on the basis of rich environmental regularities.

6.6 NEW CRITERIA FOR FEATURE BASICNESS

What is the relevance of the *Perceptual Shunt Hypothesis* for the analysis of possible constraints on perceptually relevant sensory properties? Acknowledging the existence of perceptual shunt mechanisms might considerably reorient the way in which vision science assesses the relevance of certain classes of features as input for perceptual processing.

According to the perceptual shunt hypothesis, a perceptual system “locked onto” some regularities of the environment in which it is embedded can easily use such regularities as a “proxy” to pick out and track some coinstantiated cognitively valuable properties. The structure of the environment and the internal setup of the perceiver must of course comply with some basic requirements in order for shunt mechanisms to work:

- (1) The environment (or environmental niche) in which the perceiver is embedded must be sufficiently stable and constrained to allow for a robust correlation between shunting properties and shunted properties.
- (2) The perceptual system must be equipped with sensory devices enabling the parsing of shunting properties.

If these two conditions are met, we can say that basic features can be characterized as *shunting properties*, i.e. those measurable properties of the sensory stimulation that maximize the cognitive gain and minimize the cost for the perceptual system in retrieving shunted properties. Assuming that there is a computational benefit in using low-level sensory patterns for tracking

the co-occurrence of cognitively valuable properties, one might argue that the most affordable basic features for environmentally embedded perceptual systems are low-level stimulus properties that can shunt cognitively valuable properties.

I will conclude this presentation of the perceptual shunt hypothesis in Chapter 7 by analyzing some of the major theoretical and methodological issues arising from the characterization of shunting strategies as psychologically plausible perceptual mechanisms.

Chapter 7

Issues in the study of adaptive constraints

THE AIM OF THIS LAST CHAPTER is to address a number of general theoretical issues related to the definition of adaptive constraints on the characterization of what counts as basic features. In particular, I intend to clarify what distinguishes (and what does not) the present proposal from some established research programmes in perceptual science (such as Perceptual Ecology) as well as from some major theoretical stances adopted in the study of perception (such as *Indirect Perception* or *Empiricism*). The first point that I make in this chapter is that the study of adaptive constraints on perception can shed light on a number of perceptual abilities (that are usually qualified as *inferential*) in strictly *non-inferential* terms: this has considerable consequences on the choice of the kind of properties that can be characterized as constituting a functional input for perceptual processing. In this sense, adaptive mechanisms like perceptual shunts can be considered as a paradigmatic case of strategies allowing direct detection of higher-order properties that are hardly considered as basic in mainstream perceptual science. A second point that I address is the relation between adaptive constraints and the idea of *delegation*, i.e. the individuation of cognitively affordable solutions to handle problems of overwhelming complexity. I suggest that the

study of adaptive constraints can help orient empirical research on perception towards computationally affordable mechanisms.

7.1 NON-INFERENTIALITY

A major tradition in perceptual science has regarded perception as a matter of *unconscious inference* from sensation. The origins of this hypothesis, which has enjoyed in the 20th century a considerable fortune, are usually traced back to the work of Hermann von Helmholtz. We have already characterized this view as the background assumption underlying the strong internalist view on what constitutes perceptual competence. The fundamental tenet of this hypothesis can be summarized by saying that perceiving can be considered as a matter of (implicitly) inferring correct information about the distal sources of the stimulation from incomplete sensory premises on the one hand and available knowledge and representations on the other hand. The main rationale behind this assumption is that proximal sensory stimulation provides to the perceiver impoverished information that is not sufficient *per se* to correctly represent the distal cause of the stimulus or to disambiguate the possible distal sources that produce a specific sensory pattern. It is for this reason that perception can be considered as the process of retrieving a reliable picture of the world from ambiguous or uninformative sensory stimuli: it is assumed to be *inferential* insofar as it takes the form of an (unconscious) process of derivation of a (perceptual) conclusion from a set of premises (that typically include sensation plus memory and background knowledge); it is *indirect* insofar as the derivation of certain perceptual conclusions is *mediated* by alleged inferential steps that allow the transition from the representation

of the premises to the representation of the conclusion.¹ The idea of an impoverished stimulus and the necessity of inference is perfectly illustrated in this passage by Neisser (1967):

These patterns of light at the retina are [...] one-sided in their perspective, shifting radically several times each second, unique and novel at every moment. [They] bear little resemblance to either the real object that gave rise to them or to the object of experience that the perceiver will construct. [...] Visual cognition, then, deals with the process by which a perceived, remembered, and thought-about world is brought into being from as unpromising a beginning as the retinal patterns. (pp.7-8, cit. in Michaels and Carello (1981)).

Fodor and Pylyshyn (1981) probably offered the most theoretically elaborate synthesis of the indirect perception paradigm. The critical assessment they offer of Gibson's theory (Gibson, 1966, 1979) provides some points that are crucial to the present discussion and deserves some specific attention. Among the many arguments raised against the idea that perceptually relevant properties of the layout are directly picked up, one is particularly important: the idea according to which directly detectible properties are only those that can be *transduced*. The notion of transduction is introduced through the related notion of *specification*. When there is a nomological correlation between two states of affairs an organism can use the occurrence of one to find about the other: saying that S1 specifies S2 is tantamount to saying that perceiving S2 causally depends on detecting S1. Given the structure of our sensory organs, in the case of vision we can say that the detection of specific

¹Modern formulations of the *indirect perception* paradigm can be found in Rock (1977, 1983, 1997b).

patterns of light on the retina (patterns for which we have transducers) can be used to specify the structure of a distal source. S1 specifies S2 only if the organism has transducers (or detectors) for S1. Now, Fodor and Pylyshyn argue, functioning as a detector (or a transducer) for a property S1 means being illusion-free with respect to S1, since transduction is by definition direct, i.e., not dependent on specification. Perceiving S2, on the contrary, depends on detecting S1, but since S2 cannot be directly detected it must be inferred from S1. In a nutshell, the idea put forward by Fodor and Pylyshyn is that there is only one possible way to get from detected properties of the light to perceived properties of the source: through *inference*, i.e. “by inferring the latter from the former on the basis of (usually implicit) knowledge of the correlations that connect them (p.165).

A fundamental point in Fodor and Pylyshyn’s criticism is that in the case of vision we lack detectors for properties other than patterns of light on the retina. Detectible properties are typically properties that can be described by laws, i.e. nomological (counterfactual-supporting) generalizations. So typically, *having a specific wavelength, intensity or chemical composition* are examples of detectible properties, whereas *being expensive, edible or poisonous* are not. The latter cannot be detected given the structure of our transducers, they can just be specified on the basis of detectible properties. The question whether “we could have detectors for Da Vinci’s paintings” is according to Fodor and Pylyshyn trivially false, since it is possible to fake a Da Vinci that would produce a retinal pattern that cannot be distinguished from the one produced by a real Da Vinci. There is no *law* regarding the property BEING A DA VINCI PAINTING, hence there can be no Da Vinci detectors.

7.1.1 PERCEPTUAL SHUNT: DETECTION OR DIRECT PICKUP?

Let us now see to which extent the position expressed by Fodor and Pylyshyn applies to the perceptual strategies that I described as perceptual shunt mechanisms. The constraints on the kind of properties that can be directly detected according to Fodor and Pylyshyn offer an interesting solution to show that, under certain conditions, it is perfectly legitimate to assume that there *can be* detection of properties that in the general case are not the object of *physical laws* and that, accordingly, would not be directly perceivable. The possibility is acknowledged by Fodor and Pylyshyn themselves in the following passage:

The moral is: the decision about what detectors there are is linked to the decision about what laws there are; A world in which there were laws about the property *shoe* would be a world in which there could be detectors for shoes. After all, a law about *shoe* would, presumably, connect the shoe property to other sorts of properties, and then things which have properties of these other sorts would *ipso facto* be available for service as shoe detectors (p.164).

The above example fits precisely the case of perceptual shunts in which, we assumed, a property *P* that is systematically coinstantiated with another property *Q* in a given environmental context can be considered as a sort of *perceptual-proxy* for the latter, since whenever *Q* is instantiated, *P* is also instantiated. Given this situation, we would be entitled to say that – within the considered ecological niche – there *are* robust correlations having *Q* as object, which make *Q* a directly detectable property. Obviously, though, such correlations, albeit reliable within a specific environmental context,

are *accidental* (i.e. they cannot be qualified as counterfactual-supporting generalization).

