Dynamic differences in the cued identification of orientation, color, and facial expressions

Hans-Christoph Nothdurft
Visual Perception Laboratory (VPL) Göttingen, Germany

The paradigm of *cued visual selection* (Nothdurft, 2017a, www.vpl-reports.de/6/) was tested on different features and the dynamics in the identification of orientation targets, color targets, and facial expressions were compared. In large arrays of up to 80 items, one item (the target) was cued at various delays after stimulus onset; observers had to identify this item and report its properties. Performance accuracy was measured for different presentation times, Δt, after the cue. With all features, accuracy improved with longer presentations. The data were fitted with cumulative Gaussian functions and the presentation time for 75% correct responses, Δt_{75}, was estimated for each pattern and each cue delay; the values reflect the dynamics and relative strengths of underlying neural signals. For *orientation*, discrimination was strongly accelerated at cue delays 50-100 ms after the pattern onset. With longer delays, it slowed down again. For *color*, the acceleration was stronger and more prolonged. It started earlier and reached its maximum about 400 ms after stimulus onset. With longer delays, performance deteriorated but never fell back to the level at zero delay. With *facial expressions*, the required presentation times were generally increased; faster discrimination at short delays was less pronounced and not seen in all observers. Shortest Δt_{75} values were reached 50-200 ms after stimulus onset, at similar delays as for orientation. All data were used to reconstruct the signals that provide the cued discrimination of orientation, color, and facial expressions. The link to neurophysiological studies and the evidence for different impacts of these features on the control of spatial attention are discussed. *© Author*

Citation: Nothdurft, H.C. (2020). Dynamic differences in the identification of orientation, color, and facial expressions. VPL-reports, 10, 1-27, www.vpl-reports.de/10/

INTRODUCTION

It is generally assumed that the *identification* of a not too simple visual object (but not necessarily its detection) requires the perceptual processes being focused upon this object and its properties—a phenomenon often described as shifting "focal attention" to a target (Joseph & Optican, 1996; Joseph, Chun, & Nakayama, 1997; Nothdurft, 1999; but see, e.g., Braun & Julesz, 1998). Attention can be attracted and "guided" in various ways (Wolfe, 1994; Wolfe & Horowitz, 2004; Berga, Fdez-Vidal, Otazua, Leborán, & Pardo, 2019). With salient cues, for example, attention can be directed ("shifted") to certain locations in a scene (Eriksen & St. James, 1986; Nothdurft, 2002). Numerous studies in the past have analyzed the spatial and temporal properties of such attention shifts when targets followed various cues at different delays (e.g., Posner, 1980; Posner & Cohen, 1984; Downing and Pinker, 1985; Eriksen & St. James, 1986; Benso, Turatto, Mascetti, & Umiltá, 1998). But attention may also be shifted (and targets selected) in an already visible pattern and then reveal the dynamics of selection and identification processes (Nothdurft, 2002, 2017a). Using this method of "cued visual selection" (CVS) it was found that the presentation time needed to identify a cued target may also vary with the "cue delay", now between stimulus onset and the (later) presented cue (Nothdurft, 2017a). A target line cued shortly after its onset is faster identified than a target line cued much later, although the longer visibility before the cue should have provided a much better encoding and analysis of target properties. It is important to notice that these variations do not reflect dynamics of the cuing process itself (which was identical for each delay) but likely reflect modulations of underlying neural signals after stimulus onset. One way to interpret the data is that neural signals vary in strength and must be accumulated to let observers make a reliable decision about the selected object. A target represented with a strong signal at the moment of cued selection might require a shorter accumulation time for a reliable decision than a target with a momentarily weak signal. The different presentation times to reach constant performance at various delays should then be directly related to the (variable) strength of underlying neural signals after target onset (Nothdurft, 2017a).

In recent series of experiments (Nothdurft, 2017a, b, 2018, 2019) I have measured the discrimination of *oriented lines* in various configurations and have computed the presentation time that was necessary to reach a constant performance accuracy at different delays. These behavioral data were by and large predicted by population responses of single cells in area V1 under similar conditions (Nothdurft, 2017a, b). Analysis revealed the transient character of V1 responses to oriented lines (e.g., Knierim & Van Essen, 1992; Nothdurft, Gallant, & Van Essen, 1999), the decaying efficiency of exogenous cuing (Nakayama & Mackeben, 1989), and response differences between targets in uniform and popout configurations that has been seen in single-cell recordings from area V1 (Knierim & Van Essen, 1992; Kastner, Nothdurft, & Pigarev, 1997; Nothdurft, Gallant, & Van Essen, 1999). In an additional experiment, also the timing of cued selection was studied and different delays were found for the encoding of cues and the perception of target properties, not only for orientation but also for motion, color, and luminance polarity (Nothdurft, 2018).

To learn more about the general dynamics of cued visual selection and the reconstruction of underlying neural signals, the present work was designed to expand the original studies on line orientation to other feature dimensions, in particular to color and facial expressions. While orientation differences are first represented in cortical area V1, the neural encoding of color starts very early in the visual system. Already in the macaque retina and lateral geniculate nucleus (LGN) there are color-specific cells which respond differentially to stimuli in different colors. The responses are strong and relatively sustained (see, for example, Nothdurft & Lee, 1982a, b); that is, most color-specific cells respond as long as the color stimulus is shown in their receptive fields. If the CVS method can indeed reveal the properties of neural signals in the task, the identification of cued *color* targets should reveal different dynamics than the identification of cued *orientation* targets. On the other hand, neurons in certain cortical regions, in humans the fusiform facial area (FFA), in monkeys the superior temporal sulcus (area STS), are particularly sensitive to faces and facial expressions (see, e.g., Baylis, Rolls, & Leonard, 1985; Oram & Perrett, 1992; Tong, Nakayama, Moscovitch, Weinrib & Kanwisher, 2000). In the hierarchical sequence of visual processing, these areas are activated after area V1 and likely require multiple combinations of V1 neurons to reach their specificity. It should be interesting to see if the cued discrimination of, e.g., happy and angry faces would reveal other dynamics than the cued discrimination of colors and line orientations.

In three interleaved series of experiments, observers were asked to discriminate and identify cued oriented lines, cued colored squares, and cued schematic faces with different emotional expressions (Fig. 1). Targets were presented in large arrays of randomly mixed items and were individually selected by the short presentation of a "four-dot cue" in every presentation. The usage of faces and facial expressions as visual targets was deliberate. While colors set up a different stimulus quality, oriented lines and faces are both spatial shapes and might be encoded by the same neurons in simple (orientation) or more complex representations (faces). However, instead of

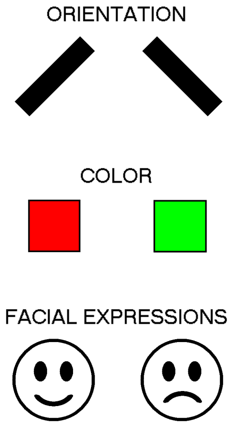

Figure 1. *Feature dimensions studied in this paper.* In three series of experiments, the speed of feature discrimination was measured after attention was directed to one of many different items. Observers had to discriminate two line orientations, red and green target colors, and happy or angry facial expressions.

testing geometric combinations of differently oriented lines, like squares, rhomboids, arrows, or even texture, facial expressions were chosen because of their known unique representation in apparently specialized areas of the visual system.

Only target-cue sequences were tested in which the cue was superimposed upon a test pattern that was already present or was switched on simultaneously with the cue. The delay between pattern and cue onsets was systematically varied, as was the stimulus presentation time ("target duration") after the cue until the test pattern was masked and finally disappeared. The dependent variable was the accuracy at which observers identified the cued target. From large test series at each tested delay, the presentation time for an accuracy of 75% correct responses was calculated. This duration is assumed to be inversely related to the strength of the underlying neural signal at this delay (Nothdurft, 2017a, 2019).

While the measurements revealed strong differences in the dynamics of tested features, a simple sequence of color first, then orientation and finally facial expressions was not strictly observed. Color identification, for example, was indeed faster than orientation discrimination, as perhaps expected, but reached its performance maximum later than orientation. The discrimination of facial expressions, on the other hand, required much longer stimulus presentations and was not always as strongly facilitated at short cue delays as the discrimination of both color and orientation. Possible causes of these differences will be discussed. Altogether, there were notable variations in detail between observers, despite their overall similar performances in these tests.

The experiments reported here were part of a larger series of experiments in which the present observers had also participated.

GENERAL METHODS

Overview

Experiments were designed to measure the presentation time needed to identify the features of selected targets that were cued at various delays after stimulus onset. Three feature dimensions (orientation, color, facial expressions) were studied in separate experiments. In each experiment, observers saw regular arrays of similar (but not identical) items (see, e.g., Fig. 3); one of these items was (later) marked by a cue and thus selected as target. Shortly after the cue presentation, all items were masked to limit the time for perceptual analysis. Two parameters were systematically varied in the course of experiments, the cue *delay* (after stimulus onset) and the stimulus presentation time (*target duration*) between the cue and the mask onsets. Performance accuracy was averaged over many similar trials.

For each feature, all data from a given cue delay (i.e., all data from different presentation times in this condition) were fitted with a Gaussian cumulative function, from which the presentation time for 75% accuracy for this delay was taken (Δt_{75}). These values represent the different speed at which a target can be identified, and thus allow for a direct comparison of cued target identification at different delays and between different features.

Stimuli

Stimuli were generated with DOS VGA techniques on a monitor in front of the observer. For Experiments 1 (orientation) and 2 (color) a 17" Sony Trinitron monitor (Sony Trinitron multiscan 17se II) was used, with frame repetition rates of 60 Hz (orientation) or 100 Hz (color). Experiment 3 (faces) was performed on a 15" ultra-high resolution monitor (Ergo-View 15; Sigma Designs Inc., Fremont, California) with 60 Hz frame rate. Viewing distances were 73 cm (Sony) and 67 cm (Ergo), respectively, with small variations (±1.5 cm) due to head size differences between observers (who had their heads conveniently leaned against the wall).

Patterns displayed lines (Exp. 1), squares (Exp. 2), or schematic faces (Exp. 3) arranged in 9 x 9 (Exp. 1 and 2) or 7 x 7 rectangular rasters (Exp. 3), with raster widths of 1.8 deg and 2.2 deg, respectively. Full stimulus patterns covered an area of approximately 15 deg x 15 deg. In each pattern, the center element of the raster was spared and instead a fixation cross (Exp. 1 and 2, 0.25 deg x 0.25 deg; green in Exp. 1, white in Exp. 2) or fixation point (Exp. 3; 0.1 deg x 0.1 deg, green) was shown.

The lines in Experiment 1 (cf. Fig. 3) were 0.8 deg x 0.2 deg. Color targets in Experiment 2 (Fig. 7) were red or green squares (0.5 deg x 0.5 deg); colors were matched for equal luminance using heterochromatic flicker photometry. The facial expressions in Experiment 3 (Fig. 11) were obtained from schematic drawings (1.3 deg diameter) in which the mouth regions were bended upwards ("happy") or downwards ("angry"). At various delays after stimulus onset, one of these items was cued (50ms), and after a variable presentation time all items were masked.

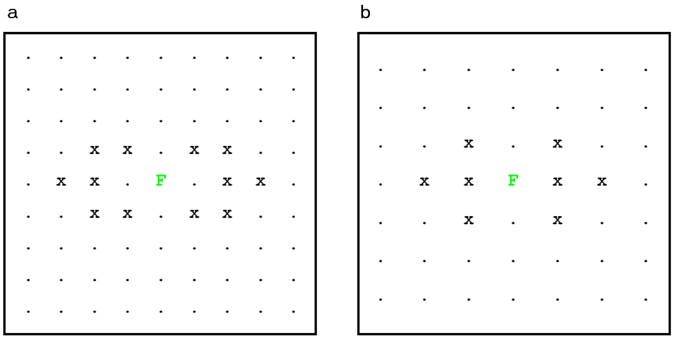

Figure 2. *Possible target locations in test patterns;* ***a,b.*** *different rasters.* Items were presented in 9x9 (*a;* Exp. 1 and 2) or 7x7 arrays (*b;* Exp. 3) with a central fixation marker *(F)*. Targets were randomly selected by the brief presentation of a cue. To avoid too large variations from crowding and attentional resolution, possible target locations were restricted to fovea-near positions *(x)*. Full stimulus patterns covered an area of about 15 deg by 15 deg.

Cues were made of four little squares around the target *(four-dot cues),* each 0.2 deg x 0.2 deg, which were located 0.6 deg (Exp. 1 and 2) or 1 deg (Exp. 3) from the target center in the four oblique directions. In a later modification of Experiment 3, same cue locations as in Experiments 1 and 2 were also tested (0.6 deg from target center); the cues then partly overlapped the target face. To reduce the large performance variations from crowding and limited attentional resolution (Intriligator & Cavanagh, 2001; Nothdurft, 2017a) but still keep the uncertainty of cued locations large enough, possible *target locations* were restricted to certain raster positions as indicated in Figure 2. Maximal target eccentricity from the point of fixation thus was 5.4 deg in Exp. 1 and 2, and 4.4 deg in Exp. 3. Subjects were not informed about this restraint.

All stimuli except the fixation markers (green in Exp. 1 and 3) and the items in Experiment 2 (color) were white on dark background. *Luminance settings* varied slightly between monitors and frame rates, and were measured as about 20 cd/m^2 and 29 cd/m^2 for lines and masks, respectively, in Experiment 1, about 20 cd/m^2 for the red and green color squares in Experiment 2, and about 7 cd/m^2 and 17 cd/m^2, respectively, for the schematic face drawings and masks in Experiment 3. The briefly shown four-dot cues were always brighter (68 cd/m^2, 58 cd/m^2, and 77 cd/m^2, for Exp. 1, 2, and Exp. 3, respectively), but note that all these luminance measures were made with large stimuli which appeared much brighter than the small blobs of the cue. This was also the case with the fixation markers (about 44 cd/m^2 in Exp. 1; 37 cd/m^2 in Exp. 2, and 40 cd/m^2 in Exp. 3). All stimuli were presented on a dark background (11 cd/m^2 and 9.5 cd/m^2 in Exp. 1 and 2, respectively; 1 cd/m^2 in Exp. 3).

Procedures

All stimuli were viewed binocularly. Trials started with a 1s presentation of the fixation point before the stimulus pattern was shown; cues (50 ms) were superimposed upon the pattern at various delays (0 ms, 50 ms, 100 ms, 150 ms, 200 ms, 300 ms, 400 ms, 500 ms, 750 ms, 1000 ms, 1500 ms, 2000 ms, 5000 ms). After the presentation time (measured from cue onset), the pattern was masked for 500 ms. Thereafter the screen was blanked and only the fixation point remained visible. Subjects could enter their responses without time pressure. After a short blank of the entire pattern, a new trial began with the 1 s presentation of the fixation point.

