Reply

Author’s response: Where to begin?

Lin Chen

Chinese Academy of Sciences, Beijing, China

The last two decades have been a long and arduous journey for the topological approach to perception (Chen, 1982). The global-to-local account is counter-intuitive and appears to conflict with common knowledge about visual perception and the biological organization of the visual system. Thus, it is a welcome breakthrough to elicit so many thoughtful commentaries, which have been either unexpectedly supportive or else critical but also stimulating and constructive. It is also encouraging to see that the commentaries span a broad and interesting range of topics, from visual perception and attention (N. Donnelly, A. Hayes, S. He, J. Todd, & J. Wolfe), through neurobiology (R. Desimone, also S. He), to temporal organization (M. Elliot & E. Pöppel), distributed cognition (Jiajie Zhang), and mathematical modelling (Jun Zhang). However, because of space limitations, I will take an easy route in my response to just address the fundamental question of “Where to begin?”, the main subject of my paper. Moreover, I will, while briefly respond to some of the comments, focus on some of the challenging issues raised in Wolfe’s commentary.

Wolfe’s guided search model (Wolfe, 2003), being a response to problems with Treisman’s influential “feature integration theory”, may be considered to represent the state of the art of the local-to-global approach. His commentary is focused on the relationship between the topological model and the guidance of visual search. However, since both the topological approach and the guided search model reflect current efforts in searching for the primitives in visual perception, I believe the following expanded discussion will be useful for addressing the fundamental question of “where to begin”. In particular, I hope the discussion will help clarify the empirical evidence for early topological perception and theoretical formulation for the topological global-to-local approach.

Please address all correspondence to: Lin Chen, Key Laboratory of Cognitive Science, Graduate School and Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing100101, China. Email: ChenL@intra.nimh.nih.gov

© 2005 Psychology Press Ltd
http://www.tandf.co.uk/journals/pp/13506285.html DOI:10.1080/13506280444000364
Wolfe agrees that topology might be pre-eminent in the early processing of visual input. As such, he seems to concur with the core idea of the global-to-local approach, that is, early topological perception. But he disagrees that topology may be pre-eminent in the guidance of visual search, and he raises several important issues, which I will address below.

**Guidance vs. where to begin**

Wolfe (2003) argues that the claims about topology and holistic processing are too sweeping, at least as relates to guided visual search. I agree with Wolfe’s concern in the sense that in establishing a single framework, supposedly to be able to cover a great variety of phenomena, great care should be taken to define its range of application. Indeed, quite some space in Part I of my paper is devoted to delimiting the concept of perceptual organization, to which the topological approach is applicable. The very notion of perceptual organization used in the present paper is defined in terms of the grouping of stimulus elements in parts or regions of the field with one another, and the “what goes with what” problem. On the one hand, the application of the topological approach to this fundamental stage of perception allows it to cover a great variety of common conditions in perceptual organization. On the other hand, this means that the topological approach is concerned mainly with the early stages of perceptual processing rather than high-level cognition.

Wolfe’s (2003) model proposes a division between bottom-up, stimulus-driven guidance and top-down, goal-driven guidance. The guided search model emphasizes the role of the top-down activation. However, with respect to the main subject of the paper, “where to begin”, the disagreement concerning top-down guidance seems not to be particularly pertinent. Even though there is evidence showing the role of topological constraints in goal-directed visual search and even higher level cognition (e.g., Jiajie Zhang in his commentary), the basic variables, the nature of the processes involved, and the mechanisms subserving the topological approach all essentially speak to bottom-up, stimulus-driven processing. My response will, therefore, further focus on the relationship of topological perception to stimulus-driven guidance of visual processing and attention.

**Feature salience vs. early topological perception**

The topological approach has two main aspects, a topological structure and a functional hierarchy in form perception. The hierarchy was put forward to clarify the relationship between topological perception and the perception of other geometrical properties. As Desimone (in his commentary) points out, although the global-to-local topological model posits that topological invariance has the greatest and most immediate effects on perceptual organization, the model allows that local features can also be extracted from the global form and utilized in
perceptual organization, albeit with reduced speed. The claim of topological primacy does not, implicitly or explicitly, exclude the role of local features in perceptual organization. Rather, the functional hierarchy implies that local features may also “pop-out” or show salience, and thus may be units of visual representation. However, since pop out effects may not be entirely stimulus driven (e.g., Wolfe, 1996; Yantis, 1996), and may involve a hierarchy of relative salience, they are not always the appropriate source of data to address “where to begin”, or to argue against the prediction of early topological perception.

