

FVS 2.0: A Unifying Framework for Understanding the Factors of Visual-Attentional Processing

Liqiang Huang

Department of Psychology, The Chinese University of Hong Kong

Across a broad range of stimulus types and tasks (16 stimulus types \times 26 tasks, 1,744 observers in total), the present study employed an individual-item differences analysis to extract the factors of visual-attentional processing. Three orthogonal factors were identified and they can be summarized as an FVS 2.0 framework: *featural*, *visual*, and *spatial strengths*. Apart from one exception (low-level motion), the FVS 2.0 framework accounts for the vast majority (95.4%) of the variances in the 25 tasks. Therefore, the three straightforward factors provide a unifying framework for understanding the relationship between stimulus types as well as those between tasks. Combining these and other related results, the role of preattentive features seems to be rather different from the traditional view: visual features are general purpose, exclusive, innate, constancy based, and keyword like. A general-purpose, exclusive, innate, constancy-based and keyword-like (GEICK) conjecture is proposed which suggests that the features are conscious-level keywords generated by the specific brain area of V4 and/or IT and then used by all other brain areas.

Keywords: visual features, visual perception, visual attention, visual working memory

The concept of *visual feature* plays a fundamentally important role in the studies of visual perception, attention, and working memory. Previous studies in experimental psychology found that these visual features, such as colors and shapes, are extracted in parallel and can be efficiently used to guide attention (e.g., Treisman & Gelade, 1980; Treisman & Gormican, 1988; Wolfe, 1994, 1998a, 1998b; Wolfe et al., 1989). For example, one can easily find a blue target that is embedded in many items in other colors. These visual features have also been discussed in neurophysiology (e.g., Hubel & Wiesel, 1962) as well as cognitive neuroscience (e.g., Livingstone & Hubel, 1988).

Factors of Visual-Attentional Processing

In the studies of visual-attentional processing in perception, attention, and working memory, visual features (e.g., colors, shapes) and other types of stimulus information (e.g., the binding of multiple features, or the spatial arrangements of multiple features) have been used as the building blocks for the visual stimulus materials of experiments. A fundamentally important question on the *factors of visual-attentional processing* has largely been neglected in these studies. Specifically, is

the processing of visual stimulus information affected by different factors in different tasks? If so, what are these factors?

The notion of “factors behind tasks” reminds us of the individual differences analysis. Individual differences analysis has often been used on topics such as executive function (e.g., Miyake et al., 2000). In visual science, researchers have also used this approach to explore different low-level vision tasks and usually found little or no common factors behind them (Bosten et al, 2015; Cappe et al, 2014; Cretenoud et al, 2019; Goodbourn et al, 2012; Grzeczowski et al, 2017; Shaqiri et al, 2019). Individual differences analysis has also occasionally been applied to study visual-attentional processing as well (e.g., see Haberman et al., 2015; Huang et al. 2012; Rosenberg et al, 2016, 2017).

Precisely speaking, these individual differences mean “differences between individual participants.” Hereinafter, these studies will be addressed as individual-participant differences analysis, so that they can be distinguished from individual-item differences analysis which will be introduced below.

The individual-participant differences analysis shows the relations between a person’s behaviors in multiple tasks. For example, if an observer performs well on a visual search task, will this observer also tend to perform well on a change detection task? This individual-participant differences analysis extracts the factors affecting the different observers’ performances on these tasks.

On the other hand, we can use the same mathematical tools (see below) to perform an individual-item differences analysis which shows the relations between the ways how an item is processed in multiple tasks. Briefly speaking, individual-item differences analysis is what happens when the roles of “participant” and “stimulus item” are *flipped* in the individual-participant differences analysis. For example, if a type of stimulus is processed efficiently in a visual search task, will it also tend to be processed efficiently in a change detection task? In other words, this individual-item differences analysis extracts the factors affecting the processing of different types of visual stimuli.

Liqiang Huang  <https://orcid.org/0000-0002-0007-9375>

The work described in this article was supported by two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (CUHK 14617615, CUHK 14128016). The writing of this article was substantially helped by the Humanities and Social Sciences Prestigious Fellowship (CUHK 34000817), also from the Research Grants Council. A part of the present data set was previously published in Huang (2015a). Data and script for data analysis are available on the Open Science Framework project page <https://osf.io/9eafp/> (Huang, 2021c).

Correspondence concerning this article should be addressed to Liqiang Huang, Department of Psychology, The Chinese University of Hong Kong, Hong Kong SAR, China. Email: lquang@psy.cuhk.edu.hk

The present study is *not* an individual-participant differences analysis, but an individual-item differences analysis. The difference between these two types will be elaborated further below.

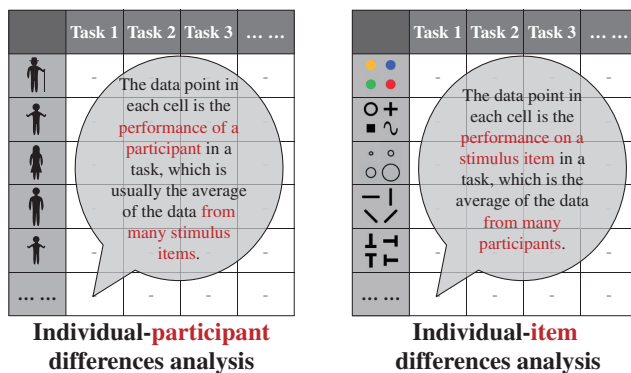
Individual-Item Differences Analysis Versus Individual-Participant Differences Analysis

The differences between individual-item differences analysis and individual-participant differences analysis are shown in Figure 1. Traditional individual-participant differences analysis can extract the factors affecting the different human individuals' performances on various tasks. This information is very useful for educational/developmental psychologists who try to understand the relations between different aspects of a person.

On the other hand, the individual-item differences analysis can extract the factors affecting the processing of different stimulus items in various tasks. This information can be very useful for cognitive psychologists who try to understand the relations between the processing of different types of stimuli.

By definition, the "factors extracted by individual-participant differences analysis" and "factors extracted by individual-item differences analysis" are different. The former is a person's trait. For a hypothetical example, perhaps an observer's performances on visual search and change detection are correlated because they are both affected by the observer's working memory capacity. On the other hand, the latter is the attribute of a type of stimulus item. For a hypothetical example, perhaps the performances on a type of stimulus item in visual search and change detection are correlated because they are both affected by the stimulus item's attentional

Figure 1
Individual-Item Differences Analysis Versus Individual-Participant Differences Analysis



Note. Briefly speaking, individual-item differences analysis is what happens when the roles of "participant" and "stimulus item" are *flipped* in the individual-participant differences analysis. Traditional individual-participant differences analysis can extract the factors affecting the different human individuals' performances on various tasks. This information is very useful for educational/developmental psychologists who try to understand the relations between different aspects of a person. On the other hand, the individual-item differences analysis can extract the factors affecting the processing of different stimulus items in various tasks. This information can be very useful for cognitive psychologists who try to understand the relations between the processing of different types of stimuli. See text for further details. See the online article for the color version of this figure.

demand. To be clear, the *factors of visual-attentional processing* studied in the present study are the latter, *not* the former.

Having made the above distinctions, only very few previous studies have adopted the individual-item differences approach to extract the *factors of visual-attentional processing*. Notably, Alvarez and Cavanagh (2004) reported that working memory capacities and search efficiencies of several stimulus types are perfectly correlated, which offered support for a single-factor account. On the other hand, Cheal and Lyon (1994) measured how several types of stimulus items are processed in texture segregation, visual search, and cueing tasks, and found that the performances do not perfectly agree with each other across these tasks. This suggested the existence of multiple factors.

This apparent lack of individual-item differences studies perhaps happens because researchers commonly (although often implicitly) assume that the processing of visual stimulus information in different tasks is governed by only *one single factor* which can be called the "stimulus discriminability" or perhaps "featural difference." To put it simply, if the stimulus difference is greater (e.g., the red–green difference when compared to the red–pink difference), then that leads to better performance on *all* tasks: It is easier to see, easier to find, and easier to remember the former than the latter. The present study examined the validity of this assumption.

Individual-Item Differences Analysis Explained

As discussed above, individual-participant differences analysis and individual-item differences analysis are fundamentally different from each other conceptually. However, they can make use of exactly the same mathematical tools. Specifically, the present study employed principal component analysis (PCA).

PCA is a statistical tool that orthogonally transforms a set of correlated variables into a set of uncorrelated variables which are called principal components. The first principal component is extracted to account for the largest possible portion of variances in the original variables, and each succeeding component is extracted to account for the largest possible portion of the remaining variances. The first several principal components usually capture the majority of the variances in the data set, achieving an effective reduction of the dimensionality of the data. Then, these principal components are rotated to become interpretable factors that supposedly reflect the independent underlying mechanisms behind these factors.

Although the same PCA is used in individual-participant differences analysis and individual-item differences analysis, the PCA takes different data as input. Specifically, in the usual individual-participant differences analysis, a sample (i.e., a set of individual participants) is drawn, and each "data point" of the input data is the performance of a participant in a task, which is usually the average of the data from many stimulus items. The PCA is based on correlations between the performances of individual participants.

However, in individual-item differences analysis, the roles of "participants" and "items" are flipped: A set of individual stimulus items are chosen,¹ and each "data point" of the input data is the

¹ The difference between the "chosen items" and the "participants randomly drawn by sampling" reflects a fundamental difference between the logic of experimental research and that of correlational research, this point will be elaborated below in the section "Experimental Nature of Individual-Item Differences Analysis."

performance on a stimulus item in a task, which is the average of the data from many participants. The PCA is based on correlations between the performances on individual stimulus items.

FVS Framework: Featural, Visual, and Spatial Strength

Recently, Huang (2015a) conducted an individual-item differences analysis to extract the factors of visual stimulus information across a range of 16 stimulus types \times 8 tasks. Huang (2015a) identified three orthogonal factors that account for the vast majority of the variance in observers' performances on these eight tasks. These three factors were terms *featural*, *visual*, and *spatial* strengths.

Featural Strength

The first of Huang's (2015a) three factors is termed *featural strength*, which describes the degree to which a type of stimulus information can be processed as a preattentive feature. In other words, it measures the efficiency of feature-based attentional guidance.

This concept originates from the notion of preattentive visual features. In the famous feature integration theory of visual attention, Treisman proposed a set of preattentive features which include colors, shapes, and orientations, etc. These preattentive visual features can be extracted in parallel and then used to guide attention. On the other hand, more complex visual stimuli need to be processed serially with focal attention (e.g., Treisman & Gelade, 1980; see also Treisman, 1996; Treisman & Gormican, 1988; Wolfe, 1994, 1998a, 1998b, 2012; Wolfe & Cave, 1999; Wolfe et al., 1989). After some follow-up debates, the consensus is that both features and feature bindings can be efficiently used for attentional selection (e.g., Huang & Pashler, 2005; Palmer, 1994; Palmer et al., 2000; Wolfe, 1994; Wolfe et al., 1989), but feature arrangements cannot (Egeth & Dagenbach, 1991; Kwak et al., 1991; Moore et al., 1996; Wolfe, 1998a, 1998b). For example, Wolfe (1998b; see also Wolfe & Horowitz, 2004, 2017) reviewed a large number of visual search studies and concluded that both feature searches and conjunction (i.e., feature binding) searches are usually fairly efficient, whereas the searches for feature arrangements are always rather laborious.²

It is important to note that, although the initial feature integration theory has compared the features and all other types of stimulus information as a parallel/serial dichotomy, the *featural strength* in the FVS framework is intended to be a continuum ranging from unambiguous preattentive features (e.g., colors, shapes) on the high featural strength end to very laborious stimulus types (e.g., color arrangements, random shapes) on the low featural strength end.

Visual Strength

The second of Huang's (2015a) three factors is termed *visual strength*, which describes how strong a visual signal is. Conceptually, visual strength and featural strength are separable factors. For example, both high featural strength stimulus (e.g., colors) and low featural strength stimulus (e.g., color arrangements) could be presented either in high-contrast format (i.e., high visual strength) or low-contrast format (i.e., low visual strength). Huang (2015a, see also Huang, 2015c; Pashler and Badgio, 1985) confirmed that the visual strength and featural strength are two dissociable factors. Specifically, in some tasks (e.g., perceptual discrimination), the

visual strength has a greater impact on the performance than featural strength does, but in some other tasks (e.g., change detection), the featural strength has a much greater impact than visual strength does.

Spatial Strength

The third of Huang's (2015a) three factors is termed *spatial strength*, which describes the degree to which a stimulus type can be represented in a global spatial structure of the stimulus items.

This concept originates from the previous studies on the Boolean map theory (Huang, 2010a, 2010b; Huang et al., 2007; Huang & Pashler, 2007) which claims that the unit of conscious access, namely what can be consciously perceived at one instant, boils down to a simple data format termed a Boolean map. The term "Boolean" originates from our opinion that conscious access (at any given instant) works by making an attentional selection: Dividing the visual field into two (i.e., Boolean) states: Attentionally selected subset versus unselected subset, and only the former, but not the latter is subject to conscious access. To put it simply, the spatial structure of the Boolean map is the set of selected locations at a given instant.³

In the original proposal (Huang, 2010a; Huang et al., 2007; Huang & Pashler, 2007), the use of the spatial structure of the Boolean map is demonstrated in the simultaneous presence of several locations, whereas colors are used as examples of "features." It is obvious that the spatial structure of the Boolean map cannot provide any useful information for a color task (top row in Figure 2), so the conceptual dichotomy between features and "locations" is absolute in these studies.

However, the dichotomy between features and locations can be less clear in some other situations. For example, a Boolean map could be created to contain multiple bars in different orientations (middle row in Figure 2) or multiple shapes (bottom row in Figure 2). An important question to ask is whether the spatial structure of the Boolean map represents the profiles of individual items, or only represents the structure consisting of the centers of these items. The initial Boolean map theory (Huang & Pashler, 2007) did not ask this question. Nevertheless, some reasonable speculations can be made.

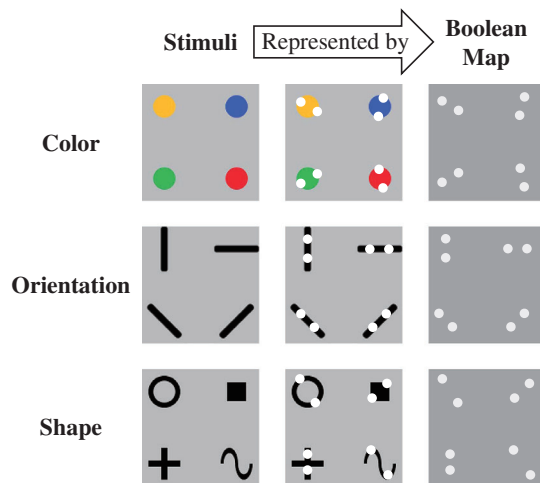
On the one hand, the basic idea of a Boolean map is a mechanism that encodes the "locations" of the currently selected set of stimuli. Some features (e.g., the orientation of a bar) are ultimately determined by the "locations" occupied by the pixels in the bar. Therefore, it seems intuitively compelling that the Boolean map structure can capture a rough sketch of the profiles of individual items. On the other hand, it is reasonable to assume that there is a capacity limit on what can be contained in the spatial structure of the Boolean map. Therefore, the spatial structure of the Boolean map will not be able to represent these profiles in detail.

Taken together, it seems the most reasonable to assume a capacity-limited Boolean map structure: A mechanism that can capture the rough sketch, but not the detailed shape, of the profiles. Figure 2

² The findings on the emergent features (e.g., Pomerantz, 2003; Pomerantz et al., 1977) may appear to conflict with this general rule. However, as the name of the concept indicates, the emergent feature involves those special situations in which the stimuli are processed as a new feature (e.g., 3D cubes or holes) rather than as arrangements of multiple individual features. So there is no substantive conflict.

³ For the present purpose, only one aspect of the Boolean map theory (i.e., multiple-location access) is elaborated. For the other claims of the Boolean map theory (e.g., single-feature access), see Huang, 2010b; Huang and Pashler, 2007.

Figure 2
The Role of the Spatial Structure of the Boolean Map



Note. This figure is reprinted from Huang's (2015a) Figure 5. It illustrates how a mimicking strategy could be applied to colors, orientations, and shapes. This mimicking strategy naturally provides no useful clue at all about the colors of stimulus items but can give very good clues about the orientations of stimulus items from as few as two locations per item. In the case of the shapes, the profiles of stimulus items are too complex to be precisely captured by one or two locations, so the mimicking strategy is ineffective. In brief, it seems that the capacity-limited Boolean map structure can capture the rough sketch of the profiles and represent the bar orientations quite well, but cannot represent the detailed shapes of individual items. See the online article for the color version of this figure.

(reproduced from Figure 5 of Huang, 2015a) illustrates the role of such a capacity-limited Boolean map structure by using eight locations⁴ to mimic a feature. This mimicking strategy naturally provides no useful clue at all about the colors of stimulus items but gives very good clues about the orientations of stimulus items. In the case of the shapes, the profiles of stimulus items are too complex to be precisely captured by one or two locations, so the mimicking strategy is ineffective. Of course, this mimicking strategy is only illustrative and the precise algorithm will need to be explored in the future.

Empirically, Huang (2015b) predicted and confirmed an *interaction* between the stimulus type (color vs. bar orientation) and the task (change detection vs. visual search). For stimulus types, bar orientations, but not colors, can be represented in the Boolean map structure. For tasks, change detection, but not visual search, can rely on the Boolean map structure. Together, a unique advantage for bar orientation in change detection is expected. Huang (2015b) confirmed this unique interaction, and Huang (2015a) showed that this factor of spatial strength, which describes the extent to which how a Boolean map structure can be used for processing a stimulus type in a task, plays a generally important role in the processing of visual stimulus information.

The Present Study

Huang (2015a) used 16 stimulus types and 8 tasks. The present study is designed to be a substantial extension of the range of tasks. In the present study, exactly the same 16 stimulus types were used, whereas a new set of 18 tasks were added, making a total of 26 tasks.

This extension on the task set is designed to achieve two goals. First, Huang (2015a) reduced a set of eight tasks into three principal components. This “8-to-3 compression ratio” is in some sense a moderate achievement and one may feel that the second and third components are accidental and may not generalize beyond these eight specific tasks. The present study was designed to test the generality of this three-factor FVS framework.

Second, FVS is a general framework for the factors of visual-attentional processing. Therefore, it has important implications for various research questions in this area. Specifically, these 26 tasks were designed to give useful clues to the nature of the featural, visual, and spatial strengths, as well as to various other questions in this research area (e.g., global selection vs. local saliency; high-level vs. low-level processing).

The present study examined this very broad range of 16 stimulus types \times 26 tasks. A major advantage of this approach is the benefit of scale. This framework allows us to pool together the ambiguous clues in individual tasks and stimulus types to come up with an unambiguous overall picture of the factors of attentional processing. Each of the individual tasks or stimulus types only adds some weak constraints on the theoretical interpretations, but the joint force of all these constraints allows us to narrow down to a very specific point in the space of possible theoretical interpretations.

The 16 stimulus types are shown in Figure 3. The 26 tasks are illustrated in Figures 4–5 and also listed in Table 1. Methodological details are skipped in the main text and can be found in Appendix A.

Results and Analysis

Open Access

Data from all 26 tasks reported in this study, including those previously published in Huang (2015a), are available on the Open Science Framework project page <https://osf.io/9eafp/>. A Matlab analysis script and the main results of the analysis (e.g., correlation matrix, extracted factors, factor loadings, & predictions) are also given for convenient references.

Reliability

The performances on the 26 tasks (16 stimulus types \times 26 tasks) are plotted in Figures 6–7. As shown in Table 1, the reliabilities of all 26 tasks (Cronbach's α)⁵ were very high (average = 0.985, in the range of 0.973–0.992). This gave a solid foundation for the interpretations of the results.

Exclusion of One Task

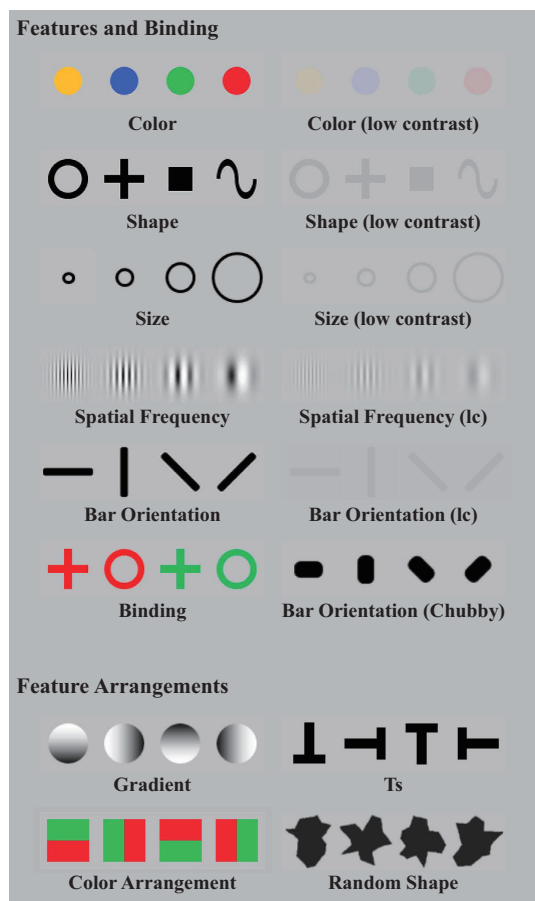
These 26 tasks were then analyzed by a PCA. The correlation matrix is shown in Figure 8.

First, a PCA was conducted on all 26 tasks. For 25 of the 26 tasks, the first 3 principal components explained a vast majority of the

⁴ In terms of number of locations, Huang (2020a) estimated that this capacity limit to be 4.9 (p. 25). This estimation is based only on the comparison of RT slopes in one situation. Therefore, it only provides us an approximate idea of the number.

⁵ In the usual individual-participant difference analysis, the Cronbach's α measures the reliability of the individual-participant differences across different items. In the present individual-item difference analysis, the roles of “participants” and “items” are flipped, so the Cronbach's α measures the reliability of the individual-item differences across different participants.

Figure 3
The 16 Stimulus Types



Note. This figure is reprinted from Huang's (2015a) Figure 1. First, to have a systematic test on the factor of *featural strength*, both items defined by a preattentive feature and those defined by feature arrangements (i.e., attention-demanding stimuli) were included. Second, to have a systematic test on the factor of *visual strength*, low-contrast versions of colors, shapes, sizes, spatial frequencies, and orientations were included. Third, to have a systematic test on the factor of *spatial strength*, both stimulus types with strong spatial structures (e.g., orientation) and those without any spatial structure (e.g., color) have been included. See the online article for the color version of this figure.

variances (average = 94.9%, ranging from 87.9% to 98.7%). However, there is one exception. For the task of low-level motion, these three principal components can only explain 54.6% of the variances, suggesting that this low-level motion task does not belong to the scope of the visual-attentional processing tasks. Therefore, this task of low-level motion was excluded and another PCA was conducted on the remaining 25 tasks.⁶ Hereinafter, the "PCA analysis" always refers to this second 25-task analysis rather than the initial 26-task analysis, and all the reported results (e.g., Table 1; Figures 9–12) are from the former rather than the latter.

