Netherlands
organization for
applied scientific
research

TNO Institute for Perception

TNO-report **IZF 1991 C-18**

VISUAL SEARCH OF TRAFFIC SCENES

J. Theeuwes

TNO Institute for Perception

TNO Defence Research

Kampweg 5
P.O. Box 23
3769 ZG Soesterberg
The Netherlands

Fax +31 3463 5 39 77
Telephone +31 3463 5 62 11

TNO-report **IZF 1991 C-18**

VISUAL SEARCH OF TRAFFIC SCENES

J. Theeuwes

Contractor:
Ministry of Transport and Public Works,
Transportation & Traffic Research Division
P.O. Box 1031, 3000 BA Rotterdam

Number of pages: 29

CONTENTS

Rep.No. IZF 1991 C-18

TNO Institute for Perception
Soesterberg, The Netherlands

Visual search of traffic scenes

J. Theeuwes

SUMMARY

The present study investigates top-down governed visual selection in natural traffic scenes. It is assumed that optimal scanning of the visual environment takes place by using appropriate search routines. In the present experiment, subjects had to search for a target object (i.e., traffic sign, other road users) which was embedded in a natural traffic scenes. Given a particular prototypical scene, the target was located either at a likely (expected) or unlikely (unexpected) position. Before search for the target object started, subjects were presented with the prototypical scene for short period which was followed by a mask. This pre-presentation was supposed to activate expectations regarding the location of the target object. The probability that a target object appeared at a likely location was varied between groups of subjects. The results show the existence of scene dependent scanning strategies: if the target object is at an expected location, search is somewhat faster than when it is located at an unexpected condition. More importantly, though, is the finding that search in the unexpected condition was significantly more error prone than search in the expected condition. This suggests that subjects strategically prepare for the upcoming stimulus and only search those locations which are likely to contain a target. If the target is not found at those likely locations, subjects tend to give a negative response. This effect remained after repeated presentation. The importance of these findings for search during actual driving is discussed.

Rap.nr. IZF 1991 C-18

Instituut voor Zintuigfysiologie TNO
Soesterberg

Visueel zoeken van verkeersscènes

J. Theeuwes

SAMENVATTING

De huidige studie onderzoekt top-down gestuurde selectie in "natuurlijke" verkeersscènes. Er wordt vanuit gegaan dat optimaal zoeken in een visuele omgeving wordt gestuurd door bepaalde zoek-routines. In het huidige experiment, dienden proefpersonen te zoeken naar objecten (verkeersborden, weggebruikers) die geplaatst waren in "natuurlijke" verkeersscènes. Gegeven een bepaalde scene, was het zoek-object geplaatst op een verwachte of onverwachte plaats. Voordat proefpersonen begonnen te zoeken naar het object, werd voor een korte tijd de verkeersscène aangeboden gevolgd door een maskeringsstimulus. Er werd vanuit gegaan dat deze voor-presentatie van de scene, bepaalde verwachtingen ten aanzien van de plaats van het object, zou induceren. De resultaten laten zien dat de voor-presentatie van de scene bepaalde zoekstrategieën oproept: zoeken naar een object op een verwachte plaats was iets sneller dan het zoeken naar een object op een onverwachte plaats. Belangrijker is de bevinding dat het zoeken naar een object op een onverwachte plaats significant meer fouten geeft dan het zoeken naar een object op een verwachte plaats. Dit suggereert dat proefpersonen zich strategisch voorbereiden op het zoeken naar het target object: alleen die plaatsen binnen een scene waar, met een grote waarschijnlijk, zich het object bevindt, worden afgezocht. Wanneer het target object niet gevonden wordt op de verwachte plaats, geeft de proefpersoon een negatieve response. Deze strategie blijft na herhaalde aanbieding gehandhaafd. De implicaties van deze bevindingen voor het zoeken tijdens autorijden worden besproken.

1 INTRODUCTION

This project is part of a larger research project sponsored by the Institute for Road Safety Research SWOV and the Ministry of Transport and Public Works. It aims at developing a theory regarding perceptual selectivity in road traffic environments.

As described previously (Theeuwes, 1989) the driver's visual performance is dependent on the direction of attention in the visual field. On the one hand, attention might *externally* be captured by objects which are highly conspicuous. On the other hand attention can *internally* be directed to particular features in the road scene environment in search for information of immediate relevance to the driver. Several laboratory studies (Theeuwes, 1989, 1990, 1991) investigating the balance between these two mechanisms of selection showed that objects that are highly conspicuous do not unintentionally capture the attention of an observer. Yet, when an object is conspicuous in its surroundings by virtue of basic properties such as brightness, color, form etc., and is relevant for the task, observers are capable of selecting these objects very efficiently, i.e., parallel across the visual field.

Although many conceptions of visual object perception (e.g., Marr, 1982) and visual search (e.g., Engel, 1977) primarily focus on the data-driven structural features (e.g., features, geons, location boundaries) of objects in their environments, it is well known that the meaning and the representation of the scene in relation to the object has an influence of speed and accuracy of both object search and identification. When considering visual selection in road scene environments, these contextual effects might be extremely important because driving - as an over-learned task - might rely on many sampling typicalities. In addition, since drivers are confronted with an enormous influx of visual information from which accurate and fast sampling is crucial, they will rely on rapid resource-inexpensive and conceptually-driven feature detection. Such type of processing will especially occur under conditions of visual and cognitive load and time pressure. Note that conceptually driven processing is only adequate if the expectations induced by the environment are correct. If expectations are incorrect and drivers rely on conceptually driven processing, severe errors might occur.

It is well substantiated that the processing of visual scenes is critically dependent on their spatial arrangement. A rearrangement of objects from their "natural" position impairs recognition of faces (e.g., Homa et al., 1976) and scenes (e.g., Biederman, 1972). On the other hand, the identification of objects is facilitated when objects are presented in a coherent scene (e.g., Biederman, 1972) but is inhibited if the objects violate their ordinary relation to the visual context (Biederman, Mezzanotte & Rabinowitz, 1982). In visual search experiments similar findings are reported: search for an object located at an "unexpected" position is impaired relatively to objects appearing at their natural position. In addition, search for an object which is not likely to appear in a scene is slow and error prone (Meyers & Rhoades, 1978). Note that in memory studies, objects which are *in*consistent with expectations are better recalled and recognized than objects consistent with expectations (Pedzek et al., 1989). Although this disadvantage for expected objects might seem incompatible with the findings on search and identification, the results can easily be explained: it is hypothesized that objects that do not fit the scene require more bottom-up analysis of local features than objects which are prototypical producing an advance for inconsistent objects at recognition tests (Friedman, 1979).

Biederman et al. (1982) give a list of five classes of relations which are sufficient to characterize the difference between a scene of unrelated objects and a well-formed scene. Support (i.e., objects appear resting on surfaces) and interposition (i.e., backgrounds should appear behind other objects) refer to general physical constraints of gravity. Probability, Position and familiar Size are *semantic* relations because they require access to the referential meaning of objects. Size is related to the size of a particular object relative to other objects appearing in the scene. Probability refers to the likelihood of a given object being in a scene. Position refers to the fact that objects which are likely to appear in a given scene often occupy specific positions in that scene.

The nature of contextual effects on the processing of objects in scenes is thought to be the result of an interaction between incoming perceptual information and higher level memory representations known as *frames* (Minsky, 1975) or *schemata* (Bartlett, 1932; McClelland & Rumelhart, 1981). For example, it has been argued that objects that are obligatory in the schema are encoded more or less automatically (with a minimum use of processing resources), whereas objects that do not fit in require more resource-expensive encoding processing involving active hypothesis testing (Friedman, 1979). Loftus et al. (1983) argue that scenes are processed in two stages. Holistic information is extracted first, followed by search for specific features. The holistic information can be assessed within a single fixation of the scene (Potter, 1975). This information is thought to activate the *scene schema* which is held in a presumed pictorial memory system (Paivio, 1971). A search is then initiated for specific objects as held in temporal storage.

