Visual Selective Behavior Can Be Triggered by a Feed-Forward Process

Rufin VanRullen and Christof Koch

Abstract

■ The ventral visual pathway implements object recognition and categorization in a hierarchy of processing areas with neuronal selectivities of increasing complexity. The presence of massive feedback connections within this hierarchy raises the possibility that normal visual processing relies on the use of computational loops. It is not known, however, whether object recognition can be performed at all without such loops (i.e., in a purely feed-forward mode). By analyzing the time course of reaction times in a masked natural scene categorization paradigm, we show that the human visual system can generate selective motor responses based on a single feed-forward pass. We confirm these results using a more constrained letter discrimination task, in which the rapid succession of a target and mask is actually perceived as a distractor. We show that a masked stimulus presented for only 26 msec—and often not consciously perceived—can fully determine the earliest selective motor responses: The neural representations of the stimulus and mask are thus kept separated during a short period corresponding to the feed-forward "sweep." Therefore, feedback loops do not appear to be "mandatory" for visual processing. Rather, we found that such loops allow the masked stimulus to reverberate in the visual system and affect behavior for nearly 150 msec after the feed-forward sweep. ■

INTRODUCTION

Over the past 40 years it has been established that the numerous cortical areas that constitute the primate visual system follow a hierarchical organization (Felleman & Van Essen, 1991; Barlow, 1972; Hubel & Wiesel, 1962). Many theoretical and computational models of visual processing in the ventral pathway suggest that object recognition can be performed in a single feed-forward pass throughout such a hierarchy (e.g., Riesenhuber & Poggio, 1999; VanRullen, Gautrais, Delorme, & Thorpe, 1998; Wallis & Rolls, 1997; Fukushima & Miyake, 1982). Yet our visual systems contain extensive feedback connections that certainly play an important role in visual perception (Bullier, 2001; Lamme & Roelfsema, 2000; Lamme, Super, & Spekreijse, 1998). Indeed, many models and algorithms for visual processing explicitly require some form of feedback or iterative loops to achieve contour integration (Li, 1998; Shashua & Ullman, 1988; Grossberg & Mingolla, 1985a), texture or figure–ground segregation (Bullier, Huppe, James, & Girard, 2001; Gove, Grossberg, & Mingolla, 1995; Heitger & von der Heydt, 1993; Grossberg & Mingolla, 1985b), object recognition (Ullman, 1989, 1995; Grossberg & Mingolla, 1985a), or visual attention (Kirkland & Gerstein, 1999; Tsotsos et al., 1995; Koch & Ullman, 1985). In order to assess the plausibility of the different classes of models, it is thus critical to establish in humans the extent and the type of computational loops participating in a given visual task: In particular, are these loops necessary, or can the task be performed at all in a feed-forward mode?

In practice, the presence of "mandatory" computational loops implies that if two stimuli are presented in close temporal succession (i.e., if the stimulus-onset asynchrony is less than the time it takes for information to travel along the feedback loops), the representation of the first stimulus will be interrupted by (or overlap with) the representation of the second before the system's response is generated. This phenomenon can thus explain "backward masking" effects (i.e., failures of the visual system to perceive a stimulus that is followed too closely in time by another one; Lamme & Roelfsema, 2000; Breitmeyer, 1984; Kahneman, 1968). Under such conditions, the masked stimulus (perceived or not) is nevertheless able to trigger selective neuronal responses, albeit brief and relatively weak, throughout the ventral pathway and up to areas of the inferotemporal cortex (Keysers, Xiao, Foldiak, & Perrett, 2001; Kovacs, Vogels, & Orban, 1995; Rolls & Tovee, 1994), and can even influence behavior (e.g., Dehaene et al., 1998, 2001; Keysers et al., 2001; Neumann & Klotz, 1994). Yet these observations do not necessarily imply that visual processing is "feed-forward." Indeed, such responses could also result from a situation in which feedback loops do not fully erase the representation of the

California Institute of Technology

stimulus, but let it overlap with that of the mask. To prove that a system can perform in a feed-forward mode, one needs to analyze not only the type but also the time course of the system's output as a function of the time course of its inputs; one needs to show that the representations of the stimulus and mask do not overlap (i.e., that a masked stimulus can trigger behavioral responses that are indistinguishable—at least for a short period—from responses to the same unmasked stimulus).

We set out to address this question using a complex natural scene categorization task, where subjects must decide as quickly as possible whether a color photograph flashed for only 26 msec contains an animal. It has been previously reported, using unmasked stimulus presentation, that visual evoked potentials in this task can reflect the subject's decision as early as 150 msec after stimulus onset (VanRullen & Thorpe, 2001a; Thorpe, Fize, & Marlot, 1996). This makes such a task a particularly good candidate for a feed-forward processing implementation (see also Thorpe & Fabre-Thorpe, 2001; Thorpe & Imbert, 1989).

RESULTS

We adopt a functional (rather than anatomical) definition of a computational loop: A neuron participates in such a loop if at some point in time it receives signals that depend on the results of its own previous computations (VanRullen, Delorme, & Thorpe, 2001; Treves, Rolls, & Tovee, 1996); the loop can be said to be "mandatory" if the performance of the system is contingent upon this neuron receiving these signals. Note that we make no assumption here as to the direction of such signals in the anatomical hierarchy of the ventral visual stream (Nowak & Bullier, 1998): A lower level neuron (say, a V1 neuron) receiving signals from a higher level area (say, V4) that are not related to this neuron's previous computations will not be said to participate in a computational loop. Note also that this definition of feedback excludes any type of top-down "preparatory" signal that could take place prior to stimulus presentation. Although there is no doubt that such signals might be critical to the observed performance, the present paradigm is unfortunately not appropriate for addressing this issue.