I assume that given such premises, the perceptual shunt hypothesis can provide an interesting alternative solution to the Fodor-Gibson controversy about the nature of direct perception. The perceptual shunt hypothesis is compatible, on the one hand, with Fodor's definition of detectible properties insofar as it assumes that the shunting property *Q* of sensory patterns is a full-fledged projectible property. On the other hand, it is compatible with Gibson's view in that it assumes that property *P* is directly picked up *without the need of inferential mechanisms*. The strategy that the shunt hypothesis adopts to avoid the Fodorian dilemma ("either a property is detected or is inferred") is to assume that the relation between *P* and *Q* can be legitimately qualified as *not* inferential, since - I argue - there is no need to assume that the organism is representing *Q* and deriving a representation of *P* on the basis of a representation *Q* as a premise.

My claim is that – within the ecological niche in which an organism is embedded and in which *P* is systematically coinstantiated with *Q* – saying that the organism is representing *P* or representing *Q* is merely a matter of *redescription*. The organism is not (implicitly or explicitly) calculating or representing *the coinstantiation of P and Q* in its niche: in virtue of its being able to detect *Q* it is also "locked onto" *P* since within the niche the two properties are extensionally inseparable.²

²In this respect one could not say that the organism is using *P* as a *cue* for *Q*, since cues are typically premisses for inferences, while in this case I assume that no inferential mechanism is in place.

7.1.2 PERCEPTUAL SHUNT: WHAT IS REPRESENTED?

I will clarify the above claim that perceptual shunt mechanisms can escape the Fodorian dilemma between inferentiality and detectability by making reference to an example proposed by [Dretske \(1986\)](#) and discussed by [Millikan \(1993a\)](#). In the context of an analysis of the problem of misrepresentation, Dretske introduces the example of a particular kind of Northern Hemisphere bacteria which orient themselves towards benign anaerobic environments by using their *magnetosome*, an inner magnetic organ which pulls towards the magnetic north pole (and hence pulls down). In the Northern Hemisphere the direction of pull of the magnetosome is contingently correlated with the direction of anaerobic environments, hence within this context the direction of pull of the magnetosome can be used as a proxy for the location of benign environments. The kind of information delivered by the magnetosome can actually be seen as a typical case of shunt mechanism whose validity is restricted to a specific environmental context. It is at this point that Dretske's analysis diverges from Millikan's. Dretske observes that since there is a purely contingent relation between location of oxygen-free environments and direction of the magnetic pole, and not a causal one, one cannot say that the magnetosome delivers reliable information on oxygen-free water. Millikan, on the other hand, suggests that magnetosome has this function precisely because of the fact that it was *selectively designed in this particular context* for that function. Proper function, she argues, is the kind of function that has been selected during the adaptive history of the organism, hence the proper function of magnetosome is certainly that of signaling oxygen-free water because this is how such information is *used* by the organism. She also argues that what the magnetosome represents is actually what the *consumer devices* require that it correspond to in order to perform the task. Hence, she concludes,

what the magnetosome represents is distal not proximal and the bacterium does not perform an inference from the value of the proximal stimulus (the magnetic field) to the existence of the represented. Applying this analysis to the formulation of a perceptual shunt hypothesis, it is certainly tempting to endorse Millikan's view: I endorse in any case the claim that there is no inferential mechanism. On the other hand, I am reluctant to embrace the strong conclusions put forward by Millikan on what is the appropriate representational content that can be described for these mechanisms. As I mentioned earlier, the fact that a shunt mechanism can be considered as delivering information about P rather than Q in the same environment is just a matter of description. On this specific issue, then, I rejoin [Dretske \(1986\)](#)'s position where he observes that there is in this case an *indeterminacy* of function regarding the description of the proper representational content delivered by such mechanisms.³

7.2 ECOLOGICAL INTELLIGENCE

A point upon which I have repeatedly insisted is that if the hypothesis I defend in this work holds, environment structure plays a major role through adaptive constraints in the selection of basic features and in the shaping of perceptual skills. The ability to use robust environmental regularities for the solution of cognitively demanding problems has been deemed by some as a form as *ecological intelligence* or *ecological rationality* ([Bullock and Todd, 1999](#); [Todd, 1999b](#)). The idea evokes some metaphors that have become common in the situated cognition literature (like that of “the world as an outside memory” - [O'Regan \(1992\)](#)). Still, many consequences of this externalist turn in the

³A short review of the debate on functional indeterminacy in the context of the literature on biological functions and teleosemantics can be found in [Neander \(2004\)](#).

study of perceptual capabilities are far from being adequately appreciated. In this section, I focus on some prospects in the study of environmentally-tailored capabilities that deserve some further analysis.

7.2.1 RECRUITMENT VS. DELEGATION

The idea of adaptive constraints on feature selection might seem to suggest that I am endorsing a strongly empiricist stance towards the kind of capabilities perceptual systems can acquire. In particular, the idea according to which environmentally salient patterns would be internalized by a organism through systematic exposure and statistical learning seems to be *prima facie* incompatible with the established view in developmental psychology according to which the human cognitive system is endowed since the very beginning with a large set of prespecified capabilities that do not depend, for their expression, on specific exposure to particularly structured environments (Spelke, 1994). The hypothesis presented in this work, regarding the existence of adaptive constraint on the selection of perceptual features is actually *neutral* with respect to the traditional issues of debate between nativist *vs.* empiricist positions. In particular it is neutral with respect to the question of *when* and *where* adaptive constraints apply to perceptual systems to select certain patterns are good properties on the basis of their match with environmental regularities. An interesting aspect of comparing innate capabilities with skills acquired through learning is the fact that adaptive constraints assumedly play in these two cases a similar, but symmetric role. We can characterize this symmetric roles by introducing two different kinds of strategies adaptive constraints can enable. Whereas adaptive constraints in the case of innate mechanisms take the form of **recruitment** strategies, in the case of learning they will take the form of **delegation** strategies. Let us try to unfold these

two ideas in more details.

Recruitment strategies are well known in the cognitive literature, especially in evolutionary psychology, where cases of high-level capabilities are derived from the exploitation of existing resources of the organism. Cases like that of graphemic parsing or face recognition are typical examples of mechanisms in which some precabled perceptual capabilities are recruited to respond to particular demanding exigences of complex perceptual tasks. Mechanisms that are selected phylogenetically to fit a specific domain are hence remapped to apply to a variety of stimuli that do not necessarily belong to the original domain. Sperber (1994), in a discussion about the notion of domain specificity, captures this distinction by introducing the idea of *actual* vs. *proper domain* of a module. There is an interesting discrepancy between the set of patterns to which a specific capacity was originally exposed to and for which it was presumably selected (the proper domain) and the actual class of patterns that are compatible with this capacity (the actual domain). This discrepancy – Sperber observes – is what allows high-level, cultural and evolved skills to parasite mechanisms that are pre-specified phylogenetically. To illustrate this case, suppose that different strategies are available to an organism to meet the demands of a complex perceptual problem and that one of such strategies exploits a relatively dumb and automatic mechanism based on arbitrary but statistically robust environmental correlations to solve the problem: we will say that if adopting this strategy results in an enhancement of the organism's benefits in terms of cognitive affordability, the organism is likely to *recruit* it over other, more expensive strategies.

Conversely, a symmetric solution can be applied in cases of perceptual learning. Suppose an organism has developed a number of strategies to deal with a specific class of perceptual problems and that it has learned, by exposure to

specific environmental regularities, that there is one of these strategies that simply requires detecting some simple property to solve the problem. We will say that if adopting this strategy yields an advantage to the organism in terms of cognitive affordability, the organism is likely to *delegate* the solution of the problem to this strategy. Perceptual learning typically consists in individuating the less expensive routine to which delegate the solution of specific problems: perceptual learning is in many cases a synonym of adopting strategies that increase the selection of highly diagnostic features (Biederman and Shiffrar, 1987; Schyns and Oliva, 1997) or reduce the dimensionality of the problem (Edelman and Intrator, 1997).