One Major Factor

The scree plot of the PCA analysis is shown in Figure 9. The scree plot is a line plot of the eigenvalues of principal components (i.e., the variance explained). The first principal component accounted for a large

portion (78.3%) of the variances in these 25 tasks. Therefore, the notion of a single dimension of a general "featural difference" (or stimulus discriminability) does account for the *main pattern* of the results. In other words, the tasks do agree with each other in the sense that, generally speaking, a stimulus type that produces better performance on one task also tends to produce a better performance on other tasks.

Three-Factor Analysis

However, it is also clear that a single-factor account is less than perfect. First, some of the correlations between the tasks were very low. For example, the correlation between monitoring and feature counting was merely 0.08 (see Figure 8), suggesting that these two tasks share almost no common variance. Therefore, a strict single-factor account can be ruled out.

Second, as shown in Figure 9, the inadequacy of a single factor was indicated by the presence of the second and third principal components which accounted for significant portions, 10.2% (*eigenvalue* = 2.5) and 6.9% (*eigenvalue* = 1.7) respectively, of the variances. Therefore, by Kaiser's criterion (i.e., *eigenvalue* > 1), they should be kept in further analysis. On the other hand, the fourth principal component is considerably less important and accounted for only 1.5% (*eigenvalue* = 0.4) of the variances, and should be dropped.

Third, although the second and third components accounted for much smaller portions of variances than the first component did, they are highly reliable because they are extracted out of many high-reliability variables. In other words, the second and third components reflect substantive mechanisms so they should be considered factors of visual-attentional processing. To show this, I conducted a split-half cross-validation. I randomly split the entire data set into two halves (i.e., split the participants for each of the 25 tasks), and attempted to see whether the components extracted from the first half also applied to explain the variances in the second half. The data set was split into random halves 1,000 times. On average, the first, second, and third components of the first half respectively accounted for 77.4%, 10.2%, and 6.9% of the variances in the first half and 77.2%, 10.0%, and 6.8% of the variances in the second half. Clearly, the second and third components can transfer between the subsets very well. Therefore, they should be kept in further analysis.

All in all, it is the most reasonable to settle with a three-factor structure. These first 3 principal components explained a very large portion of the variances of the 25 tasks (average = 95.4%, ranging from 88.4% to 98.6%). The potential mechanisms for some of the 4.6% unexplained variances will be discussed in the section "reasons for unexplained variance" in Appendix B.

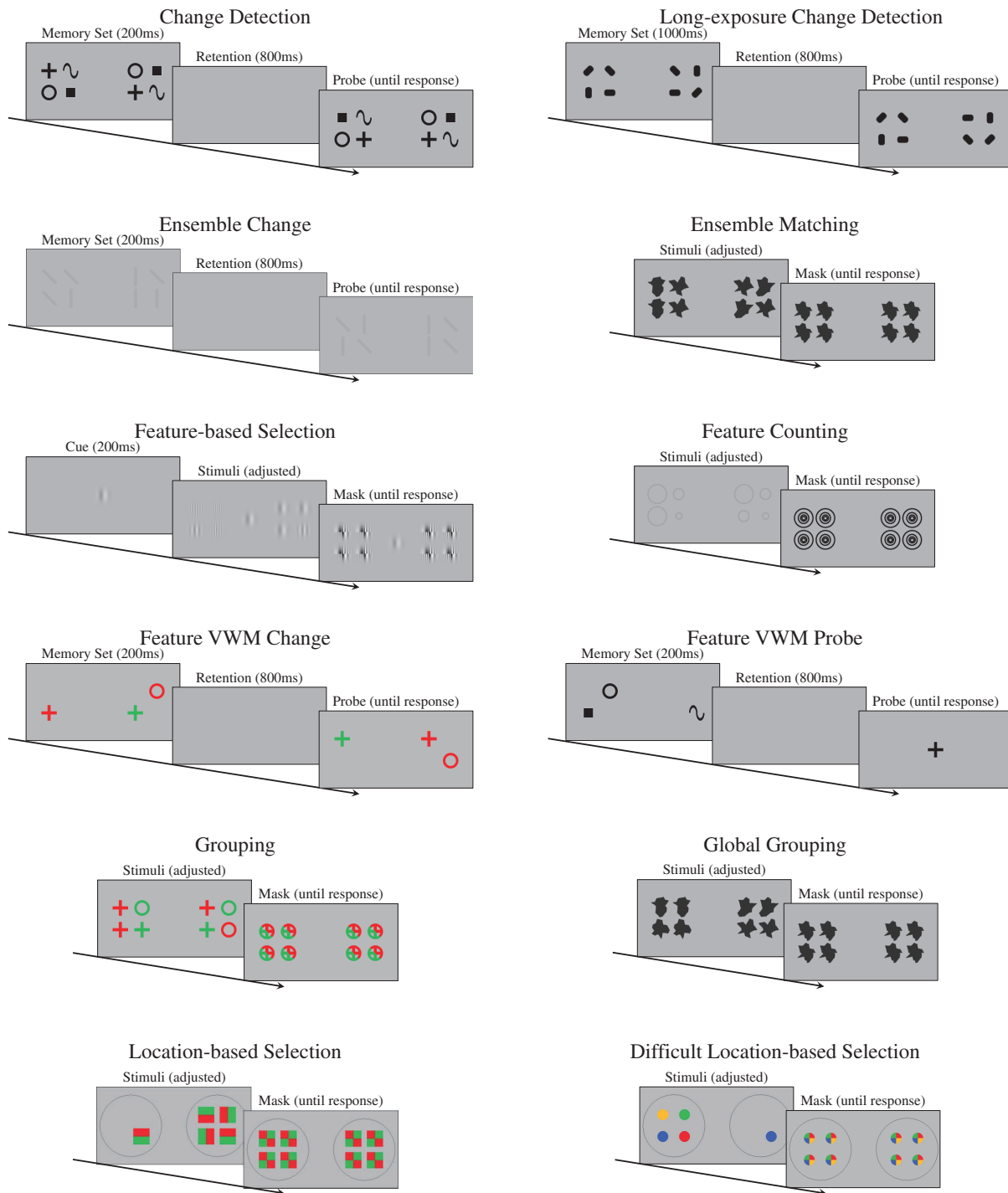
FVS 2.0 Framework

The three principal components were then orthogonally rotated⁷ to become interpretable factors. The degrees of rotations were

⁶ Following this 25-task analysis, factor loadings (and predicted performances) of low-level motion were still calculated for the purpose of comparison even if this task had not been included in this 25-task PCA which was used to extract those factors.

⁷ In PCA, an orthogonal rotation is often used so that the factors are matched with conceptually interpretable dimensions, and the actual coordinate system is not changed in this orthogonal rotation. In linear algebra, a rotation in an n -dimensional Euclidean space can be performed by multiplying an n -dimensional vector (i.e., the coordinates before rotation) by an $n \times n$ matrix (i.e., the rotation matrix) to produce another n -dimensional vector (i.e., the coordinates after rotation).

Figure 4
Method of Tasks 1–12



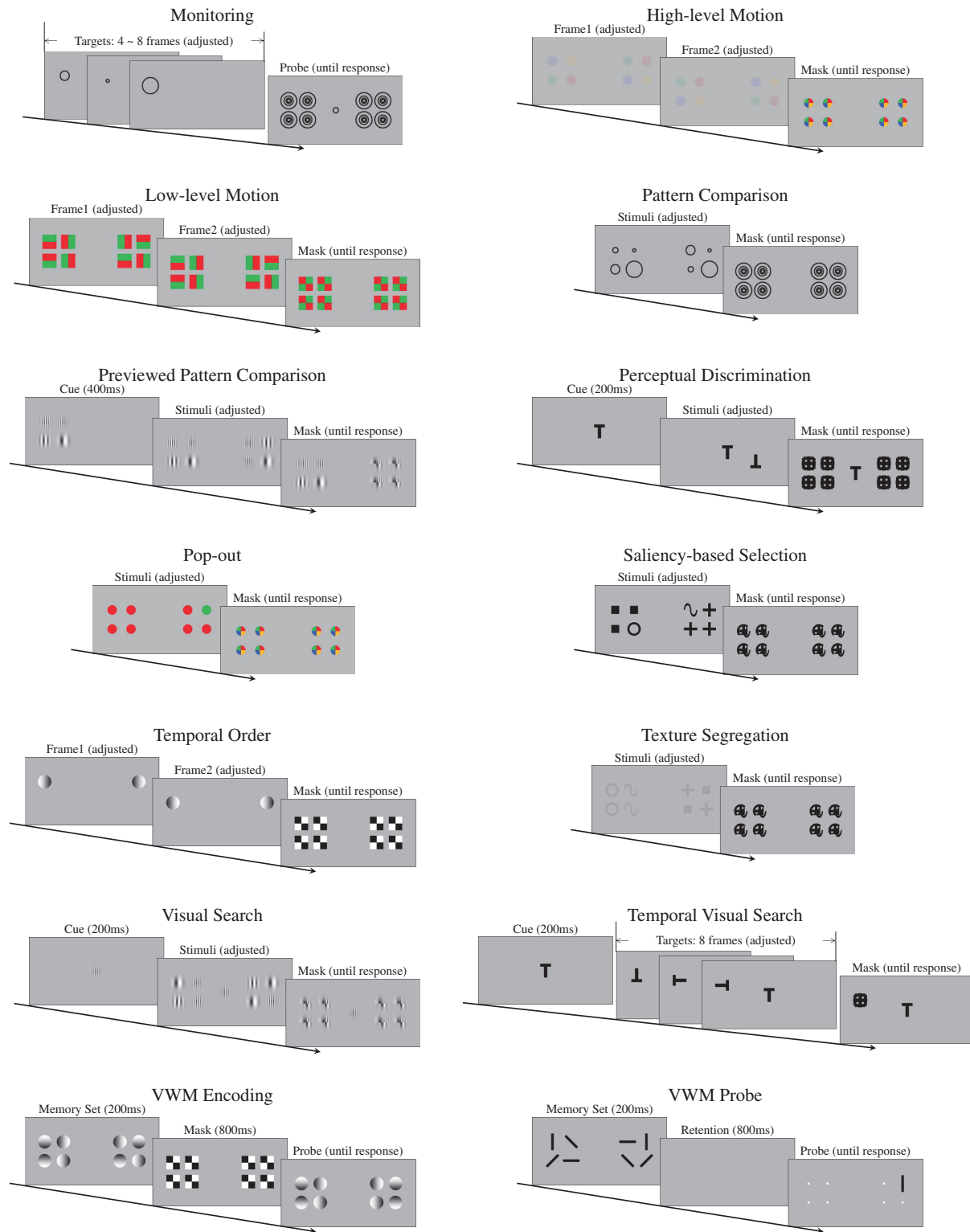
Note. See text for details. See the online article for the color version of this figure.

manually adjusted so that these factors (i.e., rotated components) were conceptually matched to the three concepts mentioned above: featural, visual, and spatial strength. The next two paragraphs will give a brief description of these factor scores. The conceptual

interpretations of these factors and other important information (e.g., how to rotate the factor axes) will be given later.

The 16 stimulus types' scores in these 3 factors are plotted in Figure 10. The first factor (i.e., featural strength) and third factor

Figure 5
Method of Tasks 13–26



Note. See text for details. See the online article for the color version of this figure.

Table 1
Summarized Information on the 26 Tasks

Task	Number of observers	Cronbach's α	Nature of measure	L_F	L_V	L_S	Total variance explained: $L_F^2 + L_V^2 + L_S^2$
1: Change detection	86	0.977	Accuracy	0.654	0.131	0.733	0.983
2: Change detection, long exposure	107	0.977	Accuracy	0.684	0.057	0.679	0.932
3: Ensemble change	81	0.99	Accuracy	0.933	0.325	-0.029	0.977
4: Ensemble matching	35	0.989	Threshold	0.981	0.089	-0.043	0.972
5: Feature-based selection	21	0.986	Threshold	0.957	0.165	-0.144	0.963
6: Feature counting	51	0.99	Threshold	0.957	-0.151	0.159	0.964
7: Feature VWM change	81	0.992	Accuracy	0.900	0.343	-0.118	0.941
8: Feature VWM probe	56	0.989	Accuracy	0.685	0.680	-0.109	0.944
9: Grouping	42	0.987	Threshold	0.797	0.475	0.154	0.884
10: Grouping, global	65	0.986	Threshold	0.774	0.417	0.383	0.919
11: Location-based selection	37	0.984	Threshold	0.797	0.581	-0.002	0.973
12: Location-based selection, difficult	92	0.986	Threshold	0.837	0.522	-0.020	0.974
13: Monitoring	83	0.979	Threshold	0.256	0.943	-0.126	0.971
14: Motion, high-level	102	0.986	Threshold	0.982	-0.030	0.063	0.969
15: Motion, low-level	56	0.975	Threshold	-0.172	0.547	-0.069	0.334
16: Pattern comparison	95	0.975	Threshold	0.870	0.208	0.398	0.959
17: Pattern comparison, previewed	104	0.983	Threshold	0.625	0.565	0.499	0.958
18: Perceptual discrimination	36	0.988	Threshold	0.455	0.794	0.222	0.887
19: Pop-out	35	0.991	Threshold	0.803	0.386	0.385	0.942
20: Saliency-based selection	29	0.986	Threshold	0.975	-0.007	0.003	0.951
21: Temporal order	44	0.987	Threshold	0.698	0.616	0.297	0.956
22: Texture segregation	51	0.992	Threshold	0.978	-0.013	0.158	0.981
23: Visual search	37	0.992	Threshold	0.955	0.273	0.013	0.986
24: Visual search, temporal	45	0.988	Threshold	0.683	0.629	-0.294	0.949
25: VWM encoding	154	0.973	Accuracy	0.733	0.461	0.454	0.957
26: VWM probe	119	0.980	Accuracy	0.948	0.045	0.225	0.952

Note. The four columns on the left side show the name of the task, the number of observers, the reliability (Cronbach's α), and the nature of the task (threshold vs. accuracy). For all tasks, the reliability was no less than 0.97. The next 3 columns show the loadings of the 26 tasks on the 3 factors (FVS: Featural, Visual, Spatial strengths). The rightmost column shows the total portion of the variance in each of the 26 tasks that are attributed to the 3 factors, which is equal to the squared sums of the 3 loadings. With one exception (low-level motion), a very large portion of the variance is attributed to the FVS 2.0 framework. VWM = Visual working memory.

(i.e., spatial strength) are respectively represented by the x -axis of the left and right panels, whereas the second factor (i.e., visual strength) is represented by the y -axis of both panels. Hereinafter, these featural, visual, and spatial strength scores will be addressed as FVS scores together. The meanings of these scores will be interpreted below.

The 95% confidence intervals of these FVS scores were estimated through a split-half reliability analysis⁸ and shown in the centers of the left-side and right-side panels of Figure 10. The values are similar for the 16 stimulus types. Therefore, one pooled estimate of standard deviation was made for featural strength scores of all 16 stimulus types, and likewise for visual and spatial strengths. Apparently, the featural strength scores are the most precise, and spatial strength scores are the least precise. This is because the featural strength has overall the largest effects on the tasks, and spatial strength the least effects on the tasks.

An FVS 2.0 Framework for Visual-Attentional Processing

Predicting the Task Performances From FVS Scores

To start our discussion on the implications of the FVS 2.0 framework, the first and most important message is its predictive power. By assigning some weights (i.e., loadings) to the 3 factors

(i.e., 16 stimulus types' FVS scores), we can predict the data from 25 tasks very well. Specifically, the prediction model is in the form of a multiple linear regression equation as follows.

Suppose the FVS scores for a type of stimulus information are F , V , and S . Then, the performance on this stimulus type in a given task can be linearly determined by a set of three loadings (L_F , L_V , & L_S).⁹

$$\text{Performance} = a + b(L_F F + L_V V + L_S S). \quad (1)$$

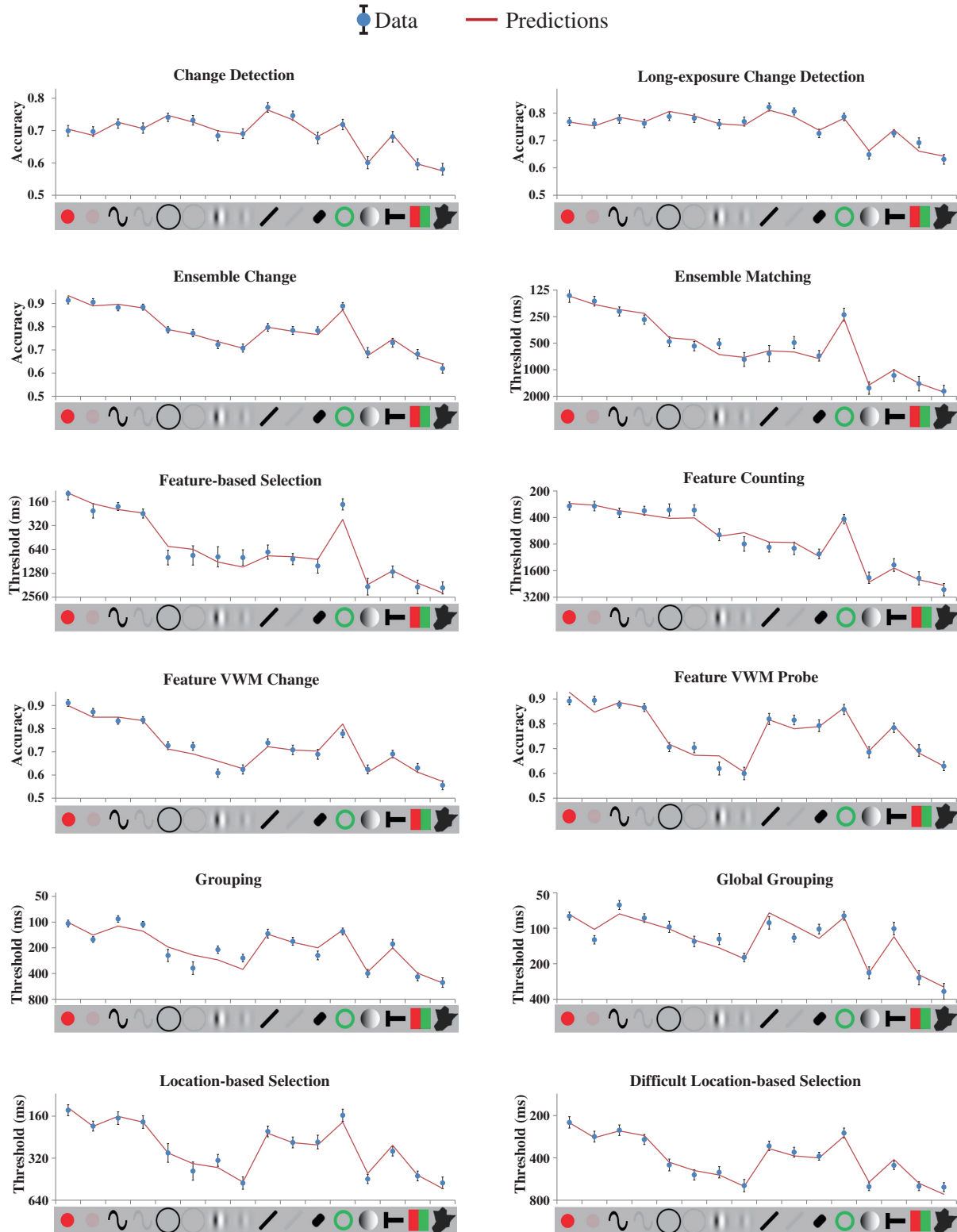
The FVS scores for the 16 stimulus types are those plotted in Figure 10, whereas the 26 sets of loadings of the 26 tasks are shown in Table 1 and also plotted in Figure 11 (see explanations below).¹⁰ Together, combining these values in Equation 1 generates $16 \times 26 = 416$ predictions. As shown in Figures 6–7, these

⁸ One half of the data were randomly chosen from that of each task and the FVS scores were calculated on the basis of these halved data. This was repeated for 1,000 times and we calculated the SD s of these 1,000 sets of halves-based FVSs. Then, the SD s of the FVS scores of the whole data were calculated as the $1/\sqrt{2}$ of the SD s of these halves-based FVSs.

⁹ In PCA, the factor loadings are the correlation coefficients between the variables and factors. In the FVS 2.0 framework, the three loadings of a task are the three correlations between the featural, visual, and spatial strengths and the performance on that task. These three loadings show the degrees to which a task is affected by each of these factors.

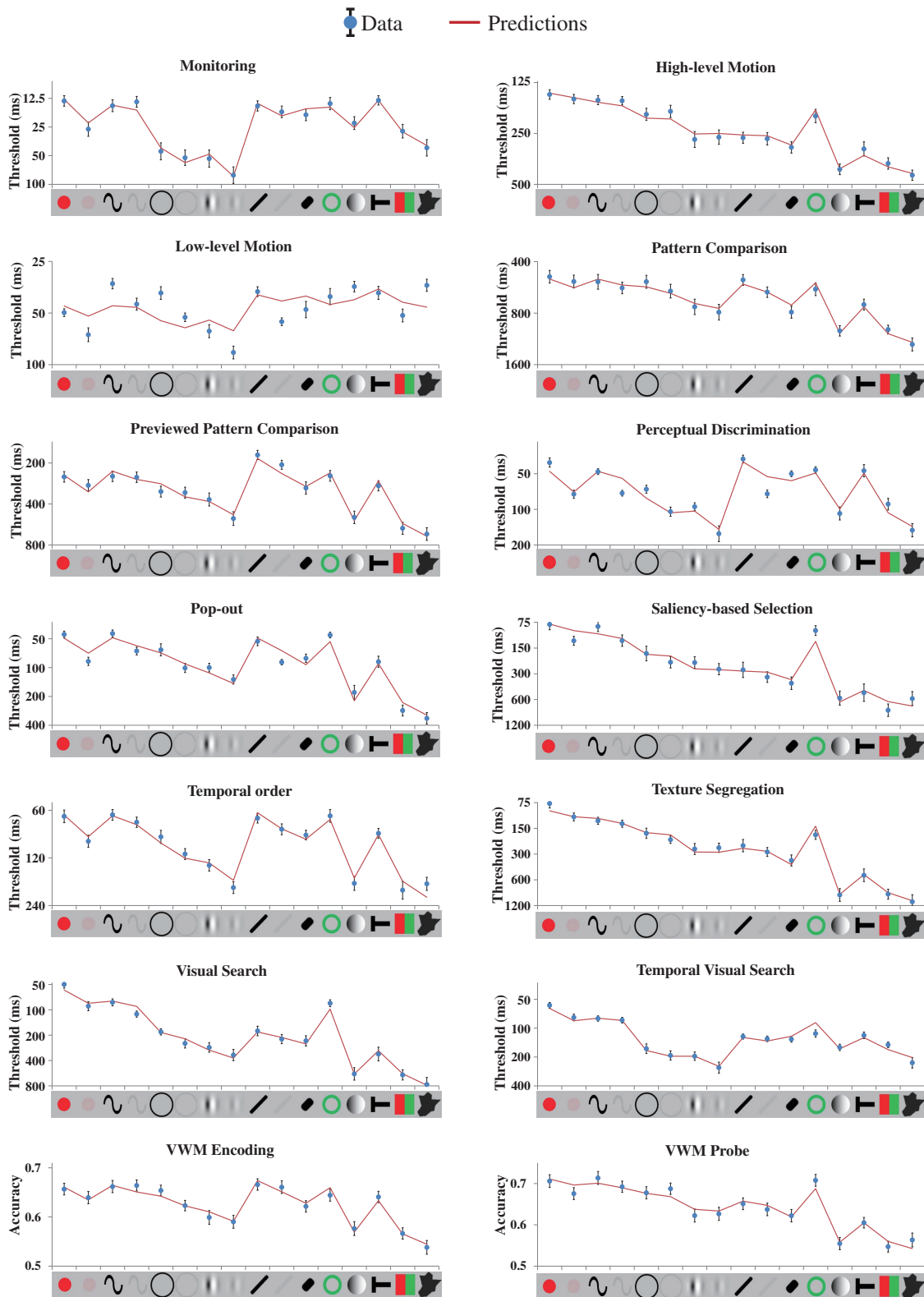
¹⁰ The a and b are parameters for linear regression. Their values depend on the scaling of the FVS scores or that of the task performances.