Our approach (illustrated in Figure 1) is to analyze the distribution of reaction times (RTs) for a particular task (using both masked and unmasked stimulus presentation), and consider two critical questions. First, the duration for which the mask can overwrite the responses to the stimulus "backward" in time (relative to the duration expected in a feed-forward model) indicates the extent and duration of "mandatory" loops (i.e., for how long stimulus information must "loop" before a reliable response can be generated; Figure 1A). Second, the duration for which the stimulus will continue to affect responses "forward" in time, even after the mask
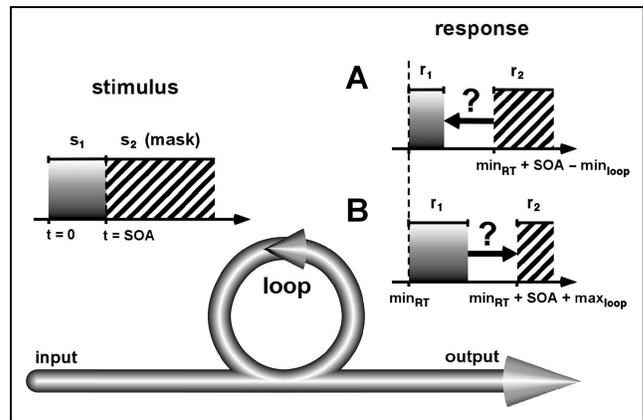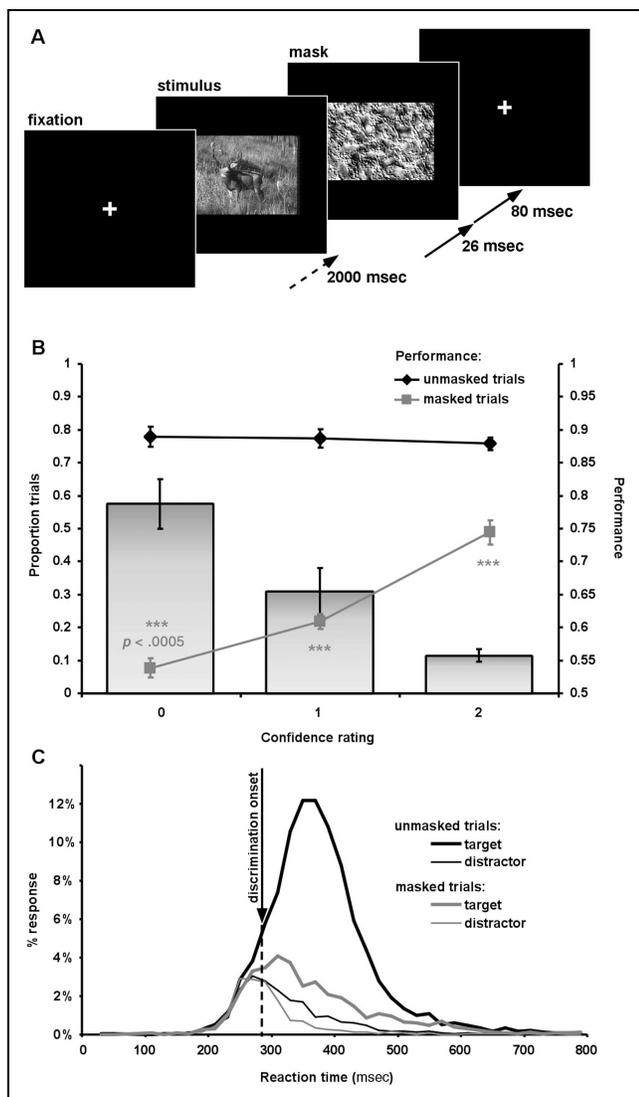

Figure 1. Processing loops in the visual system. The system's response to a succession of two stimuli (s_1 and s_2, separated in time by SOA, as shown on the left) can indicate the nature and extent of feedback loops in visual processing. In a purely feed-forward architecture with continuous inputs, the response of the system will reflect the first stimulus s_1 for a duration of exactly "SOA," then switch to reflect stimulus s_2. The system's output will directly correspond to its input, translated in time by a certain delay min_{RT} (this delay represents the minimum time at which the system is able to generate selective responses). (A) The expected duration of the normal behavioral response r_1 to stimulus s_1 is decreased by "mandatory" loops. In some systems, some form of agreement between feedback signals and the current input is required before information can be transmitted to a subsequent stage, or before a particular function can be realized (e.g., contour integration, object recognition). We refer to the minimal time during which such an agreement is necessary for processing to resume (including the time spent within the loop) as the time min_{loop}. If the information conveyed by a processing unit is replaced by a second stimulus before the feedback signal is received (formally, if SOA < min_{loop}), no agreement can be found and the processing of the first stimulus is terminated. This property is often used to explain backward masking effects. In the more general case (as shown in A), the duration of the system's normal response to the first stimulus (r_1) will be shortened by this time min_{loop} compared to the expected response duration in a purely feed-forward system (equal to SOA). (B) "Reverberation" of stimulus s_1 can affect normal behavioral response r_2 to stimulus s_2. A computational loop can imply that after a processing unit has transmitted its output and allowed the system to continue processing of stimulus s_1, it will keep receiving (through feedback connections) signals carrying information about that stimulus for a certain period. We refer to the maximal time during which information regarding a particular stimulus can "reverberate" in the system (i.e., for how long it can enter and reenter the loop before it fades away or before it is erased by a subsequent stimulation) as the time max_{loop}. In practice (as shown in B), the system's response will continue to reflect stimulus s_1 for a duration "max_{loop}" longer than the duration expected in a purely feed-forward system (i.e., SOA). A purely feed-forward system will thus be defined as: $min_{loop} = max_{loop} = 0$. When $min_{loop} > 0$, the system contains "mandatory" loops (or a minimal integration period) and thus requires a minimum time of input consistency to generate reliable responses. Finally, a feed-forward system ($min_{loop} = 0$) with reverberatory activity ($max_{loop} > 0$) is also a conceivable alternative.