Responses were made in a "modified 2AFC" task (Nothdurft, 2017b) by pressing certain keys on a computer keyboard. The modifications were: (i) Subjects could reject, and later repeat, a trial if they felt they had been inattentive during the presentation or had lost fixation (rarely used, mainly in conditions with long delays); and (ii) they could change their last response immediately after the trial if they noticed they had pressed the wrong key. For orientation (Exp. 1), the selection of response keys was intuitively clear (left-hand “<” key for targets tilted to the left; right-hand “-” key for targets tilted to the right; German keyboard layout). For all other targets, the selections ("<" for red and smiling faces; "-" for green or angry faces) had to be learned and memorized; the assignments were also sketched below the monitor screen. All observers became quickly familiar with the tasks and the keys to use.

In the main experiments, tests were blocked for target features and cuing delays; only target duration was varied within a block. The test range of target durations in blocks was individually adjusted for each observer to provide an optimal coverage between chance performance (50% accuracy) and high target identification rates (accuracy near 100%). This has led to slightly different test ranges and test resolutions for different observers. In the first session of each experiment, blocking was sometimes changed to include different cue delays in one run to obtain a quick overview of best test ranges for each observer.

Within each block, test conditions (target durations) were randomly intermixed, each with 5-10 repetitions. In the long course of an experiment, different runs were repeated in an interleaved sequence, to generate a final data base with 40-60, usually 50 repetitions of every test condition. Experiments were carried out in sessions of 2h, each covering several test runs. Subjects could pause whenever they wanted. Usually, tested features remained the same between subsequent blocks, and only the delay was (pseudorandomly) changed. In the course of a full session, however, there might have been one or two switches from one feature to the next.

All tasks in the present study were performed *under fixation.* Good fixation performance was checked by the experimenter by means of a video camera placed above the monitor and focused upon the observer's eyes. Controls were frequently made during the first sessions of every observer and regularly repeated in later sessions. All subjects had quickly learned to perform the task without moving their gaze. With short stimulus presentations after the cue (target durations < 200 ms) there was no benefit from moving the eyes (Fischer *et al*., 1993). With longer target durations, where shifting the gaze might have been tempting, observers were regularly reminded to keep the gaze strictly on the fixation marker and to skip a trial, by pressing a different key, if their eyes had moved away.

Analysis

After completion of a test series, the accuracy data at each delay (up to 10 durations with about 50 repetitions each) were fitted with a Gaussian cumulative function to calculate the target duration at 75% correct, Δt_{75}. In most cases, the standard cumulative function (in the percentage of correct responses),

$$y = 50 + 25 \cdot (1 + erf\,[\, (x - a_0) / (\sqrt{2} \cdot a_1) \,]\,),$$

with fit parameters a_0 and a_1 for the center and width of the underlying Gaussian, generated very good fits. In some cases, however, for example when observers had made too many errors so that 100% accuracy was not reliably reached, it was replaced by a general cumulative Gaussian function,

$$y = a_2 + a_3 \cdot 0.5 \cdot (1 + erf\,[\, (x - a_0) / (\sqrt{2} \cdot a_1) \,]\,),$$

in which also the level, a_2, and the amplitude, a_3, were fitted to the data. With good resolution and all repetitions, each such fit is based on 300-500 stimulus presentations at any given cue delay, for each observer. Fits typically revealed coefficients of determination, R^2, well above 0.9, often 0.98 or more.

Subjects

Altogether five observers participated in the study. Four of them (in the age of 20-23 years) were students at the Göttingen University and were paid for the time they spent in the experiments. The fifth observer was the author (68 years when the experiments began). All subjects had normal or corrected-to normal visual acuity on both eyes and, except the author, were naive about the aim of the experiments. All subjects had carried out other experiments with cued target identification before.

RESULTS

Data were collected in three series of experiments, which were run in intermingled sequence but are here split for clarity. A forth series of experiments dealing with conjunctions of color and orientation will be published in a separate paper (Nothdurft, 2020).

Experiment 1:
Orientation — left vs. right tilted lines

In an array of 80 lines randomly tilted to the left or right, one line *(the target)* was cued and had to be identified. The delay between stimulus onset and cue presentation was systematically varied, as was the target presentation time after the cue. Thereafter, all items in the stimulus pattern were masked. Examples of test and masking patterns are illustrated in Figure 3; possible target locations are indicated in Figure 2a. All five observers participated in this experiment. Similar tests with this paradigm had been carried out earlier (Nothdurft, 2017a; Exp. 2) using a coarser temporal resolution and a smaller range of delays.

Figure 4 illustrates the general performance variations with one observer at four cuing delays. For each delay, accuracy increased with increasing presentation time, but increases began at different target durations. The earliest increase is seen with delay 100 ms (red curve) and is even faster than the increase at the delay 0 ms (black curve) when cues were presented simultaneously with the stimulus. With longer delays, accuracy increased more slowly again and presentation time had to be enlarged to reach similar performance.

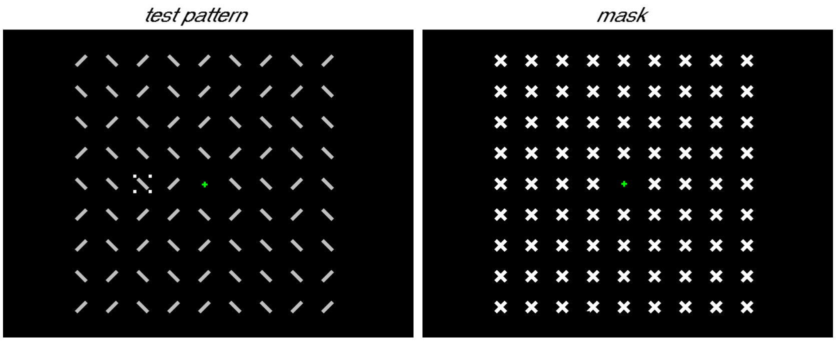

Figure 3. *Typical stimulus patterns in Experiment 1 (orientation).* The test pattern showed 80 randomly tilted oblique lines. At a certain *delay* after stimulus onset, one of these lines was marked with a 50 ms *four-dot cue* (as indicated) and thus selected as target. Observers had to indicate the target line orientation. After a short presentation time, *Δt,* after the cue onset the whole pattern was masked.

To condense the data, the presentation times for 75% correct responses, Δt_{75}, were calculated from each curve (Fig. 4b). For observer OC, there was an inflection at short delays; he needed shorter presentation times to identify half of the targets cued 50-100 ms after the pattern onset than to identify half of the targets cued simultaneously with the pattern onset (i.e., the difference between red and black curves in Fig. 4a, at 75% correct line detection). At longer delays, performance slowed down again (Δt_{75} values increase) and the observer needed longer presentations times to identify the same amount of targets. This general pattern of accuracy variations across delays was also observed with the other observers.

Δt_{75} curves from all observers are plotted in Figure 5. Despite certain differences the curves also show many similarities. Values differ in absolute numbers, indicating how fast an observer could identify the cued targets, but all curves show the same strong inflection at short delays. Targets cued 50-150ms after stimulus onset were generally much faster identified than targets cued right at the beginning of the stimulus presentation (delay 0 ms) or much later. At longer delays, curves of most observers (but not OC) reach a plateau where longer delays were not associated with a further increase of the necessary presentation time. Also the exact delays at which the Δt_{75} values became minimal, varied slightly. Observer HCN, for example, who generally needed rather short presentation times to identify the targets, revealed an inflection minimum at delay 50 ms, with a Δt_{75} value of only 10 ms. With the other observers, the minimum varied between delays 50-100 ms (OC) and 100-150 ms (LL and JP). Observer NMB showed a modulation around the minimum, which was however only partly significant. In the means of all five observers (Fig. 5f), the local

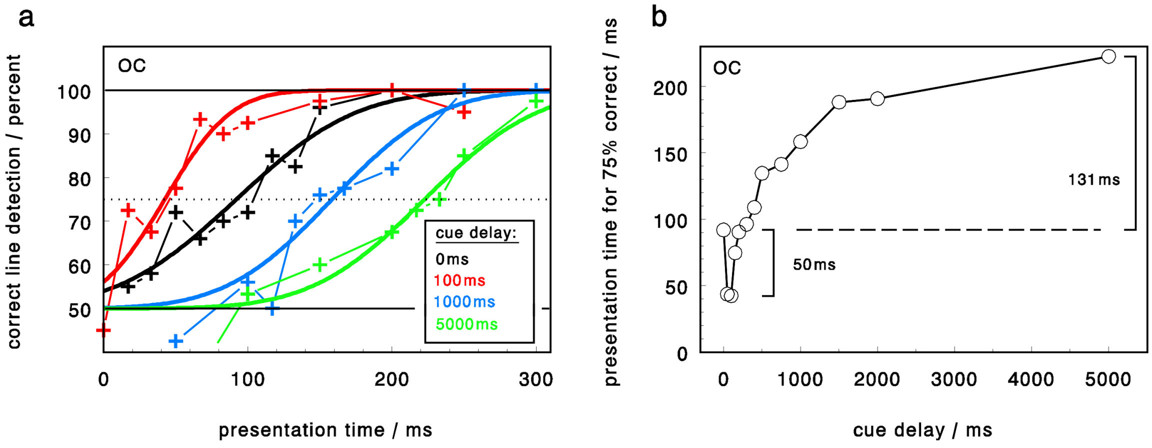

Figure 4. *Performance of one observer in Experiment 1.* ***a.*** *Accuracy variations with presentation time, at four selected delays.* Curves were fitted with accumulative Gaussian functions and the target durations for 75% correct performance, Δt_{75}, were estimated (dotted line). ***b.*** *Variations of* Δt_{75} *values over different delays.* At short delays, targets were identified faster than when cued immediately at pattern onset (compare the red and black curves in (*a*)). At longer delays (measured up to 5 seconds after stimulus onset), target identification was slowed down (blue and green curves in (*a*)).

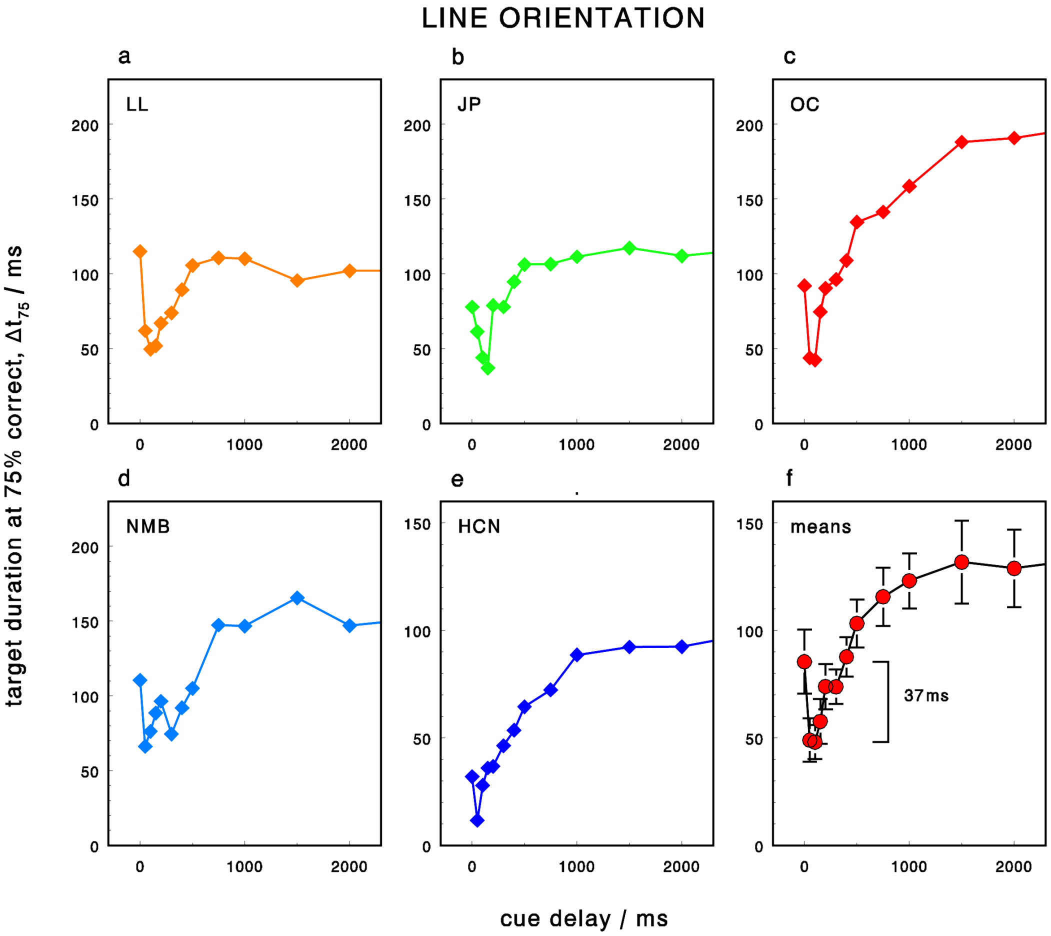

Figure 5. *Δt_{75} variations of all observers (**a-e**) with orientation discrimination (Experiment 1) and means with s.e.m. (**f**).* Please note the different scales in *(e)* and *(f)*. Despite notable deviations, all observers revealed the same characteristic modulation of performance speed with increasing cue delays. Target identification was strongly accelerated at short delays (50-300 ms) and slowed down at long delays. In order to present the individual data for comparison but keep the figure readable, axis labeling had to be made relatively coarse (and data from delay 5000 ms even be left out). For the evaluation of the exact delay at individual data points, it may therefore be helpful to remember the series of tested delays (which was the same in all experiments): 0 ms, 50 ms, 100 ms, 150 ms, 200 ms, 300 ms, 400 ms, 500 ms, 750 ms, 1000 ms, 1500 ms, 2000 ms, and 5000 ms (not shown here).

inflection (at delays 50-100 ms) is quite pronounced and strongly exceeds the standard error of the means.

Statistics

The statistical significance of Δt_{75} inflections in Figures 4 and 5 was evaluated in three ways. First, performance variations between selected delays were analyzed in the original data (like those in Fig. 4a) using the *Wilcoxon signed-rank test*. For that, accuracy measures with same target durations but different delays were compared. Such comparisons were made between the zero delay and the delay with shortest Δt_{75}, and between the shortest Δt_{75} and delay 500 ms. For observer OC, for example, this would correspond to calculating the statistical difference between the black and the red curves in Figure 4a, and between the red curve and the curve obtained for delay 500 ms (not shown in Fig. 4). With one exception, these differences were all significant, for every observer ($n \geq 6$; $W \leq W_{crit}$; $p<0.025$). The exception was the difference between delays 0 ms and 50ms for observer HCN who had revealed a particularly fast performance already at delay 0 ms (cf. Fig. 5e). In a second series of statistical tests, the Δt_{75} values at various delays (Fig. 5) were compared. A *one-way ANOVA* of all observers over all delays revealed $F(12,52) = 3.14$ and thus significant variations ($p<0.001$) of Δt_{75} values across the tested delays. The Bartlett test confirmed that variances were homogeneously distributed ($\chi^2 = 14.04$; $\alpha = 0.57$), a precondition for using ANOVA. This does, however, not identify exactly which Δt_{75} differences were significant. Therefore, a third statistical test was performed to prove that the Δt_{75} values around the inflection were significantly shorter than the Δt_{75} values at other delays. Using the nonparametric *Mann-Whitney U test,* values from delays 50-200 ms were compared with values from delays 1000-5000 ms, from all observers. These values differed significantly ($p<0.01$; $U=5.5 < U_{crit}$ ($n_1 = n_2 = 20$; $\alpha = 0.001$) $= 81$), even when the ranges of compared delays were enlarged to 50-400 ms vs. 500-5000 ms ($z=4.31$; $p<0.001$; note that for large n, here $n_1 = n_2 = 30$, the sampling distribution approaches a normal distribution and the test statistic U can be transformed into a z-score). The inflections in Figure 5 are thus significant.

