Probabilistic models of perceptual features

Jacob Feldman
Dept. of Psychology, Center for Cognitive Science, Rutgers University

To appear in:
Oxford Handbook of Perceptual Organization
Oxford University Press
Edited by Johan Wagemans

Abstract

Perceptual features—properties of objects as the visual system represents them—are a central construct of perception. Classically, features have been treated as deterministic qualities of images, assigned definite values based on image structure. But the development of probabilistic models of perception has led to a new way of understanding features, treating them as probabilistic estimates of parameters of the scene. This chapter briefly develops the probabilistic conception of features, illustrating it with examples drawn from the literature on perceptual organization. Major topics include non-accidental features and non-local features.

1. Features

A ubiquitous element in perceptual theory is that of a *feature*, meaning a measurable attribute of an object, such as its color, form, orientation, or motion. Features are a routine part of the description of experimental stimuli, and an essential component of verbal descriptions of everyday visual experience (*the black pen is on the square table*). Features play a wide variety of roles in perceptual theory. Features such as convexity and symmetry are thought to influence figure/ground interpretation (Kanizsa & Gerbino, 1976), helping to form initial representations of objects (see Peterson, this volume). Later on each object's features are bound together to form complex object representations (Treisman & Gelade, 1980; Ashby, Prinzmetal, Ivry, & Maddox, 1996). Still later each objects' features are used to classify them into larger categories (Feldman, 2000; Lee & Navarro, 2002; Ullman, Vidal-Naquet, & Sali, 2002).

But behind the simple idea of a "feature" lurks some deep theoretical issues and controversies, involving how features are defined and what motivates the choice of a particular feature vocabulary (Jepson & Richards, 1992; Koenderink, 1993). This brief chapter centers on the ongoing evolution of the feature concept from a "classical" view, in which features are deterministic attributes of objects, to a more probabilistic view, in which features are probabilistic estimates of attributes inferred from image data. The newer view has grown in prominence in conjunction with a broader probabilistic conception of perceptual inference more generally (Knill & Richards, 1996).

I am grateful to Irv Biederman, Manish Singh, Wolf Vanpaemel, Johan Wagemans, and an anonymous reviewer for helpful comments. Preparation of this article was supported by NIH EY021494. Please direct correspondence to the author at jacob@ruccs.rutgers.edu.

It is useful first to distinguish certain commonly used terms and among different notions of "feature" that are occasionally conflated. The terms *feature, property, dimension* and *attribute* are all commonly used to refer to an image characteristic that varies among visual objects. Each of these terms is sometimes used to indicate the characteristic that can vary (e.g. *size*), or to a particular value that it can take (e.g., *large*). Thus some authors refer to *color* as a feature, while others use the term to refer to specific values such as *red, white* or *blue*; and so forth. Some authors reserve one term (e.g. feature) for the variable, another for the value (e.g. property), but such usage does not seem to be consistent across the literature. The term feature is sometimes reserved for discrete qualities, meaning those that can take one of a finite number of distinct values, including discretizations of what are normally continuous-valued properties: examples include *red* vs. *green* (two discrete cases of the continuous parameter *color*) or *vertical* vs. *horizontal* (discrete cases of the continuous parameter *orientation*), and so forth (Aitkin & Feldman, 2013). Features with exactly two values, often referred to as binary or Boolean, can be understood to involve the presence or absence of some attribute (e.g. *red* vs. *non-red*).

A more subtle distinction particular to the term *feature* is that it is sometimes used to refer to localizable elements within an image, such as the facial "features"—eyes, nose, and mouth—located at various positions on a face. Researchers in stereopsis, for example, refer to correspondence between features in the left and right visual images, meaning local elements of the image with well-defined locations (Poggio & Poggio, 1984). Any visual function that involves searching for, counting, or measuring distances among features presumes this sense of the word. In contrast many "features," such as shape or color, are not localizable, but are characteristic of whole objects. The distinction between these two senses of *feature* breaks down a bit when spatially localizable elements are described in terms of their characteristics. For example, a T-junction is a spatially localizable element of a line drawing, but is also a characteristic that some line junctions have and others do not. In this review I will focus on the first sense of feature, as a characteristic that varies among objects, although the issue of localizability becomes central later when we consider local vs. global support for features.

2. Classical vs. probabilistic models of features

Historically, features have usually (and often tacitly) been defined by clear-cut criteria: e.g. feature f holds when image measurement m lies above some threshold m_0 ($m \geq m_0$) and does not hold (i.e., $\neg f$ holds) otherwise ($m < m_0$) (Fig. 1). Thus vertical lines are those between within 5° of the direction of gravity; collinear edges have orientation difference of less than 30° (e.g. Field, Hayes, & Hess, 1993); *relatable* edges have linear extensions that intersect at an acute angle (Kellman & Shipley, 1991). Such definitions have the advantage of clarity, and are often perfectly apt for experimental contexts in which stimuli are artificially constructed to either fulfill them or not fulfill them as desired for the purposes of the experiment. However, with natural stimuli this simple criterial conceptualization of features suffers from at least three problems: hard boundaries, arbitrariness, and insensitivity to context.

2.1. Hard boundaries.

Criterial definitions impose clear-cut boundaries between values of a feature, treating all instances that meet the criterion equivalently. Thus all vertical lines are equally vertical, while all non-vertical lines are equally not. Such a boundary inevitably treats many nearby cases as qualitatively different, while folding together cases that are distant in the underlying space, a distortion that rarely corresponds well to the more graded percept. With hard criteria, in-between cases do not exist; there is no way of expressing the idea of a line that is somewhat, almost, or partly vertical.

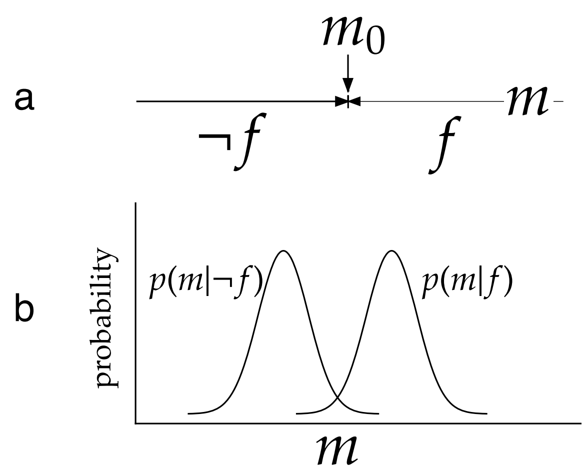

Figure 1. Schematic illustrating the difference between (a) classical features, which divide the measurement space (*m*) into clean-cut classes; and (b) probabilistic features, which are based on potentially overlapping probability distributions.