has been presented, indicates how the representation of the stimulus "reverberates" in the system (i.e., for how long it can enter and reenter feedback loops before it fades away or before it is fully erased by the mask; Figure 1B).

subjective confidence in perceiving the contents of the photographs (ranging from 0 if they had not perceived anything besides the mask itself to 2 if they were certain about some aspect, element, or object in the scene). For most images (roughly 60%), the mask rendered the scene perceptually invisible (Figure 2B). This was supported by the fact that further confidence judgments for these images could not ($p > .15$; see Methods) predict categorization performance (Kunimoto, Miller, & Pashler, 2001; Kolb & Braun, 1995). However, for all three confidence levels (including ''zero'' confidence), categorization performance was significantly above chance in masked trials (t test, $p < .0005$) and almost 90% correct for unmasked trials (Figure 2B).

Most importantly, the time at which subjects were first able to discriminate between targets and distractors (i.e., generate selective behavioral responses) was not different in the masked and unmasked conditions (290 msec, χ^2 test on cumulative RT distributions, $p < .05$ for more than three consecutive time bins; Figure 2C). This indicates that in both cases, the system must have based its responses on information collected during the first 26 msec of visual stimulation. Accordingly, the responses to masked and unmasked targets significantly differ after 300 msec poststimulus (χ^2 test on cumulative RT distributions, $p < .05$ for more than three consecutive time bins; average over three subjects), which indicates that the influence of the mask is observed rapidly, but not immediately (10 msec, i.e., one time bin) after

Figure 2. Natural scene categorization (animal vs. nonanimal) with and without a mask (three subjects). (A) Schematic timeline of a masked target trial. The natural scene, subtending roughly $10° \times 6°$ of visual angle, is presented for 26 msec, immediately followed by a strong contrast mask presented for 80 msec. Subjects must respond as fast as possible if the picture contains an animal. In a later task, subjects rate their confidence in perceiving the contents of each masked scene. (B) Proportion of trials associated with each possible confidence rating (gray bars). Zero confidence (more than 50% of the trials) corresponds to a situation where masking the scenes prevents visual awareness, as indicated by the fact that further confidence judgments for these images will fail to predict categorization accuracy. For all images including those associated with zero confidence ratings, subjects can perform the categorization task above chance (t test, $p < .0005$) under conditions of backwards masking (gray line) and at almost 90% correct in unmasked trials (black line). Error bars indicate standard error of the mean. (C) Distribution of RTs to masked and unmasked scenes (average for three subjects; 20-msec time bins). The values represent the probability of the subjects responding to a particular trial type (e.g., target or distractor, masked or unmasked) in each particular time bin. Because we use a go/no-go paradigm, the sum over all time bins is not necessarily equal to 1, but to the overall probability of responding to this trial type. Therefore, responses to different trial types in each time bin are directly comparable. RTs to targets significantly differ from RTs to distractors (χ^2 test on cumulative RT distributions, $p < .05$ for more than three consecutive time bins) at the same time in the masked and unmasked condition (290 msec, marked ''discrimination onset''). This suggests that in the general case of unmasked natural scenes, the visual system must base its first responses on the information extracted during the first 26 msec of stimulation.

Categorization of Masked Natural Scenes

Subjects were presented for 26 msec with color photographs of natural scenes, half of which contained one or more animals. These included not only mammals but also fish, birds, or reptiles and could appear at any size or location within the picture. In each condition, one-third of the images were immediately masked (stimulus onset asynchrony [SOA] 26 msec) by a strong contrast mixture of white noise at different spatial frequencies, on which a naturalistic colored texture was superimposed (Figure 2A). The subjects were asked to respond as fast as possible when the stimulus contained an animal, and withhold a response otherwise. No indication was given as to whether or not they should respond on masked trials (we thus assumed that subjects would only respond if they perceived an animal in spite of the mask—or by mistake). In separate subsequent sessions, the subjects were presented with the same ''animal'' and ''distractor'' images, always followed by a mask under the same conditions as previously, and were asked to report their

discrimination onset. In other words, the duration of any potential "mandatory" loop in this task must be extremely short, and in any case less than 20 msec. These results constitute evidence for feed-forward-based processing in a complex visual task.