7.2.2 THE VIRTUES OF DUMB MECHANISMS

There are some interesting issues related to statistical learning of low-level properties that bear a high informational value in a specific environment. Several studies have been conducted to investigate the statistical learning skills of infants exposed to environments with a number of salient regularities. Such regularities have proved essential in bootstrapping different aspects of language acquisition (Saffran et al., 1996a,b; Smith et al., 1996) as well as visual capabilities (Fiser and Aslin, 2002a,b). In a nutshell the idea is the following: infants are sensitive to some distributional patterns of low-level properties (like prosodic features) that are co-instantiated in an accidental but robust way with complex properties (like syntactic or morphological features). In virtue of the learning of these low-level properties, they manage to bootstrap the acquisition of the correlated complex properties in a way that would not otherwise be manageable. Distributional learning of low-level patterns actually warps the kind of high-level properties that infants are able to learn. The “less is more hypothesis” suggested by E. Newport and

collaborators assumes that this nice ‘fit’ between the basic morphological structures and children’s limited early capacities to perceive complex stimuli is not a lucky coincidence. The fit is no accident. But neither is it the case that the child’s early capacities were selected so as to facilitate language learning. Instead, the structure of the language was selected so as to exploit those early (and independent) limitations and not vice versa. That many smart capabilities were selected thanks to the dumbness and rigidity of the initial resources upon which they draw is a thesis defended by many. [Sperber \(2006\)](#) argues that the limitations and context-insensitivity of modular systems is what allows the simple solution of problems of overwhelming complexity: it is plausible to assume that strategies that resulted in efficient performance through the mobilization of relatively automatic and dumb mechanisms were selected as winning strategies over evolution. Whereas automatism, domain specificity and lack of flexibility were traditionally seen as marks of scarce adaptability, these considerations suggest on the contrary that they can yield a more appropriate strategy to the solution of specifically demanding problems (see [Clark and Dukas, 2003](#)).

7.2.3 PERCEPTION AS A HEURISTICAL PROCESS

An argument that can be used to trivialize the claim that perceptual shunt mechanisms can be taken as a realistic alternative to standard perceptual processing mechanisms might run as follows: contextual variability is far too large to allow the kind of exploitation of distributional regularities that shunt mechanisms draw upon. Contingent but reliable correlations like those required for perceptual shunt mechanisms to work are too rare in real-world conditions to be plausibly instantiated by perceptual mechanisms: the actual contexts - this argument may conclude- in which the validity of perceptual

shunts is guaranteed are too narrow to be of any psychological interest. I assume that this argument is valid only on the condition that we concede to our opponent that the goal of perceptual processing is to deliver information with the same degree of reliability, in any condition and for any perceptual task. The argument does not hold, instead, if we grant that a large number of perceptual routines do not actually require accurate representing any aspect of the visual scene, or depend on strictly error-free disambiguation of visual configurations (see [Rensink, 2000](#)).

More precisely, I argue that such an argument does not threaten the validity of the hypothesis according to which perceptual shunt mechanisms might be psychologically plausible candidates for perceptual processing. The psychological plausibility of shunt mechanisms should be assessed, I argue, against the extent to which they can provide *default* solutions to common perceptual problems that do not require coping with a high degree of variability or for which reliability must be weighted against computational affordability. I assume in this sense that perceptual shunt mechanisms are perceptual strategies that can coexist with other, more reliable forms of perceptual inference. The interest in studying mechanisms that are only reliable under given contextual conditions is that in such contexts specific perceptual routines can be delegated to less computationally expensive mechanisms (that we could qualify as *heuristics*) that are able to provide reliable solutions for local task demands. Typically heuristic strategies do not replace more reflexive capabilities, but are taken as default perceptual strategies whose output can be corrected or revised if required (see [Todd, 1999a](#)). For instance, if a heuristic is unreliable because it is applied in the wrong context, it might still be cognitively advantageous to the extent that it is applicable in other contexts. The interesting issue, then, (an issue, though, that is difficult to frame in explicit empirical terms or

measurable conditions) is to understand whether the massive use of devices that do not deliver information with a degree of reliability equal to $p = 1$ but with a certain degree of uncertainty, must necessarily result in a cognitive disadvantage for the organism. I assume that biological systems are systems that are constantly seeking to reduce the computational load of cognitive processing, by either optimizing single strategies to adopt the most affordable mechanisms or by using multiple strategies to arrive to the goal in the most rapid and inexpensive way. Mechanisms meeting adaptive constraints cannot be considered as optimal solutions for *any* kind of perceptual routine, but they represent a computationally plausible solution for different cases of less epistemically demanding perceptual tasks.

7.3 CONCLUSIONS

The above discussion has tried to tackle some of the main problems related to the idea that adaptive constraints might provide a realistic alternative to the characterization of what counts as the relevant input for perceptual processing. The goal of this analysis, in particular, was to frame the alleged *scope* of mechanisms tailored to environmental regularities and their *relations* to other, more traditional kinds of perceptual mechanisms that rely on inferential resources. I have argued that the main properties of adaptively-constrained perceptual mechanisms are the following:

- they do not need to rely on inferential capabilities of the organism;
- they can be accommodated with traditional requirements on detection mechanisms;
- they can be selected as effective solutions both through phylogenetic adaptation and ontogenetic development;

- they can provide computationally affordable solutions for specific kinds of perceptual routines with low requirements of epistemic reliability *without* replacing more reliable kinds of processing that draw on inferential skills;

In the general conclusions of the present work I will provide an assessment of the extent to which the proposal that I have tried to articulate in Part II can provide some new theoretical insights and substantial empirical research prescriptions on the characterization of basic visual properties.

Conclusions

WHEN THE IDEAS discussed in this work were first presented in a preliminary version at the 26th *European Conference on Visual Perception* in Paris⁴, they were received with a mix of curiosity and ill-disguised skepticism. The main reason why the topic raised an embarrassed reaction in the audience could be summarized in a single question: *why should we ever care about what basic visual features are and whether there are any?* Several replies that were formulated actually tried to articulate this skepticism along two different lines:

1. Deciding what is a basic feature is a *merely terminological issue*; the definition of what we call “feature” is instrumental to the descriptive needs of each single investigation, it is hence the result of a terminological stipulation that cannot be given *per se* any further theoretical value.
2. Deciding what is a basic feature is a *merely empirical matter*; the relevant variables for a specific kind of perceptual phenomena are to be established by looking at data, only then we will be able to see what properties and patterns are relevant for the explanation of specific kinds of perceptual performance.

Both replies – I maintain – are instances of a common strategy to deny the theoretical relevance of certain kinds of conceptual analysis. Historically, the

⁴See [Taraborelli \(2003\)](#).

same kind of skepticism was raised against attempts to consider that the notion of *module*⁵, *representation*⁶, *inference*⁷ were genuine theoretical notions relevant for empirical research and deserving a careful conceptual treatment. Today, though, the lively debate found in the literature about each of these notions witnesses that their conceptual framing is largely recognized as a fundamental step in grounding empirical research. It is then not completely surprising that raising the question of understanding whether the notion of a basic visual feature has theoretical relevance can be criticized as a non-issue. The main motivation behind this work, indeed, and its very starting point is the realization that the notion of a basic feature *is* already used, albeit implicitly, as a pre-theoretical term in current experimental work. I have tried to show through the present analysis that there are two main motivations to defend the fact that the notion of basic feature already has the *de facto* status of a quasi-theoretical notion:

- A. It is used in order to provide cross-level explanations of perceptual phenomena at different levels of description.
- B. It is used to frame the kind of properties that are relevant for perceptual processing.

In the present work, I have adopted these two motivations as the rationale orienting my analysis of the concept of basic feature. Let us try to sum up some of the conclusions.

⁵See [Hirschfeld and Gelman \(1994\)](#).

⁶See [Tarr and Black \(1994b\)](#) and the ensuing debate in the special issue.

⁷See [Fodor and Pylyshyn \(1981\)](#); [Rock \(1997a\)](#); [Ullman \(1980\)](#).

Basic features are required for cross-level explanations

The need for a robust characterization of the notion of basic visual feature is actually implicit in many of the methodological attempts to establish criteria to give cross-level explanations of perceptual capabilities. I have mentioned *Linking propositions*⁸, the *Matching-Content Doctrine*⁹, the *Psychophysical Linking Hypothesis*¹⁰ as some of the major explicit formulations of principles that are adopted to establish explanatory links between phenomena studied at the level of neural processing and properties of perceptual performance. Principles of this kinds are exemplified by cases in which one assumes that:

if you are currently paying attention to a friend discussing some point with you, neurons in area MT respond to the motion of his face, neurons in V4 respond to its hue, and neurons in auditory cortex ... respond to the words coming from his face – [Crick and Koch \(1990\)](#).

and consequently assumes that it is the firing of a certain neural population coding for feature *P* what causally explains the perceptual experience of *P*. Cross-level explanatory principles involving reference to basic features are not restricted to the neurophysiology/psychology interface but are spread at any level of explanation of perceptual phenomena. A cross-level link is, for instance, established between performance and phenomenology when it is assumed that the conscious experience of a visual object is explained by an integration of its features performed through selective attention.¹¹

I have reviewed the case study of the *Feature Binding Problem* (Chapter 3) as one of the most interesting testbeds for evaluating the problems raised

⁸[Teller \(1984\)](#); [Teller and Pugh \(1983\)](#).