The present study investigates the effect of contextual information on visual search of every-day life traffic scenes. More specifically, the study explores the effect of the object-context relation "position" as defined by Biederman et al. (1982). In contrast to earlier studies, the present study investigated quite subtle "position" violations. From an application point of view, examining this effect is particularly important because this relation might be violated in every-day life traffic situations. Perceptual errors might evolve when road users have wrong expectations regarding the location appearance of particular target objects. Thuto subs, for example, traffic signs are not perceived adequately when they are located at locations which are unlikely given a particular scene.

Since it is not yet immediately clear how search objectives evolve during actual driving (for a discussion, see Theeuwes, 1989), in the present study, at the beginning of each trial, the search objective (target name: bicycle, traffic sign, car) was given to the subject. This was followed by a short presentation of the "precue" slide consisting of a traffic scene identical to the search scene, yet, without actually containing the target object. This combination of target object and to-be-searched scene was supposed to "prepare" the subject optimally for the upcoming scene in which subjects had to search for the target object. In the to-be-searched scene, there was a target object in 50% of the trials which could - for different groups of subjects - be at an expected or unexpected location. The short glance at the precue scene in combination with the target object was supposed to activate a scene specific schema which is assumed to contain knowledge about the typical makeup and contents of a scene being viewed. This scene schema will generate expectations about the locations of objects present in that scene. These expectations will bias search behavior towards those portions of the visual field which are supposed to contain maximum information (e.g., Biederman, 1972; Meyers & Rhoades, 1978). The present experiment investigates whether, dependent on scene-induced expectancies, subjects are biased to scan certain portions of traffic scenes.

2 METHOD

2.1 Subjects

Seventy-two subjects ranging in age from 19 to 55 years participated as paid volunteers. Twenty-four subjects each were randomly assigned to the "expected", "unexpected" and "mixed" conditions. All had normal or corrected-to-normal vision and had their driving-license for at least 1 year.

2.2 Apparatus

An S-R interface with external clocks (accuracy of 1 ms) connected to an IBM AT-3 with video-digitizer (Matrox Inc.) controlled the timing of the events, generated video pictures, controlled slide projectors and recorded reaction times (RTs). The response panels consisted of left and right response keys (1 x 1 cm), which were mounted 1.5 cm apart.
The stimuli were projected by means of two Kodak carousel slide projectors (Kodak Carousel S-AV 2000) on a white screen (170 x 215 cm). Fixation point, target name and mask stimulus were projected on the same screen by means of a video projector (Barco data 400). Stimuli subtended a visual angle of about 18° in the horizontal and about 14° in the vertical direction.
Four subjects separated by wooden partitions were tested in a dimly-lit room. Subjects were seated approximately 365 cm from the screen. The center of the screen was located 185 cm above the floor of the room. An intercom was used for communication with the subject.

2.3 Stimuli

The search stimuli were 35 mm black-and-white slides of specific traffic scenes. The 44 traffic scenes used, were taken from a larger sample of scenes. Various types of traffic situations were used with the requirement that the traffic scenes were considered not to be too ambiguous or unstructured and had at least some visual clutter. With respect to clutter this implied that scenes which did not contain many elements were not used. In addition, it was required that the target object (bicycle, car, traffic sign) was a naturally occurring object in the specific scene. Target objects could either appear at an expected location or at an unexpected location (e.g. expected: traffic sign appearing at the right side of a crossing; unexpected: traffic sign appearing at the left side of the crossing). The expectancy judgement was made intuitively by two observers. Precue scenes were similar to the search scenes with the modification that the target object was not present in the scene. The masking stimulus was video-digitized by computer and consisted of a jumbled mixture of elements of various traffic scenes so that no single scene could be recognized.

The expected condition slides were - when possible - produced by presenting the photographed scenes mirror-reversed, and for the precue slide, the target object was photo-technically removed. This manipulation guaranteed that also these slides were all identical with respect to light and local conspicuity. In other cases, expectancy manipulations were accomplished by changing the location of the target object in the actual scene. The distance from the center of the visual field to the target object located at an expected and unexpected location was more or less the same. This certified that differences in search times could not be attributed to differences in distances from the center of the visual field. The precue scene (the scene without the search target) and the scene containing the search target at an expected or an unexpected location were photographed within a rather short interval from the exactly same position implying that, besides very minor light differences, the slides were identical. Because target objects were never artificially *inserted* into the scene, the expectancy manipulations did not violate any of Biederman et al. 's relations which are supposed to define an object in a coherent real-world scene. Thus target objects rested on surfaces ('support'), were solid objects which could appear behind other objects ('interposition'), had the right size ('size'), and given a particular scene were likely to appear ('probability'). The only factor manipulated was 'position' suggesting that either the target object occupied a likely position (condition: "expected") or an unlikely position (condition: "unexpected").

Of the 44 trials, subjects searched 26 times for a bicyclist, 16 times for a traffic sign and 2 times for a car. In about half of these trials a target was present i.e., subjects searched 12 times for a bicyclist, 9 times for a traffic sign and once for a car, totalling 22 "target-present" trials. For each "target-present" trial, the "expected" and "unexpected" conditions were matched implying that for *each* scene there were two slides i.e., a slide in which the target object occupied an expected position and a slide in which the target object occupied an unexpected position.

2.4 **Procedure**

The sequence of events during a trial was as follows: initially, a target name (the Dutch equivalents for either traffic sign, "verkeersbord"; bicyclist, "fietser"; car, "auto") was presented for 2000 ms at the center of the screen. These letters were printed in lowercase black letters against a white background. This was followed for 800 ms by a black fixation dot at the center of the screen. Then, the precue slide was presented consisting of a traffic scene identical to the search scene without actually containing the target object. This scene was presented for about 500 ms followed by the masking stimulus which was presented for 1600 ms. Finally, the search slide was presented for a maximum of 5 s until all 4 responses were emitted. Subjects did not receive performance feedback. Between trials there was a dark time of 3 s. The sequence of events is shown by Fig. 1.

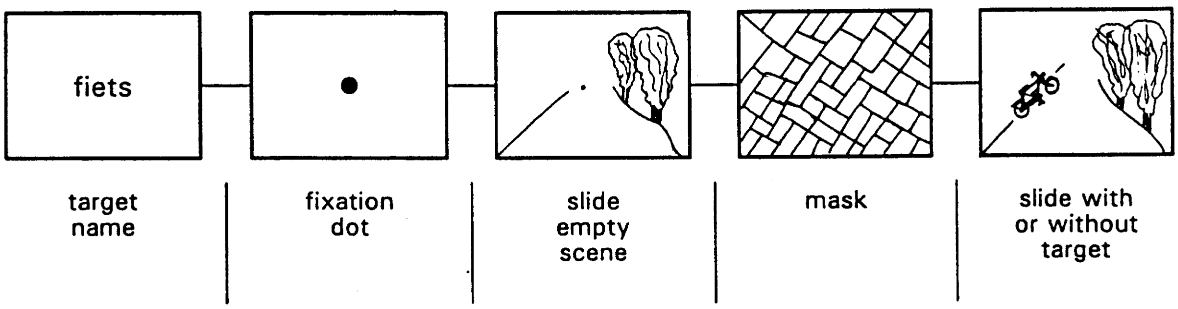


Fig. 1 Sequence of events during a trial.

Initially, two groups of 24 subjects were run. For one group of subjects the target object was consistently located at the "expected" location, whereas for the other group the target object was located at an "unexpected" location. Based on the results of these two groups, of the set of slides showing an expectancy effect (faster RTs and less errors in the expected condition), 6 slides of different traffic situations were chosen and were used in an additional run involving 24 subjects. In this run, 16 slides in which the target was located at an expected location were mixed with the 6 slides in which the target was located at an unexpected location. The "mixed" condition provides a condition in which targets are usually at expected locations (in 73% of the target-present trials) and sometimes at unexpected locations (27% of the target-present trials). The mixed condition was run to substantiate the findings of the first run and to evaluate the effects of expectancy in conditions in which most of the trials the target is positioned at expected locations.