Note however that the responses to masked targets do not stop abruptly 26 msec (i.e., the SOA) after the discrimination onset, as would be predicted by a purely feed-forward model. The mask does not fully erase the representation of previous stimuli (Keysers & Perrett, 2002), which remains available to the system until 500 msec poststimulus (last of three consecutive significant time bins, χ^2 test between masked targets and distractors, $p < .05$). This is 180 msec longer than what would be observed in a purely feed-forward architecture. It thus appears that the role of computational loops in natural scene categorization is not a "mandatory" one (i.e., feed-forward processing is indeed possible), but rather to allow the representation of a stimulus to "reverberate" in the visual system for approximately 150–200 msec, even after it has been replaced by a mask.

Our conclusions are potentially limited by two factors. First, using natural scenes makes it virtually impossible to control for the distribution of local contrasts in the stimulus and mask in any systematic manner. In many regions of the scene, the contrast might in fact be much lower than at the corresponding locations in the mask, and vice versa. This implies that stimulus and mask can potentially merge to form a "compound" stimulus (e.g., due to ganglion cells integration times in the retina; Meister & Berry, 1999). The first 26 msec of the input signal to the visual cortex would thus be "contaminated" by the mask. This could explain why masked and unmasked targets in our task differ only 10 msec after discrimination onset (and not 26 msec as expected in a system with absolutely no mandatory loops or minimal integration times). Second, we cannot easily relate the present results to the subjects' conscious perception of the masked scenes, as we do not know (in particular, in cases where the subjects claim to be confident about "some" property of the scene) what exactly was perceived and at what point in time this occurred. In order to compare feed-forward behavioral responses with the actual subjects' perception, we need to ensure that all masked targets are consciously registered as nontargets for the task. The following experiment, a much simplified but better controlled version of the rapid visual categorization task was designed to overcome these shortcomings.

Discrimination of Masked Letters

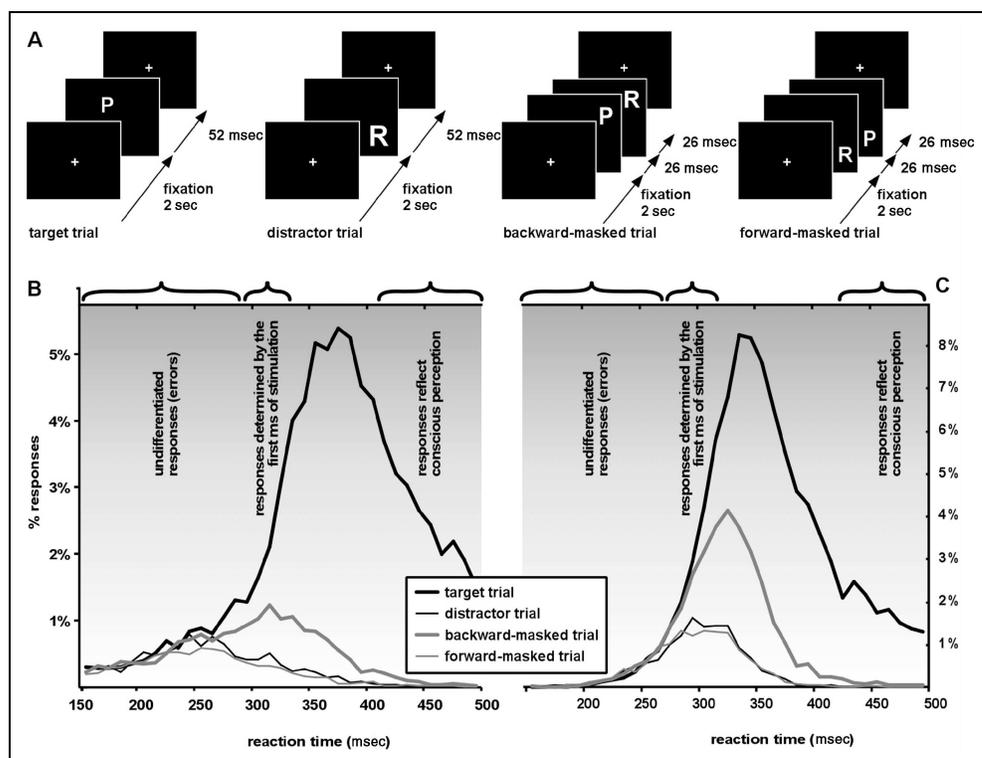In this task, the subjects had to respond as fast as possible (and within 500 msec) to the letter P (target), and ignore the letters R or B (distractors). Despite its simplicity, the main advantage of this task over the previous one was that both distractors "contained" the target, and could thus also be used as masks (Figure 3A). Letters were presented with unpredictable positions (up to $2.5°$ of visual angle away from fixation) and sizes (ranging from $0.6°$ to $1.2°$ of visual angle). In half of the trials, the target (25% of trials) or the distractors (25% of trials) were flashed for 52 msec, without a following mask. In a certain number of trials (25%), however, the target (P) was replaced after 26 msec by one of the distractors (R or B, presented at the same size and position for 26 msec) that effectively served as a mask (backward-masked trials). Additionally, an equivalent number of trials (25%) contained the same two stimuli in reverse order: one distractor (R or B) followed by the target (forward-masked trials).

In one session before and one session after they performed this discrimination task, the subjects were presented with many repetitions of these various stimulus conditions and asked to report using the keyboard what letter they had perceived. Both backward- and forward-masked trials were invariably (more than 99% of the time) reported as the component distractor letter (i.e., R or B). The target (P) was thus effectively masked, either backward or forward in time by the distractors (R or B). Even when the subjects were informed (in one final session at the end of the experiment) that some of the stimuli contained a succession of two different letters and given a chance to report two letters instead of one, they were unable to estimate above chance level the order of the two letters (Jaskowski, 1996; Neumann, Esselmann, & Klotz, 1993): On any trial where they chose to report two letters, their reports were not significantly different (χ^2 test, $p > .05$) between the backward- and forward-masked conditions.

