Running head: DIMENSION-BASED ATTENTION

Dimension-based attention in the recognition of facial identity and facial expression

Zhicheng Lin

Department of Psychology

University of Minnesota

Address correspondence to:

Zhicheng Lin,

Department of Psychology

University of Minnesota

75 East River Parkway,

Minneapolis, MN 55455, USA

Phone: 612-625-2470

FAX: 612-626-2079

E-mail: linxx443@umn.edu

Abstract

Although the human visual system is constantly flooded with sensory information, the brain is remarkable in inferring structures from the massive inputs and selectively attending to behaviorally relevant information. However, how the two processes interact remains largely unknown. Can top-down attention efficiently select the task-relevant dimension (e.g. gender) during face recognition to override interference in the task-irrelevant dimension (e.g. expression)? To address this issue, participants were asked to classify real face images according to gender or expression, which were preceded by other faces (masked priming task) or words (face-word Stroop task). Results show that face classification was 1) affected by the task-relevant but not the task-irrelevant dimension of the preceding faces, and 2) modulated by words depicting the task-relevant but not the task-irrelevant dimension of the face. These results suggest that high level dimensions such as facial expression and facial identity can serve as units of attentional selection, possibly due to the late binding of the two dimensions.

We are constantly bombarded by huge amounts of visual inputs. How does the brain translate such overwhelming inputs into fine-tuned decisions and actions in everyday life? The computational principles underlying such adaptive behaviors must rely on a divide and conquer strategy to organize and interpret selective visual inputs in order to guide actions (Neisser, 1967). This strategy is subserved by at least two critical processes: a perceptual grouping process to parse the visual scene into elementary units (Roelfsema, 2006; Wertheimer, 1923) and a selective attention process to amplify task-relevant information for further processing while suppress distracting information (Desimone & Duncan, 1995; Egeth & Yantis, 1997). A fundamental question in cognitive psychology is to understand the link between the two processes; in other words, how efficiently can humans deploy visual attention based on different organization strategies in segmenting visual scenes (e.g. organization by space, object, and dimension)?

Organization by dimension is a ubiquitous grouping principle. For example, objects in the visual environment can be represented in terms of featural dimensions such as orientation, color, motion, and depth; a special category of objects—faces—can be represented in terms of semantic dimensions such as identity and expression. Here I ask whether top-down attention can efficiently select the task-relevant dimension (e.g. gender) during face recognition to override interference in the task-irrelevant dimension (e.g. expression). This question is important for theoretical and practical reasons. First, given the constraint of cognitive capacity limits, characterizing the efficiency of attentional selection based on different units (e.g. space, object, and dimension) is crucial to understanding the bottleneck of information processing. Second, understanding the efficiency of attentional selection will provide clues as how different dimensions interact (e.g. whether attentional selection of one dimension will inevitably select another dimension), which is fundamental to perception and performance (Garner, 1974). Third, since successful communication largely depends on the ability to reason about others' mental states and act accordingly (Apperly, Samson, & Humphreys, 2005), monitoring the changes of facial expression of different individuals plays a major adaptive role (Haxby, Hoffman, & Gobbini, 2002). For instance, the extent to which we can attend to variation in a relevant dimension (e.g. monitoring potential changes of expressions) while filtering out concurrent interference within an irrelevant dimension (e.g. ignoring gender information) may be essential for the survival of humans. Forth, practically, understanding the efficiency of attentional selection will shed light on how to cope with distraction, which is known to be associated with various types of accidents such as car and workplace accidents.

Despite its importance, little is known about the efficiency of top-down selection towards facial identity and expression. Although units of attentional selection are known to include locations (Eriksen & Eriksen, 1974; Eriksen & St. James, 1986; Eriksen & Yeh, 1985; Posner, 1980; Posner, Snyder, & Davidson, 1980), objects (Baylis & Driver, 1993; Duncan, 1984; Egly, Driver, & Rafal, 1994), and featural dimensions such as motion and color (Chawla, Rees, & Friston, 1999; Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1990; Liu, Slotnick, Serences, & Yantis, 2003; Maunsell & Treue, 2006; O'Craven, Rosen, Kwong, Treisman, & Savoy, 1997; Treue & Martinez Trujillo, 1999), it remains elusive regarding the efficiency of semantic dimension-based attention. First, semantic dimension-based attention cannot be explained by space-based attention. This is because the configuration nature of the face determines the two dimensions being *integral* in physical space and mental representation (e.g. when expression is manipulated, the information that gives rise to gender is also affected). For spatially *separable* dimensions, which can be manipulated independently (e.g. the nose can be edited with the eyes exactly the same), the efficiency in suppressing concurrent variation and interference at other locations (e.g. Hoffman & Haxby, 2000) is space-based (Posner, 1980). Second, whether conclusions regarding other units of attentional selection can generalize to the recognition of real facial expression and identity remains unclear. Of particular relevance here are featural dimensions. For featural dimensions, previous studies suggest asymmetric selection efficiency for different dimensions. For instance, while attention to the relevant dimension of either hue (what we commonly referred to by the color names) or shape is effective (i.e. variation in one dimension does not affect the performance on the other dimension), attention to either hue or brightness is not (Garner & Felfoldy, 1970). Similarly, in a typical attention capture study, when searching multidimensional displays for a salient color, the presence of an element with a unique form (i.e. singleton) did *not* interfere; yet the presence of an element with a unique color did interfere with visual search for a salient form (Theeuwes, 1991).

Object-based attention theories (Baylis & Driver, 1993; Duncan, 1984; Egly et al., 1994) would predict that processing of facial expression and identity are strongly coupled and thus selection based on dimension would be difficult (e.g. paying attention to identity would inevitably process information from the whole face including expression). Dimension-based attention accounts would predict that processing of a single dimension within an object can be efficient to override interference within another dimension, even when the two dimensions are spatially intertwined, analogical to low level featural dimension-based attention (Maunsell & Treue, 2006). In this study, I investigate dimension-based attention in the recognition of facial expression and facial identity using masked priming paradigm (Experiment 1A and 1B) and face-word Stroop task (Experiment 2A and 2B). In the masked priming paradigm, in each trial participants were asked to classify a face image according to either gender (male vs. female) or expression (positive vs. negative in Experiment 1A; more specifically, happy vs. angry in Experiment 1B), which was preceded by another brief face image. Critically, the relationship between the two images in both the task-relevant and the task-irrelevant dimensions were orthogonally manipulated. In the face-word Stroop task, instead of congruency effect induced from a briefly exposed face image, a word depicting either the task-relevant dimension or the task-irrelevant dimension of the face was used. Finding a strong congruency effect in the task-relevant dimension but not the task-irrelevant dimension will support the dimension-based attention hypothesis, whereas finding a strong congruency effect in both the task-relevant dimension and the task-irrelevant dimension will lend support to the object-based attention hypothesis.