There were two random sequences of trial order presentations, which were counterbalanced between subjects. Each subject received two experimental runs of 44 experimental trials with a short brake of about 5 minutes in between runs. Each experimental run was preceded by 8 dummy trials to avoid start up effects.

Prior to the start of the experiment subjects received written instructions explaining the purpose of the experiment. Subjects were not informed about the expectancy manipulations. They were asked to search for the target object as fast as possible while minimizing errors. It was explained that the precue scene was similar to the search scene and that they should make use of this information to prepare for the upcoming search scene. When subjects thought that no target was present they pressed the "target-absent" button with their left thumb, and when they thought that a target was present they pressed the "target-present" button with their right. Before the experiment started subjects received ten practice trials.

3 RESULTS AND DISCUSSION

3.1 **Analysis of RT and error rate**

Mean RTs and error rates were computed for each subject in each condition. A separate analysis computing the mean error rate for each slide revealed that slide-number 1 in which a traffic sign was present produced error rates exceeding 85% in both conditions. Inspection of the slide showed that the sign was too unclear to be perceived. Because this extreme high error rate the data of this slide were eliminated in the data analysis. In addition, RTs faster than 250 ms were considered guessing and therefore eliminated as well.

The mean RT for correct trials was submitted to an ANOVA with repetition (first time vs second time presented), expectancy (expected vs unexpected) and target presence (present vs absent) as main factors. There were main effects on RT for repetition [$F(1,46) = 46.9$; $p < .001$], and for target presence [$F(1,46) = 70.0$; $p < .001$]. In addition, there was an interaction effect on RT of repetition and target presence [$F(1,46) = 15.3$; $p < .001$]. The data indicates that, as expected, "target-presence" search was much faster than "target-absent" search (presence mean RT: 990.3 ms; absent mean RT: 1541.7 ms) indicating that search was self-terminating, i.e., search was terminated as soon as a target was found. The interaction between repetition and target presence indicates that the second time of presentation, search times were faster for trials in which a target was present than for trials in which a target was not present. This might suggest that subjects remember those trials in which a target was present better than the trials in which a target was not present.

The hypothesis that search would be faster for those trials in which the target appeared at an expected location is not confirmed by the present data: it was expected that for "target presence" trials a target located at an expected location would have been found faster than when the object is located at an unexpected location. Note that such an effect was not to be expected for the "target-absent" trials. Therefore, it was expected that the interaction between expectancy and target presence would have been significant. The results show that, albeit a trend in the correct direction, the interaction failed to reach significance [$F(1,46) = 2.6$; $p = .11$]. Of additional importance is the finding that repetition did not

interact with expectation [$F(1,46) = .13$; $p = .72$] suggesting that repeated presentation did not alter the effect of expectation.
In order to achieve homogeneity of the error rate variance (e.g., Kirk, 1968, p. 66), the mean error rates per cell were transformed by means of an arcsine transformation. The transformed error data were entered into the same ANOVA as performed on the response data. Again, there were main effects on error rate for repetition [$F(1,46) = 17.1$; $p < .001$], and target presence [$F(1,46) = 18.2$; $p < .001$]. In addition, again the interaction between repetition and target presence [$F(1,46) = 10.4$; $p < .01$] was significant, indicating that subjects tend to make selectively make less errors when "target-present" trials are presented for the second time.
Contrary to the effects found on response times, the interaction between expectation and target presence for error rate was significant [$F(1,46) = 7.0$; $p < .05$]. Subsequent planned comparisons on the arcsine transformed error rates between expected and unexpected conditions show that there was a significant difference between the error rates of expected and unexpected conditions of the target-present condition [$F(1,46) = 6.3$; $p < .05$]. As expected this difference in error rates of the target-absent condition was not significant. The error data clearly indicate that when subjects search for a target located at an unexpected location they are more likely to respond "target not present" than when the target is located at an expected location. This suggest that subjects check those places which are likely to contain targets and when they do not find the target at the expected locations they are likely to give a negative response ("target-not-present").
The data on search times and error rates are summarized by Fig. 2. The analyses show, as evident in Fig. 2, that in the target-present condition, subjects apply different search strategies depending on the contextual cues available. There is a non-significant trend that search in the expected location condition was to some extent faster than the search in the unexpected location condition. More importantly though is the finding that in the target present condition search in the unexpected condition was significantly more error prone than search in the expected condition. This indicates that subjects strategically prepare for the upcoming stimulus and search the scene based on the available contextual cues. If the target is at the expected location search is relatively fast and accurate; if the target is at the unexpected location search is somewhat slower and more error prone. Important is that this expectancy effect remains after repeated presentation. In addition, it appears that the group of subjects in the unexpected condition do *not* adjust their strategy over trials, i.e., it appears that subjects keep on preparing for the expected location even though for that group of subjects the target object never appeared at the expected location.

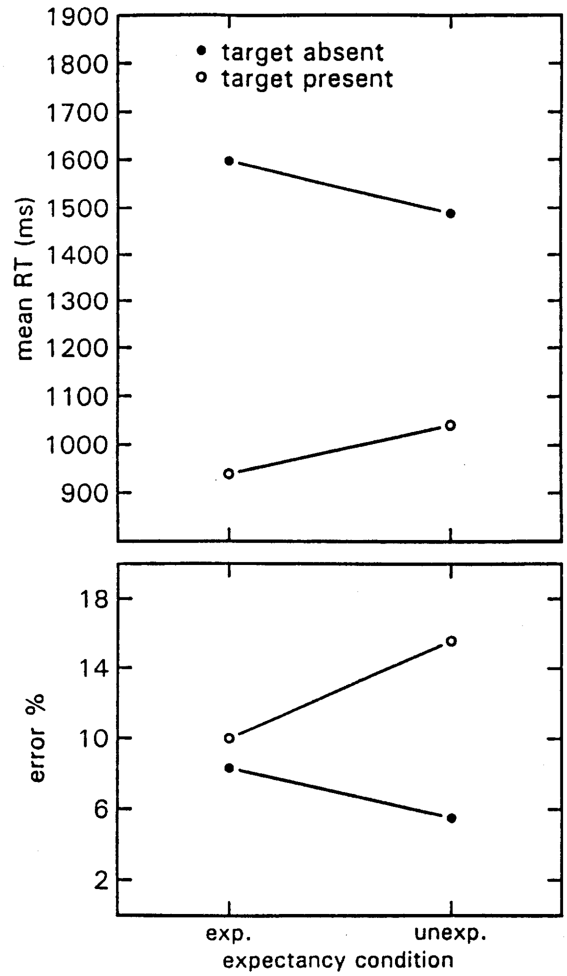


Fig. 2 Mean RT and error rate in the expected and unexpected condition.

3.2 Analysis of RT and error rates for individual slides

The mean RT and error rate collapsed over subjects was computed for each slide. Table I shows the search times and error rates in the expected and unexpected condition for each slide number. Slide number 2-22 are "target-present" slides, slide number 23-44 are "target-absent" slides. For both search times and error rates the direction of difference is indicated. A plus-sign indicates that search time or error rate is higher in the unexpected condition than in the expected condition, as could be expected by the location expectancy hypothesis.