Figure 6
The Performances on Tasks 1–12



Note. See text for details. See the online article for the color version of this figure.

Figure 7
The Performances on Tasks 13–26



Note. See text for details. See the online article for the color version of this figure.

Figure 8
Correlation Matrix for the 26 Tasks

Task	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
1: Change detection																										
2: Change detection, long exposure	0.97																									
3: Ensemble change	0.64	0.65																								
4: Ensemble matching	0.62	0.67	0.95																							
5: Feature-based selection	0.54	0.58	0.94	0.96																						
6: Feature counting	0.72	0.73	0.85	0.91	0.84																					
7: Feature VWM change	0.56	0.56	0.96	0.90	0.91	0.81																				
8: Feature VWM probe	0.48	0.45	0.89	0.74	0.77	0.54	0.90																			
9: Grouping	0.69	0.69	0.87	0.82	0.86	0.68	0.82	0.81																		
10: Grouping, global	0.81	0.77	0.84	0.77	0.76	0.75	0.76	0.73	0.90																	
11: Location-based selection	0.59	0.58	0.93	0.85	0.87	0.66	0.89	0.93	0.92	0.86																
12: Location-based selection, difficult	0.61	0.60	0.95	0.89	0.90	0.70	0.91	0.93	0.92	0.84	0.99															
13: Monitoring	0.20	0.14	0.55	0.33	0.43	0.08	0.57	0.83	0.64	0.54	0.74	0.70														
14: Motion, high-level	0.69	0.70	0.91	0.94	0.90	0.98	0.90	0.66	0.76	0.78	0.74	0.78	0.22													
15: Motion, low-level	-0.11	-0.27	-0.02	-0.19	-0.08	-0.21	0.03	0.22	0.05	0.12	0.17	0.13	0.51	-0.16												
16: Pattern comparison	0.90	0.88	0.87	0.85	0.79	0.88	0.85	0.71	0.83	0.88	0.80	0.82	0.36	0.89	-0.06											
17: Pattern comparison, previewed	0.87	0.81	0.76	0.66	0.62	0.59	0.70	0.78	0.84	0.87	0.83	0.82	0.63	0.62	0.12	0.86										
18: Perceptual discrimination	0.54	0.49	0.67	0.50	0.53	0.37	0.65	0.79	0.73	0.80	0.83	0.79	0.81	0.44	0.39	0.66	0.80									
19: Pop-out	0.84	0.80	0.86	0.80	0.79	0.77	0.78	0.73	0.90	0.98	0.87	0.86	0.52	0.79	0.11	0.91	0.89	0.80								
20: Saliency-based selection	0.63	0.65	0.89	0.95	0.96	0.92	0.83	0.63	0.80	0.78	0.79	0.83	0.25	0.94	-0.07	0.82	0.60	0.45	0.82							
21: Temporal order	0.75	0.68	0.83	0.72	0.72	0.63	0.79	0.86	0.88	0.92	0.92	0.91	0.72	0.69	0.30	0.84	0.93	0.87	0.92	0.70						
22: Texture segregation	0.75	0.78	0.90	0.95	0.90	0.96	0.87	0.64	0.81	0.81	0.77	0.81	0.21	0.97	-0.25	0.92	0.68	0.47	0.84	0.94	0.71					
23: Visual search	0.67	0.68	0.98	0.97	0.96	0.87	0.94	0.83	0.89	0.86	0.93	0.95	0.49	0.92	-0.03	0.89	0.75	0.68	0.89	0.94	0.84	0.94				
24: Visual search, temporal	0.32	0.32	0.85	0.74	0.77	0.52	0.90	0.94	0.79	0.66	0.89	0.89	0.80	0.66	0.19	0.64	0.64	0.74	0.65	0.63	0.75	0.63	0.81			
25: VWM encoding	0.88	0.83	0.83	0.73	0.69	0.72	0.80	0.78	0.87	0.93	0.83	0.83	0.58	0.76	0.10	0.94	0.95	0.77	0.92	0.69	0.93	0.79	0.82	0.69		
26: VWM probe	0.79	0.78	0.89	0.91	0.89	0.94	0.83	0.66	0.79	0.85	0.78	0.81	0.27	0.95	-0.03	0.91	0.72	0.52	0.88	0.95	0.79	0.94	0.92	0.58	0.81	

-0.3 -0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

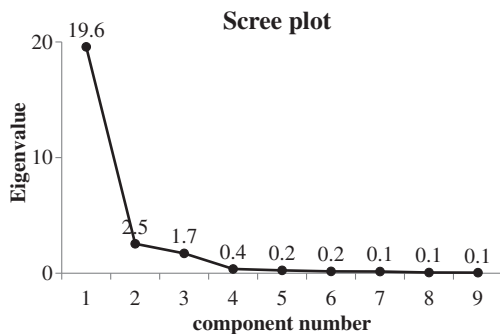
Note. See text for details. See the online article for the color version of this figure.

predictions fit excellently to the data of 25 tasks, except for those of low-level motion.

From the Three Principal Components to the FVS Scores

Next, we will explain how to convert the first three principal components generated by PCA analysis into the three factors: featural, visual, and spatial strengths.

Figure 9
The Scree Plot of the PCA Analysis



Note. The first principal component accounted for a large portion (78.3%) of the variances in these 25 tasks. Therefore, the notion of a single dimension of general “featural difference” (or stimulus discriminability) does account for the main pattern of the results. But the second and third principal components also accounted for significant portions, 10.2% and 6.9% respectively, of the variances. On the other hand, the fourth principal component is considerably less important and accounted for only 1.5% of the variances. All in all, it is the most reasonable to settle with a three-factor structure. PCA = principal component analysis.

Replication of the FVS Framework in Huang (2015a)

The first three principal components of the PCA were first manually rotated so that the factors were matched to the featural, visual, and spatial strength scores proposed in Huang (2015a). This can be readily achieved.¹¹ The correlations between the featural, visual, and spatial strengths in this middle-stage factor structure and those in Huang’s (2015a) factor structure are respectively 0.992, 0.931, and 0.964, suggesting that the factor structure can be generalized fairly well from Huang’s (2015a) 8 tasks to the current 25 tasks.

However, an important update has to be made for converting this middle-stage factor structure to the final factor structure of the FVS 2.0 framework, as will be described next.

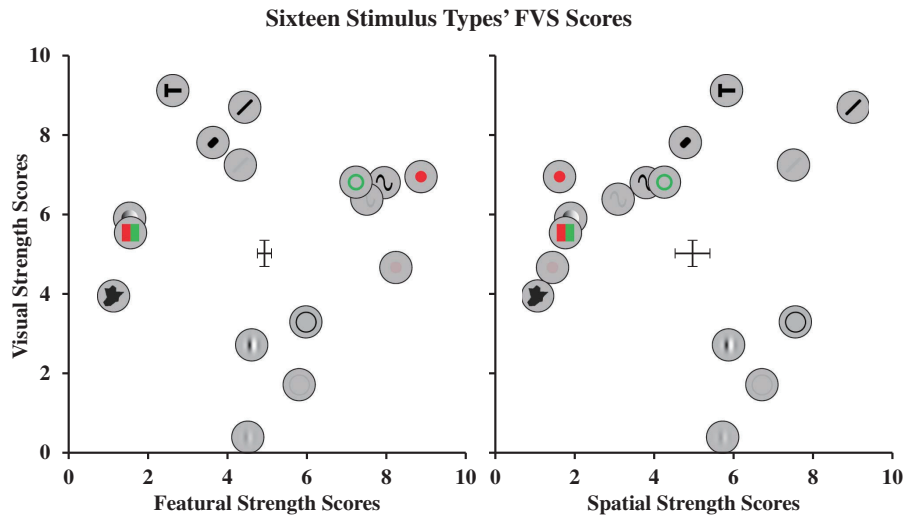
An Update on the Factor Structure

As mentioned above, In PCA analysis, the rotation of the components depends on the conceptual interpretations. Indeed, the FVS 2.0 framework in the present study has made an important revision to the interpretations of the initial FVS framework described in Huang (2015a). Specifically, as shown in Figure 12a, the axes have been rotated in the featural-spatial plane for approximately 22.9° from the green axes (which shows the middle-stage factor structure described above) to the black axes (which shows the final factor structure of the FVS 2.0 framework). As a critical consequence of this rotation, the orientations¹² scored moderately in the featural strength of FVS 2.0 so it is *no longer considered a preattentive feature*. Instead, the good performance on orientations

¹¹ The rotation matrix can be found in the Matlab script on <https://osf.io/9eafp/>.

¹² As shown in Figure 12a, there is a similar but smaller effect on sizes. For simplicity, the following discussions will only focus on the case of orientations.

Figure 10
The 16 Stimulus Types' FVS Scores



Note. The first factor (i.e., featural strength) and third factor (i.e., spatial strength) are respectively represented by the *x*-axis of the left and right panels, whereas the second factor (i.e., visual strength) is represented by the *y*-axis of both panels. The high and low featural strength stimuli (e.g., colors, shapes vs. color arrangements, random shapes) respectively appeared on the right and left sides of the left panel. Therefore, this first factor (*x*-axis of the left panel) represents *featural strength*. The high-contrast–low-contrast pairs (e.g., high-contrast color vs. low-contrast color) mainly differ on the *y*-axis of the two panels. Therefore, this second factor (*y*-axis of both panels) represents *visual strength*. Orientations (which have a clear spatial structure) and colors (which have no spatial structure component) respectively appeared on the right and left sides of the right panel. Therefore, this third factor (*x*-axis of the right panel) represents *spatial strength*. The 95% confidence intervals of these FVS scores were estimated through a split-half reliability analysis and shown in the centers of the left-side and right-side panels. The values are similar for the 16 stimulus types so one pooled estimate of standard deviation was made for featural strength scores of all 16 stimulus types, and likewise for visual and spatial strengths. FVS = featural, visual, and spatial strength. See the online article for the color version of this figure.

in certain tasks (e.g., change detection) is now attributed to the contribution of its high-spatial strength.

Why are we making this update? Although Huang (2015a, 2015b) already suggested that the good performance on orientations is likely to be largely due to the Boolean map structure (e.g., Figure 2 of Huang, 2015b), Huang's (2015a) axes of factors were still chosen so that the conceptual interpretation of featural strength is consistent with the classic understanding of the preattentive features: colors, orientations, shapes are all preattentive features (e.g., Wolfe, 1998b). However, as shown in Figures 6–7, the performance on orientations is very poor in the majority of high-level tasks (e.g., ensemble matching, feature-based selection, feature counting, high-level motion, saliency-based selection, texture segregation, visual search, Visual working memory [VWM] probe), making its good performance on change detection more of an exception than a rule. This gives us good reasons to switch from the interpretation that orientation is a basic visual feature to the interpretation that the orientation is not a preattentive feature and its uncharacteristic good performance on certain tasks is mainly due to the contribution of its high-spatial strength.

In FVS 2.0 framework's interpretations, how much has spatial strength contributed to the performance on orientation? Referring to Equation 1, the predicted accuracy for orientations in change detection is 0.764 (actual accuracy = 0.772), and this predicted accuracy would have dropped to 0.645 if we discount the contribution of spatial strength (i.e., setting the spatial strength of the orientation to

be equal to that of the color). For another example, the predicted threshold for the orientation in the pop-out task is 50 ms (actual threshold = 53 ms), and this predicted threshold would have been worsened to 99 ms if we discount the contribution of spatial strength. In both cases, spatial strength has substantially improved the performance on orientations.

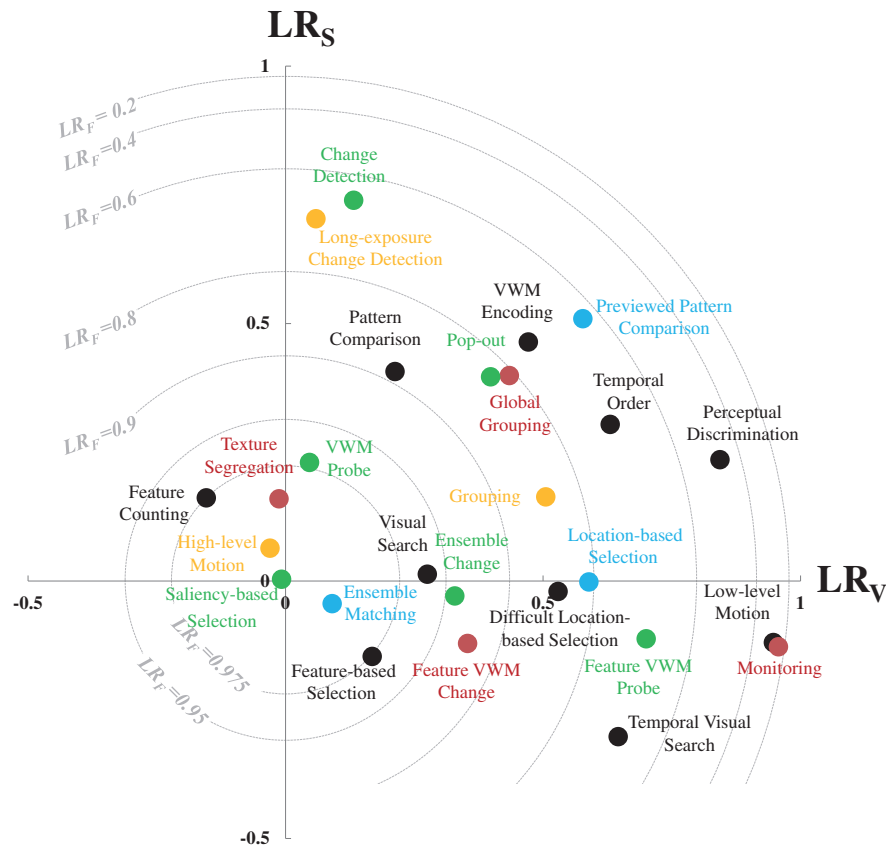
The pop-out is the classic task that has been used to identify preattentive features (Treisman, 1986; Treisman & Gelade, 1980, see also “the best tasks for studies of visual attention” section in Appendix B). The good performance on orientation in the pop-out task is probably the most important reason why the orientation has been widely considered to be a preattentive feature (see Wolfe, 1998b, for a review) even if, as mentioned above, the performances for orientation are quite poor in many other high-level tasks. Altogether, it seems reasonable to suggest that the previous single-paradigm-based conclusion (i.e., orientation is a preattentive feature) should be superseded by the present synthesis-based conclusion (i.e., orientation is not a preattentive feature).¹³

In addition to providing a better interpretation of the *F* score of orientations, the new factor structure also provides a better interpretation of the loadings. In PCA, rotation of the factor structure also leads

¹³ In the section “visual feature is more exclusive than assumed,” we will discuss additional reasons why orientation is unlikely to be a preattentive feature.

Figure 11
Relative Loadings

Loadings of the 25 tasks on the three factors



Note. This figure is a 2D representation of a 3D space. The x -axis represents the LR_V , so the tasks on the right side (perceptual discrimination, monitoring) are those that are mainly affected by the visual strength of the stimulus. The y -axis represents the LR_S . The tasks on the top (change detection) are those that clearly depend on the global comparison so that the Boolean map structure is very helpful. The LR_F is reflected by the distance from the origin of the coordinate plane, as indicated by the gray concentric circles. The tasks close to the origin (high-level motion, texture segregation, feature-based selection, ensemble matching) are those that are exclusively affected by the featural strength of the stimulus. The colors are only included as a visual aid and are *not* related to the LR_F , LR_V , or LR_S values. See the online article for the color version of this figure.

to a corresponding rotation of the loadings. Figure 12b illustrates how this rotation affects the L_F and the L_S of a few classic tasks. In Huang's (2015a) factor structure (as well as the above-mentioned middle-stage factor structure shown by green axes), the Boolean map supposedly plays no role in pattern comparison and pop-out tasks and hinders the visual search task. On the other hand, in the FVS 2.0 framework's factor structure (i.e., black axes), the Boolean map moderately helps the pattern comparison and pop-out tasks and plays no role in the visual search task. The roles assigned by the FVS 2.0 framework's factor structure make better sense: Pattern comparison and pop-out can plausibly benefit from the Boolean map structure, and it is also more plausible to conceptualize the Boolean map structure as playing no role in visual search task rather than significantly hindering it.

All in all, it seems reasonable to switch to the new factor structure as illustrated in Figure 12. Of course, such interpretation-based reasoning

is always somewhat ambiguous. Future studies will be needed to develop objective ways of determining the directions of these axes.

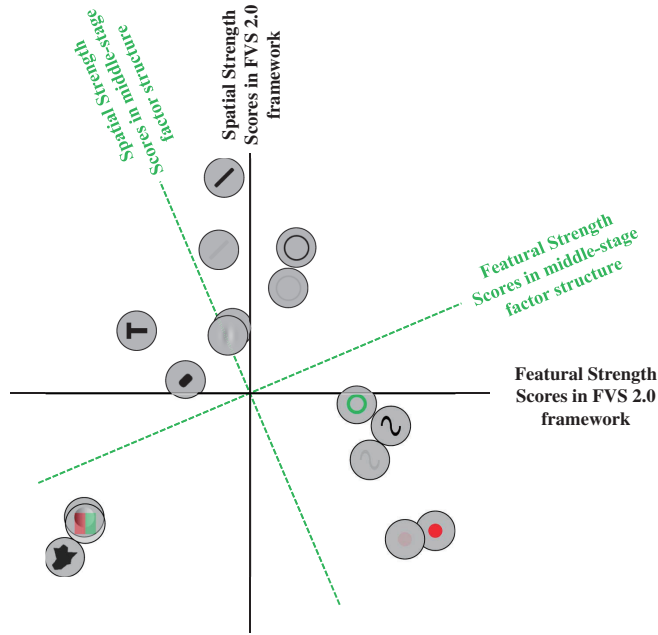
Scaling of FVS Scores

After making the above-mentioned update, the FVS scores were rescaled so that each of these fits into the range of [0, 10].¹⁴ This

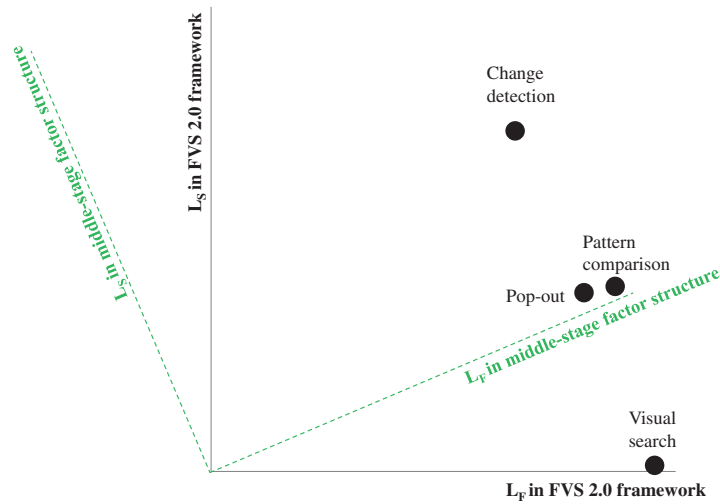
¹⁴ This scaling was implemented by multiplying the FVS scores by 2.5, and adding constants of 5, 5.5, and 4.5 to respectively featural, visual, and spatial strength scores. In future testing, FVS scores of typical visual stimuli will probably also fall in the range of [0, 10]. However, there are no theoretical upper and lower bounds for these scores. Therefore, one can imagine that FVS scores can go out of this range for very extreme visual stimulus (e.g., negative visual strength scores for very low contrast stimulus items).

Figure 12
An Update on the Factor Structure

(a) Featural and spatial strength values of the two accounts



(b) Loadings of featural and spatial strengths of the two accounts



Note. The FVS 2.0 framework has made an important revision to the interpretations of the initial FVS framework described in Huang (2015a). Specifically, as shown in *Panel a*, the axes have been rotated in the featural-spatial plane for approximately 22.9° from the green axes (the middle-stage factor structure which was matched to that of Huang, 2015a) to the black axes (the final factor structure of FVS 2.0 framework). As a critical consequence of this rotation, the orientations scored moderately in the featural strength of FVS 2.0 so it is no longer considered a preattentive feature. Instead, the good performance on orientations in certain tasks (e.g., change detection) is now attributed to the contribution of its high-spatial strength. *Panel b* illustrates how this rotation affects the loadings of a few classic tasks on the featural and spatial strengths. See text for details. FVS = featural, visual, and spatial strength. See the online article for the color version of this figure.

rescaling has *no* theoretical implication but was only made so that the scores can be understood more intuitively. For example, the color and the random shape respectively have featural strength scores of 8.9 and 1.0, so that the former is a very strong feature but the latter is a very laborious stimulus type.

Interpretation of FVS Scores

Next, we will discuss the conceptual interpretations of the FVS scores. Specifically, why do we think these factors should be matched to these concepts?

Featural Strength

The 16 stimulus types' featural strength scores are shown on the *x*-axis in the left panel of Figure 10. The high and low featural strength stimuli (e.g., colors, shapes vs. color arrangements, random shapes) respectively appeared on the right and left sides of the left panel of Figure 10. Therefore, this factor (*x*-axis of the left panel) represents featural strength.

In the FVS 2.0 framework, colors and shapes score high on featural strength and are the unambiguous preattentive features. However, the sizes, orientations, and spatial frequencies are now intermediate in the range of featural strength. As mentioned above, in the FVS 2.0 framework's interpretations, the Boolean map structure is responsible for their good performances on certain tasks (e.g., change detection).

Visual Strength

The 16 stimulus types' visual strength scores are shown on the *y*-axis of both panels in Figure 10, and they were fairly consistent with expectations. Specifically, the high-contrast–low-contrast pairs (e.g., high-contrast color vs. low-contrast color) mainly differ on the *y*-axis of the two panels, and this is consistent with the expectation that a high-contrast–low-contrast pair mainly differs on their visual strengths rather than on their featural or spatial strengths. Therefore, this second factor (*y*-axis of both panels) represents *visual strength*.