Experiment 1A

Participants were asked to classify each probe face image (hereafter probe) according to either gender (male vs. female) or expression (positive vs. negative) in different blocks by pressing buttons. The probe was preceded by a brief prime face image (hereafter prime). The prime and the probe were congruent in the task-relevant dimension on half of the trials (e.g. in the gender task, both the prime and the probe could be male faces) and incongruent on the other half (e.g. in the gender task, the prime and the probe could be a male face and a female face, respectively). Congruency in the task-irrelevant dimension was orthogonally manipulated such that the prime and the probe were congruent in the task-irrelevant dimension on half of the trials (e.g. in the gender task, both the prime and the probe could be happy faces) and incongruent on the other half (e.g. in the gender task, the prime and the probe could be a happy face and an angry face, respectively).

Method

Participants. Thirty-two volunteers (16 men, 16 women) between 18 and 35 years old from the University of Minnesota community participated in the experiment for course credit or money. All participants in this and subsequent experiments had normal or corrected-to-normal vision, and signed an Institutional Review Board approved consent form.

Apparatus and Stimuli. The stimuli were presented on a SONY Trinitron cathode ray tube (CRT) monitor (model: CPD-G200; refresh rate: 100 Hz; resolution: 1024 × 768 pixels) using the MATLAB (The Math Works Inc., Natick, MA) Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Participants sat approximately 60 cm from the monitor with their heads positioned in a chin rest in a dimly lit room while an experimenter was present.

The face images consisted of 16 grayscale frontal-view images drawn from a standard set of pictures of facial affect (Ekman & Friesen, 1976). These included 8 Caucasian individuals, half women and half men, depicting happy or angry expression. All images were edited using Adobe Photoshop CS3 (Adobe Systems, San Jose, CA) to achieve uniform luminance and contrast. To ensure that gender discrimination was done based on facial features alone, they were further cropped by a uniform oval (130 by 180 pixels) that removed both the hair and the face contour information while preserving the internal features of the face. Moreover, all the face images were edited to be as symmetrical as possible (e.g. the two eyes were placed at proximally the same distance to the nearest edge of each image), and the positions of the facial features were edited to be consistent across images. Although certain facial features (e.g. the shape of the mouth; happy faces tend to have open mouths while angry faces tend to have closed months) are of particular importance for the recognition of expression, this was not controlled in experiment 1A and 2B (but it was controlled in 1B and 2A); such salient differences of local features in different expressions should afford me a better chance to detect modulation effect of expression on gender priming, and expression priming itself. Each face image subtended roughly 5.8° high and 4.1° wide at a viewing distance of 60 cm and was displayed on a uniform grey background. Grayscale masks with the same size as the face image were arrangement of pixels with intensity randomly varying from black to white.

Design and Procedure. Participants were tested in a 2 × 2 × 2 within-subjects design: task dimension (gender vs. expression), task-relevant congruency (congruency of the task-relevant dimension between the prime and the probe: congruent vs. incongruent), and task-irrelevant congruency (congruency of the task-irrelevant dimension between the prime and the probe: congruent vs. incongruent). All factors were fully crossed, yielding 8 experimental conditions. Task dimension was blocked (4 consecutive blocks for the gender condition and 4 for the expression condition; the order was counterbalanced between participants) while task-relevant congruency and task-irrelevant congruency were randomized within each block. There were four images in each of the four trial types defined by the factorial combination of gender (male or female) and expression (happy or angry); each prime and probe was randomly selected from one of these four images with the only constraint that the prime and the probe were of different individuals. Before initiating the experiment, participants viewed all the images to indicate that they were able to differentiate the male part from the female part, and the happy part from the angry part (this generally took around 1 to 2 minutes).

The experiment began with 48 practice trials in 8 blocks followed by 784 experimental trials in 8 blocks, 96 trials each. As illustrated in Figure 1A, each trial began with a white fixation cross (1000 ms) followed by a forward mask (150 ms), a prime (50 ms), a probe (300 ms), a backward mask (150 ms), and another fixation cross (until response but up to 1000 ms), all at the center of the screen. To warn the upcoming of the target, a brief tone began after 500 ms of the first fixation cross. The trial ended as soon as a response was made, up to a limit of 1450 ms after the onset of the target. A warning beep (with a 1 s pause) was given only on each incorrect response. In gender blocks, participants were asked to determine as quickly and accurately as possible whether the target was male or female (in expression blocks: positive or negative); they responded by pressing either the “q” key with one index finger or the “]” key with the other index finger (in expression blocks: either the “h” key with the index finger or the “space” key with the thumb of the dominant hand) on a standard keyboard, which terminated the display and initiated the next trial immediately. Response keys were fully counterbalanced between participants. Note that different keys and response manner were used in gender blocks (“q” key and “]” key with two hands) and expression blocks (“h” key and “space” key with the dominant hand) to minimize motor or response priming effect between tasks. The whole experiment took about 40 min.

In short, participants performed gender and expression tasks in different blocks, with congruency between the prime and the probe manipulated such that the prime and the probe can be congruent or incongruent in the task-relevant dimension and the task-irrelevant dimension.

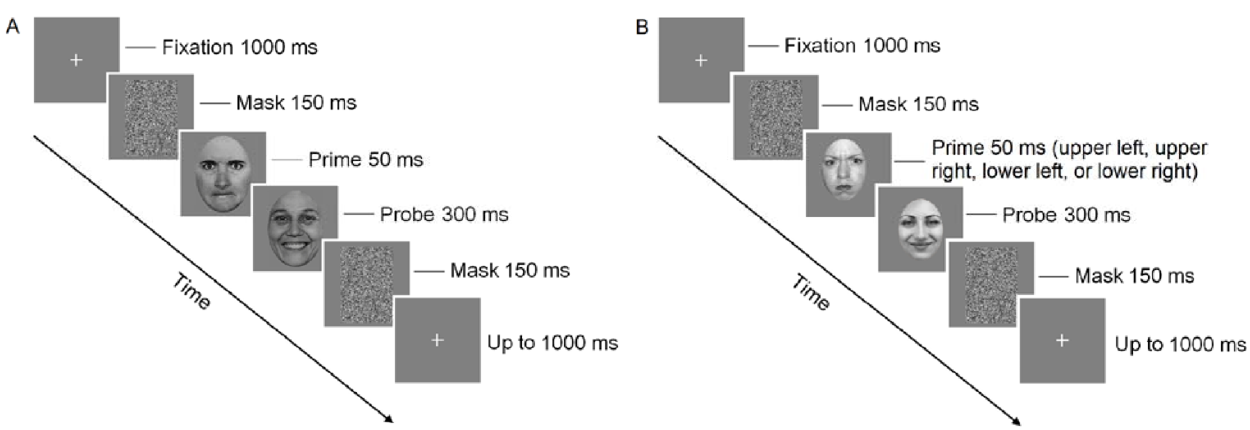

Figure 1. Temporal structure of a trial in Experiment 1. (A) In Experiment 1A, at the beginning of each trial, a white fixation appeared on the screen, followed by a forward pattern mask. Then a prime face was