Table I Mean RT and error rate for each slide and direction of difference.

	mean RT expected	mean RT unexpected	direction	mean error rate expected	mean error rate unexpected	direction
1	2019	2241		.88	.83	
2	756	982	+	.0	.08	+
3	1112	1745	+	.06	.33	+
4	1013	1177	+	.0	.08	+
5	1364	1411	+	.37	.29	-
6	788	852	+	.02	.04	+
7	821	910	+	.02	.02	0
8	1495	1349	-	.42	.46	+
9	908	1063	+	.02	.04	+
10	930	1062	+	.06	.0	-
11	726	913	+	.0	.08	+
12	1114	1323	+	.56	.83	+
13	1378	1182	-	.25	.29	+
14	875	1361	+	.0	.21	+
15	785	804	+	.0	.0	0
16	1044	1426	+	.08	.17	+
17	1212	1437	+	.06	.23	+
18	836	688	-	.06	.02	-
19	760	758	-	.04	.08	+
20	931	825	-	.02	.0	-
21	692	809	+	.0	.0	0
22	760	789	+	.04	.02	-
23	1526	1463	-	.08	.02	-
24	1200	1058	-	.0	.04	+
25	1993	1839	-	.02	.04	+
26	1534	1490	-	.02	.0	-
27	1046	1089	+	.02	.04	+
28	1708	1603	-	.04	.10	+
29	1662	1438	-	.08	.08	0
30	1821	1583	-	.08	.10	+
31	1899	1590	-	.04	.0	-
32	1028	1129	+	.66	.29	-
33	1577	1522	-	.0	.02	+
34	1873	1947	+	.31	.14	-
35	1848	1831	-	.08	.04	-
36	1735	1704	-	.14	.08	-
37	1218	1118	-	.06	.04	-
38	1791	1806	+	.10	.08	-
39	1614	1706	+	.04	.0	-
40	1700	1494	-	.06	.0	-
41	1144	1155	+	.02	.0	-
42	1271	1138	-	.02	.06	+
43	1700	1555	-	.02	.04	+
44	1422	1360	-	.06	.0	-

Table I shows that of the 21 target-present slides, only 5 slides showed effects in the opposite direction of that predicted (5 received a minus). If there would be no difference between the expected and unexpected conditions, the probability that 5 or less minuses occur for $N = 21$ is $p = 0.013$ (one-tailed sign-test). This suggest that according to a sign-test the hypothesis that there is no effect of expectancy should be rejected in favor of the hypothesis that RTs in the expected condition are faster than in the unexpected condition. Again, the error rates show that in the unexpected condition more errors are made than in the expected condition ($N = 18$; 5 minuses; $p = 0.048$). Note that the conclusions based on this non-parametric sign test are only confirmed by the ANOVA as far as errors are concerned. As shown above the interaction expectancy x target presence failed to reach significance.

It is realized that according to the sign-test, in the "target-absent" condition, subjects tend to be faster in the unexpected condition, a finding which is not easily understood. Note, though, that in comparison with the "target-present" condition the RT differences are rather small. More importantly, it should be realized that the experiment focused on the expected versus unexpected manipulation for the "target-present" condition and the "target-absent" slides only were used as filler items. Because these filler slides are not relevant at all for the hypotheses tested, they probably were chosen haphazardly, possibly giving rise to uncontrolled effects.

3.3 The mixed condition

In this condition subjects received again two times 22 target-present and 22 target-absent slides. Of the target-present slides, in 16 slides, the target was located at an expected location, whereas in 6 slides the target was located at an unexpected location (slide no. 2, 3, 9, 14, 17, 21). Mean RTs and error rates were computed for each subject averaged over the 6 slides for each of the expectancy conditions (expected, unexpected, mixed unexpected). Fig. 3 shows the results.

An ANOVA performed on the mean RTs of all groups for correct responses, for the slides 2, 3, 9, 14, 17, 21 shows significant main effects on RT for repetition [$F(1,69) = 16.2$, $p < 0.001$], and expectancy [$F(2,69) = 3.3$, $p < 0.05$]. The effects of these latter factors also interacted [$F(2,69) = 10.3$, $p < 0.001$]. An ANOVA on the arcsine transformed error rates showed a significant main effect of expectancy [$F(1,69) = 12.0$; $p < .001$].

These results confirm earlier conclusions that subjects use contextual cues to look at those places where target objects are likely to be found. Fig. 3 suggests that in the mixed conditions subjects tend to search even faster and more error prone than in the unexpected condition. This indicates that subjects search selectively for the likely location and if a target object is not found at the expected location, search is stopped and a negative response is given. Planned comparisons show, however, that this trend of increasing errors and reducing search time going from unexpected to the mixed condition failed to reach significance. This finding is in line with earlier conclusions which suggested that the expectancy effect is rather robust and does not change over repeated trial presentations. The expectancy "set" induced by the mixed condition appears to have some effect in the expected direction but does not significantly increase the already strong (extra-laboratory) expectancy effects as observed in the expected and unexpected conditions.

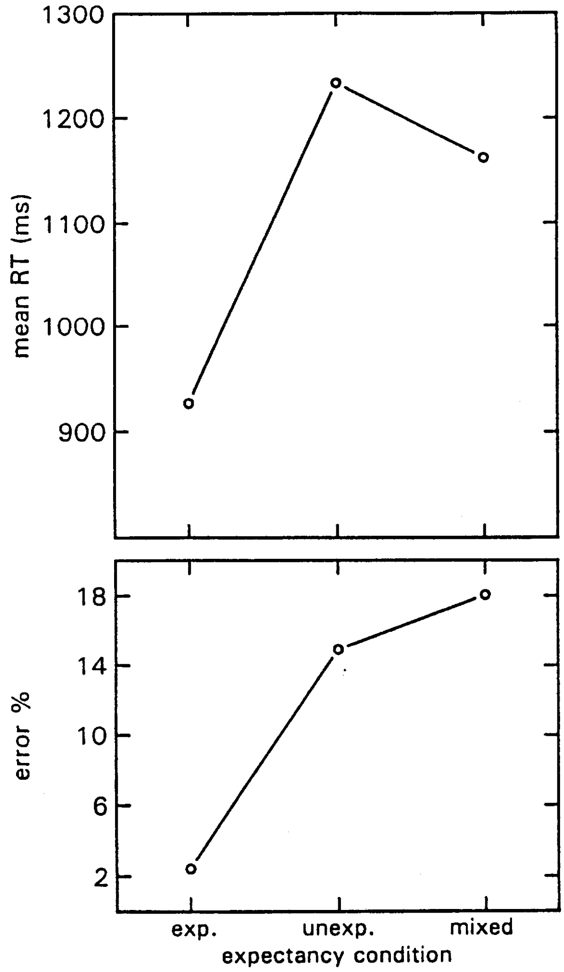


Fig. 3 Mean RT in the expected, unexpected, and mixed unexpected condition (above); error rates in the expected, unexpected, mixed unexpected conditions (below).

3.4 Qualitative evaluation of single slides

Based on the observed differences in the expected and unexpected condition (see Table I) the individual "target-present" slides will shortly be discussed. In the appendix the 22 target present slides for the expected and unexpected conditions are provided. Note that the conclusions from this paragraph are only qualitative and are not substantiated by statistical evidence. The directions and the magnitudes of the effects are discussed in global terms.

Slide 1. Search objective: traffic sign. As noted the traffic sign (reserved parking sign) was too small to be perceived and generated an extremely high error score. This was the reason to exclude the data from this slide from all analyses.

Slide 2. Search objective: traffic sign. This slide generates moderate expectancy effects both on error rate and RT. In the unexpected condition the traffic sign ("give way") is located on the left side, in the expected condition on the right side of the street.

Slide 3. Search objective: traffic sign. This slide generates large expectancy effects on RT and error rate. At the approach of the crossing the "give way" sign is either located at the right (expected) or left (unexpected) side of the street. Note the extreme large error rates for the unexpected condition suggesting that the scene induces a strong expectancy for finding a sign on the right side corner of the crossing.

Slide 4. Search objective: traffic sign. Moderate expectancy effects on both RT and error rate. The sign (warning sign for crossing children) on the right is found somewhat faster than on the left.