Reconstruction of underlying neural signals

As already described, all curves in Figure 5 have a local minimum where target identification required the shortest presentation times, and strongly increase toward longer delays. It is plausible to assume that presentation times needed to reach a constant accuracy are inversely related

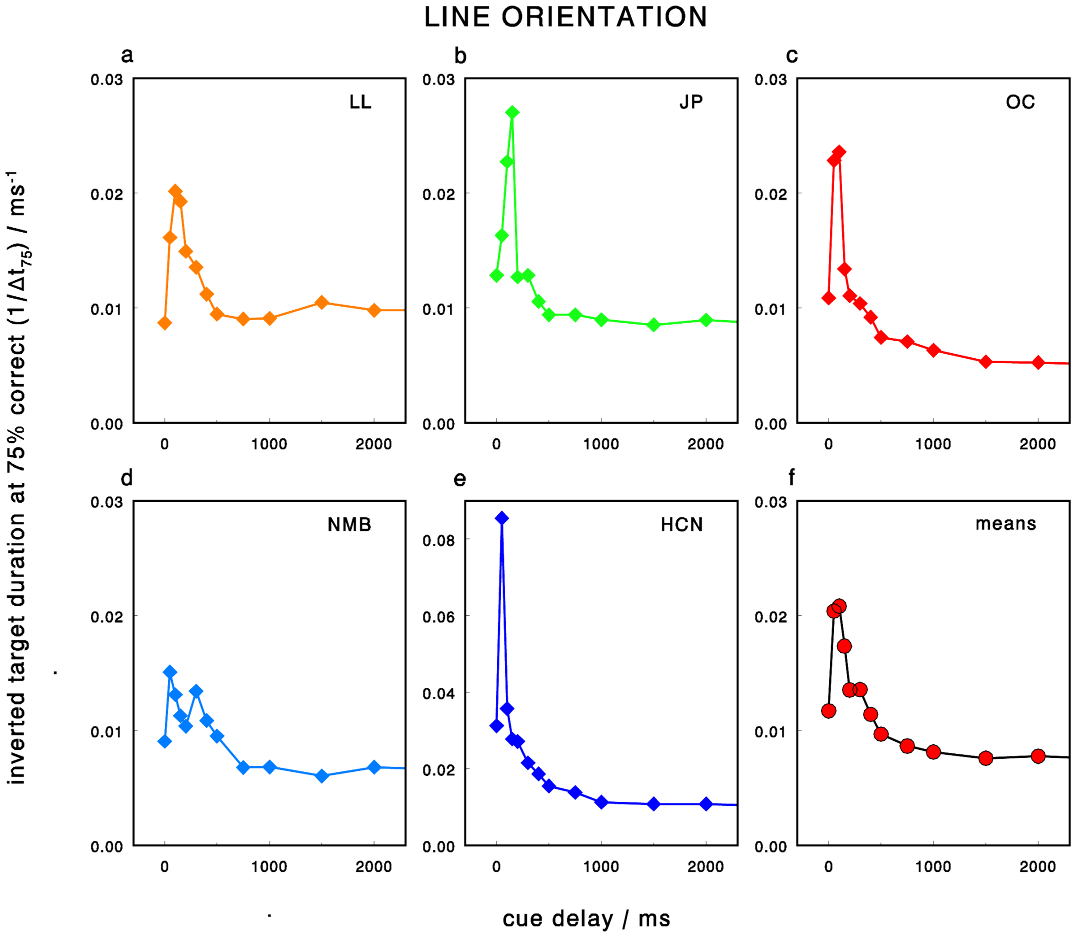

Figure 6. *Inverse $1/\Delta t_{75}$ signals for the graphs in Figure 5;* ***a-e.*** *individual observers;* ***f.****, means.* The $1/\Delta t_{75}$ values are assumed to reconstruct the neural signals underlying the performance in the target identification task. (For details, see text and Nothdurft, 2017a, b, 2019). The presumed orientation signals show a strong and transient peak at short delays; due to the increased activity at these delays, targets could be identified particularly fast. At shorter or longer delays, the signal was reduced and had to be accumulated over longer presentation times to reach the same level of performance.

to the strength of (neural) signals from which the target had to be identified (see Nothdurft, 2017a). With certain simplifications, therefore, inverted plots of the data in Figure 5 should indicate the strength of underlying neural signals in the brain. For all observers, these signals (Fig. 6) were temporarily enhanced (50-500ms), with a strong transient peak. For absolute timings relative to the stimulus onset, however, one has to take into account that orientation selection was not exactly synchronized with the cue perception (Nothdurft, 2018; see Discussions below).

Discussion

The findings of Experiment 1, illustrated in Figures 4-6, confirm earlier reports about the cued discrimination of oriented lines (Nothdurft, 2017a, b, 2019). Targets were identified particularly fast when cues occurred 50-100 ms after the stimulus onset. Earlier and later cuing increased the needed presentation time. Inverted plots of the data reveal a strong modulation of the underlying neural signal, with little strength immediately at target onset and a (delayed) transient peak with the maximum around 50-100 ms; both properties would by and large mimic the population response of neurons in area V1 (e.g., Knierim & Van Essen, 1992; Nothdurft, Gallant, & Van Essen, 1999; see Nothdurft, 2017a, b). Note however that there is a principle asynchrony between cued selection and perceived line orientation, which might have diminished the apparent delay of neural signals. With flickering stimuli, not the line presented simultaneously with the cue but a line presented 50-100 ms later is seen as cued (Nothdurft, 2018). Thus, the maximum of the truly underlying signal should, in fact, be more delayed than the maximum of the reconstructed signal. This might explain why some observers (e.g., HCN) performed seemingly so fast in this experiment. Training and experience may also have accelerated performance, which would certainly apply to the author (HCN) who had performed numerous such experiments before.

It should be interesting to compare this modulation with performance variations obtained with the cued selection and discrimination of color squares or facial expressions.

Experiment 2:
Color — red vs. green squares

The experimental outline was similar to that of Experiment 1; instead of oriented lines, test patterns contained randomly colored red or green squares (Fig. 7); by means of heterochromatic flicker minimization the colors were matched for equal luminance, individually for each observer. Since preliminary tests had shown that colors could be identified from relatively short stimulus presentations, monitor frame rate was increased to 100 Hz for a better temporal resolution (10 ms per frame). As in Experiment 1, cue delays (after stimulus onset) and target presentation time (after cue onset) were systematically varied, and observers had to indicate the color of the cued target by pressing different keys on a computer keyboard. Special care had to be taken of the masking stimulus. Early tests showed that adaptation effects could irritate observers and might have allowed some of them to evaluate target colors not only from the target but also from the strong afterimages in the mask, in particular after long cue delays. While most observers could still indicate the color they had perceived with the cue, before the mask, one explicitly reported that he had sometimes reconstructed the target color from the the mask. In a number of preliminary tests, therefore, several modifications of the masking pattern were tested to

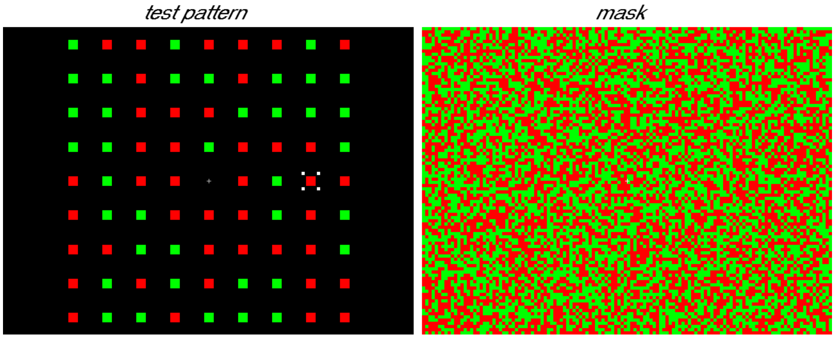

Figure 7. *Typical stimulus patterns in Experiment 2 (color).* The test pattern showed 80 randomly red or green squares. At a certain *delay* after stimulus onset, one of the squares was marked with a 50 ms four-dot cue (as shown) and observers had to indicate its color. After a short presentation time, *Δt,* the whole pattern was masked.

© christoph.nothdurft@vpl-goettingen.de ISSN:2364-3641

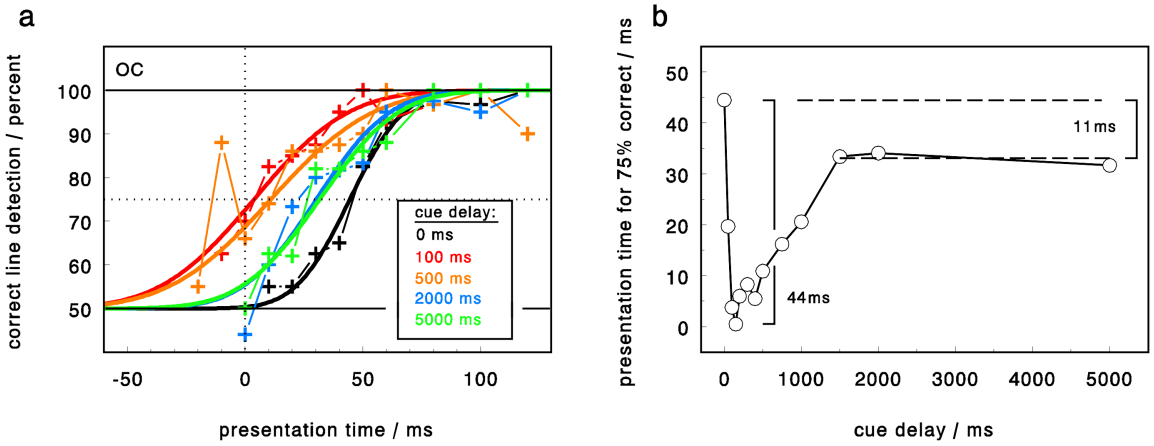

Figure 8. *Performance of the same observer as in Figure 4, now on color discrimination in Experiment 2.* ***a.*** *Accuracy variations with presentation time*, at selected delays; ***b.*** *variations of Δt_{75} values* from the curves in *(a)* and at other delays. Similar presentation as in Figure 4, but note the differences: Curves in *(a)* fall closer together; the black curve (delay 0 ms) is the right-most curve; although performance in *(b)* is similarly accelerated at short delays, it is not similarly strongly slowed down at longer delays.

minimize the effects of color adaptation upon target identification. The best solution was achieved with a red-green random dot pattern of the same colors as used in the test pattern and with a spatial configuration that was changed in every trial (Fig. 7).

A few other peculiarities with color are illustrated in Figure 8. Similar to Figure 4 for orientation, performance variations with increasing target duration are here shown for color, at four selected cue delays. Accuracy increased very rapidly (note the different abscissa scales in Fig. 8a compared to Fig. 4a) leading to almost 100% accuracy with target durations of less than 100ms. (Similar performance levels with oriented lines had been reached with target durations of 150-300ms; Fig. 4a.) It is also worth noting that color targets cued at the pattern onset (delay 0 ms) required the longest presentation times at all (the black curve is the right-most one in Fig. 8a), whereas, with orientation, target identification at long cue delays had often required further enlarged presentation times. Interestingly, the identification of color targets began sometimes before the cue and might then have already reached 75% accuracy at a target duration of 0 ms (see the red and rose curves in Fig. 8a). Thus, even when the patterns were masked at the moment when the cue occurred (zero presentation time), still half of the targets could be correctly identified. This was not only the case with long cue delays where adaptation effects might still have been present, but also at very short delays (e.g., 100 ms; red curve in Fig. 8a) where adaptation should not yet have been pronounced. To optimize the later evaluation of Δt_{75} values from the data, testing was therefore expanded into "negative" durations, at these delays. (With negative durations, the target is already masked before the cue occurs.) The expectation was that target identification should then be more difficult and performance accuracy be reduced, ideally even down to chance (50%). At certain cue delays and with certain observers this was however not the case.

An extreme was the performance of observer NMB who generally identified over 75% of the targets at all cue delays above 150 ms, irrespective of the target duration and sometimes even when the target was masked 50 ms before the cue. Cumulative functions could not reliably be fit to these data and no Δt_{75} values be computed. Even at short delays likely not yet affected by adaptation, this observer reported the cued target colors such as if she could recall them from eidetic memory where the entire test pattern was stored before the cue occurred.

The general modulation of performance with the cuing delay and the variations between observers are best seen in the Δt_{75} values plotted in Figure 9. Except for NMB who showed a strong decline until delay 150 ms (for larger delays the Δt_{75} values could not be estimated) all observers revealed a notable modulation of Δt_{75} values, starting from about 40-90 ms at delay 0 ms, going through a minimum at 0 ms or even below between delays 100 ms and 400 ms and finally slightly increasing back to medium durations at longer delays. The small fluctuations of Δt_{75} values between neighboring delays (e.g., observer JP) were not significant (Wilcoxon signed-rank test between same target durations at neighboring delays) except the difference between delays 500ms and 750ms of observer LL (W=10.5; n=13; p<0.01). The shortest presentation time for 75% target identifications was generally reached at longer delays (100-400 ms; Fig. 9) than for orientation (50-150 ms, see Fig. 5), for each individual observer. The means across observers (except NMB) reveal a broad reduction of the required presentation time at 150-400 ms (Fig. 9f).