For the purpose of the rapid discrimination task, backward-masked, forward-masked, and distractor trials were thus all consciously perceived as distractors. Yet in a feed-forward system, backward-masked trials should be interpreted as targets, at least for a certain period after discrimination onset. Accordingly, in the rapid discrimination task subjects responded significantly more (paired t test, $p < .001$) to backward-masked trials (23%) than distractors (14%) or forward-masked trials (12%).

In fact, we found that the earliest selective motor responses to unmasked and backward-masked targets showed virtually identical distributions (Figure 3B). After a period during which responses to all types of stimuli were equally likely ("undifferentiated responses," i.e., anticipations), the subjects began to discriminate targets and distractors above chance (discrimination onset between targets and distractors: 303 msec; between targets and forward-masked trials: 297 msec). However, subjects still responded to targets and backward-masked trials with the same probability for another 25 msec after

Figure 3. By using letters as stimuli we can ensure that the contrast of the mask is comparable to the contrast of the stimulus. Under these conditions, a feed-forward model predicts that fine temporal differences between stimulus and mask onset should be preserved up to the highest levels of the processing hierarchy, and observed in motor responses as well. (A) Subjects are required to respond as fast as possible when the letter P is presented and withhold responding when the letters R or B are displayed (examples are shown here only with the distractor letter R). The letters' location and size are randomized for each trial. In half of the trials, letters are flashed for 52 msec, while in the other half, two distinct letters are flashed successively for 26 msec (the target followed by a distractor in backward-masked trials, a distractor followed by the target in forward-masked trials). Under these conditions, due to backward and forward masking effects, only the distractor letter is consciously perceived. (B) Average distribution of RTs for 10 subjects (10-msec time bins). As predicted by the feed-forward model, responses to backward-masked trials follow the distribution of responses to targets for a certain period after the discrimination onset (290 msec). During this period, which lasts approximately 25 msec, behavioral responses are only determined by the first 26 msec of stimulation. After this period, the masking letter begins to affect responses, but it is only after more than 415 msec that RTs will fully reflect the subject's perception of the stimulation. (C) Individual data for one additional subject (an author) who performed more than 42,000 trials. This subject had relatively short RTs (median 355 vs. 375 msec for the other 10 subjects). The similar onset of RTs distributions for target and backward-masked trials, which was smeared by the averaging process in the previous panel, is particularly apparent here. The discrimination onset for this subject was 275 msec, and the difference between targets and backward-masked trials appeared after 305 msec (i.e., 30 msec later). Backward-masked trials went down to the level of distractors after 435 msec.

this minimum discrimination time (i.e., until 322 msec poststimulus; note that averages over 10 subjects are reported here with 1-msec precision, while individual subjects' responses were analyzed using 10-msec time bins). A similar result was obtained for an additional subject who performed more than 42,000 trials (i.e., about 10 times as much as each of the other subjects; Figure 3C). The discrimination time for this subject was 275 msec (for both the discrimination onset between targets and distractors and between targets and forward-masked trials), while the difference between targets and backward-masked trials was only visible after 305 msec (i.e., 30 msec later—here again, one should remember that the temporal resolution of this analysis is not greater than one time bin, i.e., 10 msec). Thus, for this subject as well as the other 10, the second component letter started modifying response distributions only around 25–30 msec after the onset of discrimination, as predicted by a simple feed-forward model. In other words, the processing time taken by any potential ''mandatory'' loops in this task was essentially zero.

By contrast, forward-masked trials, although they contained the same two component letters, presented for the same duration as in backward-masked trials, never triggered (i.e., in any of the 10-msec time bins considered) significantly more behavioral responses than distractors (average over all subjects, χ^2 test, $p > .05$). The preceding results thus critically depend on the order of presentation of the component letters.

As in the case of the natural scene categorization task, we must note that the masking letter in backward-masked trials did not fully overwrite the representation of the target letter. In the present task, behavioral responses to backward-masked trials were significantly higher than responses to distractors until 415 msec poststimulus (435 msec for the additional subject depicted in Figure 3C). Again, this suggests that feedback pathways and loops in the visual system can allow the representation of a masked stimulus to reverberate for 100–130 msec. From the present data, it is not possible to conclude whether this reverberation would take place at a level prior or posterior to motor decision stages (or potentially, both). In any case, however, this result is more than surprising because it implies that for nearly 150 msec after discrimination onset, the system's response is already selective, yet

can be incompatible with the subject's (subsequent) conscious perception of the stimulation.

The picture that emerges from these results is that of a feed-forward architecture in which information about the first milliseconds of visual stimulation can propagate throughout the visual system, unaffected by later changes, and can determine the earliest behavioral responses even when this information is not (because of backward masking) or not yet (in the more general case) available to consciousness. Although only a small proportion of responses will be determined solely on the basis of this feed-forward sweep (no more than 5% on average in the present experimental conditions), its effects will continue to have strong repercussions on later behavioral responses for a significant period, due to the reverberatory action of feedback loops.