Slide 5. Search objective: traffic sign. An expectancy effect on RT in the expected direction; on error rate in the unexpected direction. The sign (one way street) is located either on the right side on the to-be-approached street (expected) or on the left side (unexpected).

Slide 6. Search objective: traffic sign. An expectancy effect on RT and error rates. Again right side expected, left side unexpected. The relatively short search times suggest that this sign is rather conspicuous within the scene. Yet, albeit its conspicuity the expectancy effect remains.

Slide 7. Search objective: traffic sign. An expectancy effect on RT. The "give way" sign appears to be rather conspicuous, and is found faster on the right than on the left side.

Slide 8. Search objective: traffic sign. No consistent expectancy effect. The error rate and relatively large search times suggest that the sign ("reserved parking") is hard to find.

Slide 9. Search objective: car. The car coming from the expected (in this case right side) is found faster than a car coming from the unexpected (left) side.

Slide 10. Search objective: traffic sign. An expectancy effect on RT suggesting that the (do not enter: one way street sign) is found faster when located on the right than on the left side.

Slide 11. Search objective: bicyclist. A bicyclist coming from the expected direction in a one-way street is found faster than when he is coming from the unexpected direction.

Slide 12. Search objective: bicyclist. In the expected condition the bicyclist is riding on the same level as the main road is; in the unexpected condition he is riding on a road which is somewhat lower than the main road. Larger expectancy effects on RT and error rate. The general level of errors is relatively large.

Slide 13. Search objective: bicyclist. Reversed expectancy effects on RT. It appears that subjects find the bicyclist who is somewhat lower than the main road faster than the bicyclist who is on the same level as the main road. It appears that subjects search for the right side first.

Slide 14. Search objective: bicyclist. The bicyclist coming from the direction as indicated by the one-way street arrow is found much faster than the bicyclist coming from the opposite direction as indicated by the one way street arrow.

Slide 15. Search objective: bicyclist. A bicyclist coming from the direction as indicated by the one-way street arrow (in this case from the left) is found faster than coming from the opposite direction.

Slide 16. Search objective: bicyclist. A bicyclist coming from the right side is found much faster and more accurate than a bike coming from the left side.

Slide 17. Search objective: bicyclist. Similar to slide 16.

Slide 18. Search objective: bicyclist. Expectancy effect in the opposite direction. In appears that a bicyclist on the main road (where bikers are not allowed to bike) is found faster than a bike on the bike path.
Slide 19. Search objective: bicyclist. Similar to slide 18.
Slide 20. Search objective: bicyclist. Similar to slide 18.
Slide 21. Search objective: bicyclist. Expectancy effect in the right direction: a bicyclist at the bike path is found faster than the bike at the main road.
Slide 22. Search objective: bicyclist. A bicyclist coming from the direction as indicated by the one-way street arrow (in this case from the left) is found faster than coming from the opposite direction.

This qualitative evaluation reveals that for some slides the expectancy effect was quite small or even in the opposite direction, (i.e., the slides in which the biker was positioned on the main road. Note though, that the scenes had to chosen arbitrarily because no theory exists regarding subjective categorization of road scenes, and the expectations they induce.

4 GENERAL DISCUSSION

The present study demonstrates the existence of scene dependent scanning behavior. Dependent upon the meaning and content of the scene in combination with the search objective, search behavior is biased towards certain portions of the visual field. In daily life people operate in a traffic environment which is reasonably predictable, and appear to rely on this assumption. The assumption is reinforced by a large number of times in which this inference is correct and it is supported by a common lack of feedback and/or penalties for detection misses. For typical traffic scene, it appears that this search strategy is carried over into the laboratory and remains present after repeated presentation even in conditions in which the target object is presented consistently at unlikely positions.
The present results are in line with the findings of Meyers & Rhoades (1978). In their study, subjects searched for target objects in every-day life scene (e.g., searching for a saucepan in a kitchen scene). They demonstrated the effect of location probability: searching for an object appearing at an "out of place" position (e.g., the saucepan under the kitchen table) was much slower than when the target object was at a likely position (e.g., the saucepan on the stove). In a mixed condition in which on half of the trials, the target object appeared at a likely position and on the other half on an unlikely position, search time was relatively short (similar to the "in-place" group) with a relatively high error score. Comparing the present results to Meyers & Rhoades' findings indicate that the present subjects in the unexpected and mixed conditions had a search performance similar to subjects in the mixed condition of the Meyers and Rhoades study. In these conditions, subjects only search places that are likely to contain a target object, and when it is not found at that location, they give a negative (target-not-present) response. In Meyers & Rhoades' study, subjects in the out-of-place (unexpected) condition learned that context was not a useful aid in their search

and adapted their strategy resulting in an accurate yet slow search. The present data do not suggest such a mechanism: subjects in the unexpected condition are somewhat slower; yet, the high error rate suggests that they do not adapt their strategy and have the tendency to remain searching for the likely positions only.

The present findings demonstrating the existence of scene dependent scanning behavior should be considered as quite striking given the real-life stimulus material used. Since the present study uses real traffic scenes, the possibilities of placing objects at "out-of-place" positions are rather limited. Unlike for example the stimulus material as used by Meyers & Rhoades (1978), traffic scenes have usually a clear 3-D perspective. This perceptive limits the possibilities of placing an object of a given size in that scene. In most cases one can only put the object at, for example, the right or left side of the scene. Putting the object at any given other location, might violate other relations defining a coherent scene such as "size" and "support". In addition, to ensure that search time differences reflect expectancy effects rather than conspicuity effects is was necessary to keep local conspicuity as similar as possible. For example, if an object on the right side of the road is located against an empty background and on the left side against a cluttered background, search time differences are more likely to reflect conspicuity differences rather than expectancy differences. Also, to reduce the effects of conspicuity on search behavior the present experiment used black-and-white pictures rather than colored pictures.

Note that all these constraints were taken into account to increase the ecological validity of the present study. It should be realized that the effects of contextual driven search might be much stronger in real driving especially in conditions in which there is a relatively high visual load i.e., driving in busy traffic in urban environments, or under reduced sight conditions, i.e. driving at night, in fog and rain. The black-and-white picture used in the experiment can be considered as simulating these sub-optimal driving conditions. Especially in these situations, rapid resource-inexpensive and conceptually-driven feature detection is advantageous. The presently observed data suggest that objects at unexpected locations are not seen too late but, in most cases, not seen at all, i.e., when searching for objects at unexpected locations subjects tend to say "target-not-present". It is very likely that these type of errors also occur when searching during actual driving. In fact, accident data seem to confirm this notion: a large portion of drivers (about 37%) involved in automobile crashes do not act too late but do not act at all to avoid the collision (Sussman, Bishop, Madnick & Walter, 1985). Note that accidents do not occur often indicating that errors in visual sampling, i.e., detection misses, are not fed back to the driver. On the other hand, correct expectancies i.e., finding an object where you expected it, are consistently reinforced.

Given these considerations, it is clear that extremely dangerous situations may occur when the design of the traffic environment induces certain expectations regarding the spatial arrangement of objects in that scene, which are not correct. For example, car drivers approaching a typical crossing with a bicycle track tend to look only to the right side and stop searching when a bike is not found at that location. When at such a crossing, bikes approach from both directions, bicyclist

coming from the left are likely to be missed. Because the lay-out of the bike path induced wrong expectations, drivers tend to look to one side only. In those circumstances, it would be better that the bike path does not resemble a bike path at all so that no expectancies are induced.
In which driving situations and under what conditions, expectancies are in operation is still very much unclear. As a first instance, Riemersma (1988a, 1988b) demonstrated that road users have different and more categories to distinguish traffic environments than there are distinguished officially. Future studies should identify the circumstances under which conceptually driven detection occurs, which elements of the environment induce expectancies, and what type of expectancies can be recognized. These findings may help in the design of safer road environments.

REFERENCES

Bartlett, F.C. (1932). Remembering. Cambridge: Cambridge University Press.