This issue parallels a famous debate in the literature on cognitive categories, which following the seminal papers of Posner and Keele (1968) and Rosch (1973) evolved from a "classical" conception based on necessary and sufficient features (see Smith & Medin, 1981) to a graded and "fuzzy" view based on prototypes (Posner & Keele, 1968; Reed, 1972), exemplars (Medin & Schaffer, 1978; Nosofsky, 1986), or both (Nosofsky, Palmeri, & McKinley, 1994; Anderson & Betz, 2001), in order to account for the observation that some category instances seem to be better examples of the category than others. In recent years the modern view has been expressed via probabilistic models in which conceptual representations are probabilistic estimates of underlying generating classes (Anderson, 1991; Ashby & Alfonso-Reese, 1995; Goodman, Tenenbaum, Feldman, & Griffiths, 2008; Briscoe & Feldman, 2011). In a few famous cases, perceptual processes do seem to impose relatively hard boundaries at thresholds along continuous parameters, a phenomenon known as *categorical perception* (see Harnad, 1987). However such cases are exceptional, and in any case even they seem to involve gradations in the vicinity of the threshold.

2.2. Arbitrariness.
In the classical view a feature like *between 600 and 601 meters in height* is perfectly well-defined, even though it captures no natural kind, and may not distinguish in any useful way between objects that satisfy it and those that do not. Such features are arbitrary in that they fail to relate to real classes actually extant in the environment. A desirable property of a feature vocabulary is that it be well-tuned to the classes it use used to describe, a desideratum the classical model in no way guarantees.

2.3. Insensitivity to context.
A feature like *has a 6-cylinder engine* is perfectly well-defined for cars, but makes no sense when applied to trees, and vice versa for *evergreen*. Such features make meaningful distinctions only within a single narrow context. Indeed human subjects are known to employ different features depending on context (Blair & Homa, 2005; Schyns, Goldstone, & Thibaut, 1998; Goldstone & Steyvers, 2001) and can learn new features in new contexts (De Baene, Ons, Wagemans, & Vogels, 2008; Stilp, Rogers, & Kluender,

2010). But a classical feature vocabulary does not in any way constrain the context in which features are applied, since their definitions make reference only to image conditions satisfied or not. As with arbitrariness, the problem is that classical features allow no connection between their definitions and the properties of the environment.

The sections that follow outline a modern probabilistic conception of features that avoids each of the above defects. Probabilistic conceptions of features are certainly not new, but have grown over several decades (from roots in signal detection theory; see Green & Swets, 1966). The recent explosion in probabilistic conceptions of perception (see Kersten, Mamassian, & Yuille, 2004 or Feldman, this volume) has introduced a natural mathematical language for expressing many probabilistic ideas, including that of a perceptual feature. In what follows I attempt to lay out the basic modern idea of probabilistic features in a simple and general way.

3. Probabilistic features

From a probabilistic viewpoint, features are attributes of objects that are *estimated* from image measurements, rather than measurements of image properties per se. The assumption is that measurable image properties derive from both fixed distal properties of objects as well as random noise, and that useful features attempt to extract the signal from the noise (see Fig. 1). To formalize this, we assume that an object feature f involves a likelihood distribution over image measurements m,

$$p(m|f) \sim \mu + e, \tag{1}$$

where μ is some mean value of m conditioned on the presence of the feature, and e is an error drawn from a noise distribution with mean 0, such as a Gaussian

$$e \sim N(0, \sigma^2). \tag{2}$$

The probabilistic assignment of feature values to image structures then proceeds by Bayes' rule: an object with measurement m is assigned feature f in proportion to the posterior

$$p(f|m) \propto p(m|f)p(f), \tag{3}$$

where $p(f)$ is a prior distribution over feature values. The prior may (though need not) be uniform (e.g. $p(f) = p(\neg f)$ in the case of a Boolean feature) in which case the posterior is proportional to the likelihood. The likelihood model $p(m|f)$ is sometimes called a *generative model* because in effect it is a model of how the image was generated, describing how observables (m) are generated stochastically from the distal reality (f). The Gaussian model given above is only an example; other functional forms may be assumed, so long as they define a distribution $p(m|f)$.

For example, the feature *large* might classically have been defined via a range of permissible object sizes. But probabilistically it would be defined via a mean size μ, say 3 cm, plus some error distribution, say normal with standard deviation 1 cm. (The mean μ itself might be conditioned on other aspects of context, allowing *large* to mean different things in different contexts; see below.) In contrast with the classical view, this means that largeness is a graded quality, with some objects more likely to be regarded as large (namely, those closer to 3 cm) and others less likely. This also means that the category of large objects can actually overlap with that of small objects (see Fig. 1). That is, a given object can be described by two *contradictory* features, although generally with different probabilities.

3.1.Non-accidental features

Non-accidental features are an important class of perceptual feature that have received somewhat more careful mathematical attention. As originally defined by Binford (1981) and Lowe (1987) non-accidental features are properties of 2D configurations (e.g., cotermination of line segments in the image) that reliably occur in the presence of associated 3D configurations (cotermination of 3D line segments in the world) but are very unlikely other- wise; that is, they are unlikely to occur "by accident." Other examples include collinearity, parallelism, and skew symmetry (Wagemans, 1993). More generally, a non-accidental feature is one that has high probability if certain distal conditions are satisfied, but low probability otherwise. There is substantial, though not unalloyed, empirical evidence that the visual system is particularly sensitive to non-accidental features[1] (Wagemans, 1992; Vogels, Biederman, Bar, & Lorincz, 2001; Feldman, 2007; Amir, Biederman, & Hayworth, 2012), and they play an important role in Biederman's influential (1987) Recognition by Components (RBC) account of object recognition.