DISCUSSION

The short selective neuronal latencies observed in electrophysiological recordings of monkey inferotemporal neurons (i.e., 80–100 msec; Perrett, Rolls, & Caan, 1982; Bruce, Desimone, & Gross, 1981) and the equivalently short decision-related waves observed in event-related potentials during complex visual categorization tasks in humans (as low as 150 msec; VanRullen & Thorpe, 2001a; Thorpe et al., 1996) point to the idea that visual processing in the ventral visual pathway is sometimes so fast that it could be carried out in a feed-forward manner (Thorpe & Fabre-Thorpe, 2001; Rolls & Tovee, 1994; Thorpe & Imbert, 1989). Here, using backward masking of natural scenes or letter stimuli, we show that selective feed-forward processing is indeed possible for complex visual tasks. Let us insist again that the present results go beyond the simple demonstration that masked and potentially unseen stimuli are somehow registered by the visual system (e.g., Thompson & Schall, 1999; Kovacs et al., 1995; Rolls & Tovee, 1994) and able to influence behavioral decisions (e.g., Keysers et al., 2001; Thompson & Schall, 1999; Dehaene et al., 1998; Neumann & Klotz, 1994) or affect subsequent percepts in rather unexpected ways (e.g., "feature inheritance"; Herzog & Koch, 2001). By analyzing the time course of motor responses and showing that for a short period, selective behavioral responses are entirely determined by the first few tens of milliseconds of stimulation (i.e., are not affected by subsequent changes of the visual stimulation such as backward masking), these results allow us to quantify the extent of "mandatory" computational loops in the visual system: Under conditions where the local contrast of stimulus and mask is carefully controlled, the time taken by such loops essentially amounts to zero.

It could be argued that psychophysical measurements of RTs do not generally constitute the most direct reflection of the underlying neuronal processes. Yet this simple approach allows us here to reveal a direct link between the time course of the visual stimulation and the time course of behavioral responses, the "ultimate" output of the visual system; thus, it is not unreasonable to generalize this correspondence between the system's input and output to intermediate stages of visual analysis, recognition, and motor decisions processes. On the other hand, our psychophysical approach is necessarily limited in its temporal resolution: For the statistical tests reported here, this resolution is of one time bin (i.e., 10 msec). Averaging over a number of individual subjects clearly makes our conclusions more reliable, but does not necessarily enhance the temporal resolution. For this reason, we cannot reject here the possibility of more "local" feedback loops, within or among layers of a particular cortical area, that could take place in only a few milliseconds. Our conclusions that selective visual processing can occur without "mandatory" feedback loops is nevertheless valid in the context of more "global" feedback iterations, lasting on the order of 10 msec or more.

Of course, whether or not computational loops can be defined as "mandatory" depends on the particular function investigated and on what we regard as the visual system's output: Here, we show that a feed-forward pass can be sufficient in the case of rapid visual categorization, but the performance of other functions involving the same stimuli (in particular, higher cognitive functions such as visual awareness) might explicitly require feedback processing loops (Bullier, 2001; Lamme & Roelfsema, 2000) or long integration periods (Subramaniam, Biederman, & Madigan, 2000; Libet, 1973). Indeed, a recent study by Super, Spekreijse, and Lamme (2001) indicates that the first wave of feed-forward neural activity propagating through the visual system is not sufficient to underlie visual awareness: Failures of stimulus detection are mostly independent of this feed-forward activity, but associated with the absence of a late component of the neural response (appearing after 100 msec in primary visual cortex), believed to rely on recurrent feedback processing.

The long period of reverberatory activity (approximately 150 msec) observed immediately after the feed-forward sweep in our results could correspond to this "second" component of the visual response, during which feed-forward and feedback signals are gradually matched, compared, or combined (Keysers & Perrett, 2002). Only the outcome of such computations would eventually gain access to visual awareness, thus explaining why a masked stimulus (e.g., a backward-masked letter) can reverberate in the system for around 150 msec without inducing a subsequent conscious percept. In this sense, such reverberatory activity could be thought of as a sort of "iconic memory" (Coltheart, 1980; Di Lollo, 1977; Sperling, 1960), allowing stimulus representations to remain in the visual system after the corresponding objects have disappeared from the retina, but rapidly fading in the absence of supporting bottom-up or top-down signals.

METHODS

The subjects were seated in a dark room, 120 cm away from a computer screen connected to an SGI (O2) workstation. The monitor refresh rate was 75 Hz, allowing stimuli to be displayed with a frame duration of about 13 msec.

Categorization of Masked Natural Scenes

Three subjects, including one author, participated in this experiment. Animal and distractor images were similar to those used and described in previous experiments (e.g., VanRullen & Thorpe, 2001a, 2001b).

The categorization task was performed in three separate 1-hr sessions, during which one-third of all images were masked. These were randomized and "rotated" across sessions, so that by the end of the experiment each image had been seen only once with a mask and twice without. A session contained 1536 trials (distinct images, half of which were animals and half were distractors) distributed in 96-trial blocks. Eight different masks (all constructed in a similar manner: A combination of white noise at different spatial frequencies on which a colored naturalistic texture was superimposed) were used and chosen randomly for each trial. The image and mask subtended roughly $10° \times 6°$ of visual angle. Subjects were asked to hold down the mouse button to run the sequence of trials and to respond as fast as possible by releasing the mouse button when an animal was presented.