⁹[Noë \(2002\)](#).

¹⁰[Barlow \(1985a\)](#).

¹¹[Treisman and Schmidt \(1982\)](#).

by cross-level use of the notion of basic features. I have argued that the rationale for a FBP is flawed by the use of a notion of basic feature that takes a correlation of allegedly feature-P-related phenomena at different levels of description as an explanation of how processing of P at one level is causally explained by processing of P at the subordinate level.

The standard view against which I have argued is the one that assumes *homonymy* of features described at different levels as a straightforward cross-level explanation of their relation. The goal of the analysis of the notion of basic feature conducted in Part I has been precisely to show that there are a number of distinct and conceptually irreducible criteria behind the notion of a basic feature. The result of this analysis has shown that

- different feature basicness criteria are often conflated in the literature;
- homonymy of features at different levels of analysis is implicitly taken as evidence to assume that there are causal explanatory links between these levels;

The conclusion to which this whole section points is that if cross-level explanations are to be formulated, then a general requirement must be met: alleged correlations between different criteria that describe the “same” feature (say COLOR₁ as a phenomenal feature and COLOR₂ as a property encoded in feature maps of the visual cortex), should not be *assumed* but *explained*. Only then a principled use of the notion of basic feature will be possible in cross-level explanation of perceptual phenomena.

Basic features are required to characterize relevant perceptual variables

The second motivation of the present analysis has been the very realization that deciding what properties of the visual stimulation constitute the actual

input for perceptual processing makes crucial differences in the explanation of perceptual abilities.

As the important debate between defendants of *direct perception* and *indirect perception* witnesses (see Chapter 7), the question of deciding what counts as the relevant entry-level for perception has already been acknowledged as a genuine theoretical question. Adopting different assumptions – not only on the alleged *output* of perceptual systems, but also on their proper *input* – has immediate consequences on the understanding of the kind of processing they perform.

The goal of the analysis conducted in part II of this work has been to challenge the view according to which what counts as a proper input for perceptual processing should be decided by merely looking at internal constraints on the structure and functioning of perceptual systems. I have shown that there are reasons to assume that *if* certain environmental regularities play a role in the definition of what might count as a relevant input property for perceptual processing, *then* we should expect that a number of new perceptual mechanisms based on such properties should be found. Since the match between environmental regularities and internal properties of perceptual systems (what I called *adaptive constraints*) has not deserved so far extensive attention in the literature, I have concluded that it is likely that a number of unexplored perceptual capabilities might have been systematically neglected by empirical research.

In this respect, I have criticized in particular the standard view according to which what the senses encode (and what, accordingly, perceptual processing applies to) must be restricted to a number of properties that do not bear reliable information about the distal sources of the stimulation, and hence that the goal of perceptual processing is to correctly determine what these sources

are by using these features of the sensory stimulation as premises for perceptual inferences. By analyzing the case of potential perceptual mechanisms tailored to the adaptive needs of environmentally bounded organisms and not drawing on inferential capabilities, I have shown that a very different set of criteria for what counts as an input for perceptual processing can be defined. Providing an alternative notion of feature basicness based on plausible adaptive assumptions can hence help reorient the way in which the architecture of perception is usually studied and the way in which the functional role of sensory features is assessed. Cases like those in which it is considered “highly surprising” that the visual system of a bee might be sensitive to topological properties of the stimuli ¹² show not only that explicit criteria to define what basic features are are needed, but also that studying environmental constraints is likely to provide some fundamental insights on the definition of such criteria.

The hope is that this exploratory analysis of the characterization of feature basicness, together with the effort to disentangle the different issues that this concept raises, might draw attention in the perceptual science community on the fact that the implicit use of some core notions without a rigorous theoretical grounding is a risk, since it can lead to endorse theoretically weak explanatory principles and exclude certain phenomena from the descriptive domain of empirical research.

¹²See [Chen et al. \(2003\)](#); [Pomerantz \(2003\)](#).

Bibliography

- E. Adrian. *The Basis of Sensation*. Christophers, London, 1928. 73
- M. Arbib. Schemas for high-level vision: The problem of instantiation. In E. Schwartz, editor, *Computational neuroscience*, pages 340–351. MIT Press, Cambridge, MA, 1990. 179
- F. Attneave. Informational aspects of visual perception. *Psychological Review*, 61:183–193, 1954. 52
- F. Attneave. Apparent motion and the what-where connection. *Psychologia*, 17:108–120, 1974. 96, 118
- H. B. Barlow. Single units and sensation : a neuron doctrine for perceptual psychology. *Perception*, 1(371-394), 1972. 72
- H. B. Barlow. The twelfth bartlett memorial lecture: The role of single neurons in the psychology of perception. *Quarterly Journal of Experimental Psychology*, 37A:121–145, 1985a. 73, 76, 201
- H. B. Barlow. Cerebral cortex as model builder. In [Rose and Dobson \(1985\)](#), chapter 4. 149
- I. Biederman. On the semantics of a glance at a scene. In M. Kubovy and J. Pomerantz, editors, *Perceptual Organization*, pages 213–253. Erlbaum, Hillsdale, NJ, 1981. 179

- I. Biederman. Recognition-by-components: A theory of human image interpretation. *Psychological Review*, 94:115–148, 1987. 32, 33
- I. Biederman. Higher-level vision. In [Osherson et al. \(1990\)](#). 9, 34
- I. Biederman and M. Shiffrar. Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology : Learning, Memory and Cognition*, 13:640–645, 1987. 193
- A. Bompas, J. Clark, and J. K. O'Regan. Color perception in the sensorimotor contingencies theory. In *European Conference on Visual Perception*, Scotland, 2002. 157
- K. Briand and R. Klein. Has feature integration theory come unglued? a reply to tsal. *Journal of Experimental Psychology : Human Perception and Performance*, 15:401–406, 1989. 92, 121
- J. Broackes. The autonomy of colour. In K. Lennon and D. Charles, editors, *Reduction, explanation, and realism*. Oxford University Press, 1992. 156
- S. Bullock and P. M. Todd. Made to measure: Ecological rationality in structured environments. *Mind and Machines*, 1999. 147, 190
- D. Burr. Modular analysis - or not? *Current Biology*, 9:R90–R92, 1999. 79
- J. Campbell. *Reference and Consciousness*. Oxford University Press, Oxford, 2002. 127
- J. Campbell. Sense, reference and selective attention. *Proceedings of the Aristotelian Society*, Suppl. 71:55–74, 1997. 67, 127

- S. Carey and F. Xu. Infant's knowledge of objects: beyond object files and object tracking? *Cognition*, 80:179–213, 2001. 110
- R. Casati. Representational advantages. *Proceedings of the Aristotelian Society*, pages 281–298, 2003. 43
- R. Casati. Common sense, philosophical and theoretical notions of an object: some methodological problems. In *Barcelona ESPP-SPP Joint Conference*, 2004. 100, 110
- P. Cavanagh. The language of vision (paper presented at the 26th european conference on visual perception). *Perception*, 32 (ECVP Supplement), 2003. 102
- N. Chater. Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review*, 103:566–581, 1996. 52
- L. Chen. Perceptual organization: To reverse back the inverted (upside-down) question of feature binding. *Visual Cognition*, 8:287–303, 2001. 61
- L. Chen, S. Zhang, and M. V. Srinivasan. Global perception in small brains: Topological pattern recognition in honey bees. *PNAS*, 100(11):6884–6889, 2003. 49, 204
- N. Chomsky. *New Horizons in the Study of Language and Mind*. Cambridge University Press, Cambridge, 2000. 144
- A. Clark. *A Theory of Sentience*, chapter 2 - Qualities and Their Places. Oxford University Press, Oxford, 2000. 88, 97, 127
- A. Clark. Some logical features of feature integration. In W. Backhaus, editor, *Neuronal Coding of Perceptual Systems*, volume 9 of *Series on Biophysics and Biocybernetics*. World Scientific, New Jersey, 2001. 89, 101, 120, 127