Biederman, I. (1972). Perceiving real world scenes. Science 177, 77-80.

Biederman, I., Mezzanotte, R.J. & Rabinowitz, J.C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. Cognitive Psychology 14, 143-177.

Engel, F.L. (1977). Visual conspicuity, visual search and fixation tendencies of the eye. Vision Research 17, 95-108.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. Journal of Experimental Psychology: General 108, 316-355.

Homa, D., Haver, B. & Schwartz, T. (1976). Perceptibility of schematic face stimuli. Evidence for perceptual gestalt. Memory & Cognition 4, 175-185.

Kirk, R.E. (1968). Experimental design: procedures for the behavioral sciences. Belmont: Wadsworth, Inc.

Loftus, G.R., Nelson, W.W. & Kallman J.J. (1983). Differential acquisition rates for different types of information from pictures. Quarterly Journal of Experimental Psychology 35A, 187-198.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. San Francisco: Freeman.

McClelland, J. & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception; Part I. An account of basic findings. Psychological Review 88, 375-407.

Meyers, L.S., and Rhoades, R.W. (1978). Visual search of common scenes. Quarterly Journal of Experimental Psychology 30, 297-310.

Minsky, M. (1975). A framework for representing knowledge. In P.H. Winston (Ed.), The psychology of computer vision. New York: McGraw-Hill.

Paivio, A. (1971). Imagery and verbal processes. New York: Holt, Rinehart & Winson.

Pezdek, K., Whetstone, T, Reynolds, K. & Askari, N. (1989). Memory for real world scenes: the role of consistency with schema expectancy. Journal of Experimental Psychology: Learning, Memory and Cognition 15, 587-595.

Potter, M.C. (1975). Meaning in visual search. Science 187, 965-966.

Riemersma, J.B.J. (1988a). Zonering en herkenbaarheid; een experiment. Report IZF 1988 C-2, TNO Institute for Perception, Soesterberg, The Netherlands.

Riemersma, J.B.J. (1988b). Enkelbaans/dubbelbaans autowegen: beleving van de weggebruiker. Report IZF 1988 C-4, TNO Institute for Perception, Soesterberg, The Netherlands.

Rumelhart, D.C. (1977). Introduction to human information processing. New York: Wiley.

Sussman, E.D., Bishop, H., Madnick, B. & Walter, R., (1985). Driver inattention and highway safety. Transportation Research Record 1047, 40-48.

Theeuwes, J. (1989). Visual selection: exogenous and endogenous control; A review of the literature. Report IZF 1989 C-3, TNO Institute for Perception, Soesterberg, The Netherlands.

Theeuwes, J., (1990). Perceptual selectivity is task dependent: Evidence from selective search. Acta Psychologica 74, 81-99.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. Perception & Psychophysics 49, 83-90.