Formally, a discrete image feature M (corresponding, say, to a fixed range of some measurement m) is non-accidental with respect to a distal feature f if M has high probability in the presence of f, i.e. $p(M|f) \approx 1$, but low probability otherwise, $p(M|\neg f) \approx 0$. Jepson and Richards (1992) showed that another condition is required in order for M to reliably indicate the presence of f, namely that the prior on f be elevated relative to alternatives. That is, f must be a condition that occurs with elevated probability in the world; it must be a recurring regularity (see also Feldman, 2009).[2] As in the ubiquitous illustration of Bayesian inference in a medical context—in which reliable inference of a disease based on a positive test requires not only an accurate (sensitive and specific) test but also a high prior (e.g. see Gigerenzer & Hoffrage, 1995)—it is not sufficient that a measurement class M be likely conditioned on a world state f; the world state f must itself have a high prior.

An example of a non-accidental feature is *collinearity*, extensively studied in the literature on contour integration and completion (Hess, May, & Dumoulin, this volume; Field et al., 1993; Uttal, Bunnell, & Corwin, 1970; Elder & Goldberg, 2002; Geisler, Perry, Super, & Gallogly, 2001). In classical definitions, collinearity is defined via a criterion on the orientation difference between successive edges in a chain. In probabilistic formulations (Feldman, 1995, 1997; Feldman & Singh, 2005), collinearity is defined by a probability distribution over turning angles (usually a normal or von Mises distribution) centered on 0° (straight continuation). This distribution gives a formal definition of the graded quality the Gestaltists called "good continuation," with perfectly straight being the "best" and deviations from straight constituting progressively "worse" instances. In the probabilistic conception there is no such thing as a turning angle that is definitely collinear or definitely not; any turning angle might be an instance of the class (i.e., have been generated from a smooth contour process), though straighter ones are more likely to be.

---

[1] More precisely, there is very strong evidence that qualitative features such as non-accidental ones have special salience relative to "metric" or quantitative features (see references in text). But it is not completely clear whether non-accidentalness is the correct mathematical characterization of "qualitative" features.

[2] To see why, assume that we express the condition $p(M|f) \approx 1$ as $p(M|f) = 1 - \varepsilon$ (with ε some low nonzero probability), and similarly $p(M|\neg f) = \varepsilon$. Similarly assume f has low prior compared to alternatives, e.g. $p(f) = \varepsilon$ and $p(\neg f) = 1 - \varepsilon$ (meaning that f occurring a priori is just as unlikely an accident as M occurring without f). With these assumptions the posterior on f when M holds will be

Moreover, collinearity understood this way satisfies the requirement of elevated prior needed to guarantee statistical reliability. Collinear turning angles, generated approximately from the von Mises distribution, occur along smooth contours, but relatively rarely otherwise (only "by accident"). Smooth contours themselves are ubiquitous in the world because they occur along the boundaries of many objects (Ren, Fowlkes, & Malik, 2008). Because of this elevated probability, image conditions suggestive of collinearity generally *do* reliably signal collinearity in the world. Like a positive test for a disease that *does* have a high prior, observed collinearity reliably signals common physical origins.

3.2. Local vs. global features

A persistent issue in the definition of visual features is the size of the image region that contributes data to determining them. At one extreme, *local* features, like color, depend on data at a point or within a small neighborhood of the image. At the other extreme, more *global* features reflect properties of entire objects, large image regions, or even the entire image. Few features are perfectly local. Even nominally local image features such as motion or luminance, which are in principle well-defined at each point in the image, often require integration over substantial regions of the image in order to achieve stable estimates. Image motion, for example, is often ambiguous unless a substantial image region is considered (Ullman, 1979). The percept of luminance (perceived reflectance) can involve comparisons over large image distances (Gilchrist, 1977). Texture perception similarly requires integration across image patches (Rozenholtz, this volume; Wagemans, van Gool, Swinnen, & van Horebeek, 1993; Pizlo, Salach-Golyska, & Rosenfeld, 1997) and is even influenced by global shape (Harrison & Feldman, 2009). Many features, like figure/ground polarity along a contour, are in principle properties of individual points or small neighborhoods (Kim & Feldman, 2009), but are nevertheless determined in part by evidence from outside this neighborhood (Kogo & van Ee, this volume; Zhang & von der Heydt, 2010).

The ubiquitous dependence of local features on structure elsewhere in the image has led to a widespread recognition of the insufficiency of the classical notion of receptive field (the image region that directly influences a cell's response), as many cells are also demonstrably influenced by a much larger region (Fitzpatrick, 2000). Whether this influence is conveyed via feedback from later brain areas or via horizontal (lateral) connections is an area of ongoing debate (Angelucci & Bullier, 2003; Craft, Schutze, Niebur, & von der Heydt, 2007).

---

probability), and similarly $p(M|\neg f) = \varepsilon$. Similarly assume f has low prior compared to alternatives, e.g. $p(f) = \varepsilon$ and $p(\neg f) = 1 - \varepsilon$ (meaning that f occurring a priori is just as unlikely an accident as M occurring without f). With these assumptions the posterior on f when M holds will be

$$p(f|M) = \frac{p(M|f)p(f)}{p(M|f)p(f) + p(M|\neg f)p(\neg f)} \quad (4)$$

$$= \frac{(1-\varepsilon)(\varepsilon)}{(1-\varepsilon)(\varepsilon) + (\varepsilon)(1-\varepsilon)}$$

$$= \frac{1}{2} \quad (5)$$

That is, the probability of f in the presence of M (1/2) is no greater than the probability of $\neg f$ (also 1/2). That is, if f has low prior, then even though M is non-accidental in the standard sense, observing M does *not* actually indicate that f is particularly likely. As Jepson & Richards showed, a small "accident probability" of ε (i.e., non-accidentalness) only leads to reliable inference if the feature prior $p(f)$ is substantially greater than ε.

From a computational point of view, the difficulty posed by non-local features is the potentially enormous increase in computing complexity they pose. The larger the region of the image contributing to the determination of a feature, the more complex the computation. Partly as a result many of the most influential modern proposals for basic feature vocabularies (e.g. SIFT, Lowe, 2004, and HMAX, Riesenhuber & Poggio, 1999) rely on more sophisticated definitions of local image features and feed-forward computational architectures. But many perceptual decisions made by human observers with apparent ease depend on subtle aspects of entire objects or scenes that are difficult to specify or model (Treisman & Paterson, 1984; Biederman & Shiffrar, 1987; Pomerantz & Pristach, 1989; Wilder, Feldman, & Singh, 2011). To understand such non-local features probabilistically requires the construction of appropriate generative models, in many cases multidimensional and hierarchical ones.

