SUPPLEMENTARY MATERIALS

For

White, Palmer & Boynton: Evidence of serial processing in visual word recognition

SUPPLEMENTARY METHODS

Screen and eye-tracking details

In Experiment 1, the screen resolution was 832 x 624 pixels, with maximum luminance 104 cd/m^2. In Experiment 2, the screen resolution was 1024 x 640 pixels, with maximum luminance 90 cd/m^2. The luminance of the words was 21 cd/m^2 in Experiment 1 and 18.9 cd/m^2 in Experiment 2. Stimuli were created with Matlab software (MathWorks) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

In both experiments we monitored the right eye's gaze position with Eyelink eye-trackers (a head-mounted Eyelink 2 and an Eyelink 1000; SR Research). Fixation was established during the ITI at the start of each trial. The trial only advanced if the estimated gaze position was within 1.5 degree of visual angle (°) horizontally and 4° vertically of the fixation cross for at least 200 ms. We allowed more vertical tolerance to accommodate drifts due to pupil size changes or slippage of the eye-tracker. The gaze position averaged over the next 10 samples was taken as the current trial's fixation position. A fixation break was then defined as a deviation of gaze position more than 1° horizontally or 2° vertically from the fixation position established at the trial's start. If a fixation break occurred during the RSVP sequences in Experiment 1, or between the pre-cue offset and post-mask offset in Experiment 2, the trial was immediately terminated. The participant had to press a button to continue the next trial. Terminated trials were repeated at the end of the block, unless fewer than 3 trials remained.

Procedure

In staircase blocks, we used the "single-interval adjustment matrix" adaptive procedure (Kaernbach, 1990) to find thresholds midway between chance and perfect detection performance. The attended side (left, right) alternated across staircase blocks.

In each hour-long session after the staircase session, participants completed 12 semantic blocks (one for each category) and 12 color blocks, alternating order across

sessions. The difficulty levels *D* and *I* were adjusted as necessary between sessions to keep performance near 80% correct in the single-task conditions. We discarded and re-ran any set of 12 blocks if both single- and dual-task proportion correct were below 0.7 or above 0.9. This occurred if stimulus levels *D* or *I* missed the participant's true threshold. As a result, a total of 96 and 84 blocks were re-run in Experiments 1 and 2, respectively. Each participant completed 960 usable trials of each judgment type.

Analysis

On each trial the participant reported his/her level of confidence that a target was present on a 1-to-4 rating scale. As a bias-free measure of accuracy, we computed the area under the Receiver Operating Characteristic (ROC) curve, A_g (Pollack & Hsieh, 1969). The ROC plots hit rates (HR) as a function of false alarm rates (FR). To compute these rates from the participants' response ratings, we varied an index *i* from 0 to 4. At each index level we coded responses greater than *i* as "yes" responses. For each value of *i*, HR(*i*) is the proportion of "yes" responses on target-present trials and FR(*i*) is the proportion of "yes" responses on target-absent trials. For instance, when *i=3*, only response ratings of 4 (highest confidence) on target-present trials are considered hits, and only response ratings of 4 on target-absent trials are considered false alarms. The five pairs of HR(*i*) and FR(*i*) trace out a curve, the area under which (A_g) is a measure of accuracy.

Throughout the text we report bootstrapped 95% confidence intervals (CIs) for average measurements. To compute these, we generate a distribution of resampled means. Each of those is the mean of 10 values sampled with replacement from the original set of 10 participants. The CI is the range from the 2.5th to 97.5th percentile of the distribution of resampled means.

Models of capacity limits

In the Attention Operating Characteristic plots shown in Figure 2, we compare our dual-task accuracy data to the quantitative predictions of three models of capacity limits.

The following definitions are for "standard" versions of these models similar to those defined for accuracy measures of visual search tasks (Scharff, Palmer, & Moore, 2011). All models are based on signal detection theory: we assume that the perceptual system analyzes stimuli on each side i by computing an estimate of the evidence that a target is present, which corresponds to a random variable E_i. In all models, we assume that across trials of one condition, the E_i values for side i are independent and identically distributed Gaussian variables, and within each trial E_1 and E_2 are independent of each other.

Sensitivity (d') depends on the mean difference in E between target-present and target-absent trials, relative to the across-trial variability in E. To make a judgment about each stimulus, E is compared against three criteria $c1, c2, c3$ to determine which of the four response keys to press.

We first label the measured single-task accuracy levels for the left and right stimuli A_{L1} and A_{R1}, respectively. The models then use these single-task accuracy levels to predict the dual-task levels A_{L2} and A_{R2}.