briefly presented and followed by a target probe face. The prime and probe could be of same or different gender or expression, with all the possibilities randomized across trials. Participants were asked to discriminate the probe according to either gender or expression, followed by a backward mask and a white fixation. The fixation was displayed until response or up to 1000 ms, whichever was shorter. (B) In Experiment 1B, all aspects were the same as Experiment 1A except 1) that the prime was presented randomly in the upper left, upper right, lower left, or lower right corner to prevent sensory summation; and 2) that both happy faces and angry faces have closed mouths to minimize the contributions of local features. The stimuli in this illustration are not drawn to scale.

Results and Discussion

I ask whether gender and expression congruency between the prime and the probe affect face classification based on either the gender or expression dimension. Values with response errors or exceeding three standard deviations from the mean reaction time (RT) for each participant within each congruency condition were excluded from analysis[1]. RTs and accuracies were analyzed by a 2 × 2 × 2 repeated measures analysis of variance (ANOVA) on task (gender vs. expression), task-relevant congruency (congruent vs. incongruent), and task-irrelevant congruency (congruent vs. incongruent). Figure 2 shows the mean RTs and accuracy for the gender task (Figure 2A) and the expression task (Figure 2B).

As predicted by the dimension-based attention hypothesis, gender congruency between the prime and the probe only affected performance in the gender task (Figure 2A, upper panel, left; congruent vs. incongruent: 471.3 ms vs. 510.4 ms, $F(1,31) = 174.91$, $MSE = 48986.07$, $p < .001$) but not the expression task (Figure 2B, upper panel, left; congruent vs. incongruent: 477.2 ms vs. 479.8 ms, $F(1,31) = 0.62$, $MSE = 224.46$, $p = .44$), resulting in a task × gender congruency interaction ($F(1,31) = 73.33$, $MSE = 21289.36$, $p < .001$). Similarly, expression congruency affected performance in the expression task (Figure 2B, upper panel, right; congruent vs.

[1] Excluding outliers did not change the statistical patterns reported below.

incongruent: 467.4 ms vs. 489.6 ms, $F(1,31) = 41.03$, $MSE = 15782.43$, $p < .001$) much more than the gender task (Figure 2A, upper panel, right; congruent vs. incongruent: 489.2 ms vs. 492.5 ms, $F(1,31) = 4.87$, $MSE = 356.65$, $p = .035$), resulting in a task × expression congruency interaction ($F(1,31) = 27.11$, $MSE = 5697.04$, $p < .001$).

Analysis performed on accuracy indicated similar patterns. Gender congruency between the prime and the probe only affected performance in the gender task (Figure 2A, lower panel, left; congruent vs. incongruent: 94.2% vs. 89.0%, $F(1,31) = 34.33$, $MSE = 862.42$, $p < .001$) but not the expression task (Figure 2B, lower panel, left; congruent vs. incongruent: 93.1% vs. 92.7%, $F(1,31) = 0.39$, $MSE = 4.86$, $p = .54$), resulting in a task × gender congruency interaction ($F(1,31) = 22.83$, $MSE = 368.93$, $p < .001$). Similarly, expression congruency between the prime and the probe affected performance in the expression task (Figure 2B, lower panel, right; congruent vs. incongruent: 94.1% vs. 91.7%, $F(1,31) = 10.34$, $MSE = 180.57$, $p = .003$) but not the gender task (Figure 2A, lower panel, right; congruent vs. incongruent: 91.4% vs. 91.9%, $F(1,31) = 0.79$, $MSE = 6.18$, $p = .380$), resulting in a task × expression congruency interaction ($F(1,31) = 7.74$, $MSE = 126.79$, $p = .009$).