Many examples come from the domain of shape, a quintessentially non-local class of feature that defies easy classical definitions. Many intuitively transparent shape features, that is, lack clear qualitative definitions, but can be understood probabilistically once suitable probabilistic models are defined. For example, human observers can readily distinguish shapes with two parts from those with only one (Fig. 2), suggesting a perceptually accessible feature of *two-partedness*. But the distinction between multipart and single-part shapes is notoriously difficult to model because the decomposition of shapes into component parts does not rely on any simple attribute but instead involve a large set of non-local shape cues (Singh & Hoffman, 2001; de Winter & Wagemans, 2006). Classically, one would need to find some parameter reflecting two-partedness, and set a threshold above which a shape is considered to have two parts rather than one. But such a parameter is difficult to identify, and any threshold along it would be arbitrary. One can define a spectrum of shapes (see abscissa of Fig. 2) that vary smoothly from shapes clearly having one part (left of figure) to those clearly having two (right of figure). Exactly where along this spectrum the boundary between one and two lies is unclear.

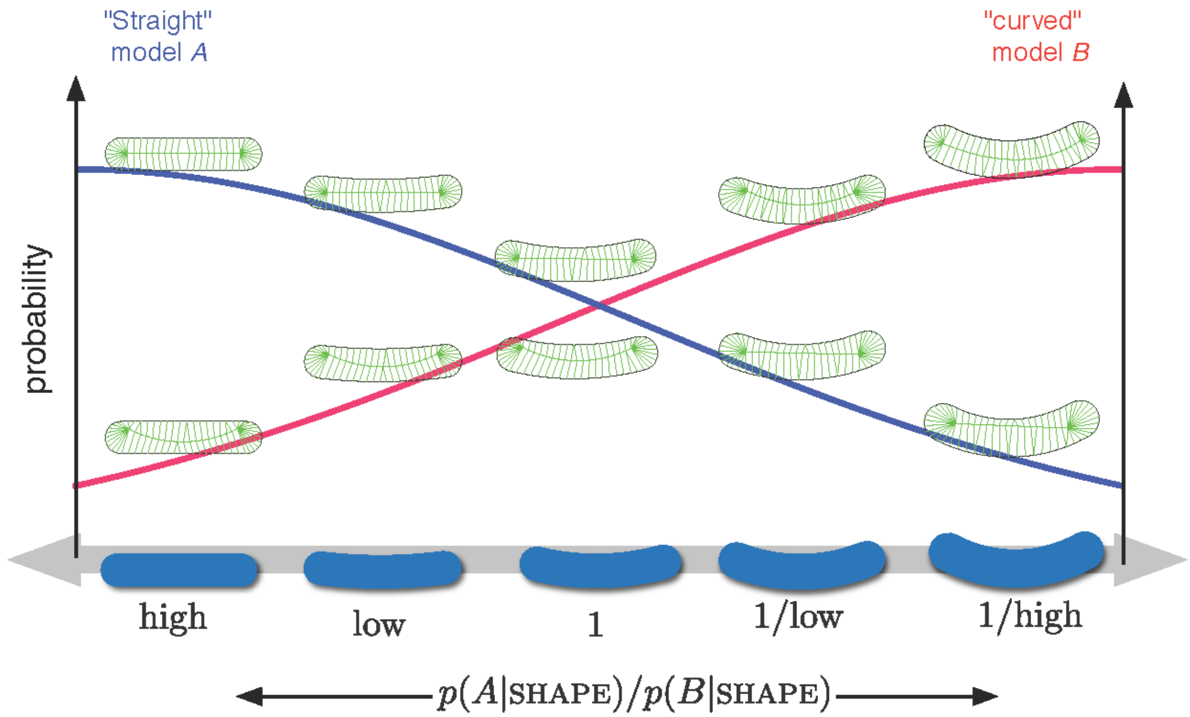

Figure 2. The shape feature *two-parts* vs. *one-part* viewed probabilistically. The figure shows a spectrum of shapes ranging from a single part (left) to two parts (right). Towards the left shapes are well fit by a two-part model and poorly fit by a one-part model; at the right, vice-versa. (Models are shown with ribs; likelihood is diminished by variance in the lengths and directions of the ribs along with several other factors.) The figure illustrates how the relative probability (posterior ratio) of the two models shifts from favoring the one-part model on the left to favoring the two-part model on the right.

Alternatively, one can understand this shape feature probabilistically by defining distinct generative models for one- and two-part shapes. In the framework of Feldman and Singh (2006), a one-part model would have a single axis (see Fig. 2) from which the shape grows laterally; this tends to yield simple elliptical shapes with random variations. Similarly, a two-part model would have two axes, one branching off the other (see Fig. 2), thus tending to generate shapes with two distinct parts. (The recursively branching aspect of this generative model makes it *hierarchical*; see Goldstone, Medin, & Gentner, 1991; Sanocki, 1999; Geisler & Super, 2000 for diverse discussion of hierarchy in perceptual representations.) Each model can generate shapes anywhere along the spectrum, but with different probabilities; the distributions overlap. Fig. 2 illustrates how the relative probability of the two models (more specifically, their posterior ratio) varies from one end of the shape space to the other, with clearly one-part shapes (left) having higher probability under the one-axis model, and clearly two-part shapes (right) having higher probability under the two-axis model, and intermediate shapes lying in between. (In the Feldman & Singh, 2006 model, variance in the lengths and angles of the "ribs" [correspondences between axis points and shape points, shown in the figure] entail poor fit between the model and shape and thus diminish likelihood. One can see by looking at the ribs in the figure how, for example, variance among the rib lengths increases as the fit between the shape and the model degrades.) Briscoe (2008; see Feldman et al., 2013) found empirical evidence for an exaggerated perceptual division between one-part and two-part shapes at about the point where the posterior ratio shifts from favoring one model to favoring the other.