In two subsequent sessions, the subjects were presented with the same 1536 images always followed by a mask and were asked to report their subjective confidence (0: "I saw nothing but the mask"; 1: "I saw something even though I do not know what it was"; 2: "I could identify at least one property or object in the scene"). For the analyses dependent on particular confidence levels, we rejected images that had been reported with opposite confidence levels (i.e., 0 and 2) by any two of the three subjects. Note that we did not reject images that had been reported with opposite confidence levels by the same subject. To assess the subjects' awareness during trials associated with zero confidence, we investigated whether further confidence judgments for these trials would be able to predict categorization performance (Kolb & Braun, 1995). As proposed by Kunimoto et al. (2001), we calculated the d' for awareness:

$$d' = z(p(\text{high confidence} \mid \text{correct}))$$
$$-z(p(\text{high confidence} \mid \text{incorrect}))$$

where z is the inverse cumulative function of the normal distribution and "correct" or "incorrect" refers to the subject's performance on this particular masked image

during the categorization task. A t test on the obtained d' values showed that they were not different from zero, both when the calculation was made according to the boundary between "zero" and "higher" confidence ($p > .15$) and when it was made according to the boundary between "full" (i.e., Level 2) and "lower" confidence ($p > .2$). The same result was obtained using a (much simpler) χ^2 test: Further confidence ratings (within the group of images already associated with zero confidence) were not correlated with categorization performance ($df = 2$, $p > .05$).

The statistical tests used for comparing RTs distributions are briefly described in the main text. The time at which two conditions significantly differ and not differ is defined as the first and last, respectively, of three consecutive significant time bins using a χ^2 test ($p < .05$) on the cumulative and noncumulative RT distributions, respectively (VanRullen & Thorpe, 2001b).

Discrimination of Masked Letters

Ten naïve subjects participated in this experiment. One additional subject, an author, performed roughly 10 times as many trials as each of the other subjects, and his results are therefore analyzed and presented separately. The experiment was performed in three 1-hr sessions, each containing approximately 18 blocks of 96 trials.

In one block at the beginning of the first session and one block around the end of the last session, the subjects were asked to report the letters that they perceived using the keyboard. Two of the 10 subjects applied instructions from another part of the experiment and performed this task in a speeded fashion. Their data were discarded before analysis.

In six separate blocks at the end of the last session, the subjects were informed that some of the stimuli actually contained a rapid succession of two letters and given a chance to report two letters (potentially the same) instead of one. One subject was discarded from the analysis after she reported that she could use the retinal fading of the second stimulus to estimate temporal order above chance level. Responses were pooled for other subjects and a χ^2 test ($p < .05$) compared the distributions of "nontrivial" responses (i.e., reports with more than one component letter) in backward- and forward-masked trials. No significant difference was observed.

In the remaining blocks, subjects performed the rapid discrimination task. They were asked to hold down the mouse button to run the sequence of trials and to respond as fast as possible by releasing the mouse button when the letter P, but not the letter R or B, was presented. The speeded response was enforced by auditory feedback, given only for target trials where the subject either did not respond or responded too slow (cutoff RT determined by the

experimenter, in general between 350 and 500 msec depending on the subjects).

The time at which two conditions significantly differ and not differ is defined as the first time bin and the time bin immediately before the first time bin, respectively, at which significance is reached and persists for all subsequent time bins, using a χ^2 test ($p < .05$) on, respectively, the cumulative RT distributions and "backward" cumulative RT distributions (i.e., a cumulative distribution starting with the last time bin and summing over bins backward in time). These discrimination onset and offset times were calculated separately for each subject. For the time at which targets differ from forward-masked trials, three subjects were discarded from the analysis because the obtained discrimination onset was not plausible (i.e., less than 160 msec). Discrimination times in different conditions (e.g., "target vs. distractor" and "target vs. backward-masked trials") were compared using a paired t test over the values obtained for all subjects.

For the additional subject depicted in Figure 3C, the number of trials was large enough to use the same statistical tests as previously, but using the distributions of RTs instead of their cumulative distributions. The results obtained with this method were comparable.

Acknowledgments

REFERENCES

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception, 1,* 371–394.

Breitmeyer, B. G. (1984). *Visual masking: An integrative approach.* New York: Oxford University Press.

Bruce, C. J., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology, 46,* 369–384.

Bullier, J. (2001). Feedback connections and conscious vision. *Trends in Cognitive Sciences, 5,* 369–370.

Bullier, J., Hupe, J. M., James, A. C., & Girard, P. (2001). The role of feedback connections in shaping the responses of visual cortical neurons. *Progress in Brain Research, 134,* 193–204.

Coltheart, M. (1980). The persistences of vision. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 290,* 57–69.

Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B., & Riviere, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience, 4,* 752–758.

Dehaene, S., Naccache, L., Le Clec, H. G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P. F., & Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature, 395,* 597–600.

Di Lollo, V. (1977). Temporal characteristics of iconic memory. *Nature, 267,* 241–243.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1,* 1–47.

Fukushima, K., & Miyake, S. (1982). Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern Recognition, 15,* 455–469.

Gove, A., Grossberg, S., & Mingolla, E. (1995). Brightness perception, illusory contours, and corticogeniculate feedback. *Visual Neuroscience, 12,* 1027–1052.

Grossberg, S., & Mingolla, E. (1985a). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review, 92,* 173–211.

Grossberg, S., & Mingolla, E. (1985b). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception and Psychophysics, 38,* 141–171.