An illustrative example comes from an early study, shown in Figure 3 of the main paper. Under the condition of very brief T-scope presentation, in a same–different task (rather than in tasks requiring judgements of shapes), performance was significantly better for the ring-disc pair than for the other two pairs, demonstrating early topological discrimination. Nevertheless, under more normal conditions of longer presentation when subjects were asked to judge which pair of stimulus figures looked more similar, subjects reported that, as an intuitive notion of similarity among round shapes would expect, the disc and ring look more similar to each other than the solid disc and the solid triangle or the solid square. This is the opposite of the results showing topological sensitivity with brief T-scope presentation (for more detailed discussion, see, e.g., Part II and Chen, 1990). Such near-threshold, “same–different” tasks, therefore, are more appropriate to use for studying “where to begin” than mere salience demonstrations, such as Wolfe’s Figures 1 and 2, since carefully designed experiments would need to take into considerations the differentiating of discriminability from visibility and of physical (spatial) connectedness from other organizations such as similarity and subjective contour (see, e.g., Part II).

A unified dimension of form properties

Wolfe’s other major concern is that in finding a hierarchy of feature salience in guided search, a topological difference (such as in holes) and a local difference (such as in orientation) are not directly comparable unless they would be matched to a common yardstick, such as the luminance contrast, by a matching method (Nothdurft, 1993). The matching method may be useful in making comparison between variables in different dimensions. However, one of the key points of the topological approach is that all geometrical properties are defined as invariants over transformations. This definition provides a unified language to characterize form properties. Thus, from the perspective of invariance perception, topological properties and other geometrical properties at different levels of stability should be considered to belong to the same dimension, that is, form properties. The empirical evidence (cited in Part II and IV) that the relative salience of different geometrical properties is remarkably consistent with the geometrical stratification, supports this unified manner of characterizing the relation between topological and other geometrical properties. Both the
empirical and theoretical considerations, therefore, lead us to conclude that it is not necessary to compare the salience between topology and other geometries (including orientation contrast) by matching them to a common yardstick, rather, for finding a hierarchy of salience in form perception, it is appropriate to directly compare them in the unified dimension of form invariants.

**Topological primacy in configurational superiority effects**

Empirically, Wolfe questions the configurational (hole) superiority effects over orientation shown in Figure 22 of the main paper, as he considers it to be “a bit surprising since orientation search is typically fast and efficient” (this issue, p. 676). The configurational superiority effects (e.g., Pomerantz, Sager, & Stoever, 1977), I think, just serve as a paradigm suitable for comparing directly relative salience of form properties in the unified dimension of form properties, including topological properties and orientation. Not to mention all the systematical results (e.g., Pomerantz et al., 1977; and the present paper), which demonstrated consistently configurational superiority effects over orientation, I will just note one of their basic results, shown by “the triangle–arrow pair” in Figure 1 here (adapted from Pomerantz et al., 1977), which controls well for orientation confound. It might be more interesting to argue (than for alternative explanations based on local features, such as symmetry, and size) that even though the RT of discrimination based on hole is the shortest among other tasks, quite a few hundred milliseconds needed for hole discrimination would be still too long to indicate early perception.

To assess this issue, we need a baseline representing the easiest discrimination task for comparison in order to understand the psychological value of the RT with hole discrimination. I found surprisingly that the RT to detect the odd quadrant in Figure 1A (hole vs. no hole) was equal to that for Figure 1B, despite the fact that Figure 1B is likely to provide a baseline as the easiest discrimination task. Is there any intrinsic property shared by the two apparently different displays, which brings about the same quickest performance? The performance with Figure 1B could be based on the detection of “something” (a large solid square) versus “nothing” (in the odd quadrant). In topology, “something or nothing” is a topological invariant, because topological transformation can neither create a new thing nor destroy an existing thing. On this view, the RTs are the same because the two discriminations take place at the same primitive level of topological perception. It would not be a surprise that topological discrimination is faster and more efficient than orientation discrimination, as topological discrimination could be performed as fastest and most efficiently as the easiest baseline task.

