Grand challenges in perception science: modeling the future

Philippe G. Schyns*

University of Glasgow, Glasgow, UK
Correspondence: p.schyns@psy.gla.ac.uk

Perception science is one of the oldest research domains of psychology, by tradition interdisciplinary. Over the years, perception has evolved to provide what is arguably the most mature test-bed for sophisticated theories. The main reasons are that the research generates complementary data sets from human and animal cognition, psychophysics, electrophysiology, neuroscience and brain imaging, and the strong tradition of perception researchers to interact with communities from statistical modeling, computer vision, philosophy of mind and the arts. Such diversity of complementary perspectives has enriched theories of perception, possibly more so than in any other domain of Psychology.

Nowadays, mechanisms of perception can be studied in their true complexity, from their early unfolding during development, maturation in adulthood and decline with age. The research platform is sufficiently mature to start specifying a computational theory (*what* is the visual information processed?), infer its abstract mechanisms (i.e. its algorithms, *how* is information transformed?) and measure with millisecond precision their neural implementation with brain measurements. Advances in functional neuroimaging provide dimensions of analysis (*where* and *when* is information processed?) that ideally complement the experimental rigor of psychophysics. For example, the researcher can use Magneto-Encephalographic (MEG) activity to reconstruct the oscillatory cortical networks associated with perception, isolate their functional nodes, and further localize network nodes in the brain with fMRI. Then, selective application of TMS can examine the contribution of each node to the overall architecture. Characterization of such information processing networks can be extended across age groups (to examine effects of development, maturation and decline), in the normal vs. impaired brain (to confirm the functional role of brain regions) and across species (to refine specificity of

e.g. primate vs. non-primate perception mechanisms, from complex to simpler brain organizations).

In the context of such a sophisticated research platform, what are, then, the Grand Challenges of perception science? The most obvious challenge is not grand, but it is a challenge nevertheless. A cursory glance at the methods section of publications in visual neuroscience, electrophysiology, neuroimaging, psychophysics and visual cognition reveals that we are far from the ideal integration of methods highlighted in the previous paragraph, implying that the research platform is not always fully exploited. The challenge is to integrate different research communities with their own experimental and methodological traditions around the same theme of perception. But another glance this time at the authors list also reveals that the research culture is at an interesting juncture, progressively evolving from individual "Master Craftsman" to interdisciplinary "Team of Experts." As this transition happens, and as training and research funding straddle disciplinary boundaries, imports of methods will optimize usage of the platform and enhance research across the field. Though real, the challenges associated with this transition are generic to any interdisciplinary venture. Sections of Biology addressed them a while back, Physics even before, now it is our turn.

Perhaps surprisingly, the Grand Challenges of perception science arise from its relative maturity. More than a century of perception research produced an articulated landscape of the main fundamental problems, a good grasp of their performance envelopes, and the development of innovative methodologies. The Grand Challenge is now to integrate behavioral and brain data using the common denominator of formal models of perceptual tasks. That is, models that formally specify input information, model how information is transformed in the brain and predict how parametric variations in the input information modify

brain processes and behavioral variables. The Grand Challenge of integrative models comprises two intertwined aspects: (a) to formalize visual information in complex perceptual tasks and (b) to model how the processing of this information in the brain predicts behavioral performance. I sketch the Grand Challenge below.

FORMALIZE VISUAL INFORMATION

Psychophysics is attractive because it typically simplifies visual information to the few dimensions of simple perceptual tasks. For example, a model observer discriminating left- from right-orientated Gabor patches near contrast threshold will require information about input orientation, contrast and spatial frequency. A parametric design involving these three dimensions would produce a thorough understanding of the performance envelope of the visual system resolving the task. The relative contribution of each dimension to perceptual decisions can be modeled within the framework of statistical decision theory and the domains of application in Psychophysics are plentiful.

Although much has been learned from this overall approach, it is questionable whether the performance of the visual system in simple tasks would generalize to the processing of more structured and objectively ambiguous information above perceptual threshold. The real visual world comprises information that is temporally and spatially structured (e.g. a street scene comprises buildings, moving cars and human beings, which in turn comprise subcomponents, which themselves comprise subcomponents and so forth), objectively ambiguous (e.g. two different 3D objects can produce the same 2D image) and often incomplete due to occlusions (see **Figure 1**). The visual system handles this complexity at an impressive rate of at least eight scenes per second.

Applying the fruitful tradition of psychophysics to complex information is the Grand Challenge. Unfortunately, unlike oriented Gabor patches, it is difficult to
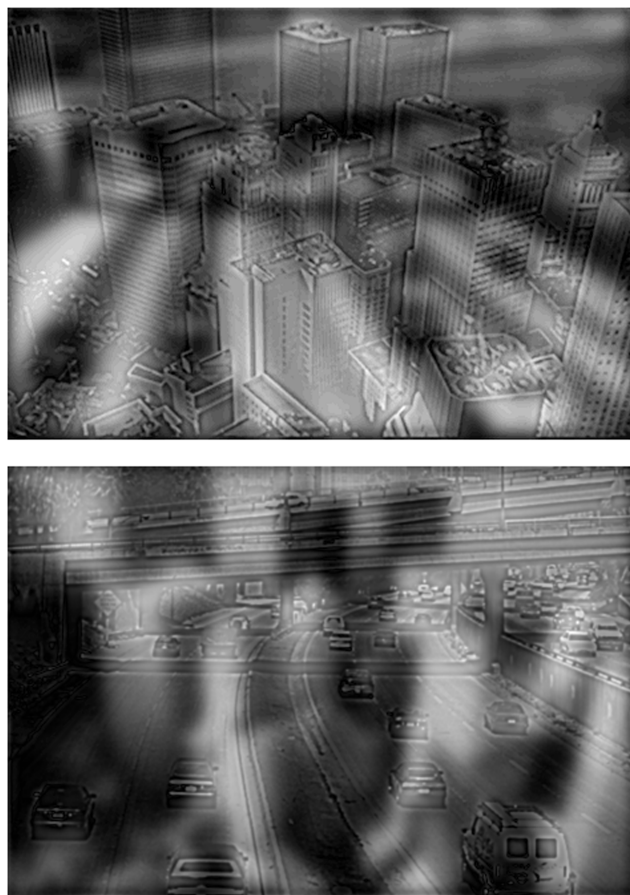
FIGURE 1 | Illustration of complex visual information and dimensionality reduction (adapted from Schyns and Oliva, 1994). The top picture illustrates the overlapped buildings making up a city. The bottom picture illustrates the road, bridges and cars characterizing a highway. Squint or step away from the picture to perform a different dimensionality reduction of the input and thereby reverse your perceptions.