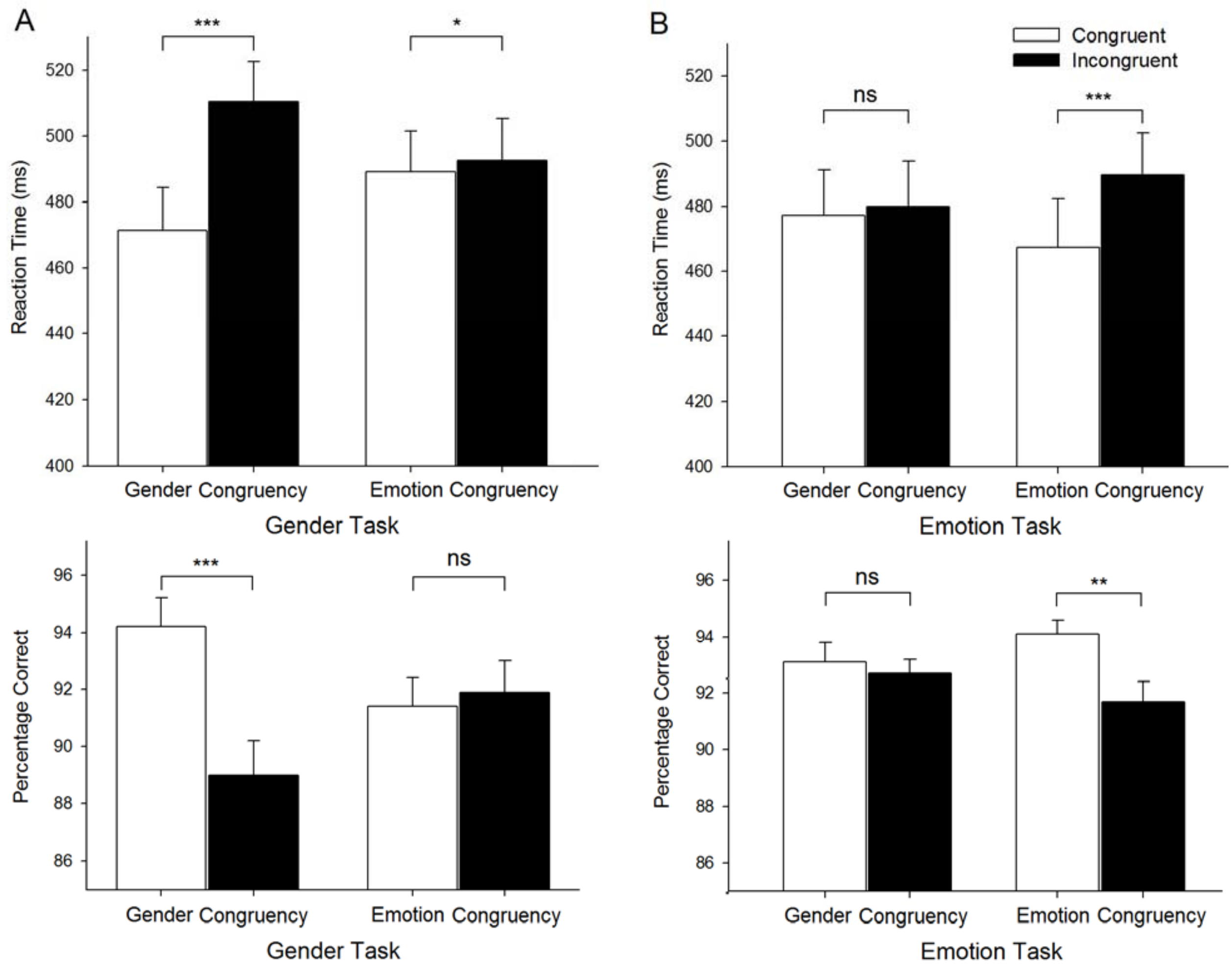

Figure 2. Mean reaction time (RT) and percentage correct as a function of congruency between the prime and the probe faces in Experiment 1A. (A) Gender task: RT (upper panel, left) was significantly shorter and accuracy (lower panel, left) was much higher when the prime and the probe were of same gender than different genders. Although expression congruency had similar effect, the RT effect (upper panel, right) was much smaller, and might be partly due to tradeoff between RT and accuracy (lower panel, right). (B) Expression task: RT (upper panel, right) was significantly shorter and accuracy (lower panel, right) was much higher when the prime and the probe were of same expression than different expressions. There was no corresponding effect in the gender dimension either in terms of RT (upper panel, left) or accuracy (lower panel, left). Error bars correspond to standard errors of means over participants. *** $p < .001$; ** $p < .005$; * $p < .05$; ns, not significant. $N = 32$.

Overall, the present experiment showed that expression classification was affected by the expression congruency effect between the prime and the probe but not affected by the gender congruency between the two. Similarly, gender classification was affected much more strongly by the gender congruency effect between the prime and the probe than by the emotion congruency between the two. These data from masked priming provide strong support to the notion that attention can efficiently select the task-relevant dimension of faces and ignore the congruency effect in the task-irrelevant dimension.

However, note that in the gender task, expression congruency still affected gender classification (Figure 2A, upper panel, right). The effect was small (3.3 ms) yet statistically significant. Although such effect seems to challenge interpretation in terms of dimension-based attention, this effect can be owing to the images used. In particular, since the happy and angry faces in this experiment contained open and closed mouths, respectively, the effect might be due to feature priming rather than expression priming. This hypothesis was tested in Experiment 1B.

Experiment 1B

Since the happy face images used in Experiment 1A contained open mouths while the angry face images possessed closed mouths, it remains possible that in the expression task participants relied on the shape of the mouths rather than the expression itself. Moreover, given that the prime and the probe were shown on the same position on the screen, critical local features such as the mouths might generate sensory summation between the prime and the probe. To rule out these factors, in Experiment 1B, I used face images with closed mouths while randomly presented the prime in four different corners.

Method

Thirty two volunteers (12 men, 20 women) participated in the experiment. The same apparatus and stimuli were used as in Experiment 1A except as follows. To tap into processing of the face itself rather than other features, as illustrated in Figure 1B, I tried to 1) reduce the contributions of sensory summation in the priming effect, by displaying the prime randomly in the upper left, upper right, lower left, or lower right corner of the screen, with the same center-to-fixation distance of 0.5°; 2) reduce the chances that participants performed the expression task based on the shape of the mouth (e.g. discriminate whether the mouth is open or closed) rather than expression itself, by including only faces with closed mouths for both happy and angry expressions; and 3) make the expression task more specific by asking the participants to discriminate whether the face was happy or angry (rather than positive or negative). Since some of these changes could potentially decrease the priming effect, I tried to compensate in two ways: 1) I increased the size of the faces from 5.8° high and 4.1° wide as in Experiment 1A to 8.4° high and 6.0° wide and decreased the viewing distance from 60 cm as in Experiment 1A to 40 cm to make the faces more salient; and 2) I decreased the total number of trials from 48 practice and 784 experimental trials as in Experiment 1A to 32 practice trials (in 8 blocks) and 512 experimental trials (in 8 blocks).

Results and Discussion

Supporting the dimension-based attention hypothesis, gender congruency between the prime and the probe only affected performance in the gender task (Figure 3A, upper panel, left; congruent vs. incongruent: 504.9 ms vs. 535.5 ms, $F(1,31) = 155.91$, $MSE = 29973.02$, $p < .001$) but not the expression task (Figure 3B, upper panel, left; congruent vs. incongruent: 553.8 ms vs. 555.5 ms, $F(1,31) = 0.25$, $MSE = 98.58$, $p = .62$), resulting in a task × gender congruency interaction ($F(1,31) = 63.05$, $MSE = 13316.84$, $p < .001$). Similarly, expression congruency between the prime and the probe affected performance in the expression task (Figure 3B, upper panel, right; congruent vs. incongruent: 546.2 ms vs. 463.1 ms, F(1,31) = 15.49, MSE = 9065.21, $p < .001$) but not the gender task (Figure 3A, upper panel, right; congruent vs. incongruent: 520.0 ms vs. 520.3 ms, F(1,31) = 0.01, MSE = 2.56, p = .92), resulting in a task × expression congruency interaction (F(1,31) = 10.95, MSE = 4381.60, p = .002). Thus, by reducing sensory featural contributions (e.g. using faces with closed mouths and presenting the primes at four different corners), the significant expression congruency effect in the gender task was now nullified.