Running the risk of redundancy, it may be worth emphasizing again the Gestalt principle demonstrated by configurational superiority effects, namely that a
pattern is more than a sum of its component parts. However, it is still not rare to see that component units in a display array are taken for granted as units of analysis but emergent configural structures created by the component units are sometimes ignored. The emergent configuration may be manifested by physically discrete but subjectively connected forms (see the concept of tolerance in Part III). To explain the phenomena that structural properties with the same individual subunits ‘could be present, then lost, then recovered’ (in Wolfe’s commentary, p. 679), one does not have to assume that early vision is dissociated from attentional guidance, but rather that emergent global properties are extracted at different levels. In the topological reformulation of Gestalt principle, namely that ‘holistic (or global) registration is prior to local analysis’, the term of ‘prior’ here implies a logical priority and a priority of time dependence of wholes determined by topological constraints. It reminds us that possible emergent configurations (such as subjectively closure vs. open structure) formed by simple component parts, such as bars, intersections, and T-junctions in stimulus arrays (e.g., Figure 3 in Wolfe’s commentary and figures in Wolfe & DiMase, 2003), rather than the single component parts per se, may need more scrutiny.

Donnelly and He both suggest new neuropsychological experiments to test early topological perception. I agree with them that the breakdown of visual processing following brain damage may provide definitive evidence for the
early topological perception. Actually, applying the paradigm of configural superiority effects, we found that the effect of brain damage on line-segment discrimination was not transferred to the perception of holes formed by the same line segments (Chen, 1989). I agree too with them that more experiments and theoretical treatments are required to refine explicitly the influence of topological perception on the perception of other local features. However, the logical priority and a priority of time dependence of topological perception implied by the topological reformulation of “global-to-local”, supported by such neuro-psychological configural superiority effects, may serve as the foundation for the refinement.

I still feel it is useful to recall the history of paradigms of configural superiority effects, which originally aimed to search for psychological evidence supporting line-based shape coding. The results obtained by the paradigms, however, lead to an important but, unfortunately, long ignored claim: “The sloped line detectors, of the variety discovered in the mammalian cortex by Hubel and Wiesel, are not the primitive of the human visual pattern recognition system” (Pomerantz, 1978). Now it is exciting to see that a similar claim is also highlighted in the commentary by Desimone, from a neuroscience perspective.

**Topological primacy in multiple object tracking**

I thank Wolfe for his attention to one our recent experiment using multiple object tracking (MOT) paradigm. In fact, this experiment was initiated by his stimulating suggestion. One of our three starting questions for the topological approach is “how to define the concept of perceptual object precisely and formally”, and one of our highlighted novel predictions is that a perceptual object can be defined as something that keeps its topological structure over time (Part I). That is, the topological approach ties a formal definition of “object” to invariance over topological transformation.

Wolfe challenged the topological definition of object by the following profound analysis (J. M. Wolfe, personal communication). An object is moving along and a hole appears in it. Would this disturb the object’s perceived continuity? Intuitively, just like that as an object is changing its colour, an observer would say “Oh, the object was red and changed to green”, an observer would say “Oh, the object was solid and now has a hole”. However, from the topological definition of an object, it follows that this topological change would disturb the perception of object continuity, while changes of shape and colour would not.

The MOT paradigm (e.g., Pylyshyn & Storm, 1988; vanMarle & Scholl, 2003) is well-suited to test directly this novel prediction. In our MOT experiments (Zhou, Lu, Hu, & Chen, 2004), in order to determine whether attended object representations survive the interruption from shape change during their motion phase, we used moving objects that remained either identical throughout
the motion or changed in shape. We indeed found that attentive tracking processes were impaired if objects changing their topology, such as changing from a solid surface to a surface with a hole. However, it did not matter if they change local features and colours. For example, when S-like figures (adapted from Figure 4 in the main paper) were moving and changing their topology to become a ring, attentive tracking processes were disrupted significantly. By contrast, tracking was preserved when they were moving and changing their local features to become a solid disc (a nontopological change). Though counterintuitive, this MOT test provided strong support that topological constraints may define a perceptual object in the first place.