Statistics

The same statistical tests as in Experiment 1 were made. Performance differences between same target durations at zero delay and the delay with the shortest Δt_{75} values (delays 100-400 ms, for different observers; delay 150 ms, for NMB) were significant (Wilcoxon signed-rank test, $W \leq W_{crit}$; p<0.01). Differences between same durations at the delay with the shortest Δt_{75} value and at delay 2000 ms were not significant in three observers (including NMB); in these observers, thus, cued target identification was not significantly slowed down at long delays, when single delays are compared. Only observers OC and HCN showed a significant increase of the needed presentation

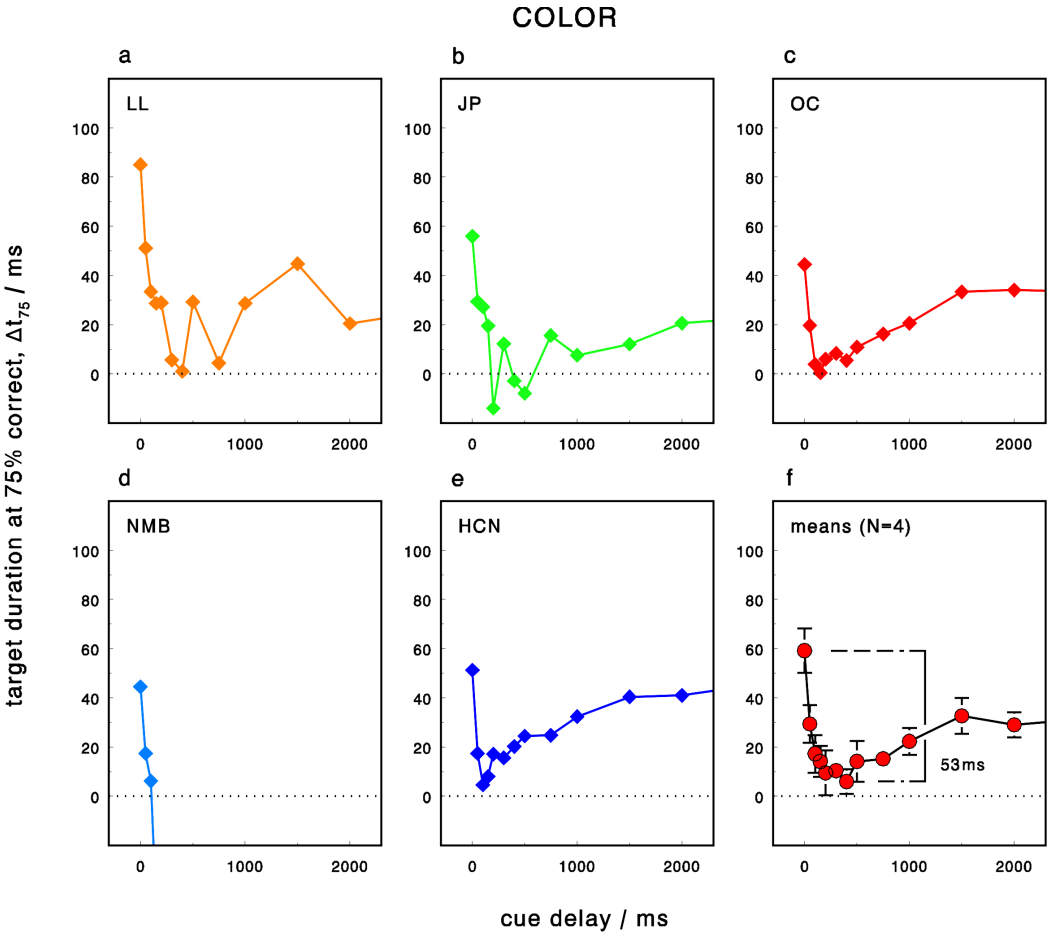

Figure 9. *Δt_{75} variations with color discrimination;* ***a-e.*** *individual observers, and* ***f.*** *means with s.e.m.* Target identification was strongly accelerated at short delays and reached even negative target durations in some observers (which should be physically impossible; see text for an explanation). Subject NMB showed a peculiarity; she could identify almost all targets that were shown long enough, irrespectively of when the cue was given. Therefore, no Δt_{75} values could be estimated at longer delays and data from observer NMB were not included in the means *(f)*.

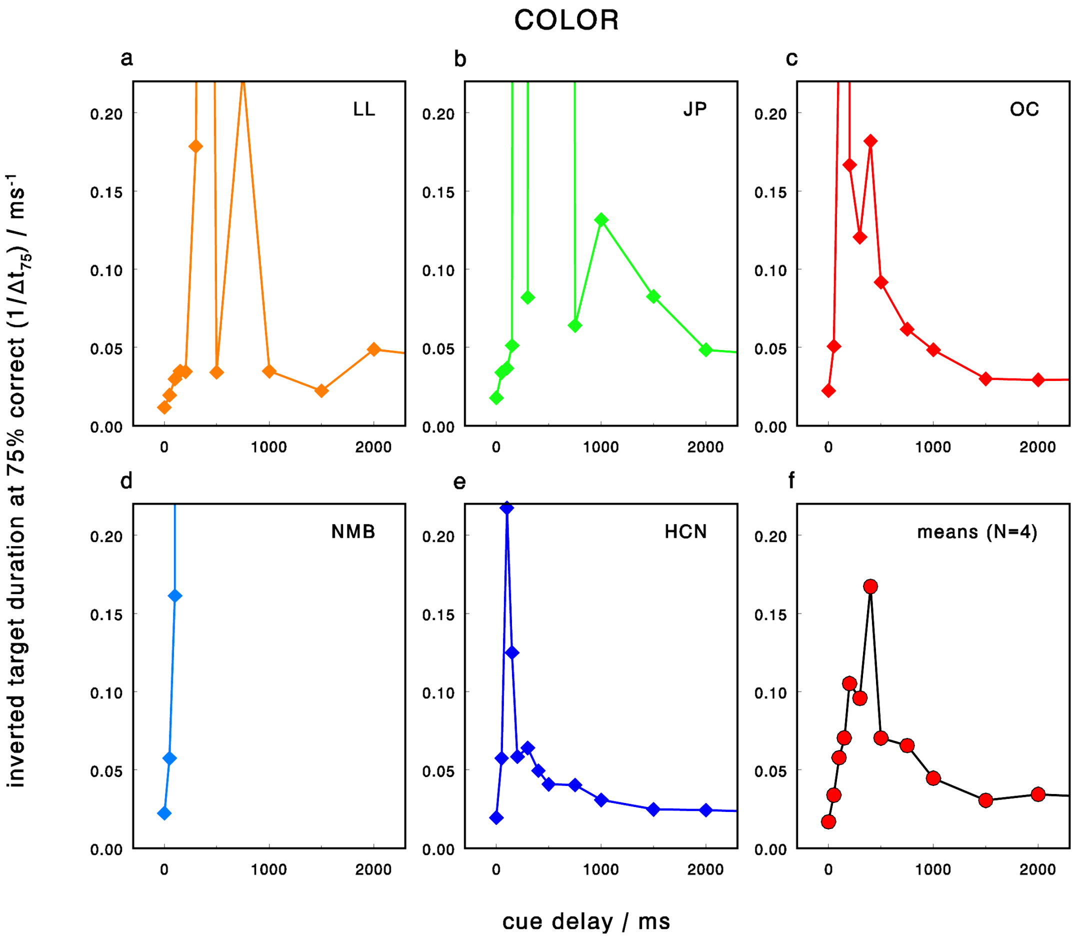

Figure 10. *Inverse 1/Δt75 signals from the graphs in Figure 9;* ***a-e.*** *individual observers;* ***f.*** *means (without NMB).* Curves are assumed to reconstruct the neural signal underlying target discrimination in the experiment. The (physically impossible) negative values in Figure 9 were set to particularly high values outside the graphs.

time between their fastest performance and that at delay 2000 ms ($W \leq 2.5$; n=9; $p<0.01$). The small *en passant* modulations in the data of LL and JP (Fig. 9) were mostly not significant in the direct comparison with neighbors. An ANOVA of Δt_{75} values from all observers (except NMB) over all delays revealed $F(12,39) = 4.86$; $p<0.0001$; the Bartlett test confirmed equally distributed variances ($\chi^2 = 7.98$; $\alpha = 0.79$). To identify the delays at which Δt_{75} values differed significantly, the Mann-Whitney U test was applied to different delay ranges. Between delays 50-200 ms and delays 1000-5000 ms, Δt_{75} values from all observers (except NMB) differ significantly ($p<0.01$; $U = 57$; U_{crit} ($n_1 = n_2 = 16$; $\alpha = 0.01$) = 66); this is also the case when the compared delay ranges are enlarged to 50-400 ms vs. 500-5000 ms ($n_1 = n_2 = 25$; $z = 2.71$; $p<0.01$). The major modulations of Δt_{75} values in Figure 9 are thus significant.

Reconstruction of underlying neural signals

Since all Δt_{75} values are rather short, inverting the data leads to very high $1/\Delta t_{75}$ values (Fig. 10) in which small (and non-significant) Δt_{75} fluctuations are exaggerated. It is unlikely that the small fluctuations in the measurements should reflect such strong variations of underlying neural signals. Another problem in reconstructing these signals are the occasional negative Δt_{75} values in Figure 9 (e.g., JP, Fig. 9b), which were arbitrarily set to high values (above the graph) in the reconstruction (Fig. 10b).

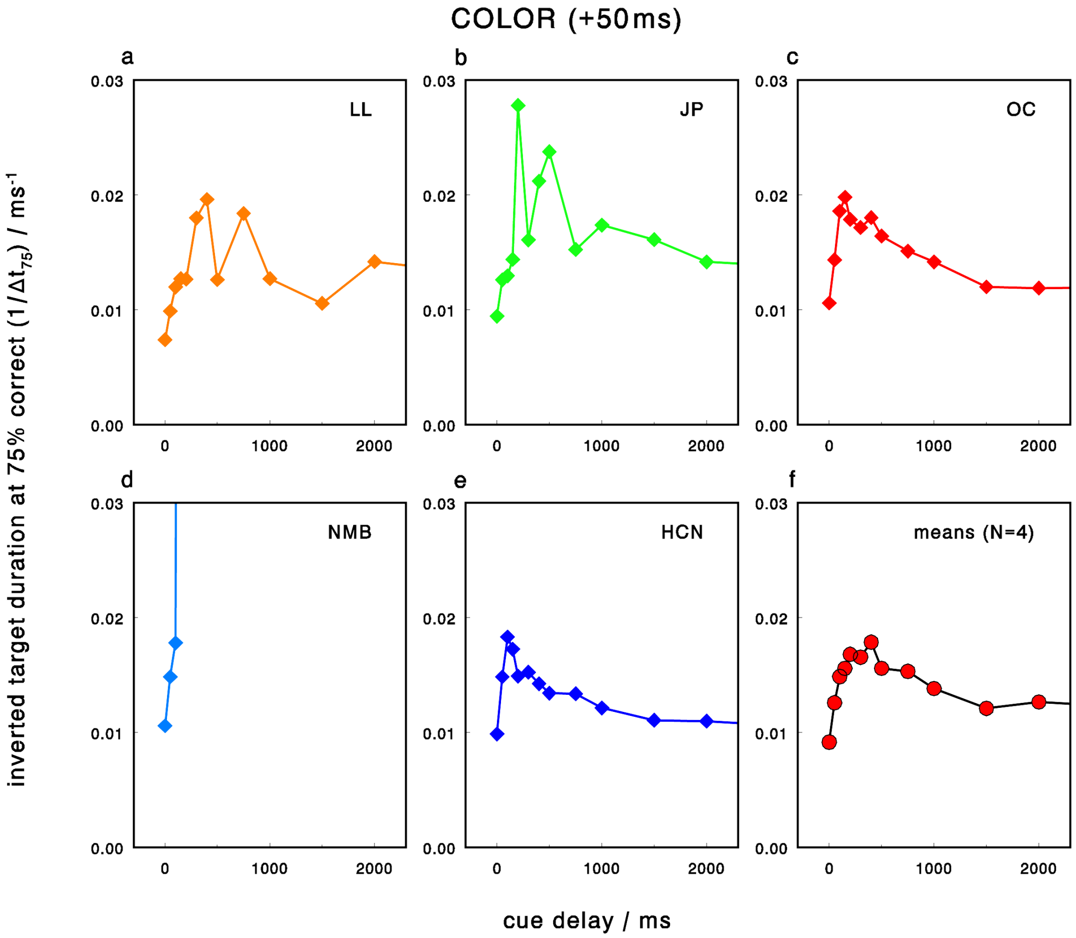

Figure 11. *Modified inverse 1/Δt75 signal reconstruction from Figure 9;* ***a-e.*** *individual observers;* ***f.*** *means (without NMB).* To overcome the problem of negative target durations, all curves in Figure 9 were first raised by 50 ms before they were inverted. This would correspond to an additional delay between color perception and cued selection which has been reported elsewhere (see text). Curves in Figures 10 and 11 indicate a strong signal underlying cued color discrimination with a maximum that is later reached than with orientation (see Figure 6) and does not decay completely to the starting level but contains a sustained response component.

Negative target durations cannot be physically obtained from the cumulation of real signals; thus inverting negative *Δt75* values would not be meaningful. (The fact that, nevertheless, negative *Δt75* values were measured at some delays indicates that target durations were incorrect. This is likely due to a systematic time shift between cued selection and color perception; Nothdurft, 2018.) A solution of the problem is shown in Figure 11 where *Δt75* values were shifted by 50 ms so that all *Δt75* values became positive before they were inverted. Physically this would correspond to a virtual cue onset 50 ms earlier than in the real experiment. This would not be implausible, since the cued target selection is delayed (by 50-100 ms) relative to color perception (Nothdurft, 2018).

While the original (very small) *Δt75* values generate huge peaks when being inverted (Fig. 10), a simple delay shift strongly reduces the amplitudes (Fig. 11; please note the different scales) without changing the general pattern of signal modulations. The presumed neural signals underlying color target identification (1/*Δt75*) were low at stimulus onset (delay 0 ms), increased to a maximum at 100-400 ms after stimulus onset and then slowly decayed back to a medium level still above that at delay 0 ms (see the means in Fig. 10f). An exception is observer NMB, for whom Δt_{75} values could not be estimated for delays > 150 ms. When adding a 50 ms delay to the measured Δt_{75} values (Fig. 11), the apparent strength of underlying signals is generally reduced (because it would now take 50 ms longer to reach the criterion for target identification) but curves reveal the same characteristic modulations as before.

Discussion

The very fast identification of target color, compared to target orientation, is compelling but perhaps not unexpected. That some observers could identify color targets even when these were already masked, may raise a number of questions, however. Could it be that target *selection* was initiated by the cue but perhaps not the decoding of the target color itself? Were certain colors perhaps "preattentively" represented in the brain? The apparently eidetic recall of target colors by observer NMB seems to support this view. The question will be more generally addressed in the General Discussion below. One must also take into account that there are latency differences between cued selection and the perception of target colors, which should have reduced the apparent delay of cued color perception by 50-100 ms; cues select the color of a later target at the cued location (Nothdurft, 2018). A similar asynchrony in the perception of color and motion has been reported before (Moutoussis & Zeki, 1997a, b; Viviani & Aymoz, 2001; Holcombe & Cavanagh, 2008; Rangelov & Zeki, 2014; McIntyre & Arnold, 2018).