Spatial Strength

The 16 stimulus types' spatial strength scores are shown on the *x*-axis in the right panel of Figure 10, and they were generally consistent with the previous findings (Huang, 2015b): orientations, which have a clear spatial structure, scored high, whereas the colors, which have no spatial structure component, scored low. Therefore, this third factor (*x*-axis of the right panel) represents spatial strength.

How the Tasks Depend on the FVS Scores

After discussing the meaning of FVS scores, another important aspect of the FVS 2.0 framework is the extent to which how each of the 26 tasks depends on these 3 factors. This task depends exclusively on featural strength, whereas that task depends heavily on spatial strength, etc. This provides a way of organizing the tasks and understanding the relationships between them. This information is reflected by the loadings of a task on each of the three factors: L_F , L_V , and L_S .

Loadings of the 26 Tasks

The Loadings of the 26 Tasks are shown in Table 1. Squared loadings reflect the portion of variances accounted for by the three factors. For the 25 tasks (excluding low-level motion), featural, visual, and spatial strengths respectively account for 66.6%, 19.5%, and 9.3% of the variances, making a total of 95.4%.

The FVS 2.0 framework accounts for a very large portion of the variances (average = 95.4%, ranging from 88.4% to 98.6%) for 25 of the 26 tasks. Therefore, for any of these 25 tasks, the three loadings L_F , L_V , and L_S essentially have only two degrees of freedom because their squared sum is close to 1. If a set of “relative loadings”¹⁵ are defined as $LR_F = L_F / \sqrt{L_F^2 + L_V^2 + L_S^2}$, $LR_V = L_V / \sqrt{L_F^2 + L_V^2 + L_S^2}$, and $LR_S = L_S / \sqrt{L_F^2 + L_V^2 + L_S^2}$, then the squared sum of these relative loadings is exactly 1: $LR_F^2 + LR_V^2 + LR_S^2 = 1$. Therefore, these three relative loadings can be presented in a 2D graph.

These relative loadings are shown in Figure 11 which is a 2D representation of a 3D space. The *x*-axis and *y*-axis respectively represent the relative loading on visual strength and that on spatial strength, whereas the relative loading on the featural strength is reflected by the distance from the origin of the coordinate plane.

The *x*-axis represents the relative loading on visual strength (LR_V), so the tasks on the right side (perceptual discrimination, monitoring) are those that are mainly affected by the visual strength of the stimulus. This is consistent with the usual understandings. These tasks are low-level tasks¹⁶ that mainly depend on the strength of visual signals regardless of whether a type of stimulus information is a preattentive feature or not.

The *y*-axis of Figure 11 represents the relative loading on spatial strength. The placement of tasks on this dimension is consistent with the conceptual understandings. The tasks on the top (change detection) are those that clearly depend on the global comparison so that the Boolean map structure is very helpful. The task that load moderately (pattern comparison, pop-out) are those in which the Boolean map structure is moderately helpful. For the tasks that require the selection of one or several items and exclusion of the other items (e.g., feature-based selection, visual search), the loadings are close to zero.

The relative loading on the featural strength is reflected by the distance from the origin of the coordinate plane in Figure 11, as indicated by the gray concentric circles. The tasks close to the origin (saliency-based selection, high-level motion, texture segregation, feature-based selection, ensemble matching, feature counting) are those that are exclusively affected by the featural strength of the stimulus. Conceptually, these are the tasks that are both attention demanding (i.e., tasks in which the performances are much better for preattentive features than for other stimulus types) and insensitive to the Boolean map structure.

Overall, Figure 11 provides the first comprehensive roadmap of the nature of tasks in this research area. It will be useful for those who wish to know the nature of a task for their experimental design or interpretation of the data. These loadings' implications for individual tasks are self-explanatory in most cases so they will

¹⁵ As shown in Table 1, the total variance explained is generally very close to 1 in 25 tasks (average = 0.954) except in low-level motion. Thus, this ratio ($\sqrt{L_F^2 + L_V^2 + L_S^2}$), which is the square root of the total variance explained, is even closer to 1 (average = 0.977 in 25 tasks). Therefore, these relative loadings are reasonably good approximations of the actual loadings in these 25 tasks.

¹⁶ This issue will be discussed further in the section “high-level versus low-level tasks” in Appendix B.

not be exhaustively repeated. In Appendix B, we will elaborate on some of the most informative cases.

Sufficiency of the FVS 2.0 Framework

As mentioned above, the FVS 2.0 framework accounts for a very large portion of the variances of the 25 tasks (average = 95.4%, ranging from 88.4% to 98.6%). The sufficiency of this framework can also be seen from the very high correlations between conceptually unrelated tasks. For example, the tasks pop-out and global grouping are not conceptually related to each other, but they happen to be close to each other in Figure 11, and the correlation between these two tasks is indeed as high as 0.979. For another example, the task visual search and ensemble change are not conceptually related to each other, but they happen to be close to each other in Figure 11, and the correlation between these two tasks is indeed as high as 0.978.

From both examples, it is clear that all that matters for the performances on these tasks are their loadings on the featural, visual, and spatial strengths (i.e., their locations in Figure 11). Therefore, to make predictions on performance on a task, one needs not to be concerned about other conceptual aspects of the task.

Application of the FVS 2.0 Framework in Future Studies

After discussing the various aspects of the FVS 2.0 framework, one may wonder whether (and how) it can be applied in future studies. Next, I will illustrate some specific ways of doing this.

Estimation Based on Three Tasks

The main strength of the FVS 2.0 framework is the broad range of stimulus types and tasks. However, that strength turns into a heavy workload if one wants to repeat the whole FVS 2.0 framework all over again. Luckily, to apply the FVS 2.0 framework to address another research question, it is unnecessary to repeat the full set of tests. In other words, although the establishment of the FVS 2.0 framework required a large amount of experimental testing as described in the present study, it provides a platform for convenient applications in future studies.

A convenient and reasonably good approximation can be achieved by running only three tasks (change detection, visual search, and perceptual discrimination) and then using the performances on the three tasks to estimate the FVS scores. Specifically, three separate linear regressions were run to determine the parameters needed to estimate the FVS scores from the performances on these three tasks (the accuracy of change detection, the log-threshold of visual search, the log-threshold of perceptual discrimination) which go as follows:

$$\begin{bmatrix} \text{Featural strength} \\ \text{Visual strength} \\ \text{Spatial strength} \end{bmatrix} = \begin{bmatrix} 1.789 \\ -1.277 \\ -33.189 \end{bmatrix} + \begin{bmatrix} 4.176 & 3.615 & -2.032 \\ -12.430 & -1.057 & 6.465 \\ 57.541 & -3.037 & 1.040 \end{bmatrix} \cdot \begin{bmatrix} \text{Accuracy}_{\text{CD}} \\ \text{Log Threshold}_{\text{VS}} \\ \text{Log Threshold}_{\text{PD}} \end{bmatrix}. \quad (2)$$

The FVS scores estimated from Equation 2 are reasonable approximations of those estimated from the full set of 25 tasks. The featural, visual, and spatial strength scores of these two estimations are correlated with the coefficients of respectively 0.992, 0.895 and 0.980 and are good enough for most research purposes.¹⁷

Application of Equation 2

What can we do with Equation 2? It will allow researchers to measure the featural, visual, and spatial strengths of novel stimulus types conveniently. Then, why would we want to do that? To see a specific reason, we can compare it with the previous way of separating the featural and visual strengths.

In the feature integration theory, the featural strength (i.e., attentional demand) of a stimulus type is measured as the slope of the RT-set size function. A flat slope (i.e., 0 ms/item) indicates parallel processing, whereas a steep slope (e.g., 50 or 100 ms/item) indicates serial processing. This straightforward approach has been challenged because the slope can also be affected by the visual quality of the stimulus. Specifically, even if a stimulus type is processed in parallel, if it is visually degraded then it will show a steep slope and appear to require serial scanning of focal attention (Palmer et al., 2000). Therefore, to determine the attentional demand of a stimulus type, one has to separate the effect of attentional demand from that of visual strength (e.g., Huang & Pashler, 2005; Palmer et al., 2005; Pashler & Badgio, 1985). Previous studies have tried to separate these two effects by building sophisticated computational models (Palmer, 1994; Palmer et al., 1993).

Needless to say, building a sophisticated computational model for each new case requires a lot of work. Now, powered by the FVS 2.0 framework, the featural and visual strengths can be separated very straightforwardly by using Equation 2.

Equation 2 has been applied in four follow-up studies, on surface orientation (Huang, 2021a), processing of motion (Zhang & Huang, 2021), learning effect on attentional processing (Huang & Pashler, 2021), as well as the relation between visual constancy and attention processing (Huang, 2021b). These studies will be published separately, but they play important roles in speculating the nature of featural strength and will be elaborated on below.

Estimation Based on Two Tasks

Although all the four above-mentioned follow-up studies have used Equation 2, one can certainly make flexible adjustments to the task set.¹⁸ For example, if one just wants to separate the

¹⁷ The fitting is relatively worse in the case of visual strength. This is probably due to the lack of control of eye fixations (see section “reasons for unexplained variance” in Appendix B). Luckily, the visual strength is rarely the theoretically critical element of a study.

¹⁸ More generally, researchers can choose a customized list of tasks. To give an anchor of visual strength, one of the two most heavily loaded tasks (monitoring or perceptual discrimination) should be chosen. Similarly, to give an anchor of spatial strength, change detection (or long-exposure change detection) should be chosen. Then, out of the many tasks that load heavily on featural strength (i.e., those close to the origin in Figure 11), researchers can choose a task that is conceptually most relevant to their research purpose. After choosing three tasks, the equations of estimations that are similar to the Equation 2 above can be determined by running three separate linear regressions to determine the parameters needed to estimate the FVS scores from the performances on these three chosen tasks. Of course, if logistically convenient, one can choose to run four or more tasks to obtain more precise measurements of the FVS scores.

contributions of featural and visual strengths and is not concerned about the spatial strength, then one can run only two tasks (perceptual discrimination and visual search) to extract the featural and visual strengths with Equation 3 as follows:

$$\begin{bmatrix} \text{Featural strength} \\ \text{Visual strength} \end{bmatrix} = \begin{bmatrix} 4.200 \\ -8.454 \end{bmatrix} + \begin{bmatrix} 3.778 & -1.950 \\ -1.540 & 6.220 \end{bmatrix} \cdot \begin{bmatrix} \text{Log Threshold}_{\text{VS}} \\ \text{Log Threshold}_{\text{PD}} \end{bmatrix}. \quad (3)$$

Nature of Featural Strength

Above, we discussed why we think these three factors should be matched to the concept of featural, visual, and spatial strengths. In this section, I will try to combine the present results with previous studies in the literature as well as related studies that will be published separately, and then try to make some broader speculations on the nature of featural strength. The next two sections will try to make speculations on the nature of visual and spatial strengths.

Visual Feature Is a General-Purpose Mechanism

As discussed above, Treisman's notion of preattentive features is a milestone in the studies of visual-attentional processing. As shown in Figure 11, the featural strength plays an important role in almost all tasks, and a dominant role in most tasks. In several of the tasks (saliency-based selection, ensemble processing, perceptual grouping, texture segregation), although one might have expected that the task is implemented by a specialized mechanism, it turns out that the general-purpose visual feature always plays a dominant role which is reflected by the lack of unexplained variances and also by the large L_F values.¹⁹

Apparently, the featural strength is even more general purpose and more ubiquitous than what is usually believed. Low-level motion is the only distinctive exception in the 26 tasks, followed by perceptual grouping which has a small but nontrivial portion (grouping: 11.6%; global grouping: 8.1%) of unexplained variances. Aside from these, there is no indication for separate sets of task-specific mechanisms.

The importance of featural strength is clearly not restricted to one specific function of cognitive processing. First, the most well-known role of visual features is the efficient selection (out of all the visual input) of the location(s) containing that feature (e.g., Wolfe, 1994). This role in the "feature-location routine" is illustrated by some of the present tasks (e.g., visual search, feature-based selection).

Second, the featural strength also defines the local saliency (i.e., the contrast between neighboring items).²⁰ This role in local saliency is illustrated by the saliency-based selection task that depends heavily ($L_F = 0.975$) and exclusively on featural strength.

Third, the featural strength also affects the efficiency of access to visual input on a prespecified location. This role in "location-feature routine"²¹ is illustrated by some of the present tasks (e.g., location-based selection, difficult location-based selection).

Fourth, the featural strength also affects the effectiveness of maintaining information in VWM. This role in VWM maintenance is illustrated by some of the present tasks (e.g., change detection,

feature VWM change). This mnemonic mechanism is conceptually separate from the processing of perceptual input, which includes all of the three aspects that have just been discussed.

Taken together, there are at least four separate functions in which the featural strength matters for visual-attentional processing. If these are implemented by four separate sets of features each specialized for an underlying function/mechanism, then it seems natural to expect that the featural strengths of different sets of features will be uncorrelated with each other. For example, we may expect that the color is used especially efficiently for one function, but the orientation is used especially efficiently for another function. However, the present results suggest that these effects all boil down to one single dimension of featural strength, and this would be a great coincidence from separate sets of features. Therefore, following the logical principle of Occam's razor, it is reasonable to conjecture that visual feature is a general-purpose code provided by one single centralized mechanism rather than implemented separately in function-specific mechanisms.

What is the nature of this general-purpose featural strength? Previous studies in support of this notion (e.g., Alvarez & Cavanagh, 2004) have not provided an explicit account. Perhaps, this general-purpose visual feature is provided by one single brain area and then used by various other brain areas to accomplish various functions. This point will be further elaborated below after all the other aspects of visual features have been discussed.

This notion of a general-purpose mechanism is consistent with some existing findings. If a feature is facilitated for one purpose of the task, then we can expect that the facilitation of this general-purpose mechanism will be generalized to other tasks. Priming of pop-out is a well-studied type of short-term facilitation of a specific target feature (Huang et al., 2004; Maljkovic & Nakayama, 1994; Meeter & Olivers, 2006; Kristjánsson et al., 2002). Yashar and Lamy (2010) reported that the priming effect was generalized between tasks, offering support for a general-purpose mechanism.

Visual Feature Is More Exclusive Than Assumed

In addition to the color and the shape, the usual list of visual features also includes the orientation and the motion (e.g., Wolfe, 1998a, 1998b; see also Wolfe & Horowitz, 2004, 2017). However, in the FVS 2.0 framework, it seems that the orientation and the motion should be removed from the list of basic preattentive features (Huang, 2021a; Zhang & Huang, 2021). Thus, the only items remaining on this list are the color and the shape. Therefore, the list of preattentive features is a lot more *exclusive* than what is usually believed.

¹⁹ The role of featural strength in these and other tasks will be discussed in the section "visual feature is a general-purpose mechanism" in Appendix B.

²⁰ For further discussions see "Global feature selection versus local saliency" in Appendix B.

²¹ It should be noted that the feature-location routine and location-feature routine are conceptually distinct from each other (see Huang & Pashler, 2007 for a detailed discussion). For an analogy, consider a telephone book that lists many names (corresponding to locations) in alphabetical order, providing a telephone number for each (corresponding to a featural value). This telephone book will support the function of "finding the phone number for a given name" which is similar to the location-feature routine. However, this telephone book, per se, will not support the function of "finding the name for a given phone number" which is similar to the feature-location routine. One will need to create a "reverse index" if one wants that additional function. Clearly, different mechanisms in the cognitive system are needed to accomplish these two computationally distinct functions.

Orientation has been widely believed to be a preattentive feature for visual-attentional processing. One classic finding in vision science by Hubel and Wiesel is that neurons in the primary visual cortex respond to the orientation of lines (e.g., Hubel & Wiesel, 1962). Subsequently, the responses of these neurons are well characterized by Gabor filters (e.g., Jones & Palmer, 1987). Consistently, it has also always been listed as one of the preattentive features for visual-attentional processing (e.g., Wolfe, 1998a, 1998b; see also Wolfe & Horowitz, 2004, 2017).

However, several studies have questioned the orientation's role as preattentive features. For example, Verghese and Nakayama (1993) showed that the performance on orientation stimulus drops substantially when there are more items, but there is almost no such drop for color stimulus. Consistently, Bilsky and Wolfe (1995, see also Wolfe et al., 1994) found that search for a whole-color-part-color target can be quite efficient, but the search for a whole-orientation-part-orientation target is always very laborious. Recently, Hulleman (2020) systematically compared the attention-guiding strengths of color, motion, and orientation by adding them to a T vs. L search. The results confirmed that color can effectively guide attention but the orientation cannot, and the motion is between them.

Besides, Nothdurft (1991, 1993; Wolfe et al., 1995; see also Inverso et al., 2016) argued that the attentional selection of orientations is mainly attributed to local saliency rather than to global feature selection, also confirming the ineffectiveness of orientation-based global selection.²²

Above, we suggested that the good performance on orientations in the change detection task is mainly attributed to the contribution of spatial strength rather than that of featural strength. The contributions of featural and spatial strengths can be separated by testing the surface orientation which, unlike the bar orientation, does not have an elongated profile so it is not expected to score high on spatial strength. Huang (2021a) used Equation 2 to estimate the FVS scores of the surface orientations and found that its featural strength score is rather low (average = 2.22 across eight spatial frequencies) and slightly lower than that of T s (2.64), so it is not a preattentive feature. Critically, the surface orientation's performance is poor in both change detection and visual search, so its featural strength will always be unambiguously low no matter how the factor structure is rotated in the featural-spatial plane.

As for motion, although many have demonstrated that the motion plays a special role in the control of attention (e.g., Franconeri & Simons, 2003), this is mainly limited to the idea that motion automatically attracts attention in a stationary environment. Zhang and Huang (2021) carried out a direct test on whether the motion is a preattentive feature in the sense that motion signals can be voluntarily selected as efficiently as other features (e.g., colors). Specifically, Zhang and Huang (2021) used Equation 2 to estimate the FVS scores of various types of motion signals. Surprisingly, none of the motion signals qualify as preattentive features.

Visual Feature Is Largely Innate

To ask whether the visual feature is innate or learnable, Huang and Pashler (2021) used Equation 2 to examine whether the featural strength of Chinese characters (and related visual stimuli) is higher in Chinese readers than nonreaders. It was found that Chinese readers perform better than nonreaders in two separate ways. First, there is a fairly large and general orientation-insensitive increase of

the visual strength for both characters and their parts, which seems to reflect a general low-level perceptual learning. Second, there is a statistically significant increase of featural strength for the upright whole character, but it can only go up to a moderate level (5.08), which is comparable to those of the size (5.97), the bar orientation (4.44), and the spatial frequency (4.61), but is much lower than those of strong features such as the color (8.86) or the shape (7.94). In other words, the visual feature is largely innate and strong preattentive features cannot be learned from experience.

Visual Feature Is Constancy Based

Although previous studies (Aks & Enns, 1996; Rensink & Enns, 1998) have suggested that attentional processing operates on postconstancy representations, there is also evidence for the effect of preconstancy representations on attentional processing (Moore & Brown, 2001). Huang (2021b) asked whether the preconstancy effects on attentional processing can be attributed to the effect of visual strength. Making use of a shadow-related color constancy phenomenon, Huang (2021b) used Equation 2 to estimate the FVS scores of physically identical color squares that are either perceived as degraded high-saturation colors or as low-saturation colors. It has been found that the featural strength of degraded high-saturation colors is always constantly high regardless of stimulus color difference, but that of low-saturation colors drops sharply when the stimulus color difference is smaller. Overall, it seems that the visual features are exclusively represented in the postconstancy level.

Visual Feature Works Like a Keyword

By definition, visual features are “keywords” extracted to reflect certain aspects of the stimulus. For example, the shape of an item is ultimately determined by the visual details which are distributed in many “pixels.” However, once extracted, the shape feature is an item-level label that can be used separately from the pixel-level information.

This implies that, for each type of stimulus information, there is a potential distinction between the stimulus variations that are picked up by the visual system as features and those that are not. In FVS 2.0 framework's terms, this means that, in each visual domain such as colors or shapes, it should be possible to distinguish between stimulus variations that are high in featural strength and those that are not.

Huang (2020b) elaborated this theoretical distinction in the case of shapes. Specifically, there is a clear distinction between preattentive shape features that can be extracted in parallel and efficiently used in guiding attention (e.g., Julesz, 1984, 1986) versus the postattentive processing of an object's shape (e.g., Biederman, 1987) that can only occur when it is in the current focus of attention.

The present study gives clear support for the distinction between “preattentive shape features” and “postattentive processing of shapes.” In this context, the “ T s” is a typical example of shape information that is *not* extracted as shape features. As shown in Figures 6–7, the performance on the shape (i.e., preattentive shape features) is much better than that on the T s in high-level²³ tasks (e.g., visual search: 81 vs. 335 ms; texture segregation: 123 vs. 531 ms; feature-based

²² However, it seems that the critical factor is not local saliency, but the homogeneity between targets. This point will be elaborated in the section “global feature selection versus local saliency” in Appendix B.

²³ A more formal definition of high-level and low-level tasks can be found in the section “high-level versus low-level tasks” in Appendix B.

selection: 185 vs. 1222 ms). But there is no such difference at all in low-level tasks (e.g., perceptual discrimination: 48 vs. 47 ms; monitoring: 15 vs. 13 ms).

These results suggest that the preattentive shape feature is not a proportional reduction of all aspects of shapes, but is a special subset of the latter set of the information. By making an analogy to the search engine of an academic database, preattentive shape features correspond to the keywords that can be used to make a search, whereas the “postattentive shape processing” corresponds to the complete content of an individual article that can only be displayed one at a time. Only the keywords, but not all details of the articles can be used to search for a target article. Moreover, the keywords of an article are not the proportional reduction of all the contents of that article, but a special subset of it.

A GEICK Conjecture for Visual Features

To summarize, the results in the present study and related studies suggest that visual features have five unique aspects as follows:

1. (*General purpose*): They are used ubiquitously for all tasks.
2. (*Exclusive*): Only the color and the shape are strong preattentive features.
3. (*Innate*): One cannot learn a preattentive feature from experience.
4. (*Constancy based*): It is exclusively in postconstancy representations.
5. (*Keyword like*): It works like a keyword.

It should be made clear that, out of these five claims, only the first one (general purpose) follows directly from the results reported in the present study. The rest four claims are more or less based on findings that will be (or have been) published separately. Besides, these claims are admittedly speculative and the validities of them should be further tested in future studies.

To summarize these five claims, we can come up with a GEICK conjecture of the underlying neural mechanism of visual features. *General-purpose* visual features that are used by many brain areas for various functions are all provided by one single centralized “feature area” in the brain. What is most remarkable among these functions is the efficiency of the feature-based selection of locations. But in general, the degree of being picked up by this feature area, namely the featural strength, also determines the efficiency of processing for all purposes including the calculation of local saliency, the access to a feature on a location, and the efficiency of working memory representations. This feature area extracts a special subset (i.e., *keyword*) of information from anatomically earlier areas. This subset is *exclusive* and is only concerned about post-*constancy* colors and (some aspects of) shapes. The other important information that is extracted by other brain areas (e.g., orientation by V1, motion by middle temporal visual area (MT), face by Fusiform face area (FFA) are not picked up by this feature area and are therefore generally inefficient in attention-demanding tasks. The content of visual features is largely *innate* and there is only a modest learning effect.