1. *Unlimited-capacity parallel processing model*: This model assumes that the distribution of E_i is independent of the condition (single-task vs. dual-task). The model therefore predicts no dual-task deficit:

   $A_{L2} = A_{L1}$, and

   $A_{R2} = A_{R1}$

2. *Fixed-capacity parallel processing model*: This model assumes that the perceptual system extracts a constant amount of information from the display regardless of the condition (single-task vs. dual-task). If that information is equally distributed among the stimuli in the dual-task condition, only half as much information is available about each stimulus compared to the single-task condition. One way to conceptualize fixed capacity is to assume that computing the estimate E_i of target presence depends on gathering sensory 'samples' from the stimulus (Shaw, 1980). All attended stimuli must share a fixed number S of samples that can be gathered from the whole display per unit time. The variability of E_i is inversely proportional to the number of

samples assigned to it, which means that reducing the number of samples for one stimulus decreases sensitivity. As the proportion q of samples given to the right stimulus increases from 0 to 1, this model's prediction traces out the black curve that connects the two single-task data points in the AOC plot. This curve is computed as follows. We first calculate the value of d' for the left and right single-task conditions:

$d'_{R1} = \sqrt{2}\,z(A_{R1})$, and

$d'_{L1} = \sqrt{2}\,z(A_{L1})$

where z is the inverse of the normal cumulative distribution function.

We then assume that in the dual-task condition, the right stimulus receives qS samples, and the left stimulus receives the remaining $(1-q)S$ samples, where $0<q<1$. From signal detection theory, receiving a portion q of samples changes d' for that stimulus by a factor $\sqrt{q}$. Therefore, d' for each side in the dual-task condition is:

$d'_{R2} = \sqrt{q}(d'_{R1})$, and

$d'_{L2} = \sqrt{1-q}(d'_{L1})$

We then convert these d-prime measures into A_g accuracy levels:

$A_{R2} = ncdf(\left.{d'_{R2}}\middle/{\sqrt{2}}\right.)$, and

$A_{L2} = ncdf(\left.{d'_{L2}}\middle/{\sqrt{2}}\right.)$

where $ncdf$ is the normal cumulative distribution function.

The parallel model can be generalized to predict less severe deficits, by assuming that the total number of samples available in the dual-task condition is more than the number (S) in the single-task condition. That is, the total number of samples shared between the two locations is aS, where $1=<a<=2$. Increasing a pushes the predicted curve into the upper right corner, eventually meeting the prediction of the unlimited capacity model when $a=2$.

3. *All-or-none serial processing model*: This model assumes the same stimulus representations and decision rule as the parallel models. Like in the unlimited-capacity parallel model, the distributions of E are identical in single- and dual-task conditions. The difference is that this model assumes that the participant only processes one stimulus and does not have time to even start processing the second,

and therefore must guess when asked about it. As the proportion of trials v in which the right side is processed increases from 0 to 1, this model's prediction traces out the diagonal black line in the AOC plot. Accuracy values along that line, which are in units of A_g but may also be viewed as the proportion of correct responses, are therefore a mixture of probabilities. Dual-task accuracy for the right stimulus therefore is:

$$A_{R2} = A_{R1} v + 0.5(1\text{-}v).$$

The second term in that equation reflects the fact that the participant must guess (with probability correct 0.5) on the (1-v) proportion of trials in which the right stimulus is not processed at all. The right and left sides trade off linearly, so dual-task accuracy for the left stimulus is:

$$A_{L2} = A_{L1}(1\text{-}v) + 0.5v.$$

The serial model can be generalized to account for less severe deficits assuming that on some fraction b of dual-task trials, *both* sides are fully processed (with the same sensitivity as in the single-task conditions). Therefore, the resulting dual-task accuracy is a mixture of trials in which only one stimulus is processed and no information is acquired about the other, and in which both stimuli are fully processed.

$$A_{R2} = bA_{R1} + (1\text{-}b)*(A_{R1}v + 0.5(1\text{-}v))$$
$$A_{L2} = bA_{L1} + (1\text{-}b)*(A_{L1}(1\text{-}v) + 0.5v)$$

In this generalized model, v can be interpreted as the proportion of trials in which the right stimulus is processed 'first'. On only b fraction of trials does the 'second' stimulus get processed at all. We were able to solve these equations for the values of b and v, given each participant's data (single-task accuracy values A_{L1} and A_{R1}, and dual-task values A_{L2} and A_{R2}).