I will not discuss the further implication of our MOT result here, but address instead Wolfe’s comment that this MOT result may indicate that “the features of early vision and those that guide attention are dissociable”. In our MOT experiments, the manageable shape changes, like abrupt onsets in attentional capture, were not explicitly related to the observer’s perceptual goals. The finding that a change in topology was disruptive of the ability to attentively track is, therefore, likely to be a stimulus-driven low-level effect. I cannot see, therefore, the reason to conclude that this MOT finding underlines the dissociation of early vision from attentional guidance. Rather, this MOT finding just supports the claim that topology may also be pre-eminent in the “stimulus-driven” guidance of visual search, like in early vision.

It is worth pointing out that Pylyshyn (2001) stressed the data-driven nature of operation involved in MOT. As Scholl (2001) said, MOT enjoys the salient properties of our pretheoretic notion of attention, while it is unclear what aspects of MOT support a lower level interpretation; if a significant portion of the operation is truly date driven, then this might be a mechanism which “gets vision off the ground”. Now our findings suggest a topologically data-driven, and low-level interpretation of the MOT phenomena.

Topological primacy in capture attention

In the study of visual attention, to differentiate the stimulus-driven mechanism from the goal-directed mechanism is particularly relevant to the topological approach. Yantis and co-workers (e.g., Yantis, 1996) raised the interesting question of whether attention can be captured in an entirely stimulus-driven fashion. They therefore pursued the important concept of abrupt visual onsets, which are considered to be a category of stimuli that behaves differently than most local features. They found that only the sudden onset of a new stimulus consistently captured attention, while other changes of highly salient features did not. Yantis (e.g., 1996) proposed a new-object explanation for why abrupt visual onsets capture attention, namely that attentional capture occurs because an abrupt onset of a stimulus establishes a new perceptual object representation.
From the topological definition of a perceptual object, a topological change should be treated by the visual system as the presence of a new object. If topology is critical in defining what constitutes a perceptual object, it follows that a topological change (e.g., a solid figure forming a hole) should, like abrupt onsets but unlike local feature changes, capture attention. This counterintuitive prediction was tested by our new experiment (Zhou et al., 2004) studying attentional capture, in which we adopted an “irrelevant feature search” paradigm. The key instruction for the subject’s task was that in order to eliminate top-down, task-relevant effects, subjects were simply asked to report if there was an item different from the rest of the items in a display, rather than being asked to report on the shape of an item. Under such entirely stimulus-driven conditions, it was found that a topological change, for example, a change of an H-like figure (containing no holes) in a placeholder array to a P-like figure (which is topologically different to H-like figure in containing a hole) in a test display, indeed captured attention. Moreover, the attentional capture effect, in terms of uniformly faster and stable (unaffected by display size) RTs, is just the same as that produced by the onset of a S-like figure (topologically equivalent to H-like figure), compared with a no-onset S-like figure. From the perspective of topological analysis, this result is not a surprise, because, as pointed out above, abrupt onsets are a type of topological change as well. On this view, the attentional capture effects of abrupt onsets and forming a hole are the same because the two events share the common nature of creating a new perceptual object by topological change.

As pointed out by Wolfe (1996), different features appear to have different abilities to attract attention from the bottom up, with the abrupt onset and/or the creation of new objects being, perhaps, the most forceful. I hope the most forceful result of topological attentional capture would be convincing for the claim that topology is pre-eminent in the early organization of the visual input in general, and in bottom-up visual search in particular.