This concept of a centralized feature area takes an approach that is quite different from the traditional framework of visual attention in which the visual features are embedded parts of the attentional

network. For example, in previous frameworks, the colors can be found efficiently because there are specialized mechanisms that connect color to locations containing that color. However, in this centralized feature area, attentional networks (or mechanisms for other purposes such as maintenance of VWM) are not tied to colors per se, but make use of whatever features (i.e., general-purpose codes) provided by the centralized feature area.

Naturally, one would ask, where is this centralized feature area? Given it is most sensitive to the color and the shape, it seems reasonable to hypothesize that it is the visual area V4 (see Burrows & Moore 2009; Mazer & Gallant 2003 for potentially relevant studies) and/or IT (inferior temporal gyrus) because of their high relevance to the processing of colors and shapes (Ghose & Ts'o, 1997; Heywood et al., 1992; Kayaert et al., 2003, 2005; Komatsu et al., 1992; Op de Beeck et al., 2001; Pasupathy & Connor, 2002; Yamane et al., 2008; Zeki, 1973, 1980). Information extracted by the other mechanisms, including the earlier areas (V1 for orientation and spatial frequencies) and some dorsal areas (e.g., MT for motion) do *not* provide preattentive visual features. Moreover, there is evidence that these areas represent postconstancy information (e.g., Zeki, 1980) and this is also consistent with the proposed function of the feature area.

One may point out that it is the dorsal pathway that is mainly responsible for visual-attentional processing. However, the idea that the operator of attentional processing is in the dorsal pathway is not in conflict with the idea that the database of the attentional processing is represented in the ventral pathway. After all, the visual feature, although commonly associated with attentional processing, is still a type of visual identity itself, so it is natural for visual features to be represented in the ventral pathway. Besides, Zhang and Huang (2021) argued that the motion, which is processed in the MT area in the dorsal pathway, is not a preattentive feature. Therefore, there is no reason to believe that there should be a dorsal privilege for featural processing.

This proposed feature area is very consistent with the reverse hierarchy theory (Hochstein & Ahissar, 2002). In the reverse hierarchy, there is a general dissociation between the feature detectors in early visual areas and preattentive features in visual-attentional processing. The conscious-level visual-attentional processing cannot readily make use of the features extracted by the anatomically early visual areas, but instead relies on the features generated by the anatomically late areas.

All in all, following the synthesis of the available clues, the GEICK conjecture seems to be the most plausible account for the role of visual features. Of course, the GEICK (and especially the neural substrate) is not a conclusion, but merely a conjecture as it is, and needs to be tested in future studies.

Nature of Visual Strength

To start, in case it is not self-evident, the visual strength of stimulus reflects the visual quality of the data, so it is *not* an innate property of the stimulus type. For example, the high-contrast and low-contrast spatial frequencies are technically the same type of stimulus, but they significantly differ on visual strength. Conceivably, visual strength will probably also be affected by other factors that are irrelevant to the definition of a stimulus type. For example, if the colors are presented in tiny objects, the visual strength of this set of color stimulus items will probably also decrease.

Having said that, this does not mean that visual strength is not important for our understanding of visual-attentional processing. Visual strength makes large or small contributions to most of the tasks. Therefore, when research questions are asked about featural or spatial strength, understanding the role of visual strength is important for at least the purpose of factoring out its contribution.

What determines the visual strength of a stimulus type? As mentioned above, the high-contrast–low-contrast stimulus pairs mainly differ in terms of their visual strength, so stimulus contrast affects visual strength. However, stimulus contrast is not the only, or even the dominant, factor affecting visual strength. The different categories of stimulus differ greatly in their visual strengths. For example, the low-contrast orientation stimulus has a much higher visual strength than even the high-contrast spatial-frequency stimulus.

What accounts for these between-category differences in visual strength? A plausible account is the distribution of spatial-frequency components. Psychophysical measurements showed that human adults are most sensitive to the spatial-frequency range of 1~5 cycles/degree. The orientation (as well as *T*s and shapes) stimuli in the present study have used line segments with spatial-frequency components that happen to be consistent with this optimal range. On the other hand, the spatial-frequency components of size and spatial-frequency stimulus are out of this optimal range. The size stimulus has used very narrow line segments, so its main spatial-frequency components are higher than the optimal range.²⁴ The spatial-frequency stimulus extends across a broad range of spatial frequencies, so two or three of the four items are outside the optimal range.

Altogether, it seems that the stimulus's variations on visual strength can be reasonably accounted for by the traditional findings on contrast sensitivity function. To fully verify this, future work will be needed to systematically vary the stimulus on these dimensions (i.e., contrast and spatial-frequency components) and try to provide a quantitative model that can calculate the visual strength of stimulus based on spatial-frequency analysis of the image information.

Nature of Spatial Strength

Among the three factors, the concept of featural strength is directly inherited from the previous theories on visual attention (e.g., Treisman & Gelade, 1980; Wolfe, 1994), and the concept of visual strength has also been well considered in these theories. However, although intuitively compelling, the concept of spatial strength (i.e., a spatial structure reflecting profiles of individual items) has hardly been considered in these theories. In these theories, orientations (or other features) are extracted from stimulus items as features, the visual details of items are usually only considered as “raw materials” for feature extraction, and the “locations of items” usually only mean a set of dots marking the centers of these items. In some sense, the spatial strength is the critical aspect on which the FVS framework differs from the previous theories of visual attention.

As mentioned above, the Boolean map structure is assumed to be the underlying mechanism for the spatial strength. Two points need to be elaborated on its nature. First, this Boolean map structure is capacity limited.

Consistent with the results of Huang (2015a), the shape's spatial strength score is rather low. One may find this counter intuitive. A shape is a spatial structure by definition. Naturally, it should have a large spatial strength score, right? As discussed above concerning Figure 2, the capacity-limited Boolean map structure can capture the

rough sketch of the profiles and represent the bar orientations quite well, but cannot represent the detailed shapes of individual items. That explains why the shape's spatial strength score is rather low. In other words, the spatial strength is *not* an index of the general “spatial analysis” but reflects the role of the capacity-limited Boolean map structure that works in the very specific way described in Figure 2.

Second, this Boolean map structure represents the locations globally rather than on the level of individual items.

The Boolean map is conceptualized as a data structure that simultaneously represents multiple locations (Huang & Pashler, 2007). This implies that, when a Boolean map is used to represent multiple bars, the locations defining individual bars are all directly represented in one global Boolean map structure rather than on the level of individual bars (e.g., see Figure 2). If this is true, then although this Boolean map structure can be effectively used to compare global patterns in a change detection task, it will be ineffective for checking whether an individual item belongs to this global pattern or not. In other words, the spatial strength-based advantage in the change detection task is expected to disappear if the memory probe involves an individual item rather than the global structure.

This prediction has indeed been confirmed by the results of the VWM probe task. As shown in Figure 11, the use of an individual-item probe leads to a substantial reduction of the L_S (0.733 in change detection \rightarrow 0.225 in VWM probe). For a representative example, the memory for orientations is much worse in the VWM probe task than in the change detection task (0.651 vs. 0.772) even if the memory for colors is no worse in the former than the latter (0.706 vs. 0.699).

This demonstration of the global nature of the Boolean map structure provides distinctive support for the present conceptual interpretation of the spatial strength. Specifically, the high-spatial-strength stimulus types (e.g., orientations) must not be special for any reason on the individual-item level because individual-item-level factors should always have equal effects on change detection and VWM probe. Therefore, the possible account of spatial strength is narrowed down to some type of global multiple-item representation mechanism, making the Boolean map structure (as illustrated in Figure 2) a very plausible candidate.

More broadly, this result is also important for the development of the Boolean map theory because it has confirmed an important claim on the nature of the Boolean map structure. Namely, the Boolean map structure is a global spatial structure that does not give direct information about the features of individual items (Huang & Pashler, 2007).

Experimental Nature of Individual-Item Differences Analysis

It should be made clear that, although the present study has used PCA which is usually used in correlational research, the present study is fundamentally still not correlational research, but experimental research. Imagine if we take only 2 out of the 16 stimulus types and 3 out of the 26 tasks, then that would be a typical 2×3 factorial design (see Huang, 2015b, for such an example). To scale this up, the present study is a 16×26 factorial design which is still

²⁴ This again illustrates the point that the visual strength is not an innate property of the stimulus type. For sizes, if the circles are presented with thicker lines, the visual strength will probably be higher.

in the category of experimental research. In this context, the FVS 2.0 framework is a parsimonious linear model to account for the IV stimulus type's effects in all tasks, and the PCA is a mathematical tool that helps us to find this parsimonious model. Certainly, the nature of the design (experimental rather than correlational) is not changed by the mathematical tool used for data analysis.

To see the implications of this experimental/correlational distinction, let's first briefly review the difference between the logic of experimental research and that of correlational research. In correlational research, all variables are *measured* on a random sample of human participants, whereas in experimental research, one or more IVs are *manipulated*.

On the one hand, experimental research has an advantage over correlational research: The IVs are manipulated, so causal relationships can be established. On the other hand, correlational research also has an advantage over experimental research. In random sampling, we can usually assume that individual participants in a sample are effectively independent and identically distributed (i.i.d.) so that the findings can be generalized to a new set of participants. But this i.i.d. *cannot* be assumed for the IVs in experimental research because those are not measured on a random sample, but manipulated by the researcher. In other words, it is not statistically guaranteed that findings can be generalized to a new set of IVs. Of course, this lack of i.i.d. of IVs does *not* mean that generalizability is unachievable in experimental research, but just it has to be achieved on the conceptual level in addition to the statistical level. For example, by making sure that an IV precisely reflects the concept it tries to capture, we can be reasonably sure that the finding can be generalized to other IVs that also precisely reflect this concept.

We can see how both the advantages and disadvantages of experimental research apply to the present study. On the one hand, causal relationships have been established: The different performances on the 416 conditions (16 stimulus types \times 26 tasks) are caused by these manipulations. On the other hand, i.i.d. cannot be assumed for the IV stimulus type so it is not statistically guaranteed that the FVS 2.0 framework can be generalized to other stimulus types.

As discussed above, in experimental research, an IV's generalizability has to be achieved on the conceptual level. In the present study, the relations between the stimulus types and the three factors (featural, visual, and spatial strengths) have been justified based on conceptual interpretations and a broad range of previous studies on visual-attentional processing. However, in this research area, there is no existing way of conceptualizing the various stimulus types in a unified framework, which is exactly what the present study is trying to achieve. Therefore, the conceptual justifications are subject to certain limitations, and it is certainly possible that new stimulus types can bring in new factors and/or change the three existing factors. Future studies will be needed to find this out.

Concluding Comments

The present study greatly expanded the scope of Huang (2015a) and revealed an FVS 2.0 framework that quantitatively determines the factors governing the ways how 16 types of visual stimulus information are processed in 25 tasks. Three factors, namely the featural, visual, and spatial strength have been identified. Featural strength seems to reflect the function of general-purpose codes that are provided by one single centralized "feature area" which is the V4

and/or IT areas in the brain. These general-purpose codes are then used by many brain areas for various functions including the global feature-based selection, calculation of local saliency, access to a feature on a location, and the efficiency of working memory representations. Only colors and (some aspects of) shapes are extracted by this feature area as preattentive features, but other important information extracted by other brain areas (e.g., orientation by V1, motion by MT) are not preattentive features. Visual strength seems to be determined by the low-level stimulus properties such as spatial-frequency components and contrasts. Spatial strength seems to reflect the role of a capacity-limited Boolean map structure that represents the locations globally.

Context of the Research

The present study is an important part of my long-term research agenda in which I am attempting to develop a general framework for understanding visual-attentional processing. In an earlier attempt, my collaborators and I proposed a Boolean map theory (Huang & Pashler, 2007; Huang et al., 2007) which starts with a set of parsimonious principles (the data format of a Boolean map and a few principles on how it is created) but has implications for a range of attention-related processes. Although progress has been made on various aspects, the Boolean map theory is defined on the conceptual level, so it is subject to the inherent ambiguities in verbal descriptions. In the FVS 2.0 framework, one important aspect of the Boolean map theory (i.e., the use of the Boolean map structure in various tasks) is quantified as the spatial strength and related to other known factors (featural and visual strengths of the stimulus information). This knowledge will be an important piece of the future framework that can hopefully quantify and connect all the conceptual elements in the Boolean map theory.

References

- Aks, D. J., & Enns, J. T. (1996). Visual search for size is influenced by a background texture gradient. *Journal of Experimental Psychology: Human Perception and Performance*, 22(6), 1467–1481. <https://doi.org/10.1037/0096-1523.22.6.1467>
- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15(3), 122–131. <https://doi.org/10.1016/j.tics.2011.01.003>
- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, 15(2), 106–111. <https://doi.org/10.1111/j.0963-7214.2004.01502006.x>
- Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, 19(4), 392–398. <https://doi.org/10.1111/j.1467-9280.2008.02098.x>
- Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences of the United States of America*, 106(18), 7345–7350. <https://doi.org/10.1073/pnas.0808981106>
- Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-Down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, 16(8), 437–443. <https://doi.org/10.1016/j.tics.2012.06.010>
- Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of vision*, 11(10), 6. <https://doi.org/10.1167/11.10.6>

- Beck, J. (1966). Perceptual grouping produced by changes in orientation and shape. *Science*, *154*(3748), 538–540. <https://doi.org/10.1126/science.154.3748.538>
- Beck, J. (1980). *Texture segregation*. University of Maryland.
- Biederman, I. (1987). Recognition-by-Components - a Theory of Human Image Understanding. *Psychological Review*, *94*(2), 115–147. <https://doi.org/10.1037/0033-295X.94.2.115>
- Bilsky, A. B., & Wolfe, J. M. (1995). Part-Whole information is useful in visual search for size x size but not orientation x orientation conjunctions. *Perception & Psychophysics*, *57*(6), 749–760. <https://doi.org/10.3758/BF03206791>
- Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., Hogg, R. E., & Mollon, J. D. (2015). A population study of binocular function. *Vision Research*, *110*, 34–50. <https://doi.org/10.1016/j.visres.2015.02.017>
- Burrows, B. E., & Moore, T. (2009). Influence and limitations of popout in the selection of salient visual stimuli by area V4 neurons. *Journal of Neuroscience*, *29*(48), 15169–15177. <https://doi.org/10.1523/JNEUROSCI.3710-09.2009>
- Cappe, C., Clarke, A., Mohr, C., & Herzog, M. H. (2014). Is there a common factor for vision? *Journal of Vision*, *14*(8), Article 4. <https://doi.org/10.1167/14.8.4>
- Cheal, M., & Lyon, D. R. (1994). Allocation of attention in texture segregation, visual search, and location-precuing paradigms. *The Quarterly Journal of Experimental Psychology*, *47*(1), 49–70. <https://doi.org/10.1080/14640749408401143>
- Chen, H., & Wyble, B. (2015). The location but not the attributes of visual cues are automatically encoded into working memory. *Vision Research*, *107*, 76–85. <https://doi.org/10.1016/j.visres.2014.11.010>
- Chen, H., & Wyble, B. (2018). The neglected contribution of memory encoding in spatial cueing: A new theory of costs and benefits. *Psychological Review*, *125*(6), 936–968. <https://doi.org/10.1037/rev0000116>
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404. [https://doi.org/10.1016/S0042-6989\(02\)00596-5](https://doi.org/10.1016/S0042-6989(02)00596-5)
- Chong, S. C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, *67*(1), 1–13. <https://doi.org/10.3758/BF03195009>
- Chong, S. C., & Treisman, A. (2005b). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, *45*(7), 891–900. <https://doi.org/10.1016/j.visres.2004.10.004>
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–185. <https://doi.org/10.1017/s0140525x01003922>
- Cretenoud, A. F., Karimpur, H., Grzeczowski, L., Francis, G., Hamburger, K., & Herzog, M. H. (2019). Factors underlying visual illusions are illusion-specific but not feature-specific. *Journal of Vision*, *19*(14), Article 12. <https://doi.org/10.1167/19.14.12>
- Donk, M., & van Zoest, W. (2008). Effects of salience are short-lived. *Psychological Science*, *19*(7), 733–739. <https://doi.org/10.1111/j.1467-9280.2008.02149.x>
- Egeth, H., & Dagenbach, D. (1991). Parallel versus serial processing in visual search: Further evidence from subadditive effects of visual quality. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(2), 551–560. <https://doi.org/10.1037/0096-1523.17.2.551>
- Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of Vision*, *11*(12), Article 3. <https://doi.org/10.1167/11.12.3>
- Fougnie, D., Cormiea, S. M., & Alvarez, G. A. (2013). Object-Based benefits without object-based representations. *Journal of Experimental Psychology: General*, *142*(3), 621–626. <https://doi.org/10.1037/a0030300>
- Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics*, *65*(7), 999–1010. <https://doi.org/10.3758/BF03194829>
- Ghose, G. M., & Ts'o, D. Y. (1997). Form processing modules in primate area V4. *Journal of Neurophysiology*, *77*(4), 2191–2196. <https://doi.org/10.1152/jn.1997.77.4.2191>
- Goodbourn, P. T., Bosten, J. M., Hogg, R. E., Bargary, G., Lawrance-Owen, A. J., & Mollon, J. D. (2012). Do different “magnocellular tasks” probe the same neural substrate? *Proceedings. Biological Sciences*, *279*(1745), 4263–4271. <https://doi.org/10.1098/rspb.2012.1430>
- Grzeczowski, L., Clarke, A. M., Francis, G., Mast, F. W., & Herzog, M. H. (2017). About individual differences in vision. *Vision Research*, *141*, 282–292. <https://doi.org/10.1016/j.visres.2016.10.006>
- Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General*, *144*(2), 432–446. <https://doi.org/10.1037/xge0000053>
- Herzog, M. H., & Fahle, M. (2002). Effects of grouping in contextual modulation. *Nature*, *415*(6870), 433–436. <https://doi.org/10.1038/415433a>
- Heywood, C. A., Gadotti, A., & Cowey, A. (1992). Cortical area V4 and its role in the perception of color. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *12*(10), 4056–4065. <https://doi.org/10.1523/JNEUROSCI.12-10-04056.1992>
- Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804. [https://doi.org/10.1016/S0896-6273\(02\)01091-7](https://doi.org/10.1016/S0896-6273(02)01091-7)
- Holcombe, A. O. (2009). Seeing slow and seeing fast: Two limits on perception. *Trends in Cognitive Sciences*, *13*(5), 216–221. <https://doi.org/10.1016/j.tics.2009.02.005>
- Holcombe, A. O., & Cavanagh, P. (2001). Early binding of feature pairs for visual perception. *Nature Neuroscience*, *4*(2), 127–128. <https://doi.org/10.1038/83945>
- Holcombe, A. O., & Judson, J. (2007). Visual binding of English and Chinese word parts is limited to low temporal frequencies. *Perception*, *36*(1), 49–74. <https://doi.org/10.1068/p5582>
- Huang, L. (2010a). What is the unit of visual attention? Object for selection, but Boolean map for access. *Journal of Experimental Psychology: General*, *139*(1), 162–179. <https://doi.org/10.1037/a0018034>
- Huang, L. (2010b). Characterizing the nature of visual conscious access: The distinction between features and locations. *Journal of Vision*, *10*(10), Article 24. <https://doi.org/10.1167/10.10.24>
- Huang, L. (2010c). The Speed of Feature-Based Attention: Attentional Advantage Is Slow but Selection Is Fast. *Journal of Experimental Psychology-Human Perception and Performance*, *36*, 1382–1390. <https://doi.org/10.1037/a0018736>
- Huang, L. (2015a). Visual features for perception, attention, and working memory: Toward a three-factor framework. *Cognition*, *145*, 43–52. <https://doi.org/10.1016/j.cognition.2015.08.007>
- Huang, L. (2015b). Color is processed less efficiently than orientation in change detection but more efficiently in visual search. *Psychological Science*, *26*(5), 646–652. <https://doi.org/10.1177/0956797615569577>
- Huang, L. (2015c). Visual features: Featural strength and visual strength are two dissociable dimensions. *Scientific Reports*, *5*, Article 13769. <https://doi.org/10.1038/srep13769>
- Huang, L. (2015d). Grouping by similarity is mediated by feature selection: Evidence from the failure of cue combination. *Psychonomic Bulletin & Review*, *22*, 1364–1369. <https://doi.org/10.3758/s13423-015-0801-z>
- Huang, L. (2015e). Statistical properties demand as much attention as object features. *PLOS ONE*, *10*(8), Article e0131191. <https://doi.org/10.1371/journal.pone.0131191>
- Huang, L. (2020a). Unit of visual working memory: A Boolean map provides a better account than an object does. *Journal of Experimental Psychology: General*, *149*(1), 1–30. <https://doi.org/10.1037/xge0000616>
- Huang, L. (2020b). Space of preattentive shape features. *Journal of Vision*, *20*(4), Article 10. <https://doi.org/10.1167/jov.20.4.10>