SUPPLEMENTARY RESULTS

Effect of response order in dual-task trials

On dual-task trials, the post-cues prompted participants to report judgments of both sides, in random order. For both judgment types in Experiment 1 and color judgments in Experiment 2, the mean accuracy of second responses was 0.01 ± 0.01 higher than the first, but that difference was not significant (all $p > 0.3$). For semantic

judgments in Experiment 2, second responses were on average 0.03 ± 0.01 A_g units more accurate (t(9)=4.04, p=0.003). Because these differences were small, all other analyses collapse across first and second dual-task responses. It is likely that these order effects were small because the order was unpredictable (Ernst, Palmer, & Boynton, 2012). For a contrasting result in a dual task, see Egeth & Pachella, 1969.

Effect of hemifield

In Experiment 1, overall semantic accuracy was on average 0.16 ± 0.02 A_g units higher for targets on the right than left side (t(9)=9.50, p<0.001; CI = [0.13 0.19]). The magnitude of that difference did not significantly vary across single-task and dual-task conditions (ANOVA interaction: F(1,9)=3.56, p=0.092). In Experiment 2, the mean hemifield difference was 0.23 ± 0.02 (t(9)=12.5, p<0.001; CI = [0.20 0.27]). That difference was greater in the dual-task (0.28 +/- 0.02) than in the single-task condition (0.15 +/- 0.02; ANOVA interaction: F(1,9)=60.47, p<0.001). However, that interaction was not significant in units of d', which is more appropriate as it is unbounded (Experiment 1: F(1,9)<1; Experiment 2: F(1,9)=3.42, p=0.097). Therefore, even when instructed to attend to only one side, participants are less accurate at categorizing words in the left hemifield are than in the right. Therefore, if an attentional bias explains the hemifield asymmetry in general, it must be difficult to voluntarily overcome.

The effect of hemifield almost disappeared for color judgments: within the single-task conditions, there was a significant interaction between judgment type and side (Experiment 1: F(1,9)=29.37, p<0.001; Experiment 2: F(1,9) = 20.11, p=0.0015). The same analyses of d' supported the same conclusions. For color judgments in Experiment 1, accuracy was on average 0.03 ± 0.01 A_g units higher on the right than the left (t(9)=2.74, p=0.023; CI = [0.01 0.05]). In Experiment 2, that mean difference was 0.04 ± 0.02 A_g units (t(9)=2.17, p=0.058; CI = [0.0 0.07]). In neither experiment did the color judgment side difference interact with pre-cue condition (single vs. dual; F<1).

Dual-task accuracy correlations

To test the prediction of the serial model that assumes across-trial switching of attention between the left and right sides, we computed linear correlation coefficients (ρ) between the accuracies of the two responses in dual-task trials of each judgment type.

Semantic judgments: across all trials of Experiment 1, the mean ρ was negative but not significantly below zero: -0.02 ± 0.01 (t(9) = -1.23, p=0.25; CI = [-0.04 0.01]. We reasoned, however, that because this task requires yes/no target detection and most participants had a conservative bias, the serial processing bottleneck might affect only the detection of targets. We therefore split our data into sets of trials depending on whether any target was present or absent. When no targets were present on either side, mean ρ was positive: 0.22 ± 0.03 (t(9) = 6.13, p<0.001; CI = [0.16 0.29]). When at least 1 target was present, the mean ρ was negative: -0.11 ± 0.02 (t(9) = 6.17, p<0.001; CI = [-0.15 -0.08]).

In Experiment 2, across all semantic trials the mean ρ was not again significantly different from zero: -0.02 ± 0.02 (t(9) = 1.28, p=0.234; CI = [-0.05 0.01]). When no targets were present on either side, the mean ρ was positive but not significantly so: 0.04 ± 0.05 (t(9) = 0.81, p=0.44; CI = [-0.06 0.14]). On trials with 1 or 2 targets, mean ρ was negative: -0.06 ± 0.03 (t(9) = 2.07, p=0.068; CI = [-0.11 -0.01]).

Although the semantic accuracy correlations were negative when at least one target was present, the basic prediction of the serial switching model was not met across all trials. This may be because the overall accuracy correlation collapses over hits and correct rejections, and over misses and false alarms. It is therefore corruptible by shifts in criterion or response bias. To investigate this possibility, we computed a measure of bias:

$$B = -0.5*(z(HR) + z(FR))$$

where z inverse of the unit normal cumulative distribution function, HR is the hit rate, and FR is the false alarm rate. In the framework of signal detection theory, B is the distance of the decision criterion from the neutral point midway between the distributions of sensory evidence on target-absent and target-present trials, in units of the standard deviation. A larger B means the participant is more conservative, needing more evidence to say yes.