Analysis performed on accuracy indicated similar patterns. Gender congruency between the prime and the probe only affected performance in the gender task (Figure 3A, lower panel, left; congruent vs. incongruent: 93.7% vs. 89.6%, F(1,31) = 17.58, MSE = 524.37, $p < .001$) but not the expression task (Figure 3A, lower panel, left; congruent vs. incongruent: 89.9% vs. 90.0%, F(1,31) = 0.02, MSE = 0.31, p = .88), resulting in a task × gender congruency interaction (F(1,31) = 33.51, MSE = 275.17, $p < .001$). Similarly, expression congruency between the prime and the probe affected performance in the expression task (Figure 3A, lower panel, right; congruent vs. incongruent: 91.0% vs. 88.9%, t(31) = 1.55, p = .16) but not the gender task (Figure 3A, lower panel, right; congruent vs. incongruent: 92.1% vs. 91.2%, t(31) = 1.55, p = .13), although the interaction between task and expression congruency was not significant (F(1,31) = 1.09, MSE = 18.30, p = .31).

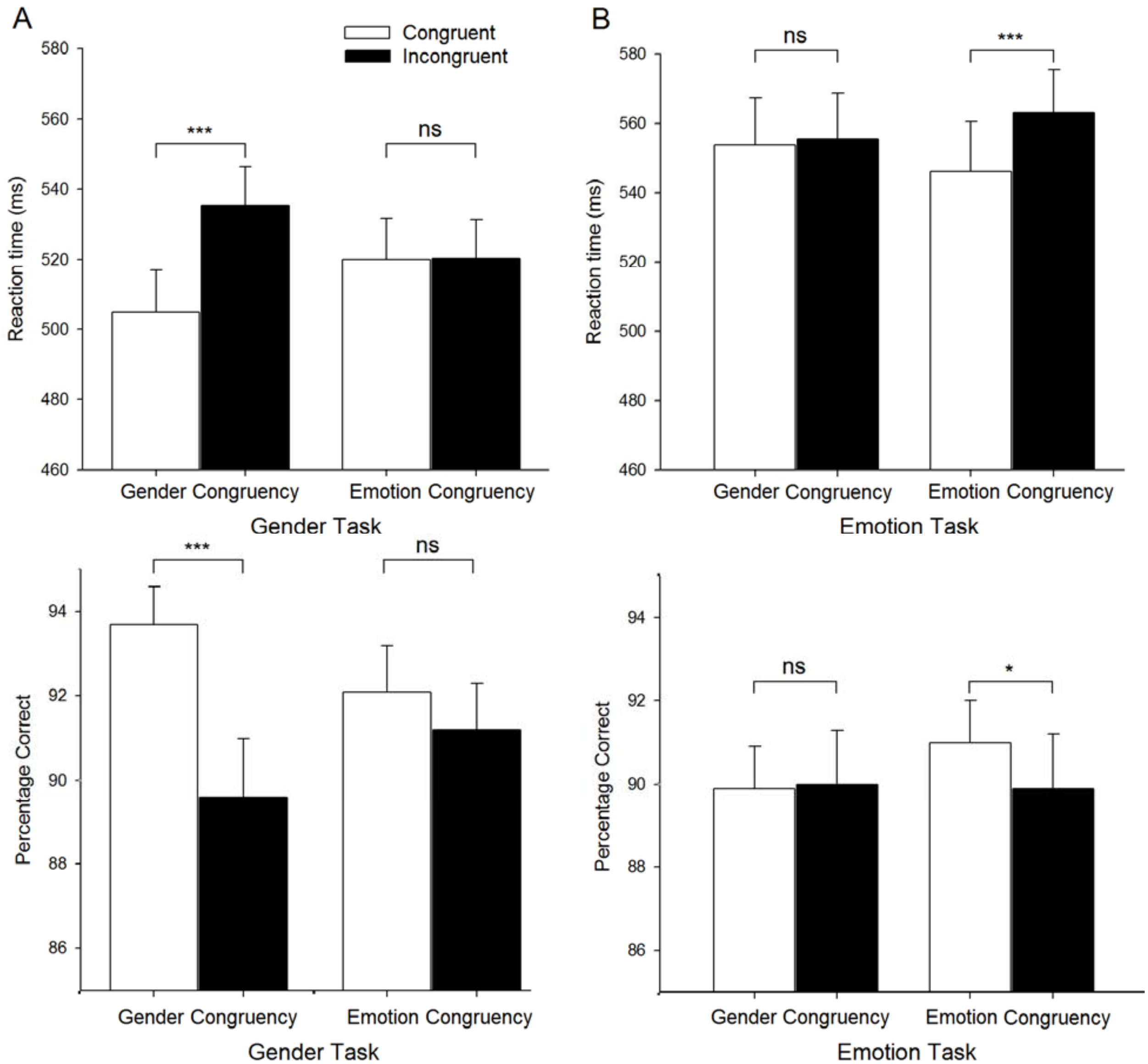

Figure 3. Mean reaction time (RT) and percentage correct as a function of congruency between the prime face and the probe face in Experiment 1B. (A) Gender task: RT (upper panel, left) was significantly shorter and accuracy (lower panel, left) was much higher when the prime and the probe were of same gender than different genders. There was no such effect in the expression dimension either in terms of RT (upper panel, right) or accuracy (lower panel, right). (B) Expression task: RT (upper panel, right) was significantly shorter and accuracy (lower panel, right) was much higher when the prime and the probe were of same expression than different expressions. There was no corresponding effect in the gender dimension either in terms of RT (upper panel, left) or accuracy (lower panel, left). Error bars correspond to standard errors of means over participants. *** $p < .001$; ** $p < .005$; * $p < .05$; ns, not significant. N = 32.

Consistent with Experiment 1A, the present experiment confirmed that performance in the expression dimension was modulated by the expression congruency but not the gender congruency between the prime and the probe. Using faces with closed mouths, this experiment strengthened the conclusion of Experiment 1A by further revealing that performance in the gender classification depended only on the gender congruency effect between the prime and the probe but not on the expression congruency between the two. Moreover, since the prime and the probe in this experiment were always presented on different locations on the screen, the priming effect observed could not be explained by sensory summation between two consecutive images. Rather, together with Experiment 1A, these findings firmly establish the power of dimension-based attention in face recognition—attention can efficiently select the task-relevant dimension of faces such as gender and inhibit the congruency effect in the task-irrelevant dimension such as expression. To generalize this notion, in Experiment 2A and 2B, I used a widely-used paradigm—the Stroop task (Stroop, 1935).

Experiment 2A

To generalize the conclusions of Experiment 1, I devised a face-word Stroop task based on the original word-face Stroop task (Stroop, 1935), as illustrated in Figure 4. The task was identical to that of Experiment 1 except that the target face was now preceded and accompanied by a salient word depicting either the task-relevant dimension (i.e. ‘‘HAPPY’’ or “ANGER” in the expression task; “MALE” or ‘”FEMALE” in the gender task) or the task-irrelevant dimension (i.e. “MALE” or “FEMALE” in the expression task; “HAPPY” or “ANGER” in the gender task) without the prime.