The observed disturbance by aftereffects in preliminary experiments, which has led to a modification of the masking pattern, is another interesting detail. It is very unlikely, however, that the particularly fast identification of color targets was due to adaptation effects and afterimages. First, with the new masking stimulus (Fig. 7b) aftereffects were not further reported and thus had likely been absent or small. In addition, one should expect that aftereffects, if they had still been present, should have been strongest at long cue delays (when the stimulus was shown for up to 5 s before the cue) and not at very short delays (50-400 ms) where target identification was particularly fast.

After the different results obtained for orientation and color, we shall now look at a quite different stimulus, the (cued) identification of facial expressions.

Experiment 3: Facial expressions — happy vs. angry

In the following experiment, observers were asked to discriminate schematic faces for happy and angry expressions. Test were performed on the Ergo 15" monitor at a frame rate of 60 Hz (see Methods). Facial expressions were randomly assigned to the individual face drawings; the distribution was refreshed in every new trial. Since the patterns were more complex than simple oriented lines (Exp. 1) or colored squares (Exp. 2), the items were slightly enlarged and the total number of items within a test pattern was reduced (Fig. 12). The range of possible target locations was adopted to the new 7 by 7 raster (Fig. 2b). With the larger target size also the size of the four-dot cue had to be enlarged (from 0.6 deg to now 1 deg from the target center). In all other aspects, the

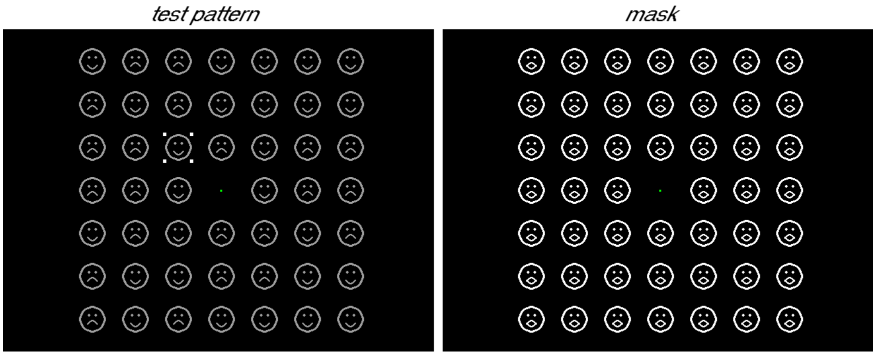

Figure 12. *Typical stimulus patterns in Experiment 3 (facial expressions).* Test patterns showed 48 face drawings with randomly happy or angry facial expressions. At certain delays after the stimulus onset, one of the faces was marked with a 50 ms four-dot cue (as shown). Observers had to identify whether the cued face looked happy or angry. After the presentation time, *Δt,* the whole pattern was masked by brighter faces with overlapping mouth regions. Note that cues in this experiment were larger than in the previous experiments. In an additional variant of the experiment, therefore, smaller cues (like in Exp. 1 and 2) were used which then, however, partly overlapped the target.

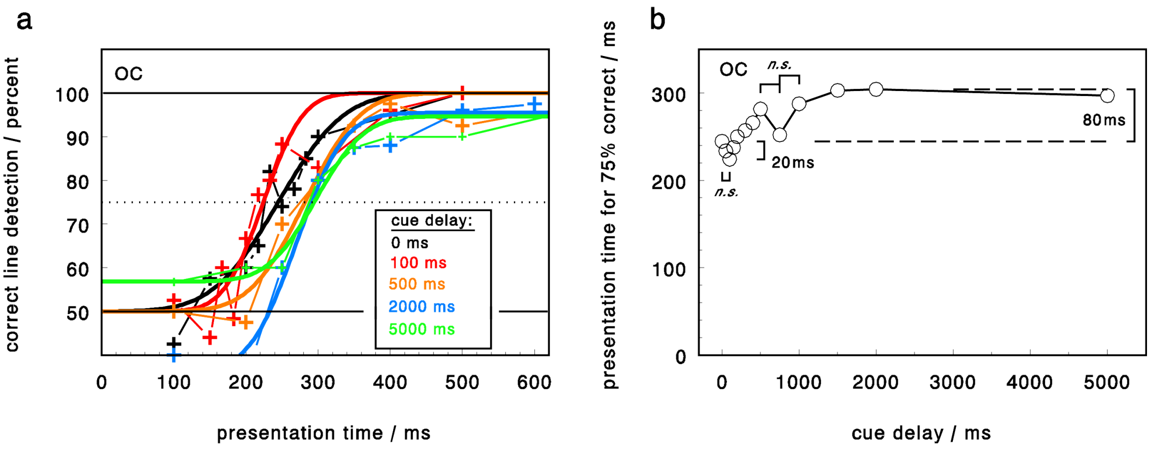

Figure 13. *Performance of observer OC (as in Figures 4 and 8), now on the discrimination of facial expressions in Experiment 3.* ***a.*** *Accuracy variations* at selected delays; ***b.*** *variations of* Δt_{75} *values*. Curves in *(a)* are generally shifted to longer target durations, compared to those from orientation or color tests before, and do partly overlap. Accordingly, the necessary presentation times in *(b)* are generally increased, and the modulation along delays is weaker than in Experiments 1 and 2.

experimental procedure was identical to that of Experiment 1. That is, after the presentation of the fixation point (1s) the test pattern was switched on and, simultaneously with the pattern onset or after a certain delay, one of the schematic faces was cued (50 ms). Patterns remained visible for variable durations, Δt, and then were masked. Masks were made by a superimposition of the two face patterns and were neutral in their expression (Fig. 12b). Observers had to identify the cued item and report the perceived facial expression by pressing different keys on the computer keyboard.

The larger size of the cue might have caused a problem, since speed and efficiency of attention shifts depend on the cue size (Eriksen & St. James, 1986; Benso, Turatto, Mascetti, & Umiltá, 1998; Nothdurft, 2002). To evaluate the magnitude of this effect, an additional test with a modified series of Experiment 3 was later run on two (still available) observers. In this series, smaller cues were used (exactly in the size of the cues used in Experiments 1 and 2), which however did now partly overlap the target.

Psychometric curves of observer OC, at selected cue delays, are plotted in Figure 13. All curves show a cumulative increase of performance accuracy with increasing target duration, but the presentation time required for good performance was generally much longer than that for oriented lines (Fig. 4) or colored squares (Fig. 8). The differences between curves for various cue delays appear to be generally less pronounced (but notice the difference scales in Figs. 4, 8, and 13).

The Δt_{75} curves of all observers are shown in Figure 14. Since global variations were hidden by small and mostly non-significant local modulations (Wilcoxon signed-rank test between same target durations at neighboring delays), data were smoothed for clarity by calculating weighted means, $y_i = 0.25 \cdot y_{i-1} + 0.5 \cdot y_i + 0.25 \cdot y_{i+1}$ (thick curves in Fig. 14). The resulting curves show small but notable reductions for observers LL, OC, and NMB, at delays 100 ms to 400 ms. For observer HCN no such reduction was seen but the curve reveals a merely continuous increase of Δt_{75} values with increasing delays (Fig. 14e, blue). This was, however, mainly due to the enlarged cues used in this experiment. When Experiment 3 was repeated with a smaller cue, which now overlapped the target, performance was accelerated and the Δt_{75} curve showed a clear modulation around delays 50-100 ms (gray). With observer LL (Fig. 14a), reduction of the cue size had no such a clear effect (gray). With this observer, apparently, the faster attention shifts from smaller cues (Eriksen & St. James, 1986; Benso, Turatto, Mascetti, & Umiltá, 1998; Nothdurft, 2002) were counterbalanced by a deteriorated target identification from the four-dot cue falling upon the target (Nothdurft, 2016a). The other observers were not available anymore and could not be tested with the smaller cue configuration.

Statistics

The same statistical analysis as in the previous experiments was performed for Experiment 3. The direct comparison of same target durations at different delays (e.g., comparisons of individual psychometric curves in Fig. 13a) revealed mostly non-significant differences (Wilcoxon signed-rank test, $W > W_{crit}$). Significant differences between single delays were seen between delays 0 ms and 200 ms for observer NMB and for the general increase of required target durations with increasing cue delays (e.g., delays 150 ms vs. 2000 ms) in observer HCN (all $p<0.05$; $W < W_{crit}$). In the later additional tests with smaller cues (gray data curves in Fig. 14), the differences between delays 0 ms and 50 ms for observer HCN were significant ($p<0.025$) and those between delays 50 ms and 2000 ms highly significant ($p<0.005$). For observer LL, accuracy differences between delays were not significant (delays 0 ms vs. 100 ms) or the significance could not be evaluated because of too few different test pairs with same target durations (delays 100 ms vs. 2000 ms). An ANOVA of the original Δt_{75} values from all observers could not be made, since the Bartlett test indicated inhomogeneous variances ($\chi^2 = 25.83$; $\alpha = 0.01$). The comparison of Δt_{75} values in different delay ranges with the Mann-Whitney U test, however, revealed significant differences across all

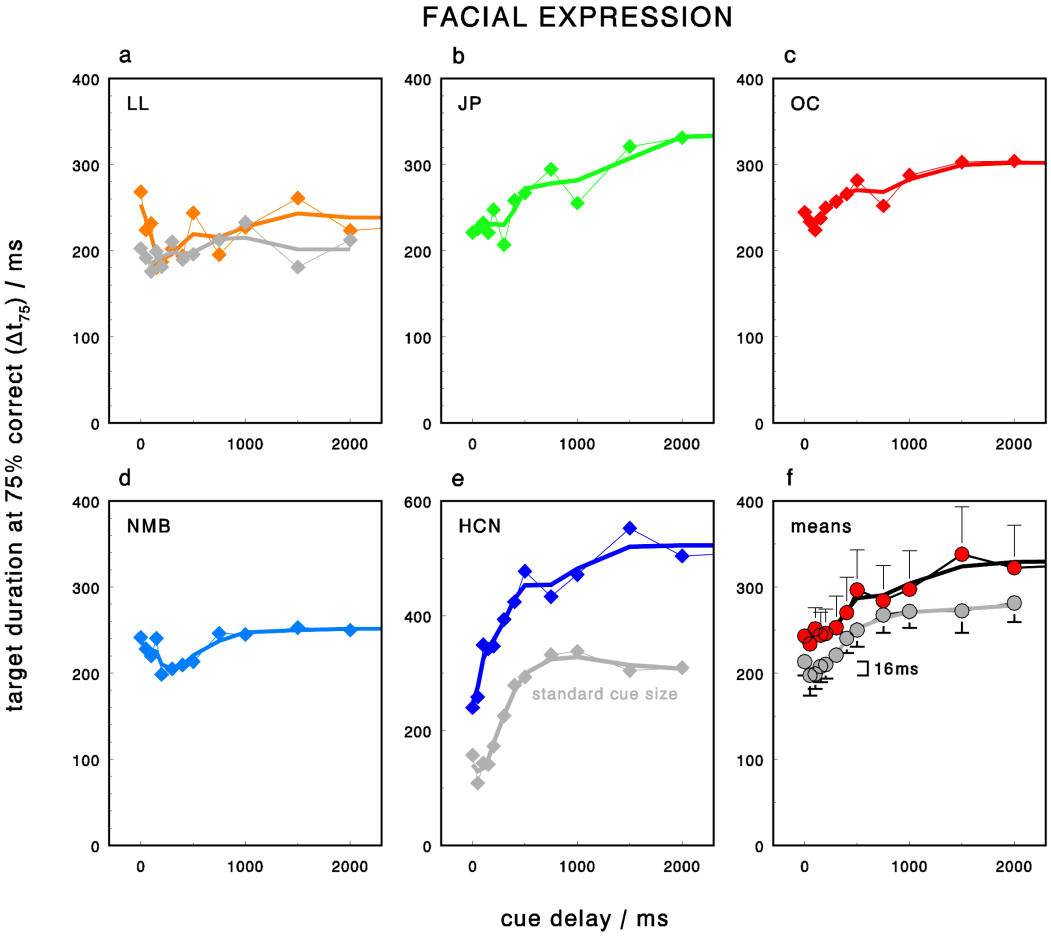

Figure 14. *Δt_{75} variations with facial expressions (Experiment 3);* ***a-e.*** *individual observers;* ***f.*** *means and s.e.m.* Note the different scale in *(e)*. Original data (symbols and thin lines) were smoothed (thick curves) to visualize performance variations with increasing cue delays. Target identification generally required longer presentation times than with orientation (Fig. 5) or color (Fig. 9). It was accelerated at delays below 500 ms but modulations were less pronounced than in previous experiments and were not seen with all observers. Gray curves and data points in *(a)* and *(e)* are from an additional experiment with the same stimuli but smaller cues (in the same size as cues in Experiment 1 and 2); these data were not obtained from all observers. The gray curve in *(f)* represents the mean data of all observers when small cue data replaced the original data of tested observers. The smaller cues strongly improved performance of observer HCN *(e)*.

observers. The Δt_{75} values for delays 50-200 ms were significantly shorter than the Δt_{75} values for delays 1000-5000 ms (p<0.01; $U = 24 < U_{crit}$ ($n_1 = n_2 = 20$; $\alpha = 0.01$) = 114); this was also the case when the delay ranges were enlarged to 50-400 ms vs. 500-5000 ms ($n_1 = n_2 = 30$; $z = 3.24$; p<0.001). Thus, while the statistical analysis had not always revealed significant differences between individual delays, the general difference between short and long delays in the cued identification of facial expressions was significant. These differences also remained significant when instead of the original data the later measurements with smaller cues were used for observers LL and HCN: delays 50-200 ms vs. 750-2000 ms (U=47 < U_{crit} ($n_1 = n_2 = 20$; $\alpha = 0.01$) = 114; p<0.01) and delays 50-300 ms vs. 500-2000 ms ($n_1 = n_2 = 25$; $z = 4.24$; p<0.001). The delay 5000 ms was not tested in the later measurements and data from this delay are not included in analysis. With the modified sample, also the one-way ANOVA could be made since variances were homogeneous (Bartlett test, χ^2 =4.46; $\alpha = 0.95$). It revealed a significant Δt_{75} modulation across delays; F(11,48) = 2.72; p<0.01.