Indeed, *B* depended on the accuracy of the other side's response. For semantic judgments in Experiment 1, participants were more conservative (higher *B*) when the other side's response was correct than incorrect (mean difference: 0.25 ± 0.05, t(9)=5.11, p<0.001; CI = [0.16 0.34]). That difference was larger when the other side had no target (0.57 ± 0.11, t(9)=4.99, p<0.001; CI = [0.38 0.82]) than when it did (0.04 ± 0.15, t(9)=0.28, p=0.78; CI = [-0.25 0.30]). The interaction (F(1,9)= 4.92, p=0.054) between other side accuracy and other side target presence corresponds to the difference in accuracy correlations between target-present and absent trials, with a significantly positive correlation when neither side had a target.

For semantic judgments in Experiment 2, participants were again more conservative when the other side's response was correct (mean difference in *B* = 0.25 ± 0.05, t(9)=4.94, p<0.001; CI = [0.14 0.34]), but that effect did not strongly differ between target-present and absent trials (F(1,9)<1). Correspondingly, the accuracy correlations in Experiment 2 differed less between target present and absent trials.

In summary, the analysis of accuracy correlations is complicated by changes in decision bias that depend on the other side's response and the other side's stimuli. We therefore rely on the analysis of bias-free accuracy (A_g) presented in the main text.

Color judgments showed a very different pattern of accuracy correlations. In Experiment 1, the correlation was overall positive (mean ρ=0.07 ± 0.02; t(9) = 3.97, p=0.003; CI = [0.04 0.10]), regardless of whether there were (0.05 +/- 0.01) or were not (0.14 ± 0.08) any color targets present. In Experiment 2, the color accuracy correlation was again overall positive (mean ρ=0.05 ± 0.02, t(9) = 3.23, p=0.010; CI = [0.02 0.08]) regardless of whether there were (0.04 ± 0.02) or were not (0.09 ± 0.04) any color targets present. These positive accuracy correlations mirror the positive effect of the other side's accuracy on A_g, as shown in the main text, and could be explained by fluctuations in overall effort or arousal across trials. The bias (*B*) of color judgments was not significantly affected by the other side's accuracy (mean effects in Experiment 1: 0.07 ± 0.05, t(9)=1.52, p=0.164, CI = [-0.01 0.16]; in Experiment 2: 0.06 ± 0.04, t(9)=1.66, p=0.131, CI = [-0.01 0.14]).

Congruency effects

8

A congruency effect is the influence of one stimulus on the participant's response to another stimulus. The classic "flanker effect" (Eriksen & Eriksen, 1974) is an example of a congruency effect: participants are instructed to discriminate a target stimulus that is flanked by irrelevant stimuli that may be congruent (correspond to the same response) or incongruent (correspond to the opposite response). Responses are typically impaired on incongruent trials, which is a sign that the flankers were not completely filtered out.

We compared accuracy on 'congruent' trials, when both sides have a target or both don't have a target, with accuracy on 'incongruent' trials, when only one side has a target. For semantic judgments, accuracy was higher on congruent than incongruent trials, by an average of 0.05 ± 0.01 A_g units in Experiment 1 (t(9)=4.23, p=0.002; CI = [0.03 0.08]), and 0.03 +/- 0.01 in Experiment 2 (t(9)=4.62, p=0.001; CI = [0.02 0.04]). Therefore, in addition to their inability to recognize both words simultaneously, participants were not perfectly able to maintain separate representations of the two stimuli and select the correct one to respond to.

Congruency effects were absent or reversed for color judgments: accuracy tended to be higher on *incongruent* trials. In Experiment 1**,** that mean difference was 0.03 ± 0.02, but not significant (t(9)=1.18, p=0.27; CI = [-0.02 0.07]). It was larger in Experiment 2: 0.06 ± 0.01 (t(9)=5.49, p<0.001; CI = [0.04 0.08]). This may be because on incongruent trials of Experiment 2, participants directly compared saturation levels across space to better detect changes from the baseline gray. In contrast, in Experiment 1, participants were perhaps more likely to compare colors of successive RSVP stimuli across time.

SUPPLEMENTARY REFERENCES

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 443–446.

Egeth, H., & Pachella, R. (1969). Multidimensional stimulus identification. *Perception and Psychophysics*, *5*, 341–346.

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

Ernst, Z. R., Palmer, J., & Boynton, G. M. (2012). Dividing attention between two transparent motion surfaces results in a failure of selective attention. *Journal of*

    Vision, *12*, 1–17.