Figs. 3 and 4 illustrate two other shape features, respectively *straight* vs. *bent* (Fig. 3) and *circular* vs. *elliptical* (Fig. 4). Again, both these shape spaces involve smoothly varying aspects of shape that, in a classical view, would require an arbitrary division between shape categories, but which are more elegantly described as varying probabilistically. Incidentally, both of these shape features (in their classical guises) are invoked in distinctions between geons in RBC (Biederman, 1987).

4. Probabilistic features and the statistical structure of the environment

Viewing features probabilistically solves the three problems of the classical model described above.

4.1. Soft boundaries.

First, and most obviously, probabilistic features avoid the hard boundaries characteristic of classical features, instead allowing smooth variation in likelihood depending on image parameters. While classical features may lump together highly dissimilar objects, or exaggerate small differences among highly similar objects, probabilistic features make categorical distinctions only in accord with the statistical evidence.

4.2. Non-arbitrarinesss.

Moreover, more subtly, probabilistic features also solve the problem of arbitrariness and context insensitivity. One of the main benefits for the probabilistic approach is that it allows us to understand and formalize the connection between the feature lexicon—the set of features used by the observer—and the statistical structure of the world (Barlow, 1961; Shepard, 1994). The world has predictable probabilistic structure: forms, scenes, and spatial relations tend to occur in systematic, reliably recurring ways. A useful feature vocabulary is one that effectively describes the probabilistic terrain.

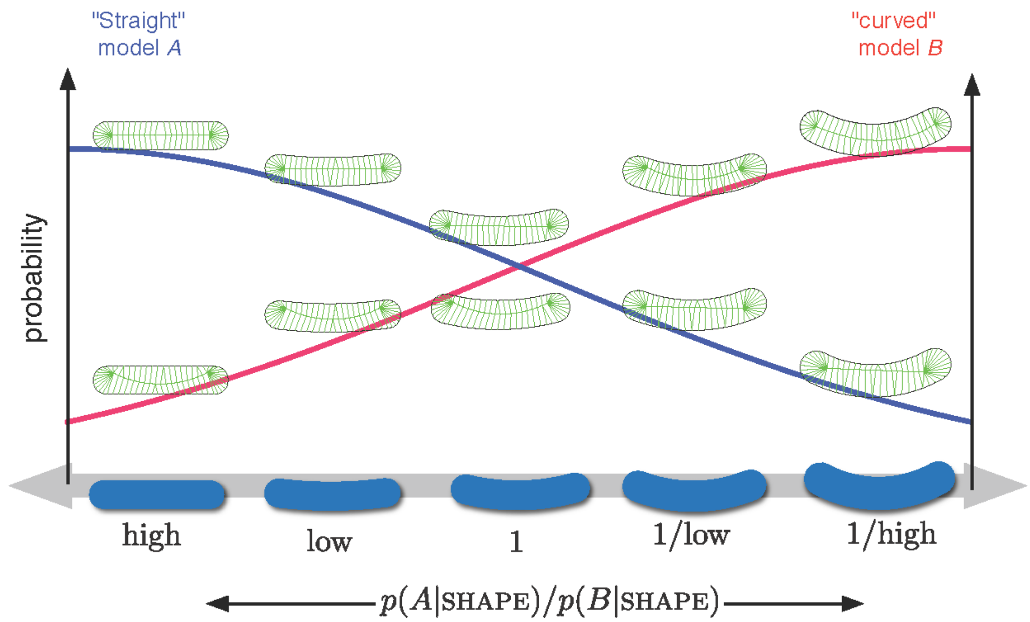

***Figure 3*.** The shape feature *bent* vs. *straight* viewed probabilistically. Straighter shapes (left) are well fit by a straight-axis model and poorly fit by a bent-axis model, while more bent shapes (right) are better fit by the bent-axis model. (Models are shown with ribs; likelihood is diminished by variance in the lengths and directions of the ribs along with several other factors.) The figure illustrates how the relative probability (posterior ratio) of the two models shifts from favoring the straight-axis model on the left to favoring the bent-axis model on the right.

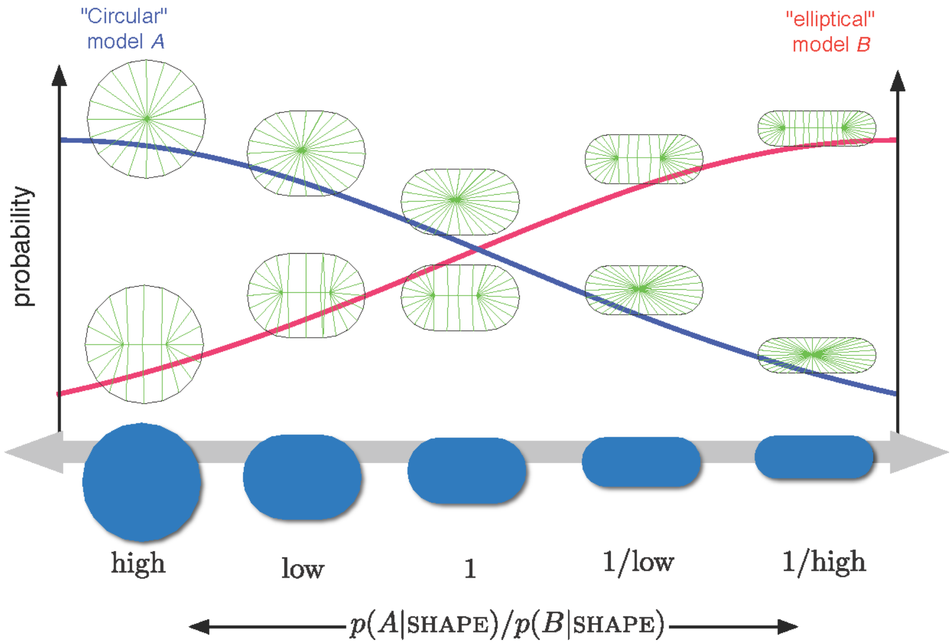

Figure 4. The shape feature *circular* vs. *elliptical* viewed probabilistically. More circular shapes (left) are well fit by a point-axis model and poorly fit by a straight-axis model, while more bent shapes (right) are better fit by the straight-axis model. (Models are shown with ribs; likelihood is diminished by variance in the lengths and directions of the ribs along with several other factors.) The figure illustrates how the relative probability (posterior ratio) of the two models shifts from favoring the point-axis model on the left to favoring the straight-axis model on the right.