Reconstruction of underlying neural signals

The relatively small modulations at long durations in Figure 14 generate relatively small peaks when the Δt_{75} values are inverted (Fig. 15). All graphs show an increased signal at cue delays below 500 ms, often with a local peak between 50 ms (HCN) and 300 ms (JP). In the means (Fig. 15f) the underlying neural signal is enhanced until

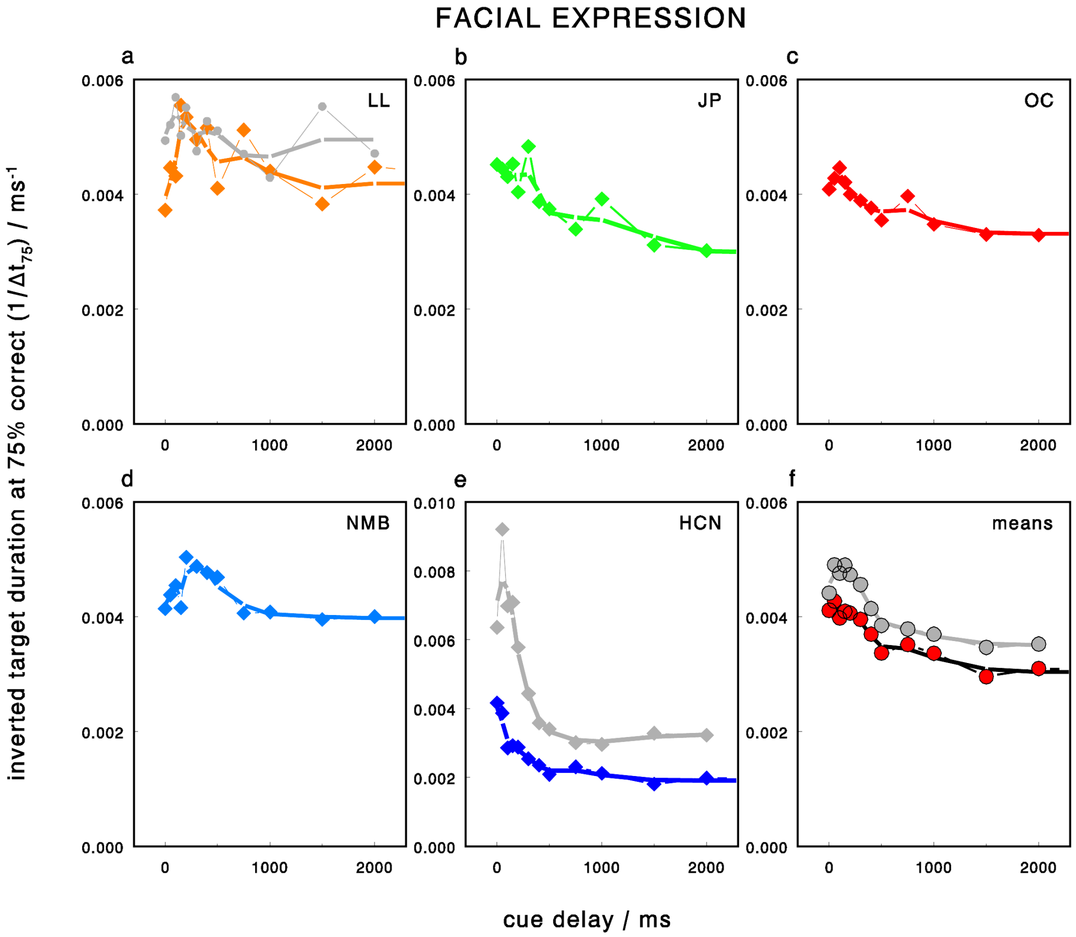

Figure 15. *Inverse 1/Δt75 signals from the graphs in Figure 14; a-e. individual observers; f., means.* Presentation as in Figure 14; notice the different scale in *(e)*. Curves show the reconstructed neural signals underlying the discrimination of facial expressions. While signals are strongly reduced compared to those for orientation discrimination (Fig. 6), curves of individual observers display similar shapes and peak locations in both tasks.

300 ms after stimulus onset and then decays towards longer delays. When for observers LL and HCN the data from the modified experiment with smaller cues are used for computation of the means, there is a pronounced signal peak at 50-300 ms (gray curve).

Discussion

The observations that the cued identification of facial expressions required longer presentation times than the discrimination of colors and oriented lines and had also generated smaller performance variations between different cue delays might be surprising. Faces are obviously more complex than simple lines or color squares, but given the importance of faces and facial expressions in social communications and the frequent use of emoticons in chats, one should not have expected such a strong difference. Shortest Δt_{75} values were 10-100 ms for orientation, less than 5 ms and even negative for color, but more than 200 ms for facial expressions. As argued above, the difference might, in part, have been due to the larger cue size used in Experiment 3. Smaller cues indeed accelerated the identification of facial expressions in one observer. But even with equal-sized cues the overall differences between experiments remained large. Lines and colors were much faster distinguished than happy and angry faces. This is in agreement with occasional reports from the observers that cued facial expressions, in contrast to line orientations or colored squares, were not recognized immediately at the cue onset but that the percepts seemed to "develop" in the stimulus after a notable delay. The overall *Δt_{75} modulations* across different delays, on the other hand, were in the same order for faces and oriented lines. For some observers, the differences between faces cued simultaneously with the pattern onset and faces cued at the delay with the fastest performance were quite strong (e.g., Fig. 14, observers LL, NMB, HCN with small cues) and not much smaller than the according modulations obtained with oriented lines (Fig. 6; please notice the different scales). But while strong signal modulations with oriented lines were observed in all observers, they were not always so pronounced with facial expressions and therefore reduced in the means (Fig. 15f).

One explanation might be that the identification of faces and facial expressions in Experiment 3 had been suboptimal, either because schematic face drawings do not represent true faces and might have poorly activated face neurons in the fusiform face area (FFA) or because faces need to be foveally inspected for the recognition of social indicators (which was not possible in Experiment 3). Neither assumption, however, is confirmed in the literature. Schematic face drawings were indeed occasionally reported to generate smaller responses in FFA than photographs of real faces (Tong, Nakayama, Moscovitch, Weinrib, & Kanwisher, 2000). But face-specific recognition is also obtained with very schematic "face" drawings, which has, in fact, led to an elegant analysis of the "face-in-the-crowd" effect, in which seemingly angry facial expressions are quickly detected from seemingly happy facial expressions even in schematic drawings (Coelho, Cloete, & Wallis, 2010; Kennett & Wallis, 2019, who carefully discuss this topic). And, yes, we usually look right into individual faces when trying to recognize people and evaluate their mood, but we may also try to avoid gaping at people and nevertheless recognize their emotion. There is plenty of evidence from search experiments (e.g., Hansen & Hansen, 1988; Hershler & Hochstein, 2005) and ensemble coding (e.g., Ji, Chen, Loeys, & Pourtois, 2018; To, Carvey, Carvey, & Liu, 2019) that faces and facial expressions can also be detected and recognized outside the fovea (but not too far in the periphery). These findings do not suggest that Experiment 3 had, in general, been ill-designed.

An interesting aspect comes from the comparison of face identification (Figs. 14-15) with that of oriented lines (Figs. 5-6). By and large, the individual dips in Δt_{75} curves, and peaks in the $1/\Delta t_{75}$ curves, occur at similar delays and even seem to have similar shapes within individual observers (see, e.g., observers NMB, HCN), although curves are shifted in amplitude between the tasks. This might indicate that the identification of faces and facial expressions (which mainly differed between local orientation components in the mouth region) was strongly, if not exclusively, controlled by the speed and the modulation of neural signals underlying the *orientation* discrimination. While the dynamics of color discrimination strongly differ from those of orientation discrimination, the dynamics of orientation discrimination and discrimination of facial expressions were surprisingly similar (although different between individual observers). This may be taken as a further support of proposals according to which the discrimination of facial expressions (happy, angry) is mainly achieved from differences in low-level visual features like orientation (Coelho, Cloete, & Wallis, 2010; Kennett & Wallis, 2019).

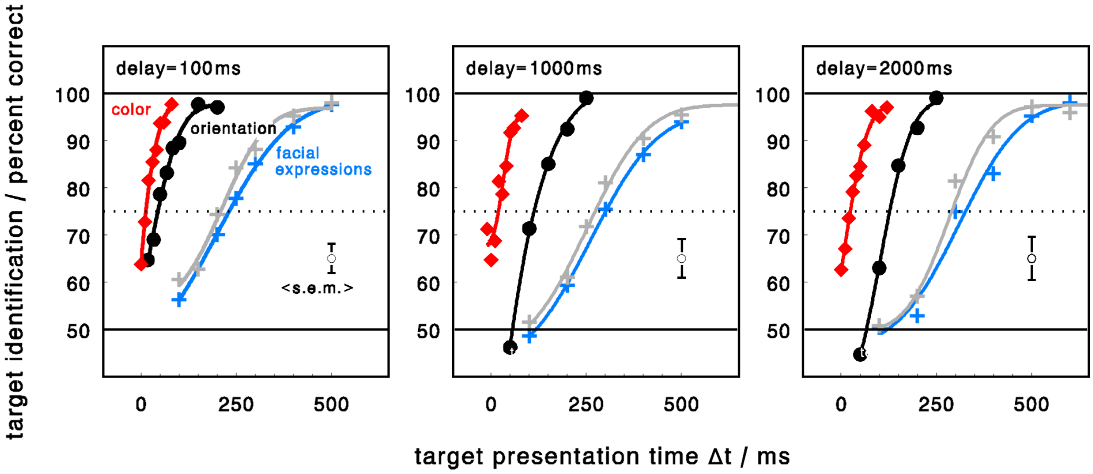

Figure 16. *Synopsis of Experiments 1-3: accuracy variations at selected delays.* Different to previous figures, data points now represent the mean performance of all observers and the s.e.m. averaged over all features and test conditions at the according delay. Curves are fitted with cumulative Gaussian functions. Data illustrate the different discrimination speed for color (red rhomboids), orientation (black circles) and facial expressions (blue and gray crosses). In all tasks, accuracy increased with stimulus presentation time. From these and all other curves at different delays, the necessary presentation times for 75% correct performance, *Δt75,* were calculated and shown in following figures. In the means for facial expressions, data of observers tested with the smaller standard cues (gray) replace their original measures (blue).

Comparison of Experiments 1 - 3

For a synopsis of the findings in this study, Figure 16 shows the *mean* accuracy variations (averaged from all observers) in the experiments. As illustrated for three cue delays, color discrimination (red) was always the fastest, followed by orientation (black) and finally face discrimination (blue and gray). That is, at any given cue delay, target color was identified from shorter presentations than target orientation, and target orientation from shorter presentations than facial expressions.

Δt75 curves derived from the *mean* accuracy data (Fig. 16) are shown in Figure 17. The curves are almost identical with the means of *Δt75* curves obtained from individual observers which are shown in Figures 5f, 9f, and 14f. This proves that all *Δt75* estimates were quite reliable and not destroyed by outliers and individual noise. The shortest *Δt75* presentation times were obtained with color, slightly longer ones with orientation, and the longest presentation times were needed to distinguish happy and angry facial expressions. But the different modulations of the *Δt75* values are also interesting. The *color* curve decays fast and reaches its minimum (with a *Δt75* of nearly 0 ms) at a delay of about 400 ms, and although values increase again towards longer delays, they do not reach the same level as at delay 0 ms. The *orientation* curve also decays fast, but reaches its minimum already at delays 50 ms and 100 ms, and then quickly increases again, faster than the color curve, to a level well above the start at delay 0 ms. The *facial expression* curves, finally, start with long presentation times at delay 0 ms, show very small inflections at delay 100 ms, and then continually increase to a higher level. Data from the modified Experiment 3 with smaller cues (tested on two observers) shift the curve towards shorter *Δt75* values.

The inverted *1/Δt75* curves for the reconstruction of underlying neural signals are too different to be plotted in one graph and are therefore re-plotted at different scales (Fig. 18). They summarize the principle differences of presumed neural signals encoding the tested features. The *color* signal (Fig. 18a) begins early and continues to rise up to 400 ms after stimulus onset; thereafter it decays to a level slightly above start where it remains over the range of tested delays. Note that the somewhat exaggerated peak will diminish in amplitude when the delay is corrected, without changing its general dynamics (see Figs. 10f and 11f). *Orientation* is also encoded fast (Fig. 18b); the reconstructed signal has a sharp peak at delays 50-100 ms and then decays within a few hundred milliseconds to a level below start. The reconstructed responses underlying

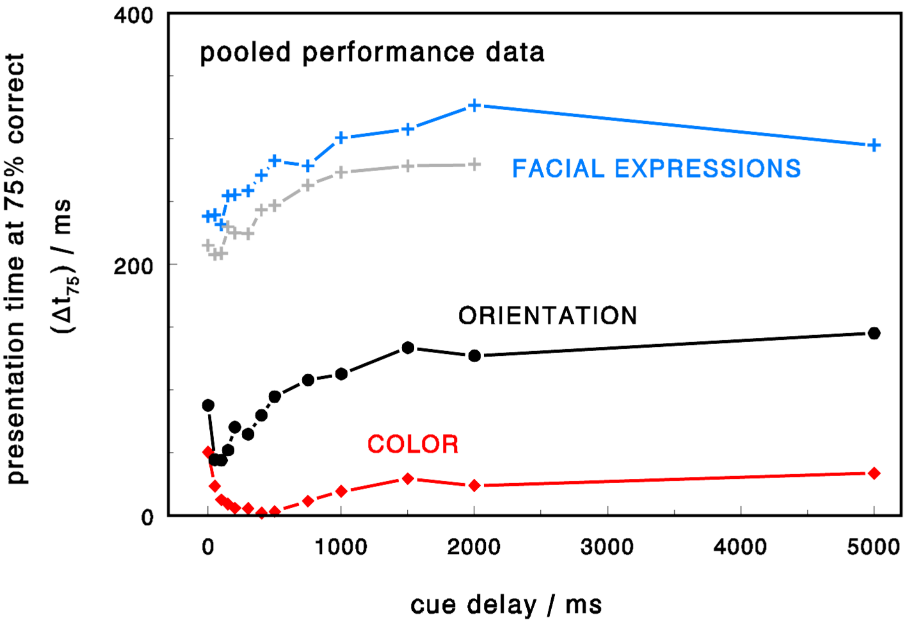

Figure 17. *Synopsis of Experiments 1-3: Δt_{75} variations with different cue delays for color (red), orientation (black), and facial expressions (blue, gray).* Orientation discrimination was strongly accelerated at short delays and notably slowed down at longer delays. Color discrimination was faster, reached the minimum later and returned afterwards to an intermediated level. The discrimination of facial expressions took notably longer and showed a similar (though less sharp) modulation as orientation discrimination. When data from the small-cue variant (where measured) replace the original data, mean performance is slightly accelerated (gray vs. blue crosses).

the discrimination of *facial expressions* show a much wider (less strongly modulated) peak at generally lower activity levels, without a sharp elevation at short delays. This elevation was more pronounced when the data from individual observers were averaged (Fig. 15f), in particular, when data from the modified tests with smaller cues are included. In shape, the peak is by and large similar to that for orientation. It shall be interesting to compare these reconstructions with the response properties of neurons in neurophysiological studies.