- A. Clark. Feature-placing and proto-objects. URL <http://www.nyu.edu/gsas/dept/philo/courses/representation/papers/Clark.pdf>. 2004. 44, 98, 129
- C. W. Clark and R. Dukas. The behavioral ecology of a cognitive constraint: limited attention. *Behavioral Ecology*, 14(2):151–156, 2003. 194
- F. Crick. The machinery of focal attention. *Perception*, 13(1):A2, 1984. 118
- F. Crick and C. Koch. Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, 2:263–275, 1990. 201
- E. De Yoe and D. C. Van Essen. Concurrent processing streams in monkey visual cortex. *Concurrent processing streams in monkey visual cortex*, 11: 219–226, 1988. 125
- M. Donk. Illusory conjunctions are an illusion : The effects of target-nontarget similarity on conjunction and feature errors. *Journal of Experimental Psychology : Human Perception and Performance*, 25:1207–1233, 1999. 121
- R. Douglas and K. Martin. Opening the grey box. *Trends in Neuroscience*, 14:286–293, 1991. 81
- F. Dretske. Misrepresentation. In R. Bogdan, editor, *Belief: Form, content and function*, pages 17–26. Oxford University Press, Oxford, 1986. 189, 190
- G. Durieux and S. Gillis. Predicting grammatical classes from phonological cues: An empirical test. In J. W. B. Höhle, editor, *Approaches to bootstrapping: phonological, syntactic and neurophysiological aspects of early language acquisition*, pages 189–232. Benjamins, Amsterdam, 2000. 145
- S. Edelman. Representation without reconstruction. *CVGIP: Image Understanding*, 60(1), 1994. 27

- S. Edelman and N. Intrator. Learning as extraction of low-dimensional representations. In P. S. D. Medin, R. Goldstone, editor, *Mechanisms of Perceptual Learning*. Academic Press, 1997. URL <http://citeseer.ist.psu.edu/article/edelman96learning.html>. 193
- J. T. Enns. Three dimensional features that pop out in visual search. In D. Brogan, editor, *Visual Search*. Taylor & Francis, London, 1990. 65
- J. T. Enns and R. A. Rensink. Influence of scene-based properties on visual search. *Science*, 247:721–723, 1990. URL <http://www.cs.ubc.ca/~rensink/publications/abs.90.1.html>. 61, 66
- J. Feldman. Does vision work? towards a semantic of perception. In E. Lepore and Z. W. Pylyshyn, editors, *What is Cognitive Science*, pages 208–229. Blackwell, Oxford, 1999. 174
- D. Felleman and D. C. Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex*, 1:1–148, 1991. 79
- J. Fiser and R. N. Aslin. Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28(3):455–467, 2002a. 193
- J. Fiser and R. N. Aslin. Statistical learning of new visual feature combinations by infants. *PNAS*, 99(24):15822–15826, 2002b. 193
- J. A. Fodor. Vicissitudes of the object concept, 2001. URL <http://rucss.rutgers.edu/pub/papers/>. 44
- J. A. Fodor and Z. W. Pylyshyn. How direct is visual perception? Some reflections on Gibson’s ”Ecological Approach”. *Cognition*, 9:139–196, 1981. 18, 19, 138, 185, 186, 200

- J. Garson. (dis)solving the binding problem. *Philosophical Psychology*, 14(4), 2001. 97, 127
- J. J. Gibson. *The Senses Considered as Perceptual Systems*. Boston, Houghton Mifflin, 1966. 18, 185
- J. J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, 1979. 18, 138, 160, 185
- G. Gigerenzer and R. Selten, editors. *Bounded Rationality*. MIT Press, Cambridge, MA, 1999. 147, 221
- G. Gigerenzer and P. Todd. *Simple heuristics that make us smart*. New York, NY: Oxford University Press, 1999. 147
- M. Green. Visual search, visual streams, and visual architectures. *Perception and Psychophysics*, 50:388–403, 1991. 57, 92, 99, 118
- C. Gross, C. Rocha-Miranda, and D. Bender. Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology*, 95: 96–1111, 1972. 9, 76
- F. Heider and M. Simmel. An experimental study of apparent behavior. *American Journal of Psychology*, 57:243–249, 1944. 177
- L. Hirschfeld and S. Gelman, editors. *Mapping the Mind: Domain Specificity in Cognition and Culture*. Cambridge University Press, Cambridge, 1994. 200, 219
- S. Hochstein and M. Ahissar. The ups and downs of conscious visual perception. *Journal of Vision*, 1(3):350, 2001. 9, 66

- S. Hochstein and M. Ahissar. View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5):791–804, 2002. 67
- D. D. Hoffman. *Visual Intelligence: How We Create What We See*. Norton & Company Inc., New York, NY, 1998. 174
- D. Hubel and T. Wiesel. Receptive fields of single neurons in the cat’s striate cortex. *Journal of Physiology*, 149:574–591, 1959. 74
- M. Hubener, D. Shoham, A. Grinvald, and T. Bonhoeffer. Spatial relationships among three columnar systems in cat area 17. *Journal of Neuroscience*, 17(23):9270–9284, 1997. 9, 77
- S. L. Hurley. *Consciousness in action*. Harvard University Press, London, 1998. 129
- H. Intraub. The representation of visual scenes. *Trends in Cognitive Sciences*, 1:217–222, 1997. 179
- F. Jackson. *Perception: A Representative Theory*. Cambridge University Press, Cambridge, 1977. 94, 121
- A. Jepson and W. Richards. What makes a good feature? In L. Harris and M. Jenkin, editors, *Spatial Vision in Humans and Robots*. Cambridge University Press, Cambridge, 1992. 149, 150, 173
- M. Johnson and J. Morton. *Biology and cognitive development: The case of face recognition*. MIT Press, Cambridge, MA, 1991. 164
- B. Julesz. A theory of preattentive texture discrimination based on first-order statistics of textons. *Biol Cybern*, 41(2):131–138, 1981. 57

- D. Kahneman, A. Treisman, and B. Gibbs. Moving objects and spatial attention. In *24th Annual Meeting of the Psychonomic Society*. San Diego, California, 1983. 103
- S. Keele, A. Cohen, R. Ivry, M. Liotti, and P. Yee. Tests of a temporal theory of attentional binding. *Journal of Experimental Psychology: Human Perception and Performance*, 14:444–452, 1988. 119
- F. Keil. Nurturing nativism. a review of cowie’s what’s within. In *A Field Guide to the Philosophy of Mind*. 2000. URL <http://www.uniroma3.it/kant/field/cowiesymp.htm>. 163
- D. Knill and W. Richards. *Perception as Bayesian Inference*. Cambridge University Press, New York, NY, 1996. 149, 173
- J. J. Koenderink. What is a ‘feature’. *Journal of Intelligent Systems*, 3(1), 1993. 32, 34, 78, 98
- K. Koffka. *Principles of Gestalt Psychology*. New York : Harcourt Brace, 1935. 50, 51
- P. M. Kuhl. A new view of language acquisition. *PNAS*, 97(22):11850–11857, 2000. 145
- P. Lennie. Single units and visual cortical organization. *Perception*, 27: 889–935, 1998. 76
- J. Y. Lettvin, H. R. Maturana, W. S. McCulloch, and W. H. Pitts. What the frog’s eye tells the frog’s brain. *Proc. IRE*, 47:1940–1951, 1959. 71, 74, 75, 166, 167
- F. F. Li, R. VanRullen, C. Koch, and P. Perona. Rapid natural scene categorization in the near absence of attention. *PNAS*, 99(14):9596–9601,

2002. URL <http://www.pnas.org/cgi/content/abstract/99/14/9596>. 67, 179
- A. Liberman and I. Mattingly. The motor theory of speech perception revised. *Cognition*, 21:1–36, 1985. 161
- A. M. Liberman, F. S. Cooper, K. S. Harris, and P. J. MacNeilage. A motor theory of speech perception. *Proceedings of the Symposium on Speech Communication Seminar*, 1963. 161
- C. Lueck, S. Zeki, K. Friston, M. Deiber, P. Cope, V. Cunningham, A. Lamertsmas, C. Kennard, and R. Frackowiak. The colour centre in the cerebral cortex of man. *Nature*, 340:386–389, 1989. 79
- D. M. MacKay. The significance of feature sensitivity. In [Rose and Dobson \(1985\)](#). 81, 82, 161
- D. Marr. *Vision: A computational investigation into the human representation and processing of visual information*. W.H. Freeman, San Francisco, 1982. 32, 70, 89
- K. A. Martin. The pope and grandmother. a frog’s-eye view of theory. *Nature Neuroscience*, 3(november 2000):1169, 2000. 71, 73
- K. A. Martin. A brief history of the ‘feature detector’. *Cerebral cortex*, 4:1–7, 1994. 71
- C. F. Michaels and C. Carello. *Direct Perception*. Prentice-Hall, Englewood Cliffs, NJ, 1981. 18, 185
- A. Michotte. *The Perception of Causality*. Basic Books, 1946/1963. 178