Kaernbach, C. (1990). A single- interval adjustment- matrix (SIAM) procedure for
    unbiased adaptive testing. *The Journal of the Acoustical Society of America*, *88*,
    2645–2655.

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

Pollack, I., & Hsieh, R. (1969). Sampling variability of the area under the ROC-curve
    and of d'e. *Psychological Bulletin*, *71*, 161–173.

Scharff, A., Palmer, J., & Moore, C. M. (2011). Extending the simultaneous-sequential
    paradigm to measure perceptual capacity for features and words. *Journal of
    Experimental Psychology: Human Perception and Performance*, *37*, 813–833.

Shaw, M. L. (1980). Identifying attentional and decision-making components in
    information processing. In R. Nickerson (Ed.), *Attention and Performance VIII* (pp.
    277–296). New York: Routledge.

Stimulus set:

animals	anatomy	clothing	food	professions	transport
bear	chin	belt	cake	aide	bike
bull	face	boot	milk	chef	boat
crow	hair	coat	rice	dean	cart
deer	hand	gown	soup	maid	jeep
frog	head	robe	stew	monk	limo
goat	skin	shoe	taco	poet	raft
lion	knee	sock	tofu	actor	ship
swan	lips	suit	wine	agent	skis
wolf	neck	veil	veal	chief	sled
seal	nose	vest	bacon	clerk	taxi
toad	shin	beret	bagel	judge	tram
worm	ankle	cloak	bread	mason	barge
cobra	bicep	clogs	curry	mayor	blimp
eagle	brain	dress	candy	miner	buggy
gecko	cheek	glove	gravy	nurse	canoe
mouse	elbow	jeans	pasta	pilot	cycle
shark	femur	pants	pizza	rabbi	ferry
sheep	heart	parka	salad	thief	kayak
skunk	lungs	scarf	sugar	tutor	plane
snake	scalp	skirt	sushi	usher	sedan
tiger	skull	shawl	toast	valet	train
whale	spine	shirt	burger	artist	truck
baboon	thigh	tunic	cheese	author	vespa
badger	thumb	visor	noodle	jockey	wagon
cougar	tooth	bikini	butter	barber	yacht
coyote	torso	blazer	cookie	butler	dinghy
falcon	waist	blouse	cereal	editor	glider
ferret	wrist	bonnet	coffee	intern	hearse
gerbil	artery	corset	omelet	lawyer	humvee
gopher	finger	fedora	pastry	notary	pickup
jaguar	muscle	jacket	pickle	priest	rocket
lizard	pelvis	mitten	sorbet	sailor	skates
monkey	spleen	sandal	sundae	tailor	sleigh
parrot	tendon	shorts	waffle	umpire	subway
weasel	throat	slacks	yogurt	warden	tanker

plants	buildings	music	household	environment	materials
fern	bank	drum	bath	cave	clay
leaf	barn	gong	crib	dune	felt
lily	fort	harp	desk	hail	gold
moss	home	jazz	lamp	hill	lace
root	jail	lute	sink	lake	lead
rose	mall	lyre	sofa	mist	silk
seed	mill	oboe	vase	pond	wool
stem	shed	song	bench	rain	zinc
tree	silo	tuba	chair	reef	alloy
twig	tomb	tune	clock	snow	caulk
vine	abbey	banjo	couch	wave	epoxy
weed	arena	bongo	divan	wind	latex
aspen	cabin	bugle	dryer	beach	nylon
birch	crypt	cello	futon	cliff	putty
bloom	igloo	chime	frame	cloud	rayon
daisy	lodge	choir	shelf	creek	satin
grass	manor	conga	stool	frost	slate
holly	motel	flute	stove	marsh	steel
lilac	shack	kazoo	table	ocean	suede
maple	tower	opera	towel	ridge	bronze
peony	villa	piano	carpet	river	burlap
petal	armory	sitar	cradle	shore	canvas
shrub	bunker	viola	drapes	sleet	carbon
stalk	castle	waltz	drawer	storm	cement
thorn	chapel	anthem	faucet	swamp	chrome
tulip	church	ballad	fridge	canyon	copper
branch	duplex	chorus	hamper	crater	cotton
dahlia	garage	cymbal	hearth	geyser	enamel
cactus	mosque	fiddle	heater	gulley	kevlar
clover	palace	guitar	lounge	island	marble
flower	prison	melody	mirror	lagoon	pewter
orchid	saloon	reggae	pillow	ravine	silver
pollen	school	rhythm	shower	stream	stucco
violet	shrine	sonata	toilet	tundra	teflon
willow	temple	violin	washer	valley	velvet