Statistics. The overall differences between features are statistically highly significant (Wilcoxon signed-rank test with large N). Pairwise comparisons of color and orientation tests (same delay, same target duration) give $|z| = 9.69$; $p<0.001$ for the pooled data from all observers ($n = 132$), and $|z| \geq 3.82$; $p<0.001$ for the individual observers ($n \geq 20$); the W statistic in the one case with $n = 20$ was $W = 0$; $W_{crit} = 21$; $p<0.001$. Pairwise comparisons of orientation and face tests give $|z| = 12.9$; $p<0.001$ for all observers ($n = 240$ pairs) and $|z| \geq 4.37$; $p<0.001$ for the individual observers ($n \geq 25$). A two factor ANOVA with replication performed on all Δt_{75} data from individual observers (not on the means in Fig. 16, from which the Δt_{75} and $1/\Delta t_{75}$ curves in Figs. 17 and 18 were derived) revealed significant modulations over the different delays, F(12,156) = 3.14; $p<0.001$, and highly significant differences between features, F(2,156) = 404.82; $p<0.0001$). When the original face data of two observers are replaced by the later collected data with smaller cues (gray in Figs. 14 and 15), the significance is further improved; F(11,144) = 7.16 ($p<0.0001$), for variations across delays; and F(2,144) = 773.7 ($p<0.0001$), for differences between features.

GENERAL DISCUSSION

The study revealed large differences in the speed at which three tested features were discriminated, and in the modulation of this speed with different cue delays. Color discrimination was faster than orientation discrimination, and both much faster than the discrimination of happy and angry faces. The identification of orientation and color properties was strongly accelerated right after stimulus onset; a similar but smaller modulation was seen in the analysis of facial expressions. When the identification speed is transformed to resemble the strength of underlying signals, modulations correspond to strong response peaks shortly after stimulus onset. Transient and partly sustained response characteristics can be distinguished. The experiments have also revealed notable variations between observers, both in the timing of performance accuracy (some observers needed longer target presentations than others to reach the same level of performance) and in the exact latency of individual performance peaks.

I will discuss these findings under three headings. First, I will describe a methodological restriction and discuss possible problems that might be considered to have affected the data. I will argue why I think they have not. Second, I will compare the data with that of other studies and, in particular, compare the reconstructed strength of underlying neural signals with neurophysiological data in the literature. Third, I will address certain assumptions made in these experiments, look at the specific role of attentional selection and discuss possible interference from attention capture by certain features themselves.

Were measurements confounded by eye movements?

The data are interpreted under the assumption that observers had performed covert attention shifts and had not moved their eyes to identify the targets foveally. All observers had quickly learned to identify cued targets without shifting their gaze, i.e. while continuously fixating the central fixation marker. When fixation performance was checked (by means of a camera looking at the observers' eyes; see General Methods), no gaze shifts during trials were detected. In addition, all observers were instructed to skip a trial if they had moved their eyes during the presentation. But there was no automatism that would have disregarded the trial in such a case. There is a number of arguments why gaze shifts should have been rare (if not absent) and should likely not have confounded the results of the study. First, observers could identify orientation and color targets from very short presentations (less than 150 ms in the means) during which they could not possibly have foveated them (Fischer *et al*., 1993). The good performance in these conditions proofs that gaze shifts had not been necessary, and likely not present, in the task. The situation is less obvious with targets that required much longer presentations for identification, like the schematic faces in Experiment 3. However, if observers had systematically shifted their gaze to the cued targets, their performance should have been nearly identical over all long delays, which clearly was not the case. There were notable performance variations even between delays larger than 250 ms, which should not have occurred when targets were always discriminated in the

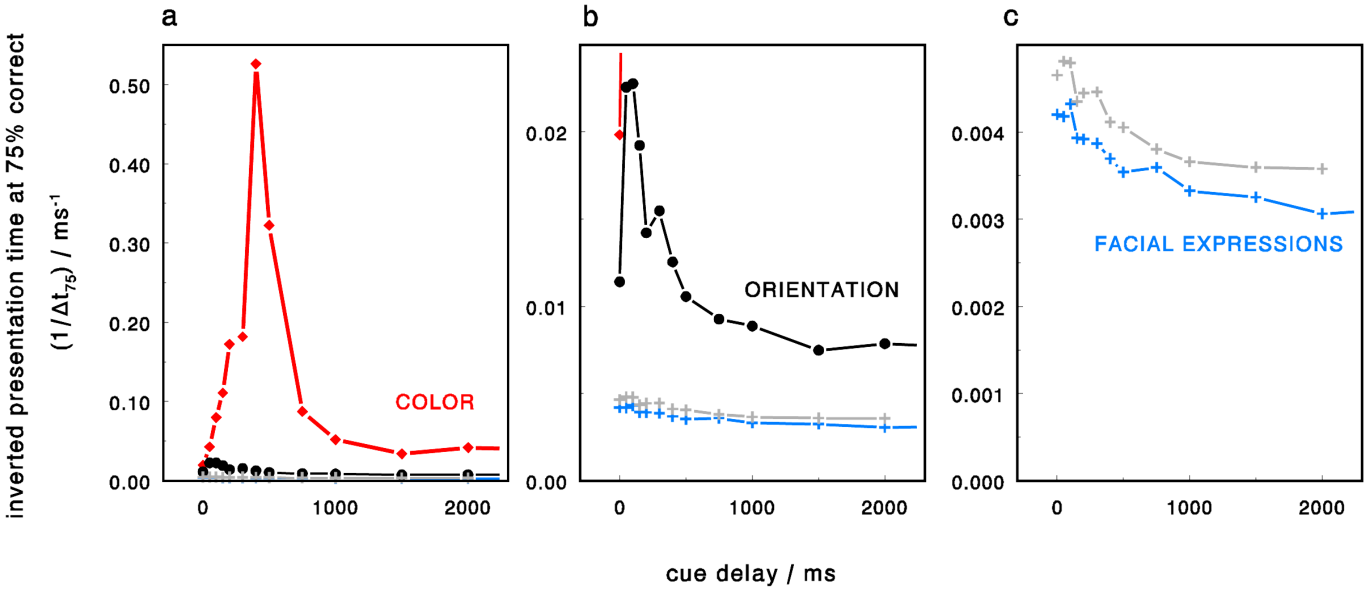

Figure 18. *Synopsis of Experiments 1-3: Reconstructed 1/Δt75 signals underlying performance in Figure 17.* Data are replotted at three different scales to visualize discrimination signals for color (red), orientation (black) and facial expressions (blue, gray).

fovea. A further argument against confounding eye movements comes from a control test that was made on two observers at the end of the study. In this test, observers were asked to deliberately perform, as fast as possible, gaze shifts to the cued target and find out if they could discriminate the facial expression before the target was masked. Both observers found deliberate gaze shifts at first difficult to perform, because they were used to hold fixation. But once they got used to it, they never reported an identifiable facial expression with presentation times of 250 ms or less, so that at least accuracy measures with target durations of 250 ms or less should likely not be confounded by eye movements. The critical duration might be even larger, since the Δt_{75} values used for analysis already imply correct identification of half of the targets. The most *interesting* Δt_{75} values at short delays in Experiment 3 were shorter. Thus, while occasional gaze shifts in a small number of trials cannot be excluded, there should be no systematic confounding of the data. In particular, gaze shifts cannot explain the strong differences in performance at short cue delays, which have produced quite different Δt_{75} curves with different features.

The reconstruction of underlying neural signals

The reduction of multiple accuracy measures to one constant-performance measure, the Δt_{75} value, has a direct impact upon the evaluation of underlying neural signals. If observers needed less, or more, time to discriminate a cued target, likely the strength of the signal they could have made use of in their decision has varied. It is plausible to assume that neural responses are accumulated until a certain signal level is reached; therefore, a strong signal should need a shorter accumulation time than a weak signal, for the same accuracy. This was the background for calculating the $1/\Delta t_{75}$ curves above. (It may be helpful to recall that all Δt_{75} variations in the present study were synchronized with target presentation, not with the cue, and hence do not reflect cuing dynamics, which should have been identical at all delays.)

There are, however, certain limitations in this reconstruction. First, all variations are relative. This is, in fact, quite obvious from the Δt_{75} curves of different observers. Some observers needed longer stimulus presentations to discriminate the targets than others. These differences may be based on two components. One component may apply to all discriminations and reflect an observer's general alertness in the performed discrimination tasks and his/her trust in "listening" and responding to small neural signals in the visual system. The other component may affect the discrimination of certain features but not others and may depend on individual settings as well as long-term and training experiences. For example, observer HCN was particularly fast in the discrimination of line orientations (Fig. 5e) but not in the discrimination of facial expressions with large cues (Fig. 14e); observer NMB could identify late-cued color targets much faster than all other observers (Fig. 9d).

Another uncertainty is the relative timing of neural signals in perception. While responses can, of course, only be accumulated after they have been generated, i.e. with a latency after stimulus onset, we do not know how perception is synchronized with the cue. Several studies have shown that different features may be encoded with different latencies so that, for example, synchronous color and movement (direction) changes are not perceived simultaneously (Moutoussis & Zeki, 1997a, b; Viviani & Aymoz, 2001; Holcombe & Cavanagh, 2008; Rangelov & Zeki, 2014; McIntyre & Arnold, 2018). Adjusting them so that the changes are perceived in synchrony would require motion changes to be presented before the color changes. Own experiments on (multiple) time windows in cued visual selection (CVS) have revealed that the cued selection is delayed relative to the perception of orientation and color, so that cues select a later target orientation and color than actually cued (Nothdurft, 2018). The differences were in the order of 50-100 ms and varied between observers. Together with the different accumulation times of individual observers, there are thus notable uncertainties in the relative delay of neural signals and the cued stimulus. A principle difficulty occurs when observers identify targets from their memory, like observer NMB apparently did with color targets (cf. Fig. 8 and 9); in that case, performance cannot be related to the cue delay at all.

With all these restrictions in mind, we may now compare the properties of reconstructed neural signals with the characteristic responses properties of neurons measured at different processing stages in the visual system. Important parameters are the response strength, in particular of population responses, and response latencies, regardless of the uncertainties mentioned above. We cannot predict *where* the reconstructed signals were generated. But it may be reasonable to start at the earliest processing levels where the according features are distinguished. For orientation this would be area V1; for color, it is the parvocellular pathway from the retina through the LGN (which may however not be directly assessed by attention and visual selection); and for faces and facial expressions it may be the *fusiform face area* (FFA) in the ventral cortex. This does not exclude that later processing stages are not also involved in the identification of cued targets, or that several areas may, in fact, interact in this analysis. It is also possible that certain response properties in early processing stages are maintained in the transfer to subsequent processing stages. On the other hand, quite a few studies have revealed notable response variations between attended and non-attended targets (e.g., Connor, Preddie, Gallant, & Van Essen, 1997); these neurons might be more strongly involved in CVS target identification than neurons in areas where differences between attended and non-attended targets are small or absent.

Orientation. For orientation, a quick comparison of $1/\Delta t_{75}$ values with the responses of orientation-sensitive neurons in area V1 was already made (Nothdurft, 2017a). The expected neural signals should be strong, with a sharp and transient peak (see Fig. 6). Such responses are widely documented in studies of the primary visual cortex. A study with partly similar stimuli as those used here is from Knierim & Van Essen (1992) on awake monkeys. The population response of 122 orientation-sensitive neurons (their Fig. 15) revealed a strong increase 40-60 ms after the stimulus onset; a peak at 60 ms with a half width of about 200 ms, and a continuous decay of firing activity from 60 ms onwards until the stimulus disappeared (500 ms); firing rates were still above spontaneous activity during that time. Similar response characteristics were described in a follow-up study on anesthetized monkeys (Nothdurft, Gallant, & Van Essen, 1999). These response patterns are, by and large, similar to the reconstructed signals in Figure 6. The short response latency of the population response, together with the perceptual asynchrony of cues and oriented targets discussed above, may account for the very fast performance of observer HCN, whose accumulation window was apparently particularly small in this task (early peak and very short Δt_{75} values). A difference occurs in the subsequent decays, which were stronger in the reconstructed signals (fast decays down to zero) than in the population responses (slower decays to a still elevated activity level). This was also observed in two earlier CVS studies (Nothdurft, 2017a, b); but there the decay varied with the actual test pattern. Signals reconstructed from targets with orthogonal surround ("popout") lasted sometimes longer than signals reconstructed from other targets (Nothdurft, 2017b; Fig. 19), so that stimulus variations might partly account for the decay differences between population responses in neural studies and the reconstructed signals in CVS experiments. In addition, decays in the present study were found to continue up to delays of 1s or more, but were measured, in the above mentioned neural studies, only up to 500 ms after stimulus onset.

Color. The neural signals encoding color should also be fast and strong. The fact that color was discriminated from much shorter presentation times than orientation (Fig. 17) suggests that the underlying color signals were even stronger than those encoding orientation but also confirms the difference of internal perceptual delays. Since color is perceived before the cue (Nothdurft, 2018) seemingly short and even negative Δt_{75} values should be possible. On the other hand, the reconstructed color signal is continuously growing up to 400 ms after stimulus onset, which suggests that the (already strong) responses continue to sum up in the population response up to this time. Finally, the in comparison to orientation slower decay of the signal, which still remains at an elevated level, indicates that the peak of the neural color response is less transient than that for orientation. All these conclusions correspond well with the primarily sustained responses of color sensitive cells in the parvocellular pathway (Dreher, Fukuda, & Rodieck, 1976; Schiller & Malpeli, 1978; Creutzfeldt, Lee, & Elepfandt, 1979; see also Nothdurft & Lee, 1982a, b).

Faces. The much longer Δt_{75} values found with facial expressions are surprising and it would be interesting to see if that could be predicted from the responses of neurons in monkey face areas. On a first view, face-specific neurons in the monkey STS region seem to have, by and large, similar latencies as orientation-specific neurons in area V1 (Baylis, Rolls, & Leonard, 1985; Rolls, Tovee, Purcell, Stewart, & Azzopardi, 1994), and peak firing rates are not or only slightly reduced. These differences are too small to account for an increase of more than 150 ms between the Δt_{75} values for oriented lines and for facial expressions in the present study (Fig. 17). In a later, quantitative analysis, however, Oram and Perrett (1992) measured the response magnitude and time course of 44 face-sensitive neurons in three Rhesus monkeys and found, in comparison to the above described population responses in macaque area V1 to oriented lines, a slightly increased mean latency (119 ms compared to 40 ms) and rise time (58 ms to 20 ms), while peak activity seemed to be about the same. This would sum to a latency difference of about 120 ms in the means, which is not too far away from the differences measured in the present study. In addition, Δt_{75} values for facial expressions were notably reduced when the cue size was changed and adopted to the same size that was used for oriented lines. Although this was tested on only two observers (and a strong effect was only seen with one of them), the finding suggests that the large differences between Δt_{75} values for orientation and Δt_{75} values for facial expression might indeed, at least partly, be explained by differences in the time course of underlying neural signals. An additional difference might come from contrast differences (the face drawings were slightly dimmer than the lines or color squares; see General Methods) which should have also affected the speed of target identification. Alternative interpretations would be that either the neural signals underlying the cued identification of facial expressions were so much smaller than those for orientation and color that the signals had to be accumulated over a much longer period of time, or that neurons in "face areas" were generally not involved in the CVS task. Both interpretations would be unlikely, as discussed above. But it is noteworthy that facial expressions and line orientations do not seem to generate own and different peaks in CVS but are similar in shape and location, for each observer. This is particularly interesting, since the only "low level" visual difference between happy and angry faces in Experiment 3 were orientation differences as measured in Experiment 1.