- Huang, L. (2021a). *Orientation is not a preattentive feature* [Manuscript submitted for publication]. Department of Psychology, the Chinese University of Hong Kong.
- Huang, L. (2021b). *What you attend is what you see: Visual features are post-constancy* [Manuscript submitted for publication]. Department of Psychology, the Chinese University of Hong Kong.
- Huang, L. (2021c, May 27). *Data and analysis for the paper "FVS 2.0: A Unifying Framework for Understanding the Factors of Visual Processing."* osf.io/9eafp
- Huang, L., Holcombe, A. O., & Pashler, H. (2004). Repetition priming in visual search: Episodic retrieval, not feature priming. *Memory & Cognition*, 32(1), 12–20. <https://doi.org/10.3758/BF03195816>
- Huang, L., Mo, L., & Li, Y. (2012). Measuring the interrelations among multiple paradigms of visual attention: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance*, 38(2), 414–428. <https://doi.org/10.1037/a0026314>
- Huang, L., & Pashler, H. (2005). Attention capacity and task difficulty in visual search. *Cognition*, 94(3), B101–B111. <https://doi.org/10.1016/j.cognition.2004.06.006>
- Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review*, 114(3), 599–631. <https://doi.org/10.1037/0033-295X.114.3.599>
- Huang, L., & Pashler, H. (2012). Distinguishing different strategies of across-dimension attentional selection. *Journal of Experimental Psychology-Human Perception and Performance*, 38, 453–464. <https://doi.org/10.1037/a0026365>
- Huang, L., & Pashler, H. (2021). *New preattentive feature can hardly be learned* [Manuscript submitted for publication]. Department of Psychology, the Chinese University of Hong Kong.
- Huang, L., Treisman, A., & Pashler, H. (2007). Characterizing the limits of human visual awareness. *Science*, 317(5839), 823–825. <https://doi.org/10.1126/science.1143515>
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1), 106–154. <https://doi.org/10.1113/jphysiol.1962.sp006837>
- Hulleman, J. (2020). Quantitative and qualitative differences in the top-down guiding attributes of visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 46, 942–964. <https://doi.org/10.1037/xhp0000764>
- Inverso, M., Sun, P., Chubb, C., Wright, C. E., & Sperling, G. (2016). Evidence against global attention filters selective for absolute bar-orientation in human vision. *Attention, Perception & Psychophysics*, 78(1), 293–308. <https://doi.org/10.3758/s13414-015-1005-3>
- Jackson-Nielsen, M., Cohen, M. A., & Pitts, M. A. (2017). Perception of ensemble statistics requires attention. *Consciousness and Cognition*, 48, 149–160. <https://doi.org/10.1016/j.concog.2016.11.007>
- Johnston, J. C., & Pashler, H. (1990). Close Binding of Identity and Location in Visual Feature Perception. *Journal of Experimental Psychology-Human Perception and Performance*, 16(4), 843–856. <https://doi.org/10.1037/0096-1523.16.4.843>
- Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6), 1233–1258. <https://doi.org/10.1152/jn.1987.58.6.1233>
- Julesz, B. (1984). A brief outline of the texton theory of human vision. *Trends in Neurosciences*, 7(2), 41–45. [https://doi.org/10.1016/S0166-2236\(84\)80275-1](https://doi.org/10.1016/S0166-2236(84)80275-1)
- Julesz, B. (1986). Texton gradients: The texton theory revisited. *Biological Cybernetics*, 54(4), 245–251. <https://doi.org/10.1007/BF00318420>
- Julesz, B., & Bergen, J. R. (1983). Textons, the fundamental elements in preattentive vision and perception of textures. *The Bell System Technical Journal*, 62(2), 1619–1645. <https://doi.org/10.1002/j.1538-7305.1983.tb03502.x>
- Kayaert, G., Biederman, I., & Vogels, R. (2003). Shape tuning in macaque inferior temporal cortex. *Journal of Neuroscience*, 23(7), 3016–3027. <https://doi.org/10.1523/jneurosci.23-07-03016.2003>
- Kayaert, G., Biederman, I., Op de Beeck, H. P., & Vogels, R. (2005). Tuning for shape dimensions in macaque inferior temporal cortex. *The European Journal of Neuroscience*, 22(1), 212–224. <https://doi.org/10.1111/j.1460-9568.2005.04202.x>
- Komatsu, H., Ideura, Y., Kaji, S., & Yamane, S. (1992). Color selectivity of neurons in the inferior temporal cortex of the awake macaque monkey. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 12(2), 408–424. <https://doi.org/10.1523/JNEUROSCI.12-02-00408.1992>
- Kristjánsson, A., Wang, D., & Nakayama, K. (2002). The role of priming in conjunctive visual search. *Cognition*, 85(1), 37–52. [https://doi.org/10.1016/s0010-0277\(02\)00074-4](https://doi.org/10.1016/s0010-0277(02)00074-4)
- Kubovy, M., & van den Berg, M. (2008). The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychological Review*, 115(1), 131–154. <https://doi.org/10.1037/0033-295X.115.1.131>
- Kwak, H.-W., Dagenbach, D., & Egeth, H. (1991). Further evidence for a time-independent shift of the focus of attention. *Perception & Psychophysics*, 49(5), 473–480. <https://doi.org/10.3758/BF03212181>
- Levinthal, B. R., & Franconeri, S. L. (2011). Common-Fate grouping as feature selection. *Psychological Science*, 22(9), 1132–1137. <https://doi.org/10.1177/0956797611418346>
- Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, 6(1), 9–16. [https://doi.org/10.1016/S1364-6613\(00\)01817-9](https://doi.org/10.1016/S1364-6613(00)01817-9)
- Li, Z., Xin, K., Lou, J., & Li, Z. (2019). Visual search may not require target representation in working memory or long-term memory. *Psychological Science*, 30(10), 1547–1555. <https://doi.org/10.1177/0956797619872749>
- Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science*, 240(4853), 740–749. <https://doi.org/10.1126/science.3283936>
- Lu, Z. L., & Sperling, G. (1995). The functional architecture of human visual motion perception. *Vision Research*, 35(19), 2697–2722. [https://doi.org/10.1016/0042-6989\(95\)00025-U](https://doi.org/10.1016/0042-6989(95)00025-U)
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279–281. <https://doi.org/10.1038/36846>
- Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, 22(6), 657–672. <https://doi.org/10.3758/BF03209251>
- Manassi, M., Sayim, B., & Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision*, 12(10), Article 13. <https://doi.org/10.1167/12.10.13>
- Mazer, J. A., & Gallant, J. L. (2003). Goal-Related activity in V4 during free viewing visual search: Evidence for a ventral stream visual salience map. *Neuron*, 40(6), 1241–1250. [https://doi.org/10.1016/S0896-6273\(03\)00764-5](https://doi.org/10.1016/S0896-6273(03)00764-5)
- McGaugh, J. L. (2000). Memory—A century of consolidation. *Science*, 287(5451), 248–251. <https://doi.org/10.1126/science.287.5451.248>
- Meeter, M., & Olivers, C. N. (2006). Intertrial priming stemming from ambiguity: A new account of priming in visual search. *Visual Cognition*, 13(2), 202–222. <https://doi.org/10.1080/13506280500277488>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100. <https://doi.org/10.1006/cogp.1999.0734>
- Moore, C. M., & Brown, L. E. (2001). Preconstancy information can influence visual search: The case of lightness constancy. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 178–194. <https://doi.org/10.1037/0096-1523.27.1.178>

- Moore, C. M., & Egeth, H. (1998). How does feature-based attention affect visual processing? *Journal of Experimental Psychology-Human Perception and Performance*, *24*, 1296–1310. <https://doi.org/10.1037/0096-1523.24.4.1296>
- Moore, C. M., Egeth, H., Berglan, L. R., & Luck, S. J. (1996). Are attentional dwell times inconsistent with serial visual search? *Psychonomic Bulletin & Review*, *3*(3), 360–365. <https://doi.org/10.3758/BF03210761>
- Müller, H. J., Geyer, T., Zehetleitner, M., & Krummenacher, J. (2009). Attentional capture by salient color singleton distractors is modulated by top-down dimensional set. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 1–16. <https://doi.org/10.1037/0096-1523.35.1.1>
- Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics*, *70*(5), 772–788. <https://doi.org/10.3758/PP.70.5.772>
- Nissen, M. J. (1985). Accessing features and objects: Is location special. In M. I. Posner & O. S. M. Martin (Eds.), *Attention and Performance XI* (pp. 205–219). Erlbaum.
- Nothdurft, H. C. (1991). Texture segmentation and pop-out from orientation contrast. *Vision Research*, *31*(6), 1073–1078. [https://doi.org/10.1016/0042-6989\(91\)90211-M](https://doi.org/10.1016/0042-6989(91)90211-M)
- Nothdurft, H.-C. (1993). The role of features in preattentive vision: Comparison of orientation, motion and color cues. *Vision Research*, *33*(14), 1937–1958. [https://doi.org/10.1016/0042-6989\(93\)90020-W](https://doi.org/10.1016/0042-6989(93)90020-W)
- Op de Beeck, H., Wagemans, J., & Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nature Neuroscience*, *4*(12), 1244–1252. <https://doi.org/10.1038/nn767>
- Palmer, J. (1994). Set-Size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research*, *34*(13), 1703–1721. [https://doi.org/10.1016/0042-6989\(94\)90128-7](https://doi.org/10.1016/0042-6989(94)90128-7)
- Palmer, J., Ames, C. T., & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(1), 108–130. <https://doi.org/10.1037/0096-1523.19.1.108>
- Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*(1), 376–404. <https://doi.org/10.1167/5.5.1>
- Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, *40*(10), 1227–1268. [https://doi.org/10.1016/S0042-6989\(99\)00244-8](https://doi.org/10.1016/S0042-6989(99)00244-8)
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*(4), 369–378. <https://doi.org/10.3758/bf03210419>
- Pashler, H., & Badgio, P. C. (1985). Visual attention and stimulus identification. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(2), 105–121. <https://doi.org/10.1037/0096-1523.11.2.105>
- Pashler, H. E. (1999). *The psychology of attention*. MIT press. <https://doi.org/10.7551/mitpress/5677.001.0001>
- Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience*, *5*(12), 1332–1338. <https://doi.org/10.1038/972>
- Pomerantz, J. R. (2003). Wholes, holes, and basic features in vision. *Trends in Cognitive Sciences*, *7*(11), 471–473. <https://doi.org/10.1016/j.tics.2003.09.007>
- Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(3), 422–435. <https://doi.org/10.1037/0096-1523.3.3.422>
- Quinlan, P. T. (2003). Visual feature integration theory: Past, present, and future. *Psychological Bulletin*, *129*, 643–673. <https://doi.org/10.1037/0033-2909.129.5.643>
- Rangelov, D., Müller, H. J., & Zehetleitner, M. (2017). Failure to pop out: Feature singletons do not capture attention under low signal-to-noise ratio conditions. *Journal of Experimental Psychology: General*, *146*(5), 651–671. <https://doi.org/10.1037/xge0000284>
- Rensink, R. A., & Enns, J. T. (1998). Early completion of occluded objects. *Vision Research*, *38*(15), 2489–2505. [https://doi.org/10.1016/S0042-6989\(98\)00051-0](https://doi.org/10.1016/S0042-6989(98)00051-0)
- Rosenberg, M. D., Finn, E. S., Scheinost, D., Constable, R. T., & Chun, M. M. (2017). Characterizing attention with predictive network models. *Trends in Cognitive Sciences*, *21*(4), 290–302. <https://doi.org/10.1016/j.tics.2017.01.011>
- Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*, *19*(1), 165–171. <https://doi.org/10.1038/nn.4179>
- Sayim, B., Westheimer, G., & Herzog, M. H. (2010). Gestalt factors modulate basic spatial vision. *Psychological Science*, *21*(5), 641–644. <https://doi.org/10.1177/0956797610368811>
- Serences, J. T., & Yantis, S. (2006). Selective visual attention and perceptual coherence. *Trends in Cognitive Sciences*, *10*(1), 38–45. <https://doi.org/10.1016/j.tics.2005.11.008>
- Shaqiri, A., Pilz, K. S., Cretenoud, A. F., Neumann, K., Clarke, A., Kunchulia, M., & Herzog, M. H. (2019). No evidence for a common factor underlying visual abilities in healthy older people. *Developmental Psychology*, *55*, 1775–1787. <https://doi.org/10.1037/dev0000740>
- Shih, S. I., & Sperling, G. (1996). Is there feature-based attentional selection in visual search? *Journal of Experimental Psychology-Human Perception and Performance*, *22*(3), 758–779. <https://doi.org/10.1037/0096-1523.22.3.758>
- Treisman, A. (1986). Features and objects in visual processing. *Scientific American*, *255*(5), 114B–125. <https://doi.org/10.1038/scientificamerica.n1186-114B>
- Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*(2), 171–178. [https://doi.org/10.1016/S0959-4388\(96\)80070-5](https://doi.org/10.1016/S0959-4388(96)80070-5)
- Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, *14*(4), 411–443. <https://doi.org/10.1080/13506280500195250>
- Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, *95*(1), 15–48. <https://doi.org/10.1037/0033-295X.95.1.15>
- Treisman, A., & Zhang, W. (2006). Location and binding in visual working memory. *Memory & Cognition*, *34*(8), 1704–1719. <https://doi.org/10.3758/BF03195932>
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)
- Utochkin, I. S., Khvostov, V. A., & Stakina, Y. M. (2018). Continuous to discrete: Ensemble-based segmentation in the perception of multiple feature conjunctions. *Cognition*, *179*, 178–191. <https://doi.org/10.1016/j.cognition.2018.06.016>
- Utochkin, I. S., & Yurevich, M. A. (2016). Similarity and heterogeneity effects in visual search are mediated by “segmentability.” *Journal of Experimental Psychology: Human Perception and Performance*, *42*(7), 995–1007. <https://doi.org/10.1037/xhp0000203>
- Verghese, P., & Nakayama, K. (1994). Stimulus discriminability in visual search. *Vision Research*, *34*(18), 2453–2467. [https://doi.org/10.1016/0042-6989\(94\)90289-5](https://doi.org/10.1016/0042-6989(94)90289-5)
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1436–1451. <https://doi.org/10.1037/0096-1523.32.6.1436>
- Watson, D. G., Maylor, E. A., & Bruce, L. A. (2005). Search, enumeration, and aging: Eye movement requirements cause age-equivalent performance in enumeration but not in search tasks. *Psychology and Aging*, *20*(2), 226–240. <https://doi.org/10.1037/0882-7974.20.2.226>
- Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*(1), 48–64. <https://doi.org/10.1037/0096-3445.131.1.48>

- Wolfe, J. M. (1992). "Effortless" texture segmentation and "parallel" visual search are not the same thing. *Vision Research*, 32(4), 757–763. [https://doi.org/10.1016/0042-6989\(92\)90190-T](https://doi.org/10.1016/0042-6989(92)90190-T)
- Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202–238. <https://doi.org/10.3758/BF03200774>
- Wolfe, J. M. (1998a). Visual search. In H. Pashler (Ed.), *Attention* (Vol. 1, pp. 13–73). University College London Press.
- Wolfe, J. M. (1998b). What can 1 million trials tell us about visual search? *Psychological Science*, 9(1), 33–39. <https://doi.org/10.1111/1467-9280.00006>
- Wolfe, J. M. (2012). The binding problem lives on: Comment on Di Lollo. *Trends in Cognitive Sciences*, 16(6), 307–308. <https://doi.org/10.1016/j.tics.2012.04.013>
- Wolfe, J. M., & Cave, K. R. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron*, 24(1), 11–17. [https://doi.org/10.1016/S0896-6273\(00\)80818-1](https://doi.org/10.1016/S0896-6273(00)80818-1)
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3), 419–433. <https://doi.org/10.1037/0096-1523.15.3.419>
- Wolfe, J. M., Chun, M. M., & Friedman-Hill, S. R. (1995). Making use of texon gradients: Visual search and perceptual grouping exploit the same parallel processes in different ways. In T. Papathomas, C. Chubb, A. Gorea, & E. Kowler (Eds.), *Early vision and beyond* (pp. 189–198). MIT Press.
- Wolfe, J. M., Friedman-Hill, S. R., & Bilsky, A. B. (1994). Parallel processing of part-whole information in visual search tasks. *Perception & Psychophysics*, 55(5), 537–550. <https://doi.org/10.3758/BF03205311>
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5(6), 495–501. <https://doi.org/10.1038/nrn1411>
- Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, 1(3), Article 0058. <https://doi.org/10.1038/s41562-017-0058>
- 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. <https://doi.org/10.1038/nn.2202>
- Yashar, A., & Lamy, D. (2010). Intertrial repetition facilitates selection in time: Common mechanisms underlie spatial and temporal search. *Psychological Science*, 21(2), 243–251. <https://doi.org/10.1177/0956797609357928>
- Yu, D., Tam, D., & Franconeri, S. L. (2019). Gestalt similarity groupings are not constructed in parallel. *Cognition*, 182, 8–13. <https://doi.org/10.1016/j.cognition.2018.08.006>
- Yu, D., Xiao, X., Bemis, D. K., & Franconeri, S. L. (2019). Similarity grouping as feature-based selection. *Psychological Science*, 30(3), 376–385. <https://doi.org/10.1177/0956797618822798>
- Zeki, S. (1980). The representation of colours in the cerebral cortex. *Nature*, 284(5755), 412–418. <https://doi.org/10.1038/284412a0>
- Zeki, S. M. (1973). Colour coding in rhesus monkey prestriate cortex. *Brain Research*, 53(2), 422–427. [https://doi.org/10.1016/0006-8993\(73\)90227-8](https://doi.org/10.1016/0006-8993(73)90227-8)
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192), 233–235. <https://doi.org/10.1038/nature06860>
- Zhang, X., Zhaoping, L., Zhou, T., & Fang, F. (2012). Neural activities in v1 create a bottom-up saliency map. *Neuron*, 73(1), 183–192. <https://doi.org/10.1016/j.neuron.2011.10.035>
- Zhang, Y., & Huang, L. (2021). *Motion is not a preattentive feature* [Manuscript submitted for publication]. Department of Psychology, the Chinese University of Hong Kong.

(Appendices follow)

Appendix A

Method

Use of Previously Published Data Set

Huang (2015a) used eight tasks (410 observers in total). In the present study, this previously published data set was reanalyzed along with the new data set, making a total of 26 tasks (1,744 observers in total). For the convenience of the readers, the method of these 8 previously published tasks are repeated here together with the 18 new tasks.

Participants

University students, all of whom had a normal or corrected-to-normal vision, participated in this study's experiments. Some participants (9 in change detection, 5 in long-exposure change detection, 5 in ensemble change, 2 in feature-based selection, 2 in feature counting, 8 in feature VWM change, 4 in feature VWM probe, 2 in grouping, 3 in global grouping, 2 in location-based selection, 4 in difficult location-based selection, 4 in monitoring, 6 in high-level motion, 3 in low-level motion, 6 in pattern comparison, 3 in previewed pattern comparison, 2 in perceptual discrimination, 1 in pop-out, 1 in saliency-based selection, 1 in temporal order, 2 in texture segregation, 1 in visual search, 2 in temporal visual search, 18 in VWM encoding, 20 in VWM probe) were excluded according to predetermined criteria (overall accuracy < 0.55 in accuracy measure tasks, and the average logarithmic threshold $> \text{Mean} + 2 \text{SD}$ in threshold measure tasks), leaving a total of 1,744 participants. The numbers of observers of all 26 tasks are shown in Table 1. These numbers were determined in advance based on pilot testing for target reliability of at least 0.97.

Apparatus

In all experiments, the stimuli were presented on a computer monitor and participants viewed the display from a distance of about 60 cm. The participants were asked to make responses by pressing one of two adjacent keys ("j" vs. "k"). They were asked to respond as accurately as possible but were under no time pressure (i.e., unspeeded responses).

Measures

How to measure all these 26 tasks? As discussed in Huang (2015a), although the response time slopes (e.g., Treisman & Gelade, 1980; Wolfe, 1994) are commonly used in previous studies, the reliability of response time slopes was terribly low in some tasks (e.g., pattern comparison). Therefore, exposure duration threshold measures were employed for the majority (19) of the tasks. For these tasks, the thresholds were always calculated and presented on the logarithmic scale. Besides, the logarithmic threshold was reversed because a lower threshold implies better performance. For example, a 1,000 ms threshold, a 500 ms threshold, and a 100 ms threshold are scored respectively as $-\ln(1) = 0$, $-\ln(0.5) = 0.6931$, and $-\ln(0.1) = 2.3026$.

For some other tasks (e.g., change detection), the threshold is inappropriate because the performances on these tasks are limited by

a working memory capacity (or some similar mechanisms). For these seven tasks, accuracy was the key measure.²⁵

The natures of measure for the 26 tasks (threshold or accuracy) are given in Table 1.

Stimuli

Examples of stimuli are shown in Figure 3. As mentioned above, a total of 16 types of stimuli were used, and each type included four feature values.

As shown in Figures 4–5, in most tasks, the stimulus items were presented in eight locations and arranged as two 2×2 arrays on the left and right sides of the display; the centers of the arrays were 2.91 cm away from the center of the display. In each array, the distance between items, both horizontally and vertically, was 1.46 cm. In several tasks (e.g., perceptual discrimination), stimulus items were presented in a subset of the eight locations of the two arrays, as will be described below.

Procedures of the 26 Tasks

The sequence of presentations is shown in Figures 4–5. In all 26 tasks, a trial started with a fixation cross. The fixation cross was presented in the center of the display for 400 ms and then followed by a gap of 400 ms, after which the stimulus display (or a cue) was presented (see below).

In all 26 tasks, each participant completed 7 blocks (96 trials per block, 6 for each stimulus type).

In tasks using threshold measures, thresholds were measured using a staircase method. The duration went down one step (5%) following a correct response but went up three steps following an incorrect response. A separate staircase was run for each of the 16 stimulus types. The first 3 blocks (18 trials per stimulus type) were regarded as the period for the staircase to stabilize, and then the threshold of each stimulus type was measured as the logarithmic average of the duration of the 24 trials in Blocks 4–7.

In tasks using accuracy measures, the first block was regarded as practice and excluded from the analysis.

Below I will turn to discussions of the individual procedure of the 26 tasks.

Change Detection

In this task, eight items were presented in the stimulus display, with each feature value presented once on each side and their arrangements randomized. The stimulus display (i.e., memory set) was presented for 200 ms, followed by a retention interval of 800 ms and then a probe display, which remained on the screen until a response was given. The probe display was identical to the stimulus display in 50% of the trials; in the remaining 50%, two

²⁵ In Huang (2015a), the performance on change detection is given in terms of the capacity estimate of memorized items (Pashler, 1988; Cowan, 2001). In the present study, the same data were presented in terms of accuracies to maintain consistency with other accuracy measures (e.g., ensemble change) that cannot be unambiguously described in terms of items.

items on the same side switched their locations. The participants were asked to report whether the stimuli and probe displays were the same or not.

Long-Exposure Change Detection

The method of this task was exactly the same as the change detection task except that the stimulus display was presented for 1000 ms.

Ensemble Change

In this task, the stimulus display (i.e., memory set) was presented for 200 ms, followed by a retention interval of 800 ms and then a probe display, which remained on the screen until a response was given. In both the stimulus display and the probe display, eight items were presented, two feature values were used, each for four times, and their arrangements were randomized. The same two values were used in the stimuli and probe displays in 50% of the trials; in the remaining 50%, only one value was shared between them. The spatial arrangements of the items were unrelated in the stimuli and probe displays in all trials. The participants were asked to report whether the same set of features were used in the stimuli and probe displays, regardless of their spatial arrangements.

Ensemble Matching

In this task, eight items were presented in the stimulus display, with two feature values used, each twice, on each side. The same two values were used in 50% of the trials; in the remaining 50%, one value was shared between the two sides. The spatial arrangements of the items were shuffled on each side. For each participant, the exposure duration of the stimulus display was adjusted for each of the 16 stimulus types to achieve a target accuracy of 75%. The stimulus display was followed by masks, which remained on the screen until a response was given. The participants were asked to report whether the items on the left and right sides of the stimulus displays were the same set or not, regardless of their spatial arrangements.

Feature-Based Selection

In this task, a cue was presented for 200 ms and then followed by the stimulus display. Eight items were presented in the stimulus display. The cued feature (i.e., the target) appeared in the stimulus display two or three times in randomized locations. The other three feature values were randomly assigned to the remaining five or six items, with the constraint that no feature was used more than twice. For each participant, the exposure duration of the stimulus display was adjusted for each of the 16 stimulus types to achieve a target accuracy of 75%. The stimulus display was followed by masks, which remained on the screen until a response was given. The participants were asked to report the number of targets (two vs. three).

Feature Counting

In this task, eight items were presented in the stimulus display. In 50% of the trials, all four feature values were used, each randomly

assigned to two items. In the remaining 50%, three of the four feature values were used, each randomly assigned two or three times. For each participant, the exposure duration of the stimulus display was adjusted for each of the 16 stimulus types to achieve a target accuracy of 75%. The stimulus display was followed by masks, which remained on the screen until a response was given. The participants were asked to report whether there were three or four types of items in the display.

