Acta Psychologica 55 (1984) 215-230 North-Holland

INITIAL MICROGENETIC STEPS IN SINGLE-GLANCE FACE RECOGNITION

Gé CALIS, Jan STERENBORG and Frans MAARSE

Accepted August 1983

Single-glance recognition of a familiar face cannot be explained as perception of a mediating image and subsequent reasoning processes to identify the image. This recognition is too fast for a normal reasoning process, neither can it be verbalised as such a process. But that does not simply mean that the face is perceived "directly". This pre-verbal high-speed identification seems only possible by means of tacit "micro-genetic" steps, which successively "actualise" the relevant information. When we present photographs of different familiar persons with film-speed on the same place, it seems possible to demonstrate this. ft was hypothesised that the first portrait "triggers" directing "schemata", capable of actualising relevant information. This "processing" of the first portrait provides general information, which specifies the presence of a face in a particular position, and perhaps even possible sets of familiar faces.

However, when a second portrait is presented after a critical interval, final identification steps, which originate from the previous phases, still have to be made. When both faces are in similar positions, these final tests are continued on the second portrait, in much the same way as must be the case with the next frame in a continuing shot of a movie picture. Especially when portraits also share common set reducing features, this interactive microgenesis initially produces a decreasing recognition function of stimulus onset asynchrony for the first portrait and simultaneously an increasing function for the second portrait.

These results are also discussed in terms of backward and forward visual masking.

The concept of " Aktualgenese" was introduced by the Leipzig-school of Gestalt psychology. It expressed the introspective notion that even single-glance perception is a hierarchical developing event over time and it was also the name for a set of rather deficient techniques to demonstrate this (cf. Smith 1957: Flavell and Draguns 1957: Navon 1977). Werner (1935) translated this concept, not wholly satisfactorily, as "Microgenesis" and invented his famous paradigm of successive presentation of a disc and an enveloping ring as a method of demonstration.

 In cognitive psychology these matters are discussed and explained in terms of a cycle (Neisser 1976) of data-driven, or bottom-up, and conceptually-driven, or top-down processing of information. Of course, especially since the publication of Gibson's (1979) final book, there is also a vehement debate about this "indirect" or "inferential" notion of perception (cf. UIlman 1980; Fodor and Pylyshyn 1981; Michaels and

CareIlo 1981; Turvey et al. 1981). But the oppositional ideas about a (apparently indeed) direct perception of, for instance, a face seen in a single glance are still resembling the old Gestalt notions in that they do not explain, and do not even focus on the crucial cognitive process (cf. Calis 1974, 1984). The conceptual framework of Gestalt psychology, much like Gibson's resonance theory, is in a sort of mixture of phenomenal terms and terms of analogue (isomorphic) processes, without much specification of the relation to discrete or categorical knowledge criteria that must guide the essentially decisive cognitive processes. At least with respect to the explanation of the perception of really complex objects (like faces) this means that they seem to be content with a gross misrepresentation of postulated " smart mechanisms" (Runeson 1977). It is not easy to imagine a "faceo-meter" on the analogy of the directly dynamo-driven automobile speedometer (not even a holographic one, cf. Haugeland 1981: 267) that would resonate to familiar faces -i.e. point to some discrete scale value -without any computation or

inference. Moreover, even if we would agree that the " pick-up" of additional energy properties in the physical context of the face might " tune" or " adjust" a specific resonator without computation, we would only describe a bottom-up device. Such a device does not yet explain the "pick-up" of the corresponding information, which Gibson therefore simply uses as a synonym of resonance. Thus we anyhow need to describe some loop from cognitive (memory) structures, that incorporates questioning, representing,

inferring, expecting, choosing and genesis. In short, we have to describe an intentional system (cf. Dennet t 1978).

 Perhaps also because of these conceptual and methodological reasons the "bidden" activity of Werner's "disrupted" disc could become the problematic phenomenon, instead of the microgenesis of what was actually seen. As such it was anyhow the origin of a deluge of research on visual masking and subliminal perception. Nevertheless Werner's technique of successively presenting separate identifiable visual stimuli might be suited for the (more adequately specified) original purpose of understanding the nature of the immediate perceptual genesis. Actually every ordinary movie picture or TV -recording presents a series of static images, that each may contribute to the specification of informational moments in a very fast genesis of some perceptual reality.

 In a first attempt to develop and adapt Werner's technique we made a film of about 50 different frames. In fact each frame of this film was a frontal portrait of a different familiar person, photographed under conditions of optimal standardisation ( eyes overlapping). When this strip of film was mounted in a close loop and projected at the normal rate of about 20 frames per second, a frontal face was perceived clearly. There was only some continuous deformation of the face as if it was seen through an oscillating and distorting screen. But it was not recognised as the face of a familiar person. Only when the projection was stopped, for example by interception of the light or closing the eyes, one of the last projected faces was recognised. To us this suggests, that some initial frame triggers a microgenetic identification procedure. When, however, the final identification step, originating from the previous phases, is "computed", the frame has been replaced by another one. Therefore this specific purposive identification is bound to fail. Of course this happens again and again. The perceiver never reaches a higher identification level than that of "some face in a frontal position".