conjecture the reduced space of structured information from which to parametrically generate complex, natural visual displays – apart from natural repeated patterns such as textures. Any single object is overloaded with a deceptively large number of different perceptual tasks, tapping into different dimensions of an infinite dimensional reality. For example, a given stimulus will be "John," "male," "happy," "trustworthy," "about 40 years of age," "handsome," "shaved," "with blue eyes and brown hair," and so forth (in addition to be seen at a 3/4 view, lighted from above) and a given scene will be "an outdoor scene," "a city," "with high density of skyscrapers," "New York," "with the Chrysler building," "with Art Deco silver tiled roof," in addition to "seen at sunset," "through a fine mist." These perceptual tasks (called "categorizations" in visual cognition) use flexibly different

dimensions of the infinite dimensional world – so called "diagnostic information." So, modeling the information of a complex visual scene is also modeling the perceptual tasks that it affords.

A modeler would acknowledge the network organization of knowledge and build a hierarchy of embedded models starting from simple cues to progressively more complex structures – e.g. starting with contours and shading at different scales and orientations, moving up to structured features such as hair, eyes, nose, mouth, chin, at some higher levels, faces at even higher levels, which together with desk, chair, books, desk-lamp produce an office scene at much higher levels. Note that each network nodes arises from the existence of a perceptual task that people can perform – e.g. discriminate oriented contours, evaluate the length or shape of a nose, identify a face and recognize an office scene.

But there is no evidence that hierarchically organized perceptual tasks (e.g. identifying a scene vs. the objects within a scene) use hierarchically organized models. Recognition studies suggest effective bypass of the hierarchy with reductions that directly access the highest, scene level–e.g. the information comprised in low-passed versions of scenes (e.g. examine **Figure 1** while squinting, or stepping away from the display). Perceptual learning also suggests that a hierarchy of models is unstable, with at least reorganization via non-linear chunking. Consequently, the information of a perceptual task is often lower in resolution than that assumed of a full hierarchical reconstruction.

This, in a few paragraphs, is a sketch of the Grand Challenge of formalizing complex visual information. Reality is infinite dimensional. A perceptual task, or a model, is a massive dimensionality reduction, but which one?

FORMALIZE INFORMATION PROCESSING MECHANISMS IN THE BRAIN

Assume that a "good" parametric model of complex information exists for a given perceptual task. Having parametrized the relevant dimensions of input information, we now turn to the recording of brain signals acquired in a parametric design to (a) model how the brain produces the relevant reduction from high-dimensional retinal inputs and (b) model how the brain progressively transforms the reduced information to reach perceptual decision and produce behavior.

Current thinking suggests that the brain comprises different cortical nodes that communicate with each other to process information. Hence, our task is then again to reduce a high-dimensional data space of brain measurements into a model whose functional nodes correlate with (a) the parametric space of the input and (b) behavior. To appreciate the challenge, consider MEG activity. A typical MEG scanner comprises ~250 sensors sampling MEG oscillatory activity on each trial at 1024 Hz from about 1000 ms before and after stimulus onset. If parametric variations of input information correlate with variance in some subspace of the high-dimensional MEG brain measurements, our challenge is to model this variance from the MEG data. For example, we

will locate the physiological sources of the variance, connect them as the nodes of an interactive network and explain how the network nodes represent the parametric variations of input information. With further testing we can also infer and test the functions that each node performs on the information parameters (e.g. gate, encode, compare, memorize) and finally deduce a flow chart of how information mechanistically moves across network nodes, between stimulus onset and behavioral decision.

But this description eschews the serious problem of measurement. Evolution has implemented perception into many neurons, their signaling properties and the intricate patterns of their interactions. Neuronal activity is nonlinear, nonstationary, chaotic, unstable and asynchronous. EEG, MEG and fMRI measure numerous correlates of this activity *en masse*; single, or multi-electrode, multi-site recordings measure correlates *in fine*. Reducing the high-dimensional brain measurements into a model of the perceptual task is not only difficult because of the high dimensionality of the space and the relative sparseness of data points. The reduction should itself be guided by data from neuroscience (e.g. an understanding of the complex communicating properties of neurons and their assemblies) but also by insightful formal thinking (e.g. from dimensionality reduction methods, to the nonlinear dynamics of complex systems). The constructive interplay of novel formalisms and data is particularly relevant to the problem of "what should we measure about the brain?"

In sum, the Grand Challenge for perception science is to integrate the computational, algorithmic and implementation levels of Marr's analysis by developing a framework that uses the best aspect of psychophysics research (its rigorous parameterization of information), generalize it to complex, naturalistic visual information and study information processing on a meaningful reduction of brain data (i.e. a model of information processing).

REFERENCE

Schyns, P. G., and Oliva, A. (1994). From blobs to boundary edges: evidence for time and scale dependent scene recognition. *Psychol. Sci.* 5, 195–200.