- R. G. Millikan. *White Queen Psychology and Other Essays for Alice*, chapter 4 - Biosemantics, pages 83–101. MIT Press, Cambridge, 1993a. 189
- R. G. Millikan. Content and vehicle. In N. Eilan, R. McCarthy, and B. Brewer, editors, *Spatial Representation*, pages 256–268. Blackwell, Oxford, 1993b. 128
- K. Neander. Teleological theories of mental content. In *Stanford Encyclopedia of Philosophy*. 2004. URL <http://plato.stanford.edu/entries/content-teleological/>. 190
- U. Neisser. *Cognitive psychology*. New York: Appleton-Century Crofts, 1967. 54, 185
- A. Noë. *Is the visual world a grand illusion?* Imprint Academic, Charlottesville, VA, 2002. 129, 201, 214
- A. Noë and E. Thompson. Neural correlates of consciousness and the matching content doctrine. In Noë (2002). 83, 84
- A. Oliva and P. G. Schyns. Diagnostic colors mediate scene recognition. *Cognitive Psychology*, 41(2):176–210, 2000. 179
- A. Oliva and P. G. Schyns. Coarse blobs or fine edges? evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, 34:72–107, 1997. 179
- J. K. O'Regan. Skill, corporality and alerting capacity in an account of sensory consciousness. In *Enactive Virtual Workshop*, 2004. URL <http://www.interdisciplines.org/enaction>. 153

- J. K. O'Regan. Solving the real mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, 46:461–488, 1992. 190
- J. K. O'Regan and A. Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5):939–1011, 2001. 97, 153, 154, 157
- D. Osherson, S. Kosslyn, and J. Hollerbach, editors. *Visual Cognition and Action. An invitation to cognitive science*. MIT Press, Cambridge, MA, 1990. 206, 219
- S. Palmer. Perceptual grouping: It's later than you think. *Current Directions in Psychological Science*, 11:101–106, 2002. 62
- S. Palmer and I. Rock. Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin and Review*, 1:29–55, 1994. 35, 43
- D. Philipona. Structure of sensorimotor contingencies. In *Enactive Virtual Workshop*, 2004. URL <http://www.interdisciplines.org/enaction>. 153, 156
- D. Philipona, J. K. O'Regan, and J.-P. Nadal. Is there something out there? inferring space from sensorimotor dependencies. *Neural Computation*, 15(9), 2003. 158
- D. Philipona, J. K. O'Regan, and O. J.-M. D. Coenen. Do sensorimotor contingencies determine the perceptual structure of color? 2005. 157
- J. R. Pomerantz. Wholes, holes and basic features in vision. *Trends in Cognitive Sciences*, 2003. 204

- J. R. Pomerantz. Are complex visual features derived from simple ones? In E. L. J. Leeuwenberg and H. F. J. M. Buffart, editors, *Formal theories of visual perception*. Wiley, Chichester, 1978. 41
- J. R. Pomerantz and M. Kubovy. Theoretical approaches to perceptual organization: Simplicity and likelihood principles. In K. R. Boff, L. Kaufman, and J. P. Thomas, editors, *Handbook of Perception and Human Performance. Vol. 2 - Cognitive Processes and Performance*, chapter 36. New York, NY: Wiley, 1986. 50
- J. R. Pomerantz, M. Portillo, S. Jewell, and A. Agrawal. The genesis of perceptual organization: Basic features in vision. In *Proceedings of the 44th Annual Meeting of the Psychonomic Society*, 2003. 9, 60, 112, 113
- M. Posner. Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32:3–25, 1980. 89
- W. Prinzmetal and B. Keysar. Functional theory of illusory conjunctions and neon colors. *Quarterly Journal of Experimental Psychology*, 118:165–190, 1989. 99
- Z. W. Pylyshyn. Situating vision in the world. *Trends in Cognitive Sciences*, 4:197–207, 2000. 104
- Z. W. Pylyshyn. Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80:127–158, 2001. 104
- Z. W. Pylyshyn. Notes for the oleron summer workshop on objects. June 2004. 107
- P. Quinlan and G. Humphreys. Visual search for targets defined by combinations of color, shape, and size: An examination of the task constraints

- on feature and conjunction searches. *Perception and Psychophysics*, 41(5): 455–472, 1987. 56
- V. Ramachandran. Perception of shape from shading. *Nature*, 331:163–166, 1988. 65
- M. Redington and N. Chater. Probabilistic and distributional approaches to language acquisition. *Trends Cogn Sci*, 1(7):273–281, 1997. 145
- M. Redington, N. Chater, and S. Finch. Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, 22: 425–469, 1998. 145
- R. A. Rensink. Seeing, sensing and scrutinizing. *Vision Research*, 40:1469–1487, 2000. 67, 97, 99, 124, 179, 195
- R. A. Rensink and J. T. Enns. Preemption effects in visual search: Evidence for low-level grouping. *Psychological Review*, 102:101–130, 1995. 64
- J. Reynolds and R. Desimone. The role of neural mechanisms of attention in solving the binding problem. *Neuron*, 24:19–29, 1999. 119, 122
- I. Rock. In defense of unconscious inference. In W. Epstein, editor, *Stability and Constancy in Visual Perception: Mechanisms and Processes*. John Wiley and Sons, New York, 1977. 19, 185
- I. Rock. *The logic of perception*. MIT Press, Cambridge, MA, 1983. 185
- I. Rock, editor. *Indirect Perception*. MIT Press, Cambridge, MA, 1997a. 200, 217
- I. Rock. The concept of indirect perception. In *Indirect Perception* [Rock \(1997a\)](#), chapter 1. 19, 185

- D. Rose and V. Dobson, editors. *Models of the visual cortex*. John Wiley and Sons, New York, 1985. 205, 213
- J. R. Saffran, R. N. Aslin, and E. L. Newport. Statistical learning by 8-month-old infants. *Science*, 274:1926–1928, 1996a. 144, 193
- J. R. Saffran, E. L. Newport, and R. N. Aslin. Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35:606–621, 1996b. 144, 193
- J. R. Saffran, A. Senghas, and J. C. Trueswell. The acquisition of language by children. *PNAS*, 98(23):12874–12875, 2001. 34
- B. J. Scholl and P. Tremoulet. Perceptual causality and animacy. *Trends Cogn Sci*, 4(8):299–309, 2000. 177, 178
- B. J. Scholl, Z. W. Pylyshyn, and S. L. Franconeri. The relationship between property-encoding and object based attention: Evidence from multiple object tracking. *Perception and Psychophysics*, submitted. URL <http://rucss.rutgers.edu/ftp/pub/papers/scholl-pylyshyn-franconeri.pdf>. 105
- P. G. Schyns and A. Oliva. Flexible, diagnosticity-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception*, 26(8):1027–1038, 1997. 193
- P. G. Schyns, R. L. Goldstone, and J.-P. Thibaut. The development of features in object concepts. *Behavioral and Brain Sciences*, 21:1–54, 1998. 35, 36
- M. Shadlen and J. Movshon. Synchrony unbound : A critical evaluation of the temporal binding hypothesis. *Neuron*, 24:67–77, 1999. 80, 122

- B. Smith. The ecological niche, 1999. 137
- L. B. Smith, S. S. Jones, and B. Landau. Naming in young children: a dumb attentional mechanism? *Cognition*, 60:143–171, 1996. 193
- E. Sober. *Simplicity*. Clarendon Press, Oxford, 1975. 49
- E. Spelke. Origins of visual knowledge. In [Osherson et al. \(1990\)](#). 43, 109
- E. Spelke. Object perception. In A. I. Goldman, editor, *Readings in Philosophy and Cognitive Science*. MIT Press, Cambridge, MA, 1993. 43
- E. Spelke. Initial knowledge: Six suggestions. *Cognition*, 50:431–445, 1994. 109, 191
- D. Sperber. Modularity and relevance: How can a massively modular mind be flexible and context-sensitive. In P. Carruthers, S. Laurence, and S. Stich, editors, *The Innate Mind*. Oxford University Press, Oxford, 2006. 194
- D. Sperber. The modularity of thought and the epidemiology of representations. In [Hirschfeld and Gelman \(1994\)](#), pages 39–67. 192
- P. Strawson. *Individuals. An Essay in Descriptive Methaphysics*. Doubleday, Garden City, NY, 1959. 88
- N. W. Swindale. How many maps are there in visual cortex? *Cerebral cortex*, 10(7):633–643, 2000. 77, 92
- N. W. Swindale. Cortical organization: Modules, polymaps and mosaics. *Current Biology*, 8:R270–R273, 1998. 79, 92
- K. Tanaka. Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities. *Cereb Cortex*, 13(1):90–99, Jan 2003. 9, 45, 48