<u>Method</u>

Thirty two volunteers (12 men, 20 women) participated in this experiment. The same apparatus and stimuli were used as in Experiment 1B except as follows, as shown in Figure 4A: 1) instead of a prime face, a word, equally likely to be ‘‘HAPPY’’, “ANGER”, “MALE”, or “FEMALE” in 32-point red bold Arial font, leaded the target face; 2) the same word stayed with the face and disappeared at the same time.

Participants were tested in a 2 × 4 within-subjects design: task dimension (gender vs. expression) and congruency (word and task-irrelevant dimension: congruent vs. incongruent; word and task-relevant dimension: congruent vs. incongruent). All factors were fully crossed, yielding 8 experimental conditions. Task dimension was blocked (2 consecutive blocks for gender condition and 2 for expression condition; the order was counterbalanced between participants); congruency was randomized within each block (64 trials each). There were four images in each of sixteen trial types defined by the factorial combination of gender (male or female), expression (happy or angry), and word (‘‘HAPPY’’, “ANGER”, “MALE”, or “FEMALE”); each face was randomly selected from one of the four images.

The experiment began with 16 practice trials (in 4 blocks) followed by 256 experimental trials (in 4 blocks). Each trial began with a white fixation cross (500 ms) followed by a word (200 ms), a target face with a word (the same as the preceding word) superimposed on it (200 ms), a backward mask (200 ms), and another fixation cross (until response but up to 1500 ms), all of which were presented at the center of the screen. Participants were asked to discriminate the faces only; the words were always task-irrelevant. Other aspects of the procedure including response mappings were exactly the same as used in Experiment 1B.

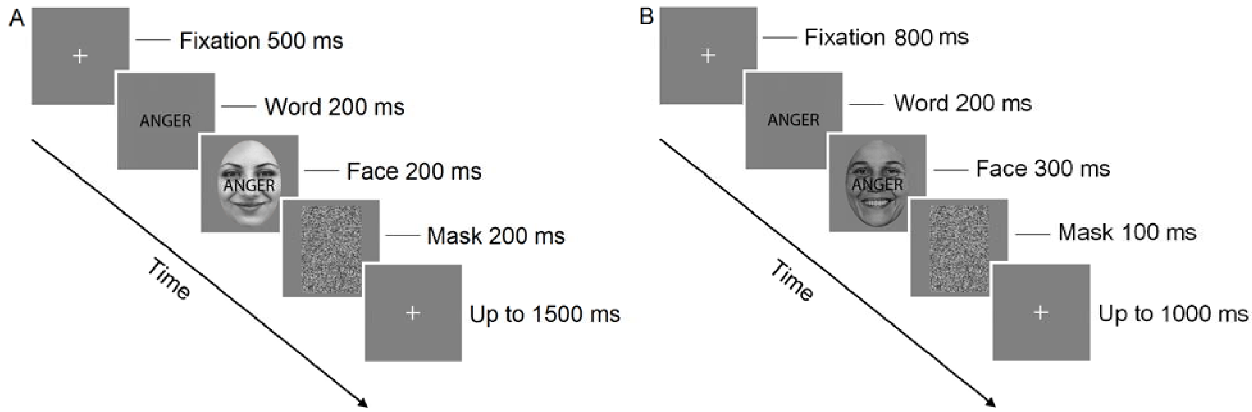

Figure 4. Temporal structure of a trial in Experiment 2. (A) In Experiment 2A, participants were asked to discriminate the face according to either gender or expression. The word preceding the face was always the same as the word on the face, which could be the same as or different from the face in either the task-irrelevant or the task-relevant dimension (i.e. in both the gender and expression tasks, the word could equally be “HAPPY”, “ANGER”, “MALE” or “FEMALE”); (2) In Experiment 2B, all aspects were the same as Experiment 2A except that 1) the word could be the same as or different from the face in the task-irrelevant dimension only (e.g. in the gender task, the word was either “HAPPY” or “ANGER”; in the expression task, the word was either “MALE” or “FEMALE”); and that 2) temporal parameters were changed. The stimuli in this illustration are not drawn to scale; in the experiments, the word was in red rather than in black.

<u>Results and Discussion</u>

Performance of face classification was expected to be modulated by the word only when the word depicted the task-relevant dimension of the face. Table 1 confirms this. In the gender task, as shown in the upper left of Table 1, RTs in the gender congruent condition was significantly faster than those in the gender incongruent condition (congruent, 523.4 ms; incongruent, 544.2 ms; $t(15) = -2.71$, $p = .01$); however, there was no difference between the expression congruent condition and the expression incongruent condition (lower left: congruent, 528.8 ms; incongruent, 531.8 ms; $t(15) = -0.42$, $p = .68$). Similarly, in the expression task, as shown in the lower right of Table 1, RTs in the expression congruent condition was much faster than those in the expression incongruent condition (congruent, 585.8 ms; incongruent, 604.6 ms; t(15) = -2.40, p = .02); however, RTs in the gender congruent condition was similar to those in the gender incongruent condition (upper right: congruent, 613.2 ms; incongruent, 616.0 ms; t(15) = -0.38, p = .71). Analysis performed on accuracy reveals similar conclusions: in the gender task, accuracy was higher in the gender congruent condition than that in the gender incongruent condition (upper left: congruent, 93.5%; incongruent, 87.7%; t(15) = 3.71, p = .001), consistent with the RTs data in the gender task. However, accuracy was even lower in the expression congruent condition than that in the expression incongruent condition (lower left: congruent, 92.2%; incongruent, 94.5%; t(15) = -2.06, p = .048), indicating that the insignificant expression congruency effect in the gender task was not due to RT-accuracy trade-off. Other contrasts were not significant. Taken together, face classification was modulated by words depicting the task-relevant dimension but not by words depicting the task-irrelevant dimension.