 With this technique it is difficult to obtain adequate quantitative results. Therefore we have actually produced a number of films of this type, each only consisting of two frames. In this way we keep the single glance as pure as possible, because a second glance is not even possible. Moreover, if we present the subject with a' simple and natural perceptual task (who do you see?) the subject not only knows, but also can respond almost immediately. He or she does not need, and, with a proper design, cannot even use some verbal reasoning strategy to meet the demands of the task. Together with the introduction of a number of experimental and control conditions this is the essence of the experimental technique which is reported with its results in the rest of this paper .

Hypothesis

We specifically want to test the hypothesis that, in the first phases of seeing an object, immediately after this "enters the visual field", there must be an identification procedure and an increasingly more specific classification. Thus at each level of this sequence it must be the classification, just reached, that, by virtue of a corresponding cognitive structure (program or schema), determines which, were and how features have to be "looked for" and found for succeeding more specific classifications.

 Now the basic assumption is that an identification procedure, started at the first photograph, will in certain conditions find a natural and logical continuation when a second photograph is introduced that spatially and meaningfully more or less combines with, or replaces, the earlier pattern. With respect to analogous movie projection one could think of a continuous registration of an event versus a shift of " shots".

Sometimes the data may change, but the identity of the object remains established and even shows further developments. Sometimes the data as well as the identity change.

 We roughly estimate the following stages in the identification process, which is started bv the input of a portrait-stimulus:

(a) Something round with specific features triggers general schema for faces.

(b) This schema appears to be appropriate and develops a general description of the portrait. A necessary and essential characteristic of this description is that it incorporates some spatial orientation of the face. lf this position description is adequate for the stimulus, it will become a basis for finding the essential features for a correct final identification.

(c) In the description is also a variety of features that triggers more specific schemata for the identification of some person. The most economical way at this moment would be that these features are still of a more general nature: for instance concerning age, gender, race, good looks, and the presence of spectacles.

(d) Finally such specific schemata are triggered that a familiar person eventually is identified. if the data fit these schemas.

If we choose the variables "Position", "Spectacles" and "Identity", we expect the corresponding perceptual steps to have this temporal order.

Method

Subjects

38 Ss, 8 females and 30 males, their ages ranging from 20 to 40 years, participated in the experiment. These were simply all available staff-members and doctoral-students of our department. They all were very familiar with the models on the photographs we describe in the next section.

Materials and design

We started by selecting six male "photo-models". All were staff-members of our department, about 35 years old, having no exceptional characteristics like very long hair, beards, birth-marks, etc.. Three of them typically wore spectacles. During the experiment these spectacles were identical. Their faces (in a neutral expression) were photographed in two standardised 3/4profile positions, in which the nose pointed to the right or to the left. These 12 photographs allowed 144 combinations of double portraits, that were arranged in the following 6 types. The reader may notice that the different types also have different numbers of pairs as shown in table 1.

Next we chose three SOAs: 0, 40 and 60 msec (see fig. 1) resulting in 3 X 144 = 432 trials. We added 12 trials in which the 12 photographs were to be presented one by one and arranged the resulting 444 trials in a random series.

Apparatus and procedure

The stimulus-series was recorded on 444 different tracks of an Ampex MD-400 video-disc, which could be monitored by a PDP-11 computer. The recordings were made in such a way that, if the two photographs were superposed on a display-scope, the frontal profiles in similar positions showed maximal overlap, while in the different positions the total external contours of the heads showed maximal overlap. A single video (European Standard, 50 Hz, 625lines) picture consists of odd and even lines. The lines are generated from top to bottom in a clustered way; the odd ones and the even ones. To generate

one complete (odd and even lines) video picture takes 40 msec.

 In case of two successive stimuli we acted on the basis of our impression that a relatively strong first and weak second stimulus produced clearer differences between

Table 1

-----------------------------------------------------------------------------------------------------------

 Same Different

 Position position

-----------------------------------------------------------------------------------------------------------

Same model 12 12

Different models, both wearing, or not wearing

Spectacles 24 24

Different models, one wearing, one not

Wearing, spectacles 36 36

----------------------------------------------------------------------------------------------------------

our experimental conditions than did equally strong stimuli. For this reason we made the first and second stimulus in successive presentations, respectively 40 and 20 msec. When we used pictures of 20 msec (see fig. 1), the picture consisted of either the odd or the even lines. In this way of course the second stimuli also have a lower resolution than the first stimuli and consequently less contour in formation. But though a favourable strong-weak relation of first and second portraits might function as a sort of catalyser, it cannot explain differentiation as such between our experimental conditions.

 In case of simultaneous (SOA = 0) presentations of the same stimulus pairs, both photographs were presented completely (40 msec, odd and even lines). However, to control brightness differences between simultaneous (for example the background would be much brighter now) and successive presentations, we only used half intensities for each of the photographs in case of simultaneous presentation. This electronic integration of the two portraits in case of simultaneous presentations might be different from the one accomplished by the eye in case of successive presentations. One can imagine a relative

emphasis on local similarities between both photographs because of a diffusion of local differences. However, again