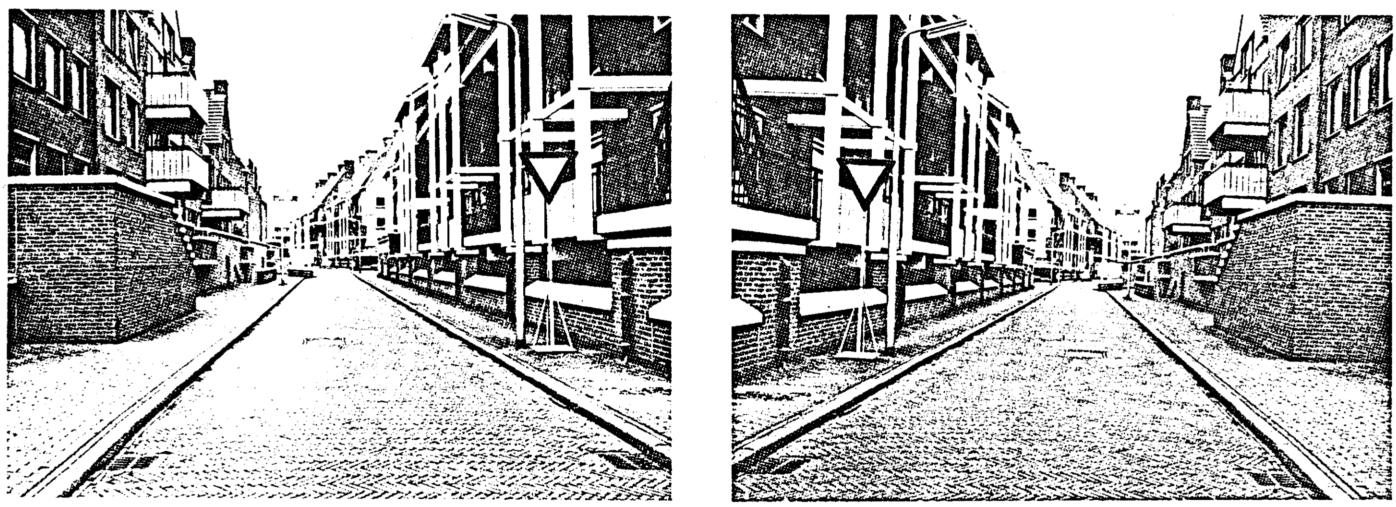
Soesterberg, August 23, 1991

Drs.ing. J. Theeuwes

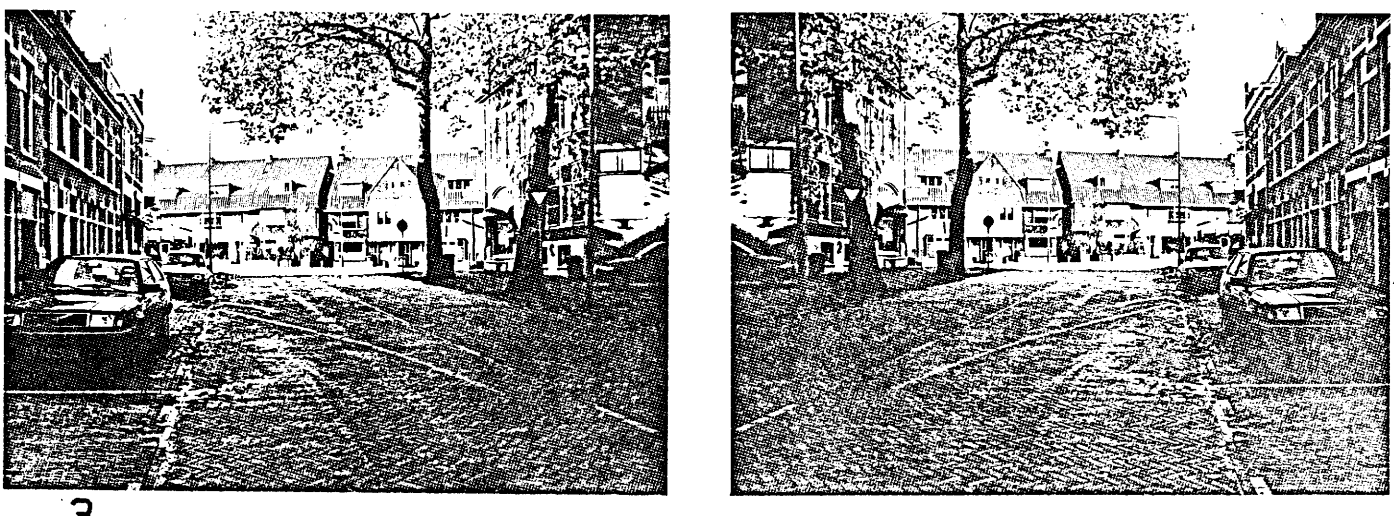
APPENDIX

expected unexpected

1

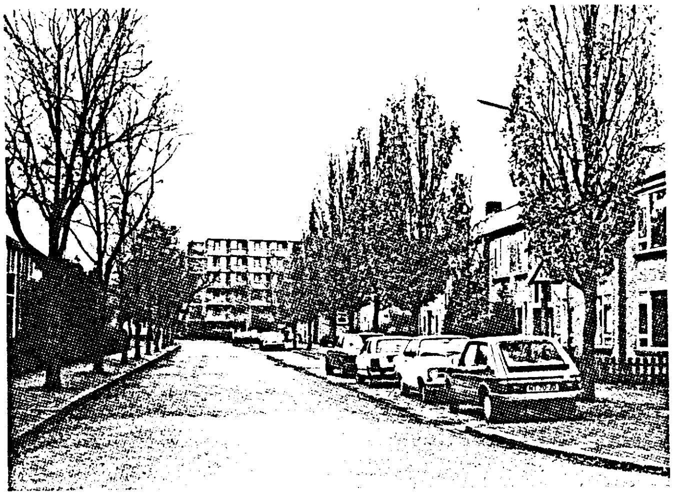


2



3

expected | unexpected




4


5


6

expected unexpected

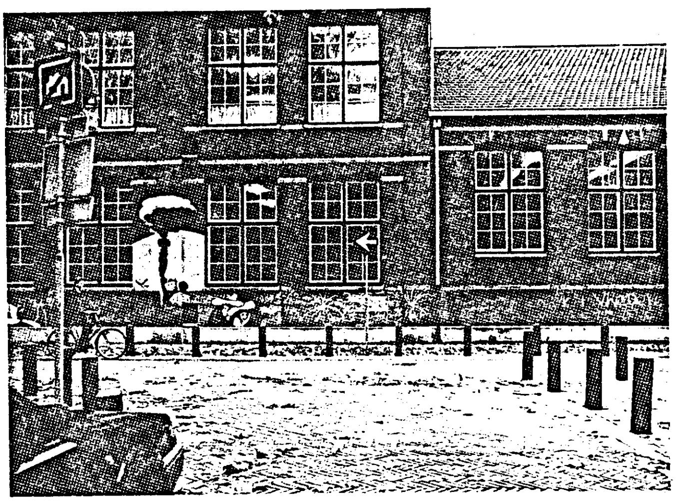
7

8

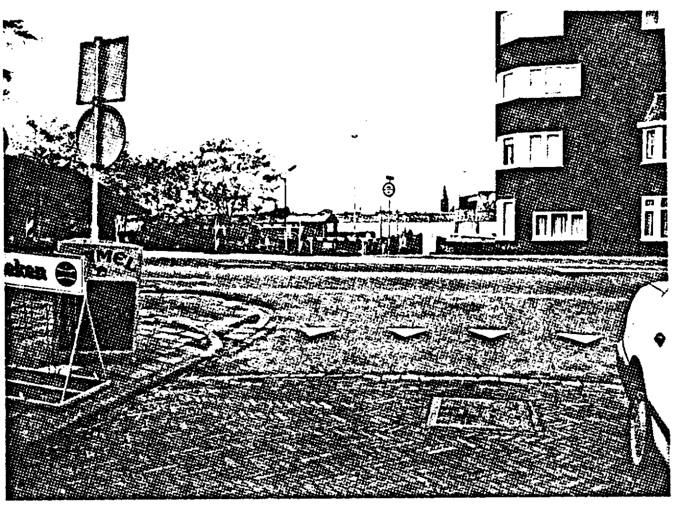
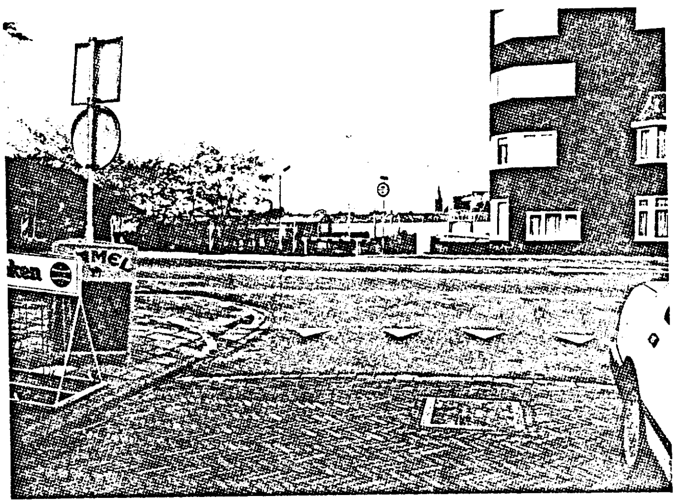
9

expected | unexpected


10

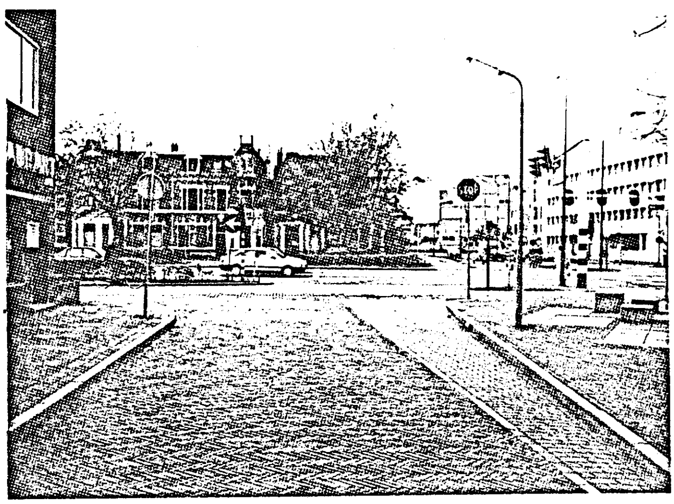
11

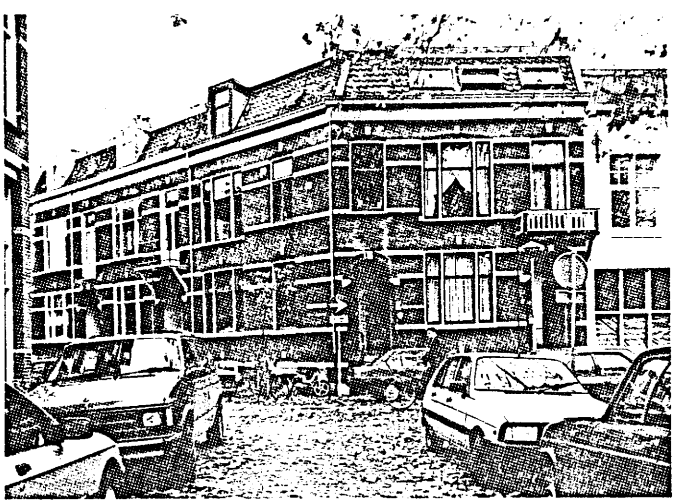
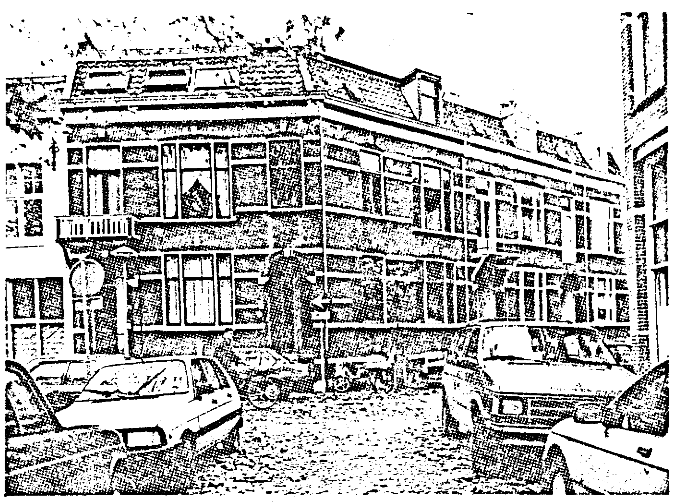
12