Feature Visual working memory (VWM) Change

In this task, the stimulus display (i.e., memory set) was presented for 200 ms, followed by a retention interval of 800 ms and then a probe display, which remained on the screen until a response was given. In both the stimulus display and the probe display, three different items were presented and their locations were randomized. The same three items (i.e., same feature values) were used in the stimuli and probe displays in 50% of the trials; in the remaining 50%, only two values were shared between them. The spatial arrangements of the items were unrelated in the stimuli and probe displays in all trials. The participants were asked to report whether the same set of features were used in the stimuli and probe displays, regardless of their spatial arrangements.

Feature Visual working memory (VWM) Probe

In this task, the stimulus display (i.e., memory set) was presented for 200 ms, followed by a retention interval of 800 ms and then a probe display, which remained on the screen until a response was given. In the stimulus display, three different items were presented and their locations were randomized. In the probe display, one single probe item was presented in the center of the screen. The probe item was the same as one of the three items in the stimulus display in 50% of the trials ("same" trials); but different from all three of them in the remaining 50% ("different" trials). The participants were asked to report this relationship (same vs. different).

Grouping

In this task, eight items were presented in the stimulus display, with each feature value used only once on each side with the exception that one item was changed to be identical to its vertical neighbor. For each participant, the exposure duration of the stimulus display was adjusted for each of the 16 stimulus types to achieve a target accuracy of 75%. The stimulus display was followed by masks, which remained on the screen until a response was given. The participants were asked to report the location of the homogeneous column (left vs. right side).

Global Grouping

In this task, eight items were presented in the stimulus display. Each of the four feature values was assigned to two items. Identical items are organized as vertical pairs in 50% of the trials (vertical grouping), or horizontal pairs in the remaining 50% of the trials (horizontal grouping). For each participant, the exposure duration of the stimulus display was adjusted for each of the 16 stimulus types to achieve a target accuracy of 75%. The stimulus display was followed by masks, which remained on the screen until a response

was given. The participants were asked to report the orientation of the grouping (vertical vs. horizontal).

Location-Based Selection

In this task, one side of the stimulus display consisted of four items (each feature value used once), whereas the other side included only one single item. The assignments of the two sides (four-item side on the left side, or the right side) were randomly determined for each trial. The feature on the single-item side matched the item on the corresponding location of the four-item side in 50% of the trials (match trials) but mismatched in the remaining 50% of the trials (mismatch trials). For each participant, the exposure duration of the stimulus display was adjusted for each of the 16 stimulus types to achieve a target accuracy of 75%. The stimulus display was followed by masks, which remained on the screen until a response was given. The participants were asked to report the above-mentioned match/mismatch relation.

Difficult Location-Based Selection

In this task, one side of the stimulus display consisted of four items (each feature value used once), whereas the other side included only one single item. The assignments of the two sides (four-item side on the left side, or the right side) were randomly determined for each trial. The feature on the single-item side matched the item on the 90° clockwise rotated location of the four-item side in 50% of the trials (match trials) but mismatched in the remaining 50% of the trials (mismatch trials). For the example shown in Figure 4, the color on the single-item side (i.e., blue) was on the bottom-right corner, so it should be compared with the color on the 90° clockwise rotated location (i.e., bottom-left corner) of the four-item side, which was also blue. So this is a match trial. For each participant, the exposure duration of the stimulus display was adjusted for each of the 16 stimulus types to achieve a target accuracy of 75%. The stimulus display was followed by masks, which remained on the screen until a response was given. The participants were asked to report the above-mentioned match/mismatch relation.

Monitoring

In this task, in each trial, one of the eight locations was chosen as the target location. A sequence of randomly chosen items was presented on the target location with the constraint that an item is always different from the preceding one. In each trial, the length of the sequence (i.e., the number of items) was randomly chosen to be 4, 5, 6, 7, or 8. For each participant, the frame duration of the sequence was adjusted for each of the 16 stimulus types to achieve a target accuracy of 75%. The last item of the sequence was followed by a probe display, which remained on the screen until a response was given. The probe display consisted of masks in all eight possible locations and a probe item on the center of the display. The central probe item was the same as the last item of the sequence in 50% of the trials. The participants were asked to report whether the central probe item and the last item of the sequence were the same or not.

High-Level Motion

In this task, two successive frames are presented, each containing eight items. In Frame 1, each feature value was presented once on each side and their arrangements were randomized. In Frame 2, the items swap their locations with their horizontal neighbors in 50% of the trials, and with their vertical neighbors in the remaining 50% of the trials. The first frame was immediately followed by the second frame. The second frame was followed by masks, which remained on the screen until a response was given. For each participant, the exposure duration of the stimulus display was adjusted for each of the 16 stimulus types to achieve a target accuracy of 75%. The participants were asked to report whether there was horizontal or vertical swapping between the two frames.

Low-Level Motion

In this task, eight items were presented in Frame 1, with each feature value presented once on each side and their arrangements randomized. In Frame 2, two items on the same side switched their locations. The first frame was immediately followed by the second frame. The second frame was followed by masks, which remained on the screen until a response was given. For each participant, the exposure duration of the stimulus display was adjusted for each of the 16 stimulus types to achieve a target accuracy of 75%. The participants were asked to report the side where the changes occur (left vs. right side, each occurring for 50% of the trials).

Pattern Comparison

In this task, eight items were presented in the stimulus display, with each feature value used once on each side. The arrangements of the items were identical on the left and right sides in 50% of the trials; in the remaining 50%, two items on one side switched their locations. For each participant, the exposure duration of the stimulus display was adjusted for each of the 16 stimulus types to achieve a target accuracy of 75%. The stimulus display was followed by masks, which remained on the screen until a response was given. The participants were asked to report whether the left and right sides of the stimulus displays were the same or not.

Previewed Pattern Comparison

In this task, eight items were presented in the stimulus display, with each feature value used once on each side. The arrangements of the items were identical on the left and right sides in 50% of the trials; in the remaining 50%, two items on one side switched their locations. In each trial, one of the two sides (left vs. right) was randomly chosen to be the cue side, whereas the other side was the target side. Items on the cue side were first presented and remained throughout the trial. The items on the target side were presented 400 ms after the cue side, remained for a certain duration, and were covered by masks. The masks remained on the screen until a response was given. For each participant, the exposure duration of the target side was adjusted for each of the 16 stimulus types to achieve a target accuracy of 75%. The participants were asked to report whether the left and right sides of the stimulus displays were the same or not.

Perceptual Discrimination

In the task, a cue was presented for 200 ms and then followed by the stimulus display. Only one item (i.e., target) was presented in the stimulus display; this item was randomly placed on one of the eight possible locations. The target was the same as the cue in 50% of the trials. For each participant, the exposure duration of the stimulus display was adjusted for each of the 16 stimulus types to achieve a target accuracy of 75%. The stimulus display was followed by masks, which remained on the screen until a response was given. The participants were asked to report whether the cue and the target were the same or not.

Pop-Out

In this task, eight items were presented in the stimulus display, all except one (i.e., the target) having the same feature value. For each participant, the exposure duration of the stimulus display was adjusted for each of the 16 stimulus types to achieve a target accuracy of 75%. The stimulus display was followed by masks, which remained on the screen until a response was given. The participants were asked to report the location of the target (left vs. right side).

Saliency-Based Selection

In this task, eight items were presented in the stimulus display. On each of the left and right sides, three (i.e., majority) items had the same feature whereas one (i.e., odd) item had a feature that is different from the rest three items. The four feature values were randomly assigned to the majority items and odd items on the left side, and those on the right side, so that all items on the left side were different from all items on the right side. The locations of the odd items were the same on the two sides in 50% of the trials but different in the remaining 50%. For each participant, the exposure duration of the stimulus display was adjusted for each of the 16 stimulus types to achieve a target accuracy of 75%. The stimulus display was followed by masks, which remained on the screen until a response was given. The participants were asked to report whether the locations of the odd items on the two sides were the same or not.

Temporal Order

In this task, the same two items were presented, in two successive frames, on one location on each side. They were always presented on one pair of symmetrical locations on the two sides (e.g., top-outside on both sides). The items on the two sides were presented in the same order in 50% of the trials but in reverse order in the remaining 50%. The exposure duration of the two frames was identical, and they were adjusted for each of the 16 stimulus types of each participant to achieve a target accuracy of 75%. The second frame was followed by masks, which remained on the screen until a response was given. The participants were asked to report whether the items on the left and right sides were presented in the same order or not.

Texture Segregation

In this task, eight items were presented in the stimulus display, with two feature values used, each twice, on each side. The same two values were used on both sides in 50% of the trials; in the remaining 50%, the set of values on one side was the complement of that on the other side. The spatial arrangements of the items were shuffled on

each side. For each participant, the exposure duration of the stimulus display was adjusted for each of the 16 stimulus types to achieve a target accuracy of 75%. The stimulus display was followed by masks, which remained on the screen until a response was given. The participants were asked to report whether the items on the left and right sides of the stimulus displays were the same set or not, regardless of their spatial arrangements.

Visual Search

In this task, a cue was presented for 200 ms and then followed by the stimulus display. Eight items were presented in the stimulus display. The cued feature (i.e., the target) appeared only once. The other three feature values were randomly assigned to the remaining seven items, with the constraint that no feature was used more than three times. For each participant, the exposure duration of the stimulus display was adjusted for each of the 16 stimulus types to achieve a target accuracy of 75%. The stimulus display was followed by masks, which remained on the screen until a response was given. The participants were asked to report the location of the target (left vs. right side).

Temporal Visual Search

In this task, in each trial, one of the eight locations was chosen as the target location. A cue was presented in the center of the display and a sequence of eight items was presented on the target location. The cue was first presented and remained throughout the trial. The first item of the sequence was presented 200 ms after the cue. The sequence of the items was randomly chosen with the constraints that an item is always different from the preceding one and that no feature value appeared more than three times in a sequence. The cued feature (i.e., the target) appeared once in the sequence in 50% of the trials, but not at all in the remaining 50% of the trials. For each participant, the frame duration of the sequence was adjusted for each of the 16 stimulus types to achieve a target accuracy of 75%. The last item of the sequence was followed by a mask, which remained on the screen until a response was given. The participants were asked to report whether the target was present or not in the sequence.

Visual working memory (VWM) Encoding

In this task, eight items were presented in the stimulus display, with each feature value presented once on each side and their arrangements randomized. The stimulus display (i.e., memory set) was presented for 200 ms, followed by a mask (duration = 800 ms) and then a probe display, which remained on the screen until a response was given. The probe display was identical to the stimulus display in 50% of the trials; in the remaining 50%, two items on the same side switched their locations. The participants were asked to report whether the stimuli and probe displays were the same or not.

Visual working memory (VWM) Probe

In this task, eight items were presented in the stimulus display, with each feature value presented once on each side and their arrangements randomized. The stimulus display (i.e., memory set) was presented for 200 ms, followed by a retention interval of 800 ms and then a probe display, which remained on the screen until a response was given. In the probe display, a probe item was

presented and the other seven items were replaced by white dots. The probe item was the same as the corresponding item (i.e., the one on the same location) in the memory set in 50% of the trials, but

different from the corresponding item in the remaining 50% of the trials. The participants were asked to report whether the probe item and the corresponding item were the same or not.

Appendix B

Further Discussions

Visual Feature Is a General-Purpose Mechanism

As shown in Figure 11, the featural strength plays an important role in almost all tasks, and a dominant role in most tasks. Here, we will elaborate on the role of featural strength in a few tasks.

Global Feature Selection Versus Local Saliency

Previous studies have distinguished between two mechanisms of attentional processing (e.g., Serences & Yantis, 2006). The first is the *global selection*: a filter that selects all the relevant items across the whole visual field (e.g., Wolfe, 1994). The second calculates the *local saliency* defined by differences between neighboring items (e.g., Li, 2002; Zhang et al., 2012). It is widely agreed that both contribute to attentional selection. But are features used in these two mechanisms different from each other? More precisely, is it possible that global selection is especially effective on some features (e.g., colors) but the local saliency is especially effective on some other features (e.g., orientations)?

Indeed, there seems to be evidence that orientation is more effective in driving local saliency than driving global selection. Nothdurft (1991, 1993; Wolfe et al., 1995; see also Inverso et al., 2016) argued that the attentional selection of orientations is mainly attributed to local saliency rather than to global feature selection. This point is illustrated in Figure B1, for the task of selecting multiple odd items, the homogeneous target condition is easier than the heterogeneous target condition in color displays, but not in orientation displays. Wolfe et al.'s (1995) interpretation is that selection of the orientation targets depends primarily on local saliency, whereas selection of the color targets depends primarily on global selection.

Several of the present tasks (e.g., visual search, feature-based selection) examined the global feature selection, whereas two tasks (i.e., pop-out, saliency-based selection) examined the local saliency. Specifically, in the saliency-based selection, the odd items on the left and right sides of the display were always different from each other, so the global selection was ineffective, and one had to rely on the local saliency to perform the task.

The FVS 2.0 framework has only a general featural strength and has not distinguished between global selection and local saliency. Therefore, if the above-mentioned special link between orientation and local saliency indeed exists, then, for orientation, there should be an obvious advantage in the saliency-based selection task, but a relative disadvantage in global selection tasks (e.g., visual search, feature-based selection). There are no such effects in Figures 6–7, throwing doubt on the link between orientation and local saliency.

More generally, as shown in Figure 11, the saliency-based selection depends heavily ($L_F = 0.975$) and exclusively on featural strength, making it clear that there is no evidence for the existence of such separate sets of features. In other words, the feature that

defines the local saliency is equivalent to the feature used in global feature selection.²⁶

Then, how do we explain the previous findings (Nothdurft, 1991, 1993; Wolfe et al., 1995)? Returning to Figure B1, it seems not implausible that the uniqueness of orientation can be interpreted by its large S score and small F score. Specifically, in color displays, the homogeneous target condition is naturally easier than the heterogeneous target condition because of the inherent difficulty of simultaneous access to multiple features (e.g., Huang et al., 2007). However, in orientation displays, the multiple orientations are represented mainly as a set of locations, rather than features, to start with. Therefore, it is not subject to this multiple-feature-induced difficulty. This possible account should be tested in future studies.

Role of Features in Ensemble Processing

A large number of studies have examined the visual processing of the statistical ensemble in the last 2 decades (for a review, see Alvarez, 2011). Some researchers believe that there are specialized mechanisms that extract ensemble or statistical properties as whole-set labels, and this specialized ensemble processing mechanism is different from the usual mechanism used to extract features of individual items. A typical account of ensemble processing mechanism (Chong & Treisman, 2003, 2005a, 2005b; Treisman, 2006) implies that it is a different way of using the same mechanism of attentional selection²⁷: Individual features are extracted when attention narrows down to individual items, but ensemble properties are extracted when attention zooms out to select a whole set. However, Myczek and Simons (2008) have argued that the ensemble processing is no more than the scanning of a few individual-item features.

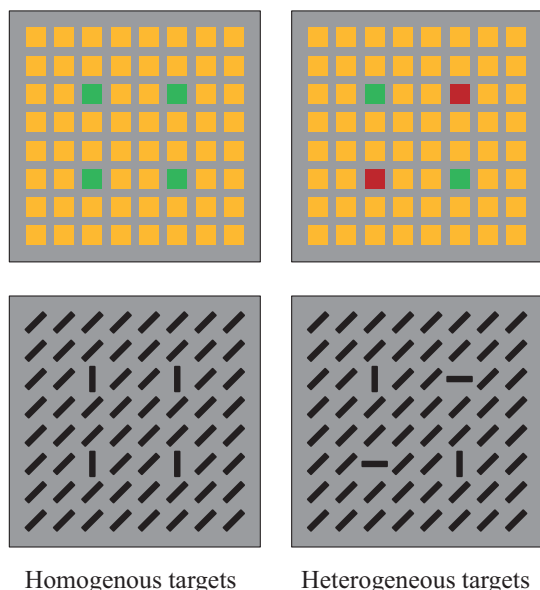
Although this conceptual distinction (whole-set labels vs. scanning of individual-item features) is important, it is difficult to tackle this distinction on an experimental basis. The present results shed novel light on this question. Specifically, two present tasks are relevant: ensemble matching examines the on-line comparison of two ensembles, whereas ensemble change examines the comparison of a memorized ensemble with a subsequent one.

As shown in Figure 11, both ensemble tasks depend critically on featural strength (ensemble change: $L_F = 0.933$; ensemble matching:

²⁶ This statement of equivalence only intends to say that local saliency is not especially effective (or especially ineffective) on making use of any of the stimulus types. Global selection and local saliency are of course different in their algorithms.

²⁷ Some researchers seem to uphold an even stronger account of specialized ensemble processing mechanism which implies that it is not subject to the usual attentional limit on the processing of individual features (e.g., Alvarez & Oliva, 2008, 2009). However, subsequent findings with better-matched comparisons are generally inconsistent with this very strong account (e.g., Huang, 2015e; Jackson-Nielsen et al., 2017).

Figure B1
Color-Orientation Difference in a Multiple-Odd-Item Task



Note. Wolfe et al. (1995; see also Nothdurft, 1991, 1993) indicated that, for the task of selecting multiple odd items, the homogeneous target condition is easier than the heterogeneous target condition in color displays, but not in orientation displays. Wolfe et al.'s (1995) interpretation is that selection of the orientation targets depends primarily on local saliency, whereas selection of the color targets depends primarily on global selection. The present results challenge this view, see text for details. See the online article for the color version of this figure.

$L_F = 0.981$). More importantly, the portions of unexplained variances are very low in both ensemble tasks (ensemble change: 2.3%; ensemble matching: 2.8%). Taken together, it is not the case that the ensemble processing is accomplished by some specialized “whole-set” mechanism which is entirely separate from the processing of individual-item features. Otherwise, we would have expected to see something comparable to the case of low-level motion: A much lower L_F and a much greater portion of unexplained variances. It may be possible that ensemble processing is achieved by scanning individual-item features like Myczek and Simons (2008) suggested. Alternatively, if there is indeed a specialized “whole-set” mechanism for ensemble processing, then this mechanism must directly depend on individual-item features in some way (e.g., using features as input).

Role of Features in Perceptual Grouping

As one of the most classic Gestalt phenomena, the perceptual grouping is usually considered to be determined by an approximately linear summation of the effects of different grouping cues (e.g., Kubovy & van den Berg, 2008). For example, an ambiguous grouping structure may be perceived as left-tilted strips based on color-based similarity and as right-tilted strips based on shape-based similarity. The final outcome depends on the probabilistic summation of these two grouping cues. This usual view of perceptual grouping is not explicitly connected to the visual features of attentional processing (or attentional processing in general). Therefore, there seems no strong theoretical reason to predict that the effectiveness of a similarity grouping cue should be related to the featural difference involved.

However, a recent set of studies provide supports for a feature selection account of similarity grouping which suggests that the similarity grouping is mediated by the feature-by-feature selection of the relevant elements (e.g., Huang & Pashler's, 2007 Figure 24; Huang, 2015d; Levinthal & Franconeri, 2011; Yu, Tam, & Franconeri, 2019; Yu, Xiao, et al., 2019; see Huang's, 2020a *Grouping by similarity is not a genuine grouping* section for a detailed discussion of this issue). If this is true, this would give strong reasons to predict that the effectiveness of a similarity grouping cue should depend on the featural difference involved.

No previous study has proposed any general account of the factors affecting the effectiveness of grouping cues. This open question can be answered by examining the present results on two tasks: grouping and global grouping.

As shown in Figure 11, both tasks load fairly highly on featural strength (grouping: $L_F = 0.797$; global grouping: $L_F = 0.774$), suggesting that the grouping process was mainly governed by the same general featural strength that is responsible for attentional processing. This result gives additional support for the feature selection account of similarity grouping.

However, it should also be mentioned that, in comparison to the other tasks, the grouping and global grouping are among the tasks with the largest portions of unexplained variances (grouping: 11.6%; global grouping: 8.1%; against an average of 4.6% for the 25 tasks). Therefore, it is also clear that the similarity grouping does have a small portion of unique variance (i.e., a unique underlying mechanism) that cannot be attributed to the visual features of attentional processing. Future work will be needed to reveal the nature of this unique underlying mechanism, perhaps on the directions of some previous findings (e.g., Herzog & Fahle, 2002; Manassi et al., 2012; Sayim et al., 2010).

Role of Features in Texture Segregation

Texture segregation is a classic task for the studies of visual features and visual attention and is especially well known for its use in identifying shape features such as intersection and closure (e.g., Beck, 1966, 1980; Julesz, 1984, 1986; Julesz & Bergen, 1983). One merit of this task is its ecological validity. This task closely resembles texture-based image segmentation which is critical to the visual processing of real-world input.

Texture segregation and visual search (pop-out in many cases) have both been used to identify visual features. However, it remains unclear whether they rely on the same set of visual features or not. In the present results, the portion of unexplained variance is very small for texture segregation (1.9%), suggesting that it uses the same set of visual features as other tasks.²⁸

²⁸ Wolfe (1992) showed that there is a difference between the features that give effortless texture segregation and those that give parallel visual search. The example given for “parallel search without texture segregation” involves the different task demand on binding stimulus (conjunction in visual search vs. disjunction in texture segregation), a point that will be elaborated below. The example given for “texture segregation without parallel search” involves two types of targets that have identical overall shapes. Therefore, if the task is actually accomplished by processing the shapes, the “texture segregation” task essentially involves repetition of identical targets, so perhaps the performance has been improved simply by the presence of redundant targets. All in all, Wolfe's (1992) findings does not seem to undermine the present finding that, generally speaking, texture segregation uses exactly the same set of visual features as other tasks.

Role of Features in Temporal Limit

The temporal limit of perception can reveal important information about the ways how the visual system functions. Holcombe (2009; see also Holcombe & Cavanagh, 2001; Holcombe & Judson, 2007) made an important claim about a dichotomy between a fast group and a slow group of temporal limits. Similar to the low-level and high-level motion perception: The fast group consists of specialized mechanisms such as those used to calculate motion direction, whereas the slow group is mediated by high-level processes. Based on Holcombe's (2009) claim, it seems reasonable to expect that this slow group will be fully mediated by features.

In the present results, there is only very little unexplained variance for the temporal order task (4.4%), and it loads significantly on the featural strength ($L_F = 0.698$). This gives further support to Holcombe's (2009) claim about the existence of a slow group of temporal limit which is built on high-level processing and further demonstrates that this high-level process is equivalent to the usual feature-by-feature attentional selection.

Role of Features in Feature Counting

Previous studies have shown that a sequential checking strategy is used in determining the number of unique features present in a display (Watson et al., 2005). In the present results, the feature counting task loads almost exclusively on the featural strength ($L_F = 0.957$). This gives further support to Watson et al.'s (2005) claim that there is no specialized "feature-based subitizing" mechanism that is similar to the subitizing of locations of objects, and further demonstrate that Watson et al.'s (2005) sequential checking mechanism is equivalent to the usual feature-by-feature attentional selection.

Location-Based Selection

The location-based selection task is included to assess the difference between location-based and feature-based selection. Presumably, if visual features indeed can guide attention, then we expect to see a greater L_F for feature-based selection than for location-based selection. This is indeed the case (0.957 vs. 0.797). Nevertheless, location-based selection's L_F is still significantly higher than that of perceptual discrimination (0.797 vs. 0.455). Therefore, as discussed above, the featural strength also affects the efficiency of access to visual input on a prespecified location. In other words, even the pure location-based selection is still conducted more efficiently for preattentive features than for laborious stimulus types.