One way to characterize the probabilistic structure in the world is by describing its "modes," meaning statistical peaks in the probability distribution that describes the world. A simple example is the mean-plus-error definition of feature $f = \mu + e$ given above, which defines a mode $p(m|f)$ in the measurement space m. A simple assumption is that image structure contains a set of such modes, each corresponding to a distinct naturally occurring class; in this case the underlying distribution is the union of such modes, called a *mixture distribution* (see McLachlan & Basford, 1988). An effective feature, then, would be one that distinguishes "natural modes" (Richards & Bobick, 1988; Feldman, 2012). Just as a single probabilistic feature separates one modal distribution from another (see again Fig. 1), a set of features is useful when it distinguishes the variety of modes extant in the world from each other. That is, a feature set is meaningful when it "carves nature at its joints"—and the probabilistic formulation allows us to specify where the joints are. Probabilistic features viewed this way are both non-arbitrary and context-dependent.

Probabilistic features are non-arbitrary because their utility depends on the statistical structure of the world they are used to describe, and a model of this statistical structure is part of the theory supporting them. Classical features, by contrast, are defined *ex nihilo*; their definitions need not in any way relate to the world. A classical definition of *large*/*small* might adopt an arbitrary size cutoff; a probabilistic definition hinges on modal size categories in the world, and thus would be different for spoons (one mode about 10cm, the other about 12cm, say) vs. cars (one mode about 4m, the other about 5m).

4.3. Context-sensitivity.

Similarly, probabilistic features are potentially sensitive to context, because the nature of the modes to which they are attuned can change subject to the probabilistic structure of the world (that is, the joint probability distribution $p(m_1, m_2, ...)$ of all image measurements). A feature may usefully distinguish modes in one context (i.e. conditioned on the value of another feature) but not in another (just as *has a six-cylinder engine* makes a useful distinction among cars but not among trees). Fig. 5 illustrates a simple joint probability distribution (that is, a model of a world)—a mixture of three modes—in which one feature f_1 distinguishes modes for one value of another feature f_2, but not for the other value of f_2—an admittedly simplistic but useful illustration of context-sensitivity.

Gestalt perceptual features, like proximity, good continuation, and closure, are infamous for their vague definitions. The probabilistic formulation suggests that these features are difficult to define because they mean different things in different contexts; a rich probabilistic description of the world is required to specify exactly what they mean in the diverse situations in which they are used. Creating such generative models is, of course, a substantial scientific challenge which has not yet been met in many cases.

5. Conclusion

Perceptual features are involved in virtually all aspects of vision science, but are still treated in a variety of divergent ways. Behavioral experiments still often use features defined by intuitively simple criteria. At the same time, an enormous neuroscientific literature has established sophisticated feature concepts based on the response properties of cells in visual cortical areas. Early in the processing stream, these include such well-established properties as orientation, motion, and stereoscopic disparity. Later in the stream, these include increasingly non-local properties such as contour curvature (Pasupathy & Connor, 2002), medial axis structure (Hung, Carlson, & Connor, 2012; Lescroart & Biederman, 2012), aspects of 3D shape (Yamane, Carlson, Bowman, Wang, & Connor, 2008), and other less easily verbalized aspects of global shape (Op de Beeck, Wagemans, & Vogels, 2001; David, Hayden, & Gallant, 2006; Cadieu et al., 2007). An important common theme to many modern proposals is that the visual system's choice of features is in some way optimized to the statistical structure of the visual world (Field, 1987; Olshausen, 2003; Geisler, Najemnik, & Ing, 2009). Indeed, there is a growing consensus that the underlying neural code is inherently probabilistic (Rieke, Warland, de Ruyter van Steveninck, & Bialek, 1996; Yang & Shadlen, 2007). However, a fully developed probabilistic model of visual features, in particular one that extends beyond early representations to incorporate non-local features such as form, shape, and spatial relations, does not yet exist. Such a model must be considered one of the main goals of the next decade of research in the visual sciences.