Table 1. Mean reaction times (RT) and percentage correct rates (% C) for gender and expression tasks as a function of congruency in the task-relevant and the task-irrelevant dimensions in Experiment 2A (standard errors of means in parentheses; t tests two-tailed; N = 32)

	Task			
Congruency	Gender		Emotion	
	RT (ms)	C %	RT (ms)	C %
Gender				
Congruent	523.4 (23.6)	93.5 (1.2)	613.2 (23.5)	87.0 (1.3)
Incongruent	544.2 (22.1)	87.7 (1.7)	616.0 (22.5)	89.2 (1.4)
t(15)	-2.71	3.71	-0.38	-1.27

p	.01	.001	.71	.22
Emotion				
Congruent	528.8 (20.4)	92.2 (1.4)	585.8 (24.0)	89.5 (1.2)
Incongruent	531.8 (20.8)	94.5 (1.1)	604.6 (23.5)	89.3 (1.3)
t(15)	-0.42	-2.06	-2.40	0.13
p	.68	.048	.02	.90

These findings thus extend evidence from masked priming task to Stroop task. In particular, by using salient words presented ahead of the target face, this experiment assures that the word was fully processed; yet similar null effect of cross-dimensional interference was found as in Experiment 1.Together with Experiment 1, these results further confirm the dimension-based attention hypothesis: humans are able to focus their attention on one particular dimension of faces such that concurrent conflict in other dimensions won't affect the processing of that particular dimension. Since the words could predict the wrong motor response (e.g. in the gender task, a "MALE" word with a female face), participants might develop a mindset to actively inhibit the words. Thus, one might argue that such inhibition strategy may explain why the task-irrelevant words did not influence the classification of the face. To provide a stronger test of the dimension-based attention hypothesis, in Experiment 2B the words were totally task-irrelevant.

Experiment 2B

To generalize the conclusions of Experiment 2A, I created a situation where the words were not actively inhibited by using words that depict only the task-irrelevant dimension (i.e. ‘‘HAPPY’’ or "ANGER" in the gender task; "MALE" or ‘"FEMALE" in the expression task), as illustrated in Figure 4B. This afforded me a better chance to detect the influence of the word on the face, if such effect did exist.

Method

Sixteen volunteers (7 men, 9 women) participated in this experiment. The same apparatus and stimuli were used as in Experiment 2A except as follows, as illustrated in Figure 4B: 1) within each block, the word (in prominent red letters in 27-point Arial font) depicted only the task-irrelevant dimension of the face (i.e. ‘‘HAPPY’’ or “ANGER” in the gender task; “MALE” or “FEMALE” in the expression task); 2) trial number was increased to 48 practice trials (in 8 blocks) and 784 experimental trials (in 8 blocks); 3) temporal parameters were changed using 800 ms fixation, 200 ms word, 300 ms face, 100 ms mask, and up to 1000 ms response time. In short, the main difference is that now participants were tested in a 2 × 2 within-subjects design: task dimension (gender vs. expression) and congruency (word and task-irrelevant dimension: congruent vs. incongruent).

Results and Discussion

Performance in face classification was expected not to be modulated by the congruency effect between the face and the word depicting the task-irrelevant dimension of the face. Table 2 confirms this. In the gender task (Table 2 left), RTs in the expression congruent condition was similar to those in the expression incongruent condition (congruent, 482.9 ms; incongruent, 483.8 ms; $t(15) = -0.29$, $p = .77$). Similarly, in the expression task (Table 2 right), RTs in the gender congruent condition did not differ from those in the gender incongruent condition (congruent, 467.8 ms; incongruent, 467.5 ms; $t(15) = 0.16$, $p = .88$). Analysis performed on accuracy reveals no significant contrasts (gender task: $t(15) = 0.55$, $p = .59$; expression task: $t(15)$

= 0.83, p = .42), indicating that the insignificant RTs effects were not due to trade-off between RTs and accuracy.

Table 2. Mean reaction times (RT) and percentage correct rates (% C) for the gender and expression tasks as a function of congruency in the task-irrelevant dimensions in Experiment 2B (standard errors of means in parentheses; N = 16)

	Gender task		Expression task	
Measure	Expression congruent	Expression incongruent	Gender congruent	Gender incongruent
RT (ms)	482.9 (20.7)	483.8 (21.1)	467.8 (18.4)	467.5 (18.3)
% C	92.5 (1.1)	92.9 (1.1)	93.5 (0.7)	92.9 (0.9)

Strengthening the findings of Experiment 2A, the present experiment confirmed that face classification was not modulated by the task-irrelevant words, even when the words were 1) perceptually salient, 2) presented ahead of the face, and 3) totally task-irrelevant so that active inhibition is unnecessary. Results from the four experiments reveal the existence of dimension-based attention—attention can efficiently select the task-relevant dimension of faces such as gender and inhibit the congruency effect in the task-irrelevant dimension such as expression.

General Discussion

This study sought to examine the efficiency of top-down attention in selecting the task-relevant dimension during face recognition to override interference in the task-irrelevant dimension. The current findings show that, in masked priming, gender classification and expression classification are affected by gender congruency and expression congruency between the prime and the probe, respectively, but not the other way around (Experiment 1A and 1B). Such dimension-based attention effect is not due to sensory summation between the prime and the probe, nor is it owing to priming effect from local features such as mouths (Experiment 1B). Moreover, this effect can be extended to word-face Stroop interference (Experiment 2A and 2B), with performance of face classification modulated by words depicting the task-relevant dimension but not the task-irrelevant dimension of the face (Experiment 2A), even when the words are not actively inhibited (Experiment 2B). The present findings thus demonstrate that attention selection towards a specific dimension in face images can be efficient even though task-relevant and task-irrelevant dimensions are spatially intertwined within the same images.

This effect found in high-level dimensions (i.e. facial expression and facial identity) is in line with feature-based attention theories, although these theories are advocated to deal with low-level features such as motion and orientation (Maunsell & Treue, 2006; Treue & Katzner, 2007). For example, a hallmark of feature-based attention is that attending to a feature (e.g. motion direction) globally enhances the responsiveness of neurons that prefer that feature, including even those whose respective fields are outside of the attentive locations. Such global spread of feature-based attention has been established in monkey neurophysilogical recordings (for a review, see Maunsell & Treue, 2006), behavioral tests (Katzner, Busse, & Treue, 2006; Saenz, Buracas, & Boynton, 2003; Tzvetanov, Womelsdorf, Niebergall, & Treue, 2006) and brain imaging studies (Liu, Larsson, & Carrasco, 2007; Saenz, Buracas, & Boynton, 2002; Serences & Boynton, 2007). Consistent with this line of research, the current study demonstrate a location-independent property of dimension-based attention, as the two dimensions—facial identity and facial expression—are spatially overlapping. More generally, in parallel with the spatial attention tradition, with the spotlight metaphor and the zoom lens models suggesting locations as units of attentional selection (Eriksen & Eriksen, 1974; Eriksen & St. James, 1986; Eriksen & Yeh, 1985; Posner, 1980; Posner et al., 1980), and the object-based attention theories (Baylis & Driver, 1993; Duncan, 1984; Egly et al., 1994; O'Craven, Downing, & Kanwisher, 1999; Roelfsema, Lamme, & Spekreijse, 1998; Scholl, 2001), these findings make a strong case for the role of dimension in selective attention.