expected | unexpected


13

14

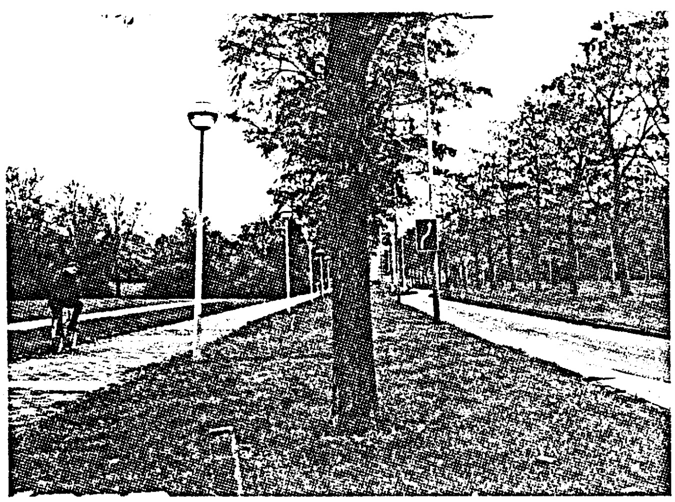
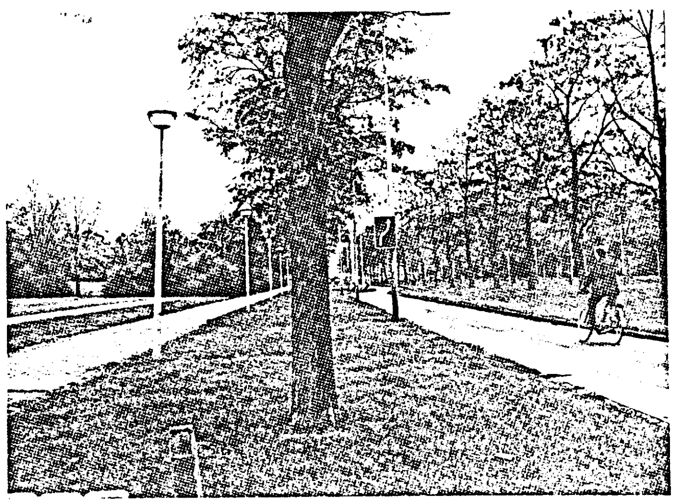



15

expected unexpected


16

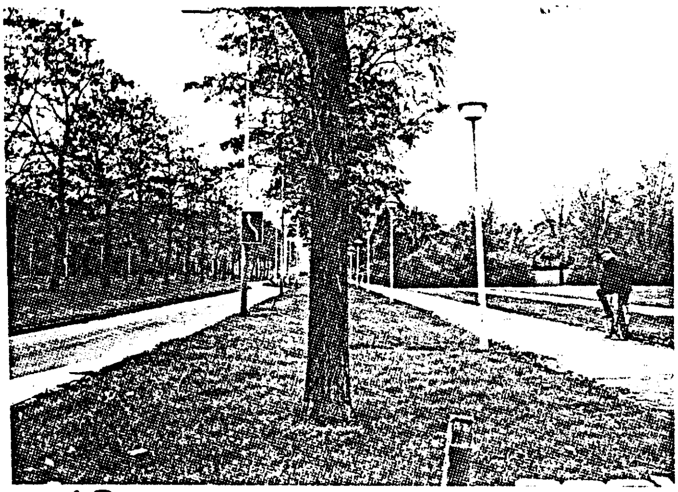

17

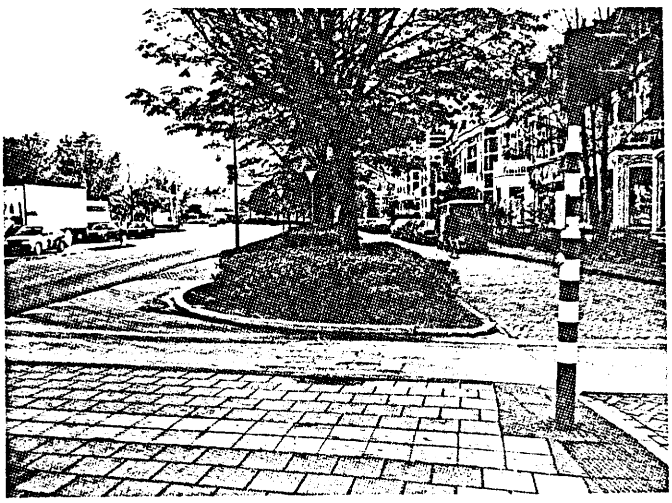
18

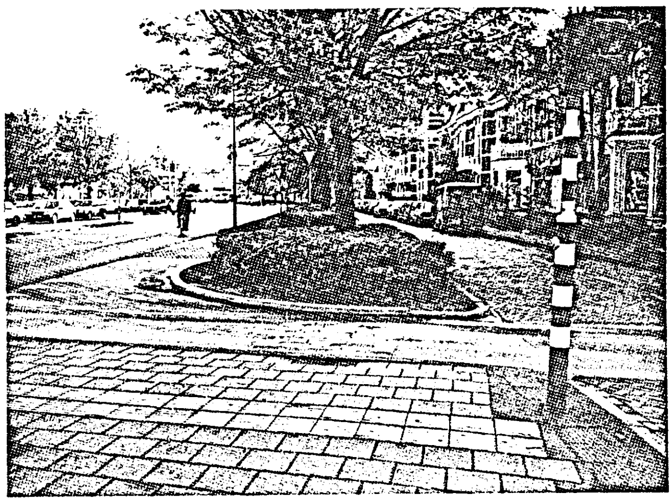
expected

unexpected

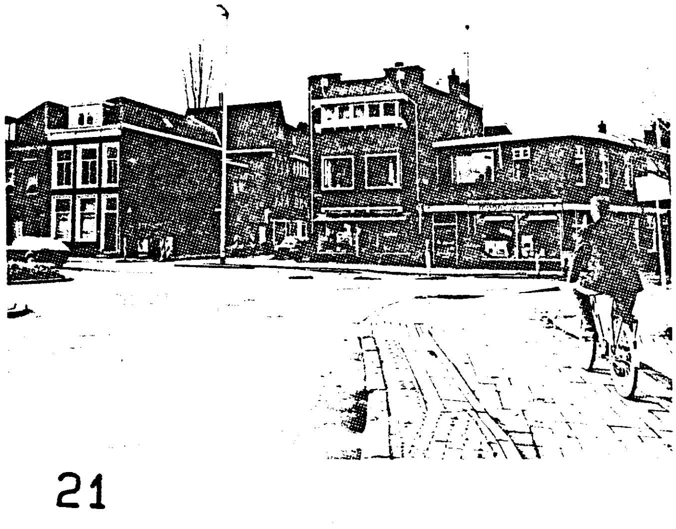


19





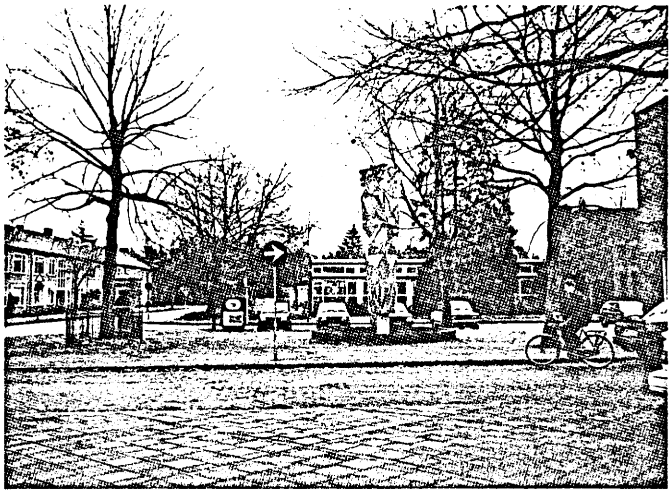
20



21

expected

unexpected



22