Heitger, F., & von der Heydt, R. (1993). A computational model of neural contour processing: Figure–ground segregation and illusory contours. *Proceedings of the Fourth International Conference on Computer Vision, 4,* 32–40.

Herzog, M. H., & Koch, C. (2001). Seeing properties of an invisible object: Feature inheritance and shine-through. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 4271–4275.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's striate cortex. *Journal of Physiology, 160,* 106–154.

Jaskowski, P. (1996). Simple reaction time and perception of temporal order: Dissociations and hypotheses. *Perceptual and Motor Skills, 82,* 707–730.

Kahneman, D. (1968). Method, findings, and theory in studies of visual masking. *Psychological Bulletin, 70,* 404–425.

Keysers, C., & Perrett, D. I. (2002). Visual masking and RSVP reveal neural competition. *Trends in Cognitive Sciences, 6,* 120–125.

Keysers, C., Xiao, D.-K., Foldiak, P., & Perrett, D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience, 13,* 1–12.

Kirkland, K. L., & Gerstein, G. L. (1999). A feedback model of attention and context dependence in visual cortical networks. *Journal of Computational Neuroscience, 7,* 255–267.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4,* 219–227.

Kolb, F. C., & Braun, J. (1995). Blindsight in normal observers. *Nature, 377,* 336–338.

Kovacs, G., Vogels, R., & Orban, G. A. (1995). Cortical correlate of pattern backward masking. *Proceedings of the National Academy of Sciences, U.S.A., 92,* 5587–5591.

Kunimoto, C., Miller, J., & Pashler, H. (2001). Confidence and accuracy of near-threshold discrimination responses. *Consciousness and Cognition, 10,* 294–340.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neuroscience, 23,* 571–579.

Lamme, V. A. F., Super, H., & Spekreijse, H. (1998). Feed-forward, horizontal, and feed-back processing in the visual cortex. *Current Opinions in Neurobiology, 8,* 529–535.

Li, Z. (1998). A neural model of contour integration in the primary visual cortex. *Neural Computation, 10,* 903–940.

Libet, B. (1973). Electrical stimulation of cortex in human subjects, and conscious memory aspects. In A. Iggo (Ed.), *Handbook of sensory physiology, II* (pp. 743–790). Berlin: Springer-Verlag.

Meister, M., & Berry, M. J., II. (1999). The neural code of the retina. *Neuron, 22,* 435–450.

Neumann, O., Esselmann, U., & Klotz, W. (1993). Differential effects of visual–spatial attention on response latency and temporal-order judgment. *Psychological Research, 56,* 26–34.

Neumann, O., & Klotz, W. (1994). Motor responses to non-reportable, masked stimuli: Where is the limit of direct parameter specification? In C. Umilta & M. Moscovitch (Eds.), *Attention and performance* (vol. 15, pp. 123–150). Cambridge: MIT Press.

Nowak, L., & Bullier, J. (1998). The timing of information transfer in the visual system. In J. H. Kaas, K. Rockland, & A. Peters (Eds.), *Cerebral Cortex* (pp. 205–241). New York: Plenum.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research, 47,* 329–342.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2,* 1019–1025.

Rolls, E. T., & Tovee, M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proceedings of the Royal Society of London, Series B: Biological Sciences, 257,* 9–15.

Shashua, A., & Ullman, S. (1988). Structural saliency. *Proceedings of the International Conference on Computer Vision, 1,* 482–488.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74,* (whole issue).

Subramaniam, S., Biederman, I., & Madigan, S. (2000). Accurate identification but no priming and chance recognition memory for pictures in RSVP sequences. *Visual Cognition, 7,* 511–535.

Super, H., Spekreijse, H., & Lamme, V. A. (2001). Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nature Neuroscience, 4,* 304–310.

Thompson, K. G., & Schall, J. D. (1999). The detection of visual signals by macaque frontal eye field during masking. *Nature Neuroscience, 2,* 283–288.

Thorpe, S. J., & Fabre-Thorpe, M. (2001). Seeking categories in the brain. *Science, 291,* 260–263.

Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381,* 520–522.

Thorpe, S. J., & Imbert, M. (1989). Biological constraints on connectionist models. In R. Pfeifer, Z. Schreter, F. Fogelman-Soulié, & L. Steels (Eds.), *Connectionism in perspective* (pp. 63–92). Amsterdam: Elsevier.

Treves, A., Rolls, E. T., & Tovee, M. J. (1996). On the time required for recurrent processing in the brain. In V. Torre & F. Conti (Eds.), *Neurobiology: Proceedings of the International School of Biophysics, XXIII course, May 1995* (pp. 371–382). New York: Plenum.

Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence, 78,* 507–545.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition, 32,* 193–254.

Ullman, S. (1995). Sequence seeking and counter streams: A computational model for bidirectional information flow in the visual cortex. *Cerebral Cortex, 5,* 1–11.

VanRullen, R., Delorme, A., & Thorpe, S. (2001). Feed-forward contour integration in primary visual cortex based on asynchronous spike propagation. *Neurocomputing, 38–40,* 1003–1009.

VanRullen, R., Gautrais, J., Delorme, A., & Thorpe, S. J. (1998). Face processing using one spike per neuron. *Biosystems, 48,* 229–239.

VanRullen, R., & Thorpe, S. J. (2001a). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience, 13,* 454–461.

VanRullen, R., & Thorpe, S. J. (2001b). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception, 30,* 655–668.

Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology, 51,* 167–194.