Besides the implications for attentional selection, results from this study provide some clues regarding the neural mechanisms of face recognition. In particular, although it is widely accepted that recognition of facial identity and facial expression involves distinct functional (Bruce & Young, 1986) and neural (Haxby, Hoffman, & Gobbini, 2000) routes, yet at what level of analysis does the identity route bifurcate from the expression route remains strongly debated. Specifically, on the one hand, dominant conceptualization and empirical investigations suggest that the functional (Bruce & Young, 1986) and neural (Haxby et al., 2000) routes in the recognition of facial identity and facial expression are distinct and parallel, with the dissociation occurring immediately after the structural encoding stage with distinct visual representations afterwards. On the other hand, such strong independence proposal has been recently challenged by image-based analysis techniques such as principal components analysis (Calder & Young, 2005), which proposes that the dissociation between facial expression and identity is 1) late, occurring after a common representation of both identity and expression, and 2) partial, with some dimensions (principal components) coding both identity and expression. The current findings favor the early dissociation account. In the masked priming tasks, significant priming effect was observed in both the gender and expression tasks, but such priming effect was not modulated by the congruency of the task-irrelevant dimension (i.e. expression congruency in the gender task and gender congruency in the expression task).

That expression and identity can serve as units of attentional selection suggest that binding of gender and expression occurs in a later stage of information processing because a strong early coupling would lead to an object-based attentional selection, which is not the case in this study. More generally, with 50 ms primes, these are unlikely to be subliminal (although participants were not formally tested about the visibility of the primes, a majority of them reported seeing "something" before the probe and some could even tell that the prime was a face), suggesting that strong binding of facial identity and expression does not necessarily occur under limited aware condition (for an argument of binding without awareness, see Lin & He, submitted). More convincing evidence can be obtained by monkey neurophysiological recordings using similar paradigms.

In this study, natural face images rather than schematic face images were used for two reasons: 1) an ultimate goal of vision science is to understand how humans perform natural tasks with natural images (Yuille & Kersten, 2006), and 2) knowledge gained from artificial, parametric stimuli may not generalize to natural stimuli (Felsen & Dan, 2005). This study thus bears strong ecological grounds in predicting face recognition in real life situations. In particular, a high-level dimension-based attention would predict that in real life we can efficiently monitor the changes of facial emotion, which is important for social interaction, while filtering out concurrent interference within facial identity. It will be interesting to know the costs of such dimension-based attention such as to miss important identity information while monitoring expression information, which can be addressed using change blindness paradigms (Rensink, O'Regan, & Clark, 1997; Simons & Levin, 1997; Simons & Rensink, 2005).

In sum, by using masked priming tasks and face-word Stroop tasks, I demonstrated that high-level dimensions such as facial expression and facial identity can serve as units of attentional selection, and that such selection can be so efficient that concurrent interference within another spatially overlapping dimension can be filtered out. Together with studies from feature-based attention, these results suggest that dimension is a useful concept in visual attention research by unifying low level dimensions such as motion, color, and orientation, and high level dimensions such as facial expression and facial identity in the framework of dimension-based attention.

Acknowledgments

I thank M. Bao, M. Bindemann, T. Egner, S. Engel, S. He, Y. Jiang, W. Koutstaal, C. J. Marsolek, and J. M. Wolfe for helpful discussion and/or comments; M. Bindemann, T. Egner, and R. Yu for providing some face stimuli; K. Schneider and P. Zhang for technical assistance; and Sheng He for the use of his laboratory space and for his support throughout this project. Supported by Graduate School Fellowship, Graduate Research Partnership Program Fellowship, and Student Research Award from the University of Minnesota (Z.L.).

References

Apperly, I. A., Samson, D., & Humphreys, G. W. (2005). Domain-specificity and theory of mind: evaluating neuropsychological evidence. *Trends in Cognitive Sciences, 9*(12), 572-577.

Baylis, G. C., & Driver, J. (1993). Visual attention and objects: evidence for hierarchical coding of location. *Journal of Experimental Psychology: Human Perception and Performance, 19*(3), 451-470.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433-436.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305-327.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, *6*(8), 641-651.

Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience*, *2*(7), 671-676.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science, 248*(4962), 1556-1559.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193-222.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113*(4), 501-517.

Egeth, H. E., & Yantis, S. (1997). Visual attention: control, representation, and time course. *Annual Review of Psychology, 48*, 269-297.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*(2), 161-177.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*, 143-149.

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics, 40*, 225-240.

Eriksen, C. W., & Yeh, Y. Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 11*(5), 583-597.

Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature Neuroscience, 8*(12), 1643-1646.

Garner, W. R. (1974). *The processing of information and structure*. New York: Wiley.

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology, 1*, 225-241.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*, 223-233.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry, 51*(1), 59-67.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience, 3*(1), 80-84.

Katzner, S., Busse, L., & Treue, S. (2006). Feature-based attentional integration of color and visual motion. *Journal of Vision, 6*(3), 269-284.

Lin, Z., & He, S. (submitted). Seeing the invisible: The scope and limits of unconscious processing in binocular rivalry.

Liu, T., Larsson, J., & Carrasco, M. (2007). Feature-based attention modulates orientation-selective responses in human visual cortex. *Neuron, 55*(2), 313-323.

Liu, T., Slotnick, S. D., Serences, J. T., & Yantis, S. (2003). Cortical mechanisms of feature-based attentional control. *Cerebral Cortex, 13*(12), 1334-1343.

Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences, 29*(6), 317-322.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature, 401*(6753), 584-587.

O'Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A., & Savoy, R. L. (1997). Voluntary attention modulates fMRI activity in human MT-MST. *Neuron, 18*(4), 591-598.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*(4), 437-442.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3-26.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology, 109*(2), 160-174.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*, 368-373.

Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature, 395*(6700), 376-381.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience, 5*(7), 631-632.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2003). Global feature-based attention for motion and color. *Vision Research, 43*(6), 629-637.

Scholl, B. J. (2001). Objects and attention: the state of the art. *Cognition, 80*(1-2), 1-46.

Serences, J. T., & Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron, 55*(2), 301-312.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1*, 261-267.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: past, present, and future. *Trends in Cognitive Sciences, 9*(1), 16-20.

Stroop, J. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception & Psychophysics, 50*(2), 184-193.

Treue, S., & Katzner, S. (2007). Visual attention: of features and transparent surfaces. *Trends in Cognitive Sciences, 11*(11), 451-453.

Treue, S., & Martinez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature, 399*(6736), 575-579.

Tzvetanov, T., Womelsdorf, T., Niebergall, R., & Treue, S. (2006). Feature-based attention influences contextual interactions during motion repulsion. *Vision Research, 46*(21), 3651-3658.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences, 10*(7), 301-308.