Finally, I respond to Wolfe’s comment on “the great divide”. Which idea is more holistic than others seems not to be a well-defined question, as its answer obviously depends on how to define holistic or “global vs. local” relations. Particularly, as pointed out in Part V, the fact that there are a variety of terms related to the concept of “global vs. local”, such as “large-scale vs. small-scale”, “compounds vs. components”, “whole vs. parts”, and “lower vs. higher spatial frequency”, indicates the lack of a clear understanding of what “global vs. local” means. Instead of discussing which theory is more holistic, it may be more useful to provide a formally precise and psychologically valid treatment on the definition of “global vs. local”.

The functional hierarchy has been established as a formal and systematical definition of “global vs. local” relations: A property is considered more global (or stable) the more general the transformation group is, under which this
property remains invariant; relative to geometrical transformations, the topological transformation is the most general and hence topological properties are the most global. Such understanding of “global vs. local” relations turns the question of “where to begin” into the more precise question of whether visual processing starts from topology (connectivity and holes) at the most global level or it starts from Euclidean geometry (distance) at the most local level. As soon as the ambiguity of understandings of “global vs. local” is removed, “the great divide” between local-to-global and global-to-local processes cannot be simply addressed by vague intuitions such as that they may occur in parallel manner, but must be answered by experimental evidence. “The great divide” does not in any sense, as Wolfe (in his commentary) criticizes, generally imply that there is “a single path from input to image understanding” (p. 681), and reject that “[a] number of parallel streams of information appear to contribute” (p. 681). Rather, it leads to the specific claim that, regarding “where to begin”, visual processing starts from the most global level of topological organization (based on physical connectivity) rather than from the most local level of feature analysis of Euclidean geometry (based on distance).

Now, I turn to the “spatial envelope” model (Oliva & Torralba, 2001), which Wolfe thinks is an example of an extreme holistic approach. The “spatial envelope” model is, as considered by their authors, a model at the medium level of visual processing specifically designed address “scene” perception. The empirical support to the model came from studies with strong goal-directed instructions—the subjects were explicitly told not to use criteria related to the objects in the scene. The major criteria concept represented by the model is “naturalness” in comparison to the concept of “artificial” scene. From an evolutionary perspective, the visual system has a much longer history than human-made artifacts. Not to mention the problem of other species, even for the human visual system it is hard to conceive how visual function would start from perceiving naturalness in comparison with artificial scene. A major mechanism underlying the degree of naturalness is the distribution of straight horizontal and vertical lines, which are still local features but represented by spectrums in the model. The model emphasizes the bypassing of processing of individual objects. But, to make computing the input possible, the model just proposes that a scene is processed like an individual object at a large spatial scale. In this sense, it is unlikely to bypass completely the processing of objects. There is no doubt that it is interesting to use statistic properties of natural images for computer vision. It is also possible that such statistic properties are used by human visual system. However, the above analyses all suggest that the spatial envelope model does not address “where to begin”, being a model at a large spatial scale but not a model with a holistic nature in the sense of global invariance perception.

I would like to take the opportunity to comment briefly on the relationship between the topological approach and models based on spatial frequency analysis, which are related to the “spatial envelope” model. My basic point is that
Fourier analysis provides no mathematical structure for describing geometrical invariants, particularly topological invariants. It is simply the case that there is no significant difference in spatial frequency components between, for example, the triangle and arrow pair in Figure 1 here, and the S-like figure and the O-like figure in Figure 4 of the main paper (and see, e.g., Chen, Zhang, & Srinivasan, 2003, which represent the most stable structural difference. Moreover, the finding of visual illusions without low spatial frequencies by Carlson, Moeller, and Anderson (1984) using so-called “balanced dots” free of low spatial frequencies, indicated that geometrical illusions are not primarily a consequence of low pass filtering. Unfortunately, the more general implication of the finding has been ignored in the literature. Their findings, I believe, indicated proximity organization without low spatial frequencies: In their figures made up of “balanced dots”, the lack of low spatial frequencies had no impact on organization based on proximity. As emphasized in Part III, from the point of view of global (tolerance) topological perception, proximity is the most basic organizational factor. Thus, this observation suggests that low spatial frequencies do not look likely to be critical to perceptual organization in general.

REFERENCES

He, S. (this issue). Beyond “local to global” and “global to local”. Visual Cognition, 12, 671–674.