Besides, the difficult location-based selection task is designed to be a more difficult version of location-based selection, and it is indeed significantly more difficult as intended (average threshold = 405 ms in difficult location-based selection vs. 270 ms in location-based selection). As shown in Figure 11, the L_F of location-based selection and difficult location-based selection are similar to each other (0.797 vs. 0.837), suggesting that the reliance on features is an innate portion of the location-based selection which plays a relatively constant role regardless of the difficulty of selection.

Visual working memory (VWM) of Features

Two tasks (feature VWM change vs. feature VWM probe) both require the memorizing of a set of features without the need of

knowing their locations. However, there is one critical difference between the ways how the memory is tested in these two tasks. In the feature VWM change task, the observers need to compare the content of VWM with up to three probe items. But in the feature VWM probe task, the observers only need to compare the content of VWM with one single probe item.

As shown in Figure 11, this task demand for multiple comparisons leads to a significant shift from the loading on visual strength to the loading on featural strength: namely a decrease of L_V (0.680 in feature VWM probe vs. 0.343 in feature VWM change) and an increase of L_F (0.685 in feature VWM probe vs. 0.900 in feature VWM change). Specifically, by directly comparing the accuracies of these two tasks, we can know that the accuracy of the feature VWM change task was as good as that of the feature VWM probe task for stimulus types that are relatively high on featural strength and relatively low on visual strength (e.g., colors, shapes, sizes, and spatial frequencies), but are considerably worse for those that are relatively low on featural strength and relatively high on visual strength (e.g., random shapes, color arrangements, T_s , orientations).

Why is it so? Perhaps the VWM representations can be created based on either features or raw (i.e., pixel level) information. The effectiveness of the former depends on the featural strength, whereas the effectiveness of the latter depends on the visual strength. Then, the former is stable and is not affected by the multiple comparisons, but the latter is fragile and can be significantly disrupted by the multiple comparisons that are required in the feature VWM change task.

Encoding of a Single Feature

Both the temporal visual search task and monitoring task require perceiving the items of a visual stream. However, these two tasks differ on one critical point. In temporal visual search, observers need to compare the items in the stream with a template that is previewed and presented (i.e., cue). However, in the monitoring, observers need to try their best to see and memorize the most recent item and compare that to a template that will be given subsequently (i.e., probe).

As a result, the threshold is generally much higher in temporal visual search than in monitoring (average threshold = 130 ms in temporal visual search vs. 25 ms in monitoring). Besides, the temporal visual search task loads fairly substantially on the featural strength ($L_F = 0.683$; $L_V = 0.629$), but the monitoring task loads almost entirely on visual strength ($L_F = 0.256$; $L_V = 0.943$).

Why is it so? Perhaps observers can perceive and memorize an item in its raw (i.e., pixel level) information very quickly and can keep updated efficiently, and this quick raw-information processing depends entirely on visual strength and is hardly affected by featural strength. However, this raw-information representation cannot be directly used to accomplish even the most straightforward task of one-to-one comparison and needs to be converted to features for that purpose.

Adding together, this comparison (temporal visual search vs. monitoring) and the above comparison (feature VWM change vs. feature VWM probe) consistently indicates that raw-information visual representations can be used fairly well in the simplest situations (i.e., keeping one or two of them for immediate use), but feature representations are more stable and are more effective than raw-information in tasks that are only slightly more complex.

Other Implications of the Present Results

Nature of Pattern Comparison

Pattern comparison, namely checking whether two patterns are the same or not, is a task that is commonly conducted in the real-world environment. It is what we do in the popular children's game "spot the difference" which is sometimes claimed by publishers to reflect or even train children's attention.

However, this task has rarely been used in scientific studies of visual attention. One possible reason for this apparent lack of studies is perhaps the ambiguity of the underlying mechanisms. Unlike visual search which reflects the attentional selection or change detection which reflects the VWM maintenance, it is not immediately clear what the pattern comparison reflects. Specifically, it is unclear how much of a role VWM plays in the comparison between the two patterns. On the one hand, the items are always available in the display so one may say there is no need to rely on durable mnemonic representations. On the other hand, perhaps it is a more effective strategy to compare the two sides through the middleman of VWM: memorizing all (or part) of the items on one set and compare the memorized content to the other set.

As shown in Figure 11, the pattern comparison is a high-level task that loads only slightly on visual strength. It is half-way between the change detection and visual search, load heavily on featural strength but also significantly on spatial strength. This seems to suggest that the pattern comparison is a task that involves half attentional selection and half VWM.

The previewed pattern comparison task was included to have a more direct assessment of the contribution of VWM to pattern comparison. On average, the threshold was significantly reduced when observers could preview one of the two sides (672 ms in pattern comparison \rightarrow 325 ms in previewed pattern comparison). For this substantial advantage to occur, it must be convenient for observers to compare the new side with the memorized information of the previewed side. This does not guarantee, but does make it plausible, that observers in the pattern comparison task largely compare the two sides through the middleman of VWM.²⁹

Nature of Visual working memory (VWM) Encoding

Some previous studies have claimed to examine the consolidation of VWM (e.g., Vogel et al., 2006; Zhang & Luck, 2008). The concept of VWM consolidation is borrowed from long-term memory studies which suggest that new memories initially exist in a fragile state and need to be consolidated over time (e.g., McGaugh, 2000). To apply this notion of consolidation on visual working memory, Vogel et al. (2006) seems to imply that, when many visual stimuli are presented in a brief period, the observers can instantly (e.g., in a 100 ms period) perceive all or most of them and create many fragile memory representations but it takes a much longer period (e.g., 500 ms) to consolidate them for subsequent reports.

However, these above-mentioned studies (e.g., Vogel et al., 2006) have technically examined the limit of encoding information in brief exposure which may or may not be due to the difficulty of VWM consolidation. In these studies, although the consolidation of VWM is certainly not an implausible factor, it is empirically indistinguishable from the attentional limit on conscious perception (i.e., access, e.g., see Huang, 2010a, pp. 176–177 for a detailed discussion). Specifically, instead of failing to consolidate the fragile

memory representations of some items, observers may fail to consciously perceive these items in the first place.

In the present study, the VWM encoding task has been included to look for evidence that can potentially distinguish between the memory-based account and the attention-based account illustrated above. If memory encoding involves some unique "consolidation" mechanism that works differently from the usual attentional limit on conscious perception, then we will see some unique mechanism that cannot be attributed to the FVS framework. However, there is no evidence for such a unique mechanism because there is only a little unexplained variance for the VWM encoding task (4.3%). This does not directly prove, but does make it plausible, that the VWM encoding can be attributed to the same attentional limit on conscious perception which also governs processes such as visual search or location-based selection.

High-Level and Low-Level Motion Perception

Previous studies suggested that there is a dichotomy between low-level and high-level motion perception (e.g., Lu & Sperling, 1995). Low-level motion perception (first- and second-order motion in Lu & Sperling, 1995) is implemented by specialized modules in the brain which conduct pixel-by-pixel comparisons between the frames, whereas the high-level motion perception tracks the features and is mediated by the distribution of attention (e.g., third-order motion in Lu & Sperling, 1995). The present study has included two motion tasks to assess the distinction between low-level and high-level motion perception.

The present study has confirmed a very clear dichotomy. The low-level motion task load moderately on the visual strength ($L_V = 0.547$), but not on featural ($L_F = -0.172$) or spatial strength ($L_S = -0.069$). More importantly, there is a large portion of unexplained variances (66.6% as against an average of 4.6% for the other 25 tasks), suggesting that low-level motion is indeed implemented by a specialized mechanism of its own. On the other hand, unlike low-level motion, the performance on high-level motion can be fully explained by FVS 2.0 framework (unexplained variance = 3.1%). Besides, as shown in Figure 11, the high-level motion loads extremely low on visual strength ($L_V = -0.030$) but extremely high on featural strength ($L_F = 0.982$), suggesting that it is entirely a high-level task that depends on the visual features rather than on the raw stimulus signals. To summarize, the present study supports the clear dichotomy between low-level and high-level motion perception and confirms that they are entirely different: the high-level, but not the low-level motion is calculated based on features.

It is well established from previous studies that the low-level motion is processed by a specialized mechanism (e.g., Lu & Sperling, 1995). Therefore, this exception to the FVS 2.0 framework comes with a good conceptual reason. In a way, the low-level motion can be seen as a "distractor" in the present set of tasks and it is indeed correctly rejected by the present analysis. The "correct rejection" of low-level motion makes sure that the other 25 tasks are "hits" rather than "false alarms."

²⁹ A recent study argued that VWM plays little role for processing of presented items (Li et al., 2019), and appears to be in conflict with the present finding which suggests that VWM plays an important role. However, the task is different and the concept of VWM is operationalized in very different ways in these studies, so this apparent discrepancy may reflect different aspects of VWM and/or their unequal involvement in different tasks. Future work will be needed to bridge these results.

High-Level Versus Low-Level Tasks

Above, it is suggested that the factor of visual strength helps significantly in low-level tasks but not in high-level tasks. Although the terms “high-level tasks” and “low-level tasks” are commonly used in the literature, it has never been possible to precisely define these terms. The loading of a task on visual strength (L_V) offers a way of precisely defining the level of processing.

As umbrella terms, the high-level and low-level processing have conflated concepts that are potentially separable from each other (see Awh et al., 2012 for a similar point). Generally speaking, the low-level tasks are thought to be automatic, stimulus-driven, and primarily determined by the sensory input of the raw stimulus in a bottom-up mode, whereas high-level tasks are thought to be voluntary, goal-driven, and primarily determined by top-down factors such as attentional operations and working memory. Future studies will be needed to examine how well the L_V can fit each of these previously associated concepts and this can hopefully lead to greater clarity of the high-level versus low-level distinction.

For now, it is worth making clear that the L_V cannot be fully attributed to simple indices such as the durations of thresholds. Donk and van Zoest (2008) showed that the effects of visual saliency, which is highly relevant to the visual strength of the present study, diminish in longer exposure durations. Therefore, one may speculate that the L_V of a task is simply determined by the average exposure duration of that task. To assess this possibility, the relationship between a task’s average threshold and its L_V is plotted in Figure B2. Although there is an apparent correlation, this relationship is far from perfect. For example, the difficult location-based selection is a moderately lower level task than pop-out ($L_V = 0.522$ vs. 0.386), but the average threshold of difficult location-based selection is much longer than that of pop-out (405 vs. 92 ms). For another example, previewed pattern comparison is a much lower level task than texture segregation ($L_V = 0.565$ vs. -0.013), but the average threshold of previewed pattern comparison is slightly longer than that of texture segregation (325 vs. 266 ms). From these examples, it is clear that a higher-level task is *not* simply a task with longer duration thresholds.

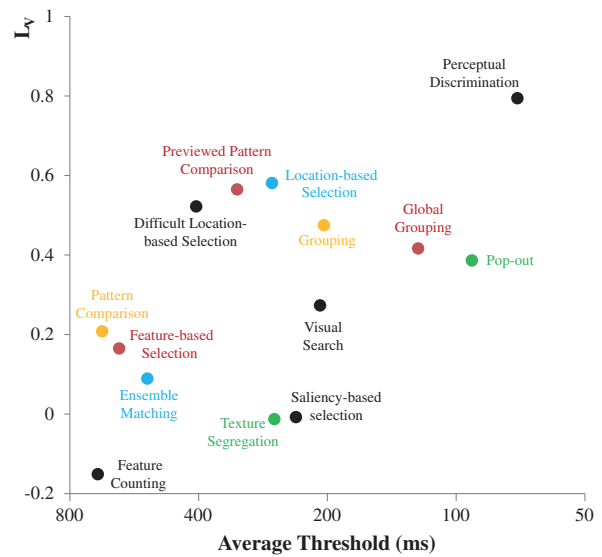
The distinction between the level of processing and the length of the threshold has also been demonstrated by Huang (2015c). Huang (2015c) showed that featural strength has a substantial effect on a high-level task but little effect on a low-level task, whereas visual strength has a substantial effect on a low-level task but little effect on a high-level task. Experiments 1 and 2 in Huang (2015c) compared the effects of thresholds in a high-level task (pattern comparison) and a low-level task (perceptual discrimination), and, like in the present study, the thresholds of pattern comparison were generally much longer than those of perceptual discrimination (450 ms vs. 40 ms). Nevertheless, Experiments 4 and 5 in Huang (2015c) showed a consistent pattern of results when the thresholds of the tasks were forced to become much closer to each other (100 ms vs. 70 ms). Therefore, the length of the threshold, per se, is not critical to the pattern of results and does not fully define the concept of “level of processing.”

The Best Tasks for Studies of Visual Attention

In the preceding section, we know that the higher-level tasks are not necessarily the more laborious tasks (in the sense of thresholds). This allows us to find the best combination to study attentional

Figure B2

Relationship Between a Task’s Average Threshold and Its L_V



Note. The accuracy tasks were not included in this analysis. Five threshold tasks (monitoring, high-level motion, low-level motion, temporal order, temporal visual search) were also excluded because they involve multiple frames and it is not self-evident whether a single frame or the sum of all frames should be counted for this purpose. For the remaining 14 tasks, although there is an apparent correlation between a task’s average threshold and its L_V , this relationship is far from perfect. From these examples, it is clear that a higher-level task is not simply a task with longer duration thresholds. The colors are only included as a visual aid and are not related to either the average threshold or L_V . See the online article for the color version of this figure.

processing, namely a task that is near the right-bottom corner of Figure B2: It is a high-level task that measures attentional processing rather than low-level vision, but the task is also fairly effortless for the participants to perform.

As shown in Figure B2, the pop-out and texture segregation are among the winners. This perhaps accounts for their popularity in the traditional studies of visual attention. They are the classic tasks for identifying preattentive features (e.g., Treisman, 1986; Wolfe, 1998b) and are widely used to study various research questions on visual attention (For recent use of pop-out, see Müller et al., 2009; Rangelov et al., 2017; For recent use of texture segregation, see Huang, 2020b; Utochkin et al., 2018; Utochkin & Yurevich, 2016).

Encoding Duration for Visual working memory (VWM)

Although traditional studies of VWM usually employed fairly brief exposure durations (200 ms or less, e.g., Luck & Vogel, 1997; Treisman & Zhang, 2006; Wheeler & Treisman, 2002; Zhang & Luck, 2008), studies in recent years start to employ long-exposure durations (~1000 ms, e.g., Fougny & Alvarez, 2011; Fougny et al., 2013). To justify this change, researchers showed that VWM encoding is insufficient in the brief displays (e.g., Bays et al., 2011) and argued that the long-exposure design is a better measurement of the VWM storage than the short-exposure design is.

The present results of the change detection and long-exposure change detection tasks gave a systematic comparison between these

two tasks. First, the accuracies in the long-exposure change detection task are generally and substantially higher than that of the change detection task (0.751 vs. 0.690 for the averages of all 16 stimulus types), confirming the previous finding that the short exposure is indeed insufficient for VWM encoding.

Second, one may potentially expect that the L_F and L_S may be affected by this difference in exposure duration. For example, there would be a higher L_S and lower L_F for long-exposure change detection than change detection if the Boolean map structure kicks in more slowly than the processing of visual features. The results revealed no indication of such differences. In other words, the Boolean map structure and processing of visual features (and more generally all the other potentially related underlying mechanisms) do not differ in their time courses. Therefore, although the overall performance is affected by the exposure duration, we can be reassured that the findings from long- and short-exposure experiments are otherwise comparable and can be merged seamlessly with each other.

Third, one may suspect that there may be some artificial strategies that can be applied in the long-exposure change detection task. For example, a usual suspect would be verbal coding which can be readily used for colors. In Figure 6, we can see that the worst prediction error of the long-exposure change detection occurred on color arrangement. The accuracy is 0.692 and is better than both what is predicted from change detection performance (0.662) and what is predicted from the FVS 2.0 framework (0.661). Perhaps, the color arrangement has benefited from an artificial strategy of attending to only red (or only green) segments, and it takes time for this artificial strategy to fully unfold. Nevertheless, the prediction error is fairly small even in this worst-case and is generally negligible in all other cases. Colors, the above-mentioned usual suspect of verbal coding, do not enjoy any special benefit in the long-exposure change task. Therefore, we can be reassured that the colors are unlikely to have been artificially inflated by verbal coding in studies that have used long-exposure durations.

Spatial Nature of the Attentional Selection

An important finding of visual attention is its spatial nature. In other words, the feature-based selection does not directly help the processing of the target feature but works by guiding spatial attention to the locations containing the target feature (Pashler, 1999, p. 98; Quinlan, 2003, p. 660). For example, attending to “red” does not directly make red objects more perceivable, but is mediated by guiding spatial attention to the locations of red objects. This point has been confirmed from a few different perspectives. Nissen (1985); see also Chen & Wyble, 2015, 2018; Johnston & Pashler, 1990) showed that reporting a feature depends on reporting its location. Moore and Egeth (1998) found that, in very brief displays, knowing the feature (color or size) of a target does not help to detect it. Huang and Pashler (2007) suggested that feature-based selection is inherently spatial because the format of visual awareness is a map. Huang (2010c) studied the time-course of feature-based selection and found that there is a temporal gap between the selection of target-feature locations and the emergence of attentional advantage: The selection of target-feature locations can be done very quickly and well within a 50 ms display, but it takes 100 ms for attentional advantage to start to emerge on those locations.

Most relevant to the present study, Shih and Sperling (1996) compared the conditions in which an “attended feature”

provides temporal or spatial information of a target. It was found that spatial information, but not temporal information, improves the accuracy of reporting the target. Therefore, attention is implemented by spatial selection. A comparison between the visual search and temporal visual search task in the present study provides additional support to this notion. In comparison to (spatial) visual search, temporal visual search has a greater L_V (0.629 vs. 0.273) but a smaller L_F (0.683 vs. 0.955), suggesting that temporal visual search depends much less on the use of visual features. This offers additional support for the notion that attention is implemented by the spatial selection of the task-relevant locations.

Reasons for Unexplained Variance

Although the FVS 2.0 framework has accounted for the majority (95.4%) of the variances in 25 tasks, there is still a small portion (4.6%) of unexplained variances. Below, I will list two likely reasons for some of the unexplained variances.

The Fixation

By comparing Figure 10 with Figure 4a of Huang (2015a), we can see the effect of stimulus contrast (i.e., the difference between high-low-contrast pairs) plays a smaller role in FVS 2.0 framework than in the initial FVS framework described in Huang (2015a). It seems that this occurs because of an expansion of the anchors of visual strength. Specifically, an anchor of visual strength means a task that loads heavily on visual strength and therefore plays an important role in defining the visual strength scores of the stimulus types. In the initial FVS framework, the visual strength is anchored mainly by perceptual discrimination ($L_V = 0.858$ in Huang, 2015a), whereas in FVS 2.0 framework, the visual strength is anchored jointly³⁰ by monitoring ($L_V = 0.943$) and perceptual discrimination ($L_V = 0.794$). By inspecting Figures 6–7, we can see that, in the FVS 2.0 framework, stimulus contrast has a larger-than-predicted effect in perceptual discrimination, but an overall slightly smaller-than-predicted effect in monitoring. Therefore, this expansion of anchor tasks reduces the importance of the stimulus contrast in defining visual strength.

Then, why do these tasks differ in terms of their susceptibility to the stimulus contrast? One plausible reason is observers’ fixations. In all tasks of the present study, observers were instructed to start a trial by fixating on the center of the display but were free to make eye movements during a trial. In the perceptual discrimination task, the stimulus durations are generally very brief, so it is reasonable to expect that the cue item is always fixated whereas the target item is perceived in the periphery. However, in the monitoring task, observers probably fixated on the stream of items.³¹ As we know, the stimulus contrast (in this range) has a greater effect on the periphery than in the fovea. Therefore, this inconsistency on fixation locations probably explains the reduced role of stimulus contrast in the FVS

³⁰ This is a simplified statement for the convenience of description. Actually, other tasks have also contributed, but with smaller weights because of their smaller L_V values.

³¹ After noticing this in data analysis, I have informally tried a few observers on the monitoring tasks. They indeed tended to fixate on the stream of items. When they were asked to perform the task without fixating on the stream of items, they reported that it was effortful and unnatural for them to do so.

2.0 framework than in the initial FVS framework described in Huang (2015a).

This fixation (or eye movement) factor is certainly beyond what is explained in the FVS 2.0 framework, and it is probably the reason why the percentage of unexplained variance for the perceptual discrimination task has increased dramatically from 2.8% in Huang (2015a) to 11.3% in the present study. There is no easy solution for this problem: strict enforcement on the fixation in monitoring and other tasks is effortful for the observers and will be ecologically invalid.

While fully admitting this methodological imperfection, this throws no doubt on the theoretical implications of the present study. It just means that the portion of unexplained variances of the FVS 2.0 framework will be even lower than what is reported here if ways can be found to exclude the effects of eye fixations.

Binding

The stimulus type “binding” is not a single feature or aspect of the stimulus, but the binding of two features: colors and shapes. The fairly good fitting of its performances on these tasks suggests that, for most tasks, binding can be simplified and treated as single-dimension variations. Nevertheless, previous studies already showed that binding is unique in some ways and this does account for some of the unexplained variances in the present results.

Specifically, previous studies on binding features showed a substantial asymmetry between the conjunction and disjunction of features. For example, Huang and Pashler (2012) examined the ability to perceive a spatial structure in multi-item subsets of a display that were defined either conjunctively or disjunctively. The observers could readily perceive conjunctively defined subsets but had great difficulty with disjunctively defined subsets.

For another example, Wolfe (1992) gave an example for “parallel search without texture segregation.” In this example, the visual

search task always includes one single type of target and is technically a conjunction task, and indeed it is easy. On the other hand, the texture segregation task always includes the simultaneous selection of multiple disjunctively defined items, so it is a disjunction target, and indeed it is difficult.

Similar conjunction/disjunction asymmetry certainly also played a role in the present study. For example, if the task requires selecting “red +” and ignoring all other types of items, that will be a conjunction task. On the other hand, if the task requires selecting both “red +” and “green O,” but ignoring “red O” and “green +,” that will be a disjunction task. We expect that there will also be an obvious advantage for conjunction trials over disjunction trials.

When the stimulus type binding is used in the present study, the portions of conjunction and disjunction trials, when applicable, differ from task to task. On the one extreme, the feature-based selection task consists exclusively of conjunction trials and there is indeed a unique advantage for binding in this task: the threshold (174 ms) is much better than what is predicted from FVS 2.0 framework (269 ms). On the other extreme, the texture segregation task includes a considerable portion (1/3) of disjunction trials (e.g., “red +” & “green O” on one side vs. “red O” & “green +” on the other side) and there is indeed a unique disadvantage for binding in this task: the threshold (179 ms) is worse than what is predicted from FVS 2.0 framework (142 ms). Further analysis showed that the accuracy of conjunction trials (0.832) was much better than that of disjunction trials (0.632). In other words, the worse-than-predicted overall threshold was indeed caused by the 1/3 disjunction trials.

Received February 2, 2020

Revision received April 14, 2021

Accepted May 26, 2021 ■