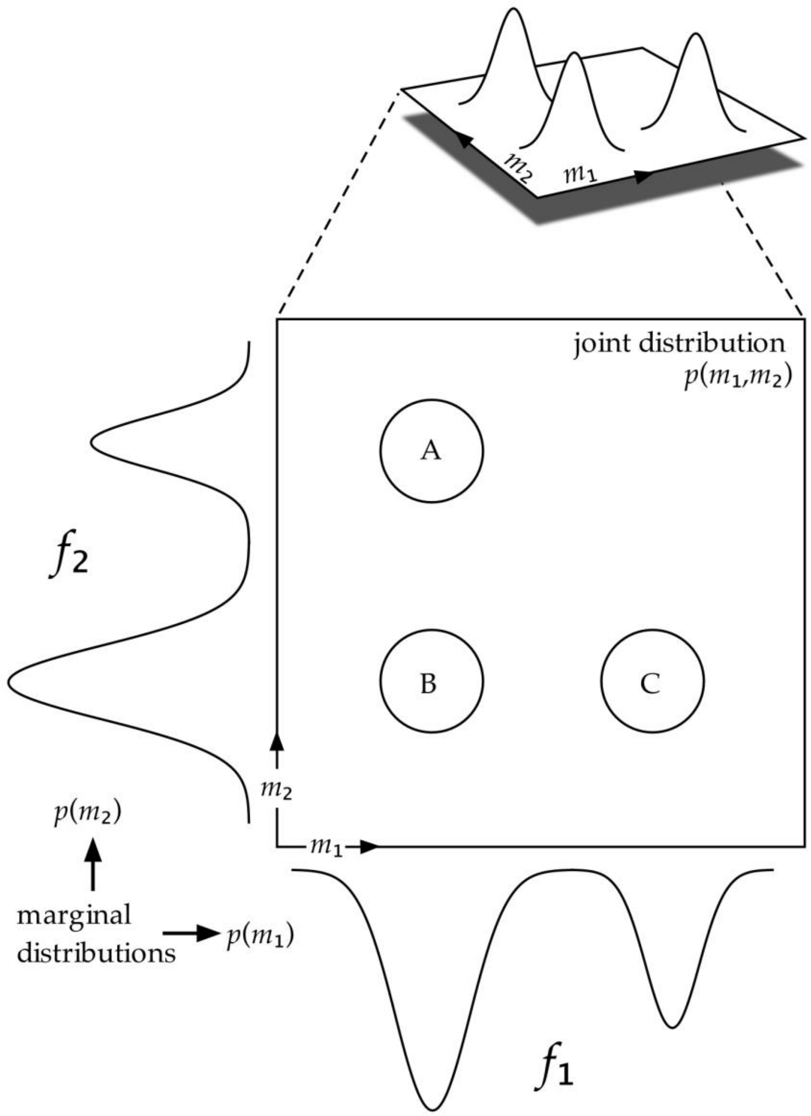

Figure 5. Context-sensitivity in probabilistic features. Because of the shape of the joint distribution $p(m_1, m_2)$ (shown in inset and as contour plot in main figure), feature f_2 is well-defined for one value of f_1 (where it distinguishes mode A from mode B) but not for the other value of f_1, which has only one mode (C).

6. References

Aitkin, D., & Feldman, J. (2013). *Discretization of continuous features by human learners.* (Manuscript under review)

Amir, O., Biederman, I., & Hayworth, K. J. (2012). Sensitivity to nonaccidental properties across various shape dimensions. *Vision Research*, *62*, 35–43.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409-429.

Anderson, J. R., & Betz, J. (2001). A hybrid model of categorization. *Psychonomic Bulletin & Review*, *8*(4), 629–647.

Angelucci, A., & Bullier, J. (2003). Reaching beyond the classical receptive field of V1 neurons: horizontal or feedback axons? *Journal of Physiology Paris*, *97*(2–3), 141–154.

Ashby, F. G., & Alfonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology*, *39*, 216–233.

Ashby, F. G., Prinzmetal, W., Ivry, R., & Maddox, W. T. (1996). A formal theory of feature binding in object perception. *Psychological Review*, *103*, 165–192.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith (Ed.), (pp. 217–234). Cambridge: M.I.T. Press.

Biederman, I. (1987). Recognition by components: a theory of human image understanding. *Psychological Review*, *94*, 115–147.

Biederman, I., & Shiffrar, M. (1987). Sexing day-old chicks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 640-645.

Binford, T. (1981). Inferring surfaces from images. *Artificial Intelligence*, *17*, 205–244.

Blair, M., & Homa, D. L. (2005). Integrating novel dimensions to eliminate category exceptions: when more is less. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*(2), 258–271.

Briscoe, E. (2008). *Shape skeletons and shape similarity*. Unpublished doctoral dissertation, Rutgers University.

Briscoe, E., & Feldman, J. (2011). Conceptual complexity and the bias/variance tradeoff. *Cognition*, *118*, 2–16.

Cadieu, C., Kouh, M., Pasupathy, A., Connor, C. E., Riesenhuber, M., & Poggio, T. (2007). A model of V4 shape selectivity and invariance. *Journal of Neurophysiology*, *98*, 1733–1750.

Craft, E., Schutze, H., Niebur, E., & von der Heydt, R. (2007). A neural model of figure-ground organization. *Journal of Neurophysiology*, *97*(6), 4310–4326.

David, S. V., Hayden, B. Y., & Gallant, J. L. (2006). Spectral receptive field properties explain shape selectivity in area V4. *Journal of Neurophysiology*, *96*, 3492–3505.

De Baene, W., Ons, B., Wagemans, J., & Vogels, R. (2008). Effects of category learning on the stimulus selectivity of macaque inferior temporal neurons. *Learning & Memory*, *15*, 717–727.

de Winter, J., & Wagemans, J. (2006). Segmentation of object outlines into parts: A large-scale integrative study. *Cognition*, *99*(3), 275–325.

Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision*, *2*(4), 324–353.

Feldman, J. (1995). Perceptual models of small dot clusters. In I. J. Cox, P. Hansen, & B. Julesz (Eds.), *Partitioning data sets* (pp. 331–357). (DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 19)

Feldman, J. (1997). Curvilinearity, covariance, and regularity in perceptual groups. *Vision Research*, *37*(20), 2835–2848.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, *407*, 630–633.

Feldman, J. (2007). Formation of visual "objects" in the early computation of spatial relations. *Perception & Psychophysics*, *69*(5), 816–827.

Feldman, J. (2009). Bayes and the simplicity principle in perception. *Psychological Review*, *116*(4), 875–887.

Feldman, J. (2012). Symbolic representation of probabilistic worlds. *Cognition*, *123*, 61–83.

Feldman, J. (this volume). Bayesian models of perceptual organization. In J. Wagemans (Ed.), *Handbook of perceptual organization.*

Feldman, J., & Singh, M. (2005). Information along contours and object boundaries. *Psychological Review*, *112*(1), 243–252.

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Science*, *103*(47), 18014–18019.

Feldman, J., Singh, M., Briscoe, E., Froyen, V., Kim, S., & Wilder, J. D. (2013). An integrated Bayesian approach to shape representation and perceptual organization. In S. Dickinson & Z. Pizlo (Eds.), *Shape perception in human and computer vision: an interdisciplinary perspective.* Springer.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J. Optical Society of America A*, *4*(12), 2379–2394.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field". *Vision Research*, *33*(2), 173–193.

Fitzpatrick, D. (2000). Seeing beyond the receptive field in primary visual cortex. *Current Opinion in Neurobiology*, *10*, 438–443.

Geisler, W. S., Najemnik, J., & Ing, A. D. (2009). Optimal stimulus encoders for natural tasks. *Journal of Vision*, *9*(13), 1–16.

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, *41*, 711–724.

Geisler, W. S., & Super, B. J. (2000). Perceptual organization of two-dimensional patterns. *Psychological Review*, *107*(4), 677–708.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704.

Gilchrist, A. L. (1977). Perceived lightness depends on perceived spatial arrangement. *Science*, *195*, 185–187.

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the non independance of features in similarity judgments. *Cognitive Psychology*, *23*, 222–262.

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology*, *130*(1), 116–139.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32*(1), 108–154.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Harnad, S. (1987). *Categorical perception: the groundwork of cognition*. Cambridge: Cambridge University Press.