Attentional control of target selection

To understand the pitfalls and potentially new findings from this study, it may be worth to recall the assumed processes in the task. First, the stimulus pattern is switched on; from this moment on, all items (and item features) are processed and encoded in various regions of the visual system. As long the cue is not yet shown, however, no particular item is yet selected. The second step is target selection, which is started with the presentation of the cue. It is assumed that the cue attracts attention and that attention may then "read out" the already processed item information. It was a surprising finding in the first CVS experiments that the early presentation of items did not generally facilitate the later target analysis (Nothdurft, 2017a). Except for certain cue delays soon after stimulus onset, target identification required about the same time whether or not targets had been visible before. In fact, targets cued very long after their onset might have required even much longer presentations for recognition. This suggests that feature analysis (e.g., of target orientation) always begins with target *selection*, not target presentation. On the other hand, the cued selection of oriented lines did not seem to work in the memory; lines that had already disappeared when being cued could generally not be identified. (A demo of the findings can be found at http://www.vpl-goettingen.de/cvs/.) Apparently, this was not always the case with color in the present study. Here, targets could be identified by one observer even when they were cued quite a while after presentation (see Fig. 9d).

There are two implicit assumptions in this second step of CVS, however. One is that attention is only attracted by the cue; should attention be attracted by any of the items in the pattern before the cue is presented, the timing would change. The other assumption is that targets can only be identified under (selected) focal attention and not, as apparently in the identification of color items by observer NMB, without focal attention (i.e., "preattentively" everywhere in the stimulus). All patterns in the present study displayed feature contrast, which can be a powerful salience key (Nothdurft, 2005, 2015; see also Gao, Mahadevan, & Vasconcelos, 2008); orientation contrast (Exp. 1) and color contrast (Exp. 2) were stronger and likely more salient than the relatively small contrast between happy and angry faces. Patterns with singular items displaying high feature contrast have been used as attention attractors before (Cheal & Lyon, 1994). In the present study, however, feature contrast was randomly spread all over the pattern and should not have attracted attention to one particular item before the cue occurred. But this might have changed if attention was directly attracted by certain features, for example by red squares or happy faces. It was reported that color (and, in particular, red) is very effective in attracting and guiding visual attention (Cheal & Lyon, 1992, 1994; Dunai, Castiello, & Rossetti, 2001; Zhuang & Papathomas, 2011; Kibbe, Kàldy, & Blaser, 2017; Kasten & Navon 2018) and might be recognized even when attention is not explicitly directed there by means of other cues. In visual search, red targets may be preattentively grouped for similarity (Nothdurft, 1992), which is usually not the case with orientation unless item orientations line up to global figures (Moraglia, 1989; Nothdurft, 1992). The findings of the present Experiment 2 confirm this exception of color features with some but not all observers.

With faces and facial expressions the findings are controversial. I have previously reported (Nothdurft, 1993) that, with schematic face drawings, facial expressions do not "pop out" (and thus do not seem to attract attention). Certain faces in an array of other faces are only quickly found when they display special features (like, e.g., an open mouth) which by themselves are quickly found even when the face percept is disturbed by scrambling the components or by turning the faces upside down. The popout failure of facial expressions was confirmed in several studies (e.g., Coelho, Cloete, & Wallis, 2010; Kennett & Wallis, 2019) and put into question by others (e.g., Hansen & Hansen, 1988). Apparently, a face among non-faces (cars, houses, etc.) is quickly found (Hershler & Hochstein, 2005), but this was not the stimulus used in Experiment 3. Several studies have claimed that angry, threatening, and occasionally happy faces pop out from other faces nearby (e.g., Hansen & Hansen, 1988; Fox *et al.*, 2000). This "face-in-the-crowd" effect was extensively studied during the last 30 years and finally plausibly explained by low-level differences associated with these stimuli (Coelho, Cloete, & Wallis, 2010; Kennett & Wallis, 2019). However, if happy (or angry) faces had indeed attracted attention in the present experiments, that should have been widely dispersed over the stimulus before one item was finally cued.

CONCLUSIONS

Using the technique of cued visual selection (CVS) the study has shown that different features are indeed processed differently, with quite different timings and distinct modulations of the underlying neural signals. This is interesting both for understanding the control of spatial attention and for analyzing the differences in the strength and temporal dynamics of encoded feature information. The study has shown that color is generally faster processed than orientation and facial expressions, and that the neural encodings of all three features follow, to some extent, their own dynamics. In the context with earlier CVS studies, the present work has also shown where these experiments might be expanded and further measurements be included. For example, it should be helpful to know the exact relative perceptual timing of individual observers to synchronize cued target selection and perceived feature properties. Finally, the study has also revealed strong differences between observers both in their sensitivity to various features and in the speed at which they could identify the targets. Although all data showed the same characteristic performance variations, a detailed look at the individual variations and timing properties may improve the understanding of underlying mechanisms.

REFERENCES

Baylis, G. C., Rolls, E. T., & Leonard, C. M. (1985). Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. *Brain Research, 342,* 91-102.

Benso, F., Turatto, M., Mascetti, G. G., & Umiltá, C. (1998). The time course of attentional focusing. *European Journal of Cognitive Psychology, 10,* 373-388.

Berga, D., Fdez-Vidal, X. R., Otazua, X., Leborán, V., & Pardo, X. M. (2019). Psychophysical evaluation of individual low-level feature influences on visual attention. *Vision Research, 154,* 60-79.

Braun, J., & Julesz, B. (1998). Withdrawing attention at little or no cost: Detection and discrimination tasks. *Perception & Psychophysics, 60 (1),* 1-23.

Cheal, M. L., & Lyon, D. R. (1992). Benefits from attention depend on the target type in location-precued discrimination. *Acta Psychologica, 81,* 243-267.

Cheal, M. L., & Lyon, D. R. (1994). Allocation of attention in texture segregation, visual search and location-precuing paradigms. *The Quarterly Journal of Experimental Psychology, 47A (4),* 49-70.

Coelho, C. M., Cloete, S., & Wallis, G. (2010). The face-in-the-crowd effect: When angry faces are just cross(es). *Journal of Vision, 10(1):7,* 1-14, https://doi.org/10.1167/10.1.7.

Connor, C. E., Preddie, D. C., Gallant, J. L., & Van Essen, D. C. (1997). Spatial attention effects in macaque area V4. *Journal of Neuroscience, 17(9),* 3201-3214.

Creutzfeldt, O. D., Lee, B. B., & Elepfandt, A. (1979). A quantitative study of chromatic organisation and receptive fields of cells in the lateral geniculate body of the Rhesus monkey. *Experimental Brain Research, 35,* 527-545.

Downing, C.J. and S. Pinker, 1985. The spatial structure of visual attention. In: M.I. Posner and O.S.M. Marin (eds.), *Mechanisms of attention: Attention and performance XI* (pp. 171-187). Hillsdale, NJ: Erlbaum.

Dreher, B., Fukuda, Y., & Rodieck, R. W. (1976). Identification, classification and anatomical segregation of cells with X-like and Y-like properties in the lateral geniculate nucleus of old-world primates. *Journal of Physiology, 258,* 433-452.

Dunai, J., Castiello, U., & Rossetti, Y. (2001). Attentional processing of colour and location cues. *Experimental Brain Research, 138,* 520-526. doi: 10.1007/s002210100740.

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* (4), 225-240.

Fischer, B., Weber, H., Biscaldi, M., Aiple, F., Otto, P., & Stuhr, V. (1993). Separate populations of visually guided saccades in humans - reactions-time and amplitudes. *Experimental Brain Research, 92,* 528-541.

Fox, E., Lester, V., Russo, R., Bowles, R. J., Pichler, A., & Dutton, K. (2000). Facial expressions of emotion: Are angry faces detected more efficiently? *Cognition & Emotion, 14* (1)*,* 61-92.

Gao, D., Mahadevan, V., & Vasconcelos, N. (2008). On the plausibility of the discriminant center-surround hypothesis for visual saliency. *Journal of Vision, 8(7):13,* 1-18, http://journalofvision.org/8/7/13/, doi: 10.1167/8.7.13.

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology*, *54*, 917-924.

Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Research, 45,* 1707-1724.

Holcombe, A. O., & Cavanagh, P. (2008) Independent, synchronous access to color and motion features. *Cognition, 107,* 552-580.

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43***,** 171-216. doi: 10.1006/cogp.2001.0755.

Ji, L., Chen, W., Loeys, T., & Pourtois, G. (2018). Ensemble representation for multiple facial expressions: Evidence for a capacity limited perceptual process. *Journal of Vision, 18(3):17,* 1-19, https://doi.org/10.1167/18.3.17.

Joseph, J. S., Chun, M. M., & Nakayama, K. (1997). Attentional requirements in a 'preattentive' feature search task. *Nature, 387,* 805-807.

Joseph, J. S., & Optican, L. M. (1996). Involuntary attentional shifts due to orientation differences. *Perception and Psychophysics, 58,* 651-665.

Kasten, R., & Navon, D. (2008). Is location cueing inherently superior to color cueing? Not if color is presented early enough. *Acta Psychologica, 127,* 89-102.

Kastner, S., Nothdurft, H. C., and Pigarev, I. N. (1997). Neuronal correlates of pop-out in cat striate cortex. *Vision Research, 37,* 371-376.

Kennett, M. J., & Wallis, G. (2019). The face-in-the-crowd effect: Threat detection versus iso-feature suppression and collinear facilitation. *Journal of Vision, 19(7):6,* 1-24, https://doi.org/10.1167/19.7.6.

Kibbe, M. M., Kàldy, Z., & Blaser, E. (2017). Rules infants look by: testing the assumption of transitivity in visual salience. *Infancy,* 1-17. International Congress of Infant Studies (ICIS), Wiley Blackwell. doi: 10.1111/infa.12219.

Knierim, J. J. & Van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology, 67,* 961-980.

McIntyre, M. E., & Arnold, D. H. (2018). Synchronous and asynchronous perceptual bindings of colour and motion following identical stimulations. *Vision Research, 146-147,* 41-47.

Moraglia, G. (1989). Display organization and the detection of horizontal line segments. *Perception & Psychophysics, 45,* 265-272.

Moutoussis, K., & Zeki, S. (1997a). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society of London B, 264,* 393-399.

Moutoussis, K., & Zeki, S. (1997b). Functional segregation and temporal hierarchy of the visual perceptive systems. *Proceedings of the Royal Society of London B, 264,* 1407-1414.

Nothdurft, H. C. (1992). Feature analysis and the role of similarity in pre-attentive vision. *Perception and Psychophysics, 52,* 355-375.

Nothdurft, H. C. (1993). Faces and facial expressions do not pop out. *Perception, 22,* 1287-1298.

Nothdurft, H. C. (1999). Focal attention in visual search. *Vision Research, 39,* 2305-2310.

Nothdurft, H. C. (2002). Attention shifts to salient targets. *Vision Research, 42,* 1287-1306.

Nothdurft, H. C. (2005). Salience of feature contrast. In: Itti, L., Rees, G., and Tsotsos, J. K. (Eds.) *Neurobiology of attention,* Elsevier, San Diego, CA, pp. 233-239.

Nothdurft, H. C. (2015). Feature contrast in salience and grouping: luminance and disparity. *VPL-reports, 3,* 1-32. www.vpl-reports.de/3/

Nothdurft, H. C. (2017a). Cued visual selection – a tool to study the dynamics of neural processes in perception? *VPL-reports, 6,* 1-24. www.vpl-reports.de/6/

Nothdurft, H. C. (2017b). Cued visual selection of targets with and without orientation contrast. *VPL-reports, 7,* 1-22. www.vpl-reports.de/7/

Nothdurft, H. C. (2018). Time window(s) of cued visual selection. *VPL-reports, 9,* 1-31. www.vpl-reports.de/9/

Nothdurft, H. C. (2019). Location-cued visual selection — placeholder dots improve target identification. *Journal of Vision, 19(13):16,* 1-15, https://doi.org/10.1167/19.13.16.

Nothdurft, H. C. (2020). Cued visual selection of conjunction targets — no evidence of additional attentional requirements for the binding of color and orientation. *VPL-reports, 11,* 1-14. www.vpl-reports.de/11/

Nothdurft, H. C., and Lee, B. B. (1982a). Responses to coloured patterns in the macaque lateral geniculate nucleus: Pattern processing in single neurones. *Experimental Brain Research, 48,* 43-54.

Nothdurft, H. C., and Lee, B. B. (1982b). Responses to coloured patterns in the macaque lateral geniculate nucleus: Analysis of receptive field properties. *Experimental Brain Research, 48,* 55-65.

Nothdurft, H. C., Gallant, J. L., & Van Essen, D. C. (1999). Response modulation by texture surround in primate area V1: Correlates of "popout" under anesthesia. *Visual Neuroscience, 16,* 15-34.

Oram, M. W., & Perrett, D. I. (1992). Time course of neural responses discriminating different views of the face and head. *Journal of Neurophysiology, 68 (1),* 70-84.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32,* 3-25.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In: Bouma, H. & Bowhuis, D. (Eds.) *Attention & Performance X (*pp. 531-556), Hillsdale, N.J.: Erlbaum.

Rangelov, D., & Zeki, S. (2014). Non-binding relationship between visual features. *Frontiers in Human Neuroscience, 8: 749,* 1-11. doi: 10.3389/fnhum.2014.00749

Rolls, E. T., Tovee, M. J., Purcell, D. G., Stewart, A. L., & Azzopardi, P. (1994). The responses of neurons in the temporal cortex of primates, and face identification and detection. *Experimental Brain Research, 101,* 473-484.

Schiller, P. H., & Malpeli, J. (1978). Functional specificity of lateral geniculate nucleus laminae of the Rhesus monkey. *Journal of Neurophysiology, 41* (3), 788-797.

To, M. P. S., Carvey, K. M., Carvey, R. J., & Liu, C. H. (2019). Averaging sets of expressive faces is modulated by eccentricity. *Journal of Vision, 19(11):2,* 1-14, https://doi.org/10.1167/19.11.2.

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (2000). Response properties of the human fusiform face area. *Cognitive Neuropsychology, 17:1-3,* 257-280, doi: 10.1080/026432900380607.

Viviani, P., & Aymoz, C., (2001). Colour, form, and movement are not perceived simultaneously. *Vision Research, 41,* 2909-2918.

Wolfe, J. M. (1994). Guided Search 2.0: a revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202-238.

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,* 1-7.

Zhuang, X., & Papathomas, T. V. (2011). Cue relevance effects in conjunctive visual search: Cueing for location, color, and orientation. *Journal of Vision, 11(7):6,* 1-13, http://www.journalofvision.org/content/11/7/6, doi:10.1167/11.7.6.