- D. Taraborelli. Feature binding and object perception. does object awareness require feature conjunction? In *Annual Meeting of the European Society for Philosophy and Psychology, Lyon 2002*, 2002. URL <http://jeanNicod.ccsd.cnrs.fr/documents/disk0/00/00/03/11/ijn-00000311-01/ijn-00000311-01.pdf>. 97, 120, 127
- D. Taraborelli. What is a 'feature'? a meta-analysis of selection criteria for salient visual properties (presented at the 26th european conference on visual perception). *Perception*, 32 (ECPV Supplement), 2003. 199
- D. Taraborelli and M. M. Mossio. A sensorimotor hypothesis for spatial competence (forthcoming). 2005. URL <http://dartar.free.fr/space>. 154
- M. J. Tarr and M. J. Black. Reconstruction and purpose. *CVGIP: Image Understanding*, 60(1):113–118, 1994a. 27
- M. J. Tarr and M. J. Black. A computational and evolutionary perspective on the role of representation in vision. *CVGIP: Image Understanding*, 60(1):65–73, 1994b. 200
- D. Teller. Linking propositions. *Vision Research*, 24:1233–1246, 1984. 129, 201
- D. Teller and E. Pugh. Linking propositions in color vision. In J. Mollon and L. Sharpe, editors, *Colour vision*, pages 577–589. Academic Press, London, 1983. 83, 129, 201
- P. M. Todd. Simple inference heuristics versus complex decision machines. *Mind and Machines*, 1999a. 195

- P. M. Todd. Fast and frugal heuristics for environmentally bounded minds. In [Gigerenzer and Selten \(1999\)](#). 147, 190
- A. Treisman. Focused attention in the perception and retrieval of multi-dimensional stimuli. *Perception and Psychophysics*, 22:1–11, 1977. 102, 118
- A. Treisman. Properties, parts and objects. In K. R. Boff, L. Kaufman, and J. P. Thomas, editors, *Handbook of Perception and Human Performance. Vol. 2 - Cognitive Processes and Performance*, chapter 35. New York, NY: Wiley, 1986. 16, 32, 71, 101
- A. Treisman. Features and objects : The fourteenth bartlett memorial lecture. *Quarterly Journal of Experimental Psychology*, 40A:201–237, 1988. 16, 56, 125
- A. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980. 54, 120
- A. Treisman and S. Gormican. Feature analysis in early vision: evidence from search asymmetries. *Psychological Review*, 95(1):15–48, Jan 1988. 56
- A. Treisman and D. Kahneman. The accumulation of information within object files. In *24th Annual Meeting of the Psychonomic Society*, 1983. 103
- A. Treisman and N. Kanwisher. Perceiving visually presented objects :recognition, awareness and modularity. *Current Opinion in Neurobiology*, 8: 218–226, 1998. 27
- A. Treisman and S. Sato. Conjunction search revisited. *Journal of Experimental Psychology : Human Perception and Performance*, 16:459–478, 1990. 58

- A. Treisman and H. Schmidt. Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14:107–141, 1982. 96, 121, 123, 201
- A. Treisman and J. Souther. Search asymmetry: a diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology*, 114 (3):285–310, Sep 1985. 56
- Y. Tsal. Do illusory conjunctions support the feature integration theory ? a critical review of theory and findings. *Journal of Experimental Psychology: Human Perception and Performance*, 15:394–400, 1989. 121
- S. Ullman. Against direct perception. *Behavioral and Brain Sciences*, 3: 373–415, 1980. 200
- S. Ullman, M. Vidal-Naquet, and E. Sali. Visual features of intermediate complexity. *Nature Neuroscience*, 50, 2002. 41
- A. Van der Heijden. Modularity and attention. *Visual Cognition*, 2:269–302, 1995. 92, 97, 118, 127
- D. Van Essen, C. Anderson, and D. Felleman. Information processing in the primate visual system: an integrated systems perspective. *Science*, 255: 419–423, 1992. 125
- C. Von der Malsburg. Binding in models of perception and brain function. *Current Opinion in Neurobiology*, 5:520–526, 1995. 132
- C. Von der Malsburg. The what and why of binding : The modeler’s perspective. *Neuron*, 24:95–104, 1999. 123, 132
- P. White and A. Milne. Phenomenal causality: impressions of pulling in the visual perception of objects in motion. *American Journal of Psychology*, 110:573–602, 1997. 178

- P. White and A. Milne. Impressions of enforced disintegration and bursting in the visual perception of collision events. *Journal of Experimental Psychology : General*, 128:499–516, 1999. 178
- E. O. Wilson. *Consilience: The Unity of Knowledge*. Knopf, 1998. 131
- J. Wolfe. What can 1,000,000 trials tell us about visual search? *Psychological Science*, 9(1), 1998a. 58
- J. Wolfe. Visual search. In H. Pashler, editor, *Attention*. University College London Press, London, 1998b. 54, 59
- J. Wolfe and K. Cave. The psychophysical evidence for a binding problem in human vision. *Neuron*, 24:11–17, 1999. 9, 55
- J. Wolfe, K. Cave, and S. Franzel. Guided search : an alternative to the feature integration model for visual search. *Journal of Experimental Psychology : Human Perception and Performance*, 15:419–433, 1989. 58
- J. Wolfe, S. Friedman-Hill, M. Stewart, and K. O’Connell. The role of categorization in visual search for orientation. *Journal of Experimental Psychology : Human Perception and Performance*, 18:34–49, 1992. 66
- J. M. Wolfe. Asymmetries in visual search: an introduction. *Percept Psychophys*, 63(3):381–389, Apr 2001. 56
- P. Wolff. Position of code and code for position: From isomorphism to a sensorimotor account of space perception. *Visual Cognition*, 11(2-3): 137–160, 2004. 92, 97, 127
- S. Zeki and L. Marini. Three cortical stages of colour processing in the human brain. *Brain*, 121:1669–1685, 1998. 125, 156

- S. Zeki and S. Shipp. The functional logic of cortical connections. *Nature*, 325:311–317, 1988. 79
- S. Zeki, J. Watson, C. Lueck, K. Friston, C. Kennard, and R. Frackowiak. A direct demonstration of functional specialization in human visual cortex. *Journal of Neuroscience*, 11(3):641–649, 1991. 79

List of abbreviations

The page number refers to the first occurrence of the abbreviation.

A-primitive	Atomic (or unanalyzable) primitive, 39
C-primitive	Compositional primitive, 31
FBP	Feature Binding Problem, 96
FING	Visual item attracting a visual index, 44
FINST	Visual index, 44
FIT	Feature Integration Theory, 56
L-primitive	Low-level primitive, 70
MDL	Minimum Description Length, 52
MOT	Multiple Object Tracking, 104
P-primitive	Processing primitive, 54
S-primitive	Simple primitive, 49

Index

- adaptation
 - adaptive history, 186–188
 - adaptive strategies, 188–190
 - constraints, 132–194
- affordance, 136, 138
- attention, 52–61, 96–97
 - space-based, 89
- binding, *see Feature Binding Problem*
- cognitive advantage, 138, 153, 163, 172, 179, 192
- color
 - and sensorimotor learning, 154–156
 - cross-level characterizations, 124
- complexity
 - Kolmogorov complexity, 52
- compositionality, 31–38
- computational economy, 137, 145, 162, 179
 - delegation strategies, 190
 - recruitment strategies, 189–190
 - rigidity, 191
- configurational superiority, 43, 104
- connectivity, 78
- constraints
 - adaptive, 20, 132–194
 - ascribability, 87–115
 - internal, 19, 22–132
 - locality, 88–100
 - functional, 93
 - intrinsic, 91
 - primitiveness, 29–85
 - sensorimotor, 151–160
- context-dependence, 147, 172–175, 185
- cross-level explanations, 197–198
 - bridge locus, 82
 - conflations, 127–129, 198
 - consilience strategies, 129–132, 198
 - homonymy, 130, 198
 - linking proposition, 82, 197
- cue, 65, 141–144, 186
- distal source, 26, 187, 199
- earliness, 52–68
- Early Selection Theories*, 61

- Ecological Approach to Perception*, 18,
136–139
- ecological niche, *see* environment, ecological niche
- Ecological Rationality*, 144–146
- empiricism, 188–190
- and language acquisition, 142
 - and perceptual learning, 162
 - and prior knowledge, 172
 - enlightened, 162
- encoding, 75
- vs.* detecting, 91
 - and object-files, 103, 105
 - economy, 50, 51
 - spatial, 92
- environment, 135–149
- distributional regularity, 141, 146,
161–166
 - ecological intelligence, 188–193
 - ecological niche, 136, 179
 - environmental regularity, 137, 199
 - non-accidental, 147
- feature
- ascribable, 87–115
 - basic, 15–22, 195–200
 - binding, *see Feature Binding Problem*
 - bundle, 102
 - conjunction, 55, 94
 - illusory, 96, 121
 - detection, 57, 70–77, 159, 183–186
 - bug detector, 73, 164
 - diagnostic, 190
 - dynamic, 98, 158–160
 - effective, 75
 - emergent, 41
 - extensional characterization, 122
 - feature placing, 88
 - global, 66, 98
 - integration, *see Feature Integration Theory*
 - local, 60, 68, 88–100
 - map, 55, 76, 91, 125
 - multilocal, 98
 - non-local, 97–99
 - non-objectual, 111–114
 - objectual, 100–115
 - preattentive, 53, 177
 - punctual, 39, 98
 - scene-based, *see* gist
 - shunttable, 175–177
 - shunting, 179–180
 - unbound, 120

- Feature Binding Problem*, 55, 96, 117–132
- Feature Integration Theory*, 56, 94, 104, 125
- FING, 44
- FINST, *see* vision, visual indexicality
- functional anatomy, 73
- functional indeterminacy, 188
- Ganzfeld*, 112
- geon, 33
- Gestalt*, 42
- gist, 66, 177
- granularity, 40
- heuristics, 144, 172, 192–193
- inference, *see* perception, perceptual inference
- language
- acquisition, 141
 - motor theories, *see* Motor Theory of Speech Perception
- Late Selection Theories*, 61
- learning, 190
- bootstrapping, 43, 142, 153, 162, 191
 - sensorimotor, 152–153
 - statistical, 141–144
 - likelihood, 50
 - low-levelness, 68–85
 - and early processing, 61
 - magnetosome, 186–188
 - Many Property Problem*, 94, 121
 - Matching-Content Doctrine*, 83, 197
 - Minimum Description Length*, 52
 - modularity
 - actual *vs.* proper domain, 190
 - consumer device, 187
 - domain specificity, 161
 - encapsulation, 41, 191
 - impenetrability, 70
 - neural segregation, 127
 - representational format, 40
 - Motor Theory of Speech Perception*, 159
 - multi-dimensional analysis, 74–77
 - nativism, 188–190
 - and language acquisition, 142
 - and prior knowledge, 172
 - natural constraint, 172
 - Neuron Doctrine*, 71–74
 - non-inferentiality, 182–188
 - object
 - Multiple Object Tracking*, 104

- object disruption properties, 110
- object fixation properties, 44, 106
- object preservation properties, 107
- object-encodable properties, 106
- object-file, 101–111
- preattentive, 112
- recognition, 33
 - faces, 162
 - Spelke Object*, 43, 109
 - superiority, 43, 104
- objecthood, 100–115
 - and cognitive development, 43, 109
 - commonsensical notions, 111
- pattern
 - patterned domain, 34, 50, 144, 153
 - recognition, 31
 - sensorimotor, 152–153
- perception
 - and epistemic reliability, 27, 132, 139, 145, 147, 172, 192, 199
 - Bayesian, 146–148, 166–172
 - direct *vs.* indirect, 18, 182–186
 - face recognition, 162
 - general-purpose, 26, 139, 145
 - perceptual agency, 175–176
 - perceptual animacy, 175–176
 - perceptual causality, 176–177
 - perceptual error, 172, 192
 - perceptual illusions
 - causality, 176
 - illusory conjunctions, *see* feature conjunction, illusory
 - perceptual inference, 25, 147–148, 172–175, 182–188, 199
 - perceptual proxy, 164, 178, 179
 - perceptual reference, 89
 - processing, 15, 27, 53
 - hierarchical, 35, 78
 - limits, 96, 145
- Perceptual Shunt, 161–180
 - and sensorimotor constraints, 178
 - and representation, 186–188
 - Perceptual Shunt Hypothesis*, 164
 - shunter, 166–172
- phenomenology, 100
 - phenomenological saliency, 35, 82–83, 108, 115
 - phenomenological unity, 120
 - internalization, 128
- Poverty of the Stimulus*, 27, 182
- Prägnanz*, 50
- preemption, 64
- primitive

- A-primitive, *see* primitive, unanalyzable
 C-primitive, *see* primitive, compositional
 L-primitive, *see* primitive, low-level
 P-primitive, *see* primitive, early
 S-primitive, *see* primitive, simple
 atomic, *see* primitive, unanalyzable
 compositional, 31–38
 dynamic, 36
 early, 52–68
 lexicon, 31, 38
 low-level, 68–85
 simple, 44
 static, 36
 unanalyzable, 38–44
 primitiveness, 29–85
 proto-predication, 89, 97, 102
Psychophysical Linking Hypothesis, 72, 197
 rationality
 bounded, 145
 ecological, *see Ecological Rationality*
Recognition by Components Theory, 33
 regularity
 and redundancy, 50–52
 distributional, *see* environment, distributional regularity
 ecological, *see* environment, environmental regularity
 prosodic, 191
 representation
 episodic, 103
 flexibility, 36
 in shunt mechanisms, 186–188
 representational format, 40
 retinotopy, 76, 92
Scaled Channels Hypothesis, 42
 segregation, *see* modularity
 sensorimotricity, 151–160
 and color perception, 154–156
 and spatial perception, 156–157
 sensorimotor contingency, 152
 sensorimotor coupling, 152–153
 sensory organ, 18
 simplicity, 44–52
 sketch, 32
 teleosemantics, 186–188
 texture
 segregation, 57
 texton, 57, 58

transduction, 183–186

unanalyzability, 38–44

vision

- at a glance, 66

- front-end, 98

- low-level, 69

- preattentive, 54

- visual grouping, 42

- visual indexicality, 44, 105

- visual pathway, 78

- visual search

 - asymmetry, 56

 - guided, 57

 - parallel, 54

 - serial, 55

- with scrutiny, 66

ACKNOWLEDGMENTS

This work was accomplished during my staying as a PhD student at the Institut Jean Nicod in Paris. As I argue in the second part of my dissertation, environment *does* matter: I have benefited from the chance to let these ideas grow and ripen in an ideal context, where philosophical and theoretical investigations on cognition meet empirical research. The opportunity to work at IJN was made possible thanks to a 3-year research grant (*Allocation de recherche*) from the French Ministry of Research and Education and the Doctoral School 3C *Cerveau, Cognition, Comportement*. Financial support for the last part of my work has been provided by a grant from the ENACTIVE Network of the 6th European Framework Programme, for which I have been working as research assistant based at the Ecole des Hautes Etudes en Sciences Sociales since March 2005.

I am indebted to so many people at IJN that it would be too risky to explicitly acknowledge each of them without being guilty of omissions. My first personal debt is towards Roberto Casati, who supervised my work at IJN, for his strenuous and energetic supervision, his scientific and human support throughout these years. I have learnt very much from him and this work would not have seen the light without his severe coaching. I would like to thank the members of the Jury of my *présoutenance*, Pierre Jacob, Jérôme Dokic and Pascal Mamassian, for insightful suggestions and criticisms on an earlier version of this work that helped me improve some of its main weaknesses. Several parts of this work have been inspired by discussions and joint work with several colleagues and friends over the last years, among whom I should mention in particular Benjamin Sylvand and Ángeles Eraña (for issues related to object-related features), Matteo Mossio (for the sensorimotor hypothesis), Christophe Heintz (for issues related to modularity and cognitive affordability). I owe to each of them much more than this work can witness. Kevin O'Regan during my first years and Dan Sperber during the last have helped me unpack some of the ideas with which I have been struggling. Last but not least, I wish to thank my DEUG students at Paris 7 who have suffered an introductory course to cognitive science during spring term 2005 more biased towards these ideas than to pedagogic reasonableness. I thank them for their patience and support in helping me understand that simplicity is the most difficult and important goal to attain.

The one thanks to whom all this has finally become possible and without whom I could not even imagine to see the last word is Emmanuelle: thank you, *olla vogala* they are damn right.