Harrison, S., & Feldman, J. (2009). Influence of shape and medial axis structure on texture perception. *Journal of Vision*, *9*(6), 1–21.

Hess, R. F., May, K. A., & Dumoulin, S. O. (this volume). Contour integration: Psychophysical, neurophysiological and computational perspectives. In J. Wagemans (Ed.), *Handbook of perceptual organization.*

Hung, C. C., Carlson, E. T., & Connor, C. E. (2012). Medial axis shape coding in macaque inferotemporal cortex. *Neuron*, *74*(6), 1099–1113.

Jepson, A., & Richards, W. A. (1992). What makes a good feature? In L. Harris & M. Jenkin (Eds.), *Spatial vision in humans and robots* (pp. 89–125). Cambridge: Cambridge University Press. Kanizsa, G., & Gerbino, W. (1976). Convexity and symmetry in figure-ground organization. In M. Henle (Ed.), *Vision and artifact.* New York: Springer.

Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, *23*, 141–221.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271—304.

Kim, S.-H., & Feldman, J. (2009). Globally inconsistent figure/ground relations induced by a negative part. *Journal of Vision*, *9*(10), 1–13.

Knill, D. C., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference*. Cambridge: Cambridge University Press.

Koenderink, J. J. (1993). What is a "feature"? *Journal of Intelligent Systems*, *3*(1), 49–82.

Kogo, N., & van Ee, R. (this volume). Neural mechanisms of figure-ground organization: Border-ownership, competition and perceptual switching. In J. Wagemans (Ed.), *Handbook of perceptual organization.*

Lee, M. D., & Navarro, D. J. (2002). Extending the ALCOVE model of category learning to featural stimulus domains. *Psychonomic Bullein & Review*, *9*(1), 43–58.

Lescroart, M. D., & Biederman, I. (2012). Cortical representation of medial axis structure. *Cerebral Cortex*.

Lowe, D. G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, *31*, 355–395.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110.

McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: inference and applications to clustering*. New York: Marcel Dekker.

Medin, D. L., & Schaffer, M. M. (1978). Context model of classification learning. *Psychological Review*, *85*, 207–238.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–61.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*(1), 53–79.

Olshausen, B. (2003). Principles of image representation in visual cortex. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (p. 1603-1615). Cambridge: M.I.T. Press.

Op de Beeck, H., Wagemans, J., & Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nature Neuroscience*, *4*(12), 1244–1252.

Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience*(12), 1332–1338.

Peterson, M. (this volume). Low-level and high-level contributions to figure-ground organization. In J. Wagemans (Ed.), *Handbook of perceptual organization.*

Pizlo, Z., Salach-Golyska, M., & Rosenfeld, A. (1997). Curve detection in a noisy image. *Vision Research*, *37*(9), 1217–1241.

Poggio, G. F., & Poggio, T. (1984). The analysis of stereopsis. *Annual reviews of neuroscience*, *7*, 379–412.

Pomerantz, J. R., & Pristach, E. A. (1989). Emergent features, attention, and perceptual glue in visual form perception. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(4), 635–649.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3), 353–363.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, *3*, 382–407. (Pages 401-402 missing. Torn from book)

Ren, X., Fowlkes, C. C., & Malik, J. (2008). Learning probabilistic models for contour completion in natural images. *International Journal of Computer Vision*, *77*, 47–63.

Richards, W. A., & Bobick, A. (1988). Playing twenty questions with nature. In Z. Pylyshyn (Ed.), *Computational processes in human vision: An interdisciplinary perspective* (pp. 3–26). Norwood, NJ: Ablex Publishing Corporation.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1996). *Spikes: exploring the neural code*. Cambridge: M.I.T. Press.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*, 1019–1025.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*, 328–350.

Rozenholtz, R. (this volume). Texture perception. In J. Wagemans (Ed.), *Handbook of perceptual organization.*

Sanocki, T. (1999). Constructing structural descriptions. *Visual Cognition*, *6*(3/4), 299–318.

Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and brain Sciences*, *21*, 1–54.

Shepard, R. N. (1994). Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin & Review*, *1*(1), 2–28.

Singh, M., & Hoffman, D. D. (2001). Part-based representations of visual shape and implications for visual cognition. In T. Shipley & P. Kellman (Eds.), *From fragments to objects: segmentation and grouping in vision, advances in psychology, vol. 130* (pp. 401–459). New York: Elsevier.

Smith, E., & Medin, D. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Stilp, C. E., Rogers, T. T., & Kluender, K. R. (2010). Rapid efficient coding of correlated complex acoustic properties. *Proceedings of the National Academy of Sciences*, *107*(50), 21914–21919.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Treisman, A., & Paterson, R. (1984). Emergent features, attention, and object perception. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(1), 12–31.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: M.I.T. Press.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature neuroscience*, *5*(7), 682–687.

Uttal, W. R., Bunnell, L. M., & Corwin, S. (1970). On the detectability of straight lines in visual noise: an extension of French's paradigm into the millisecond domain. *Perception & Psychophysics*, *8*(6), 385–388.

Vogels, R., Biederman, I., Bar, M., & Lorincz, A. (2001). Inferior temporal neurons show greater sensitivity to nonaccidental than to metric shape differences. *Journal of Cognitive Neuroscience*, *13*(4), 444–453.

Wagemans, J. (1992). Perceptual use of non-accidental properties. *Canadian Journal of Psychology*, *46*(2), 236–279.

Wagemans, J. (1993). Skewed symmetry: a nonaccidental property used to perceive visual forms. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(2), 364–380.

Wagemans, J., van Gool, L., Swinnen, V., & van Horebeek, J. (1993). Higher-order structure in regularity detection. *Vision Research*, *33*(8), 1067–1088.

Wilder, J., Feldman, J., & Singh, M. (2011). Superordinate shape classification using natural shape statistics. *Cognition*, *119*, 325–340.

Yamane, Y., Carlson, E. T., Bowman, K. C., Wang, Z., & Connor, C. E. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nature Neuroscience*, *11*(11), 1352–1360.

Yang, T., & Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature*, *447*, 1075–1082.

Zhang, N., & von der Heydt, R. (2010). Analysis of the context integration mechanisms underlying figure-ground organization in the visual cortex. *Journal of Neuroscience*, *30*(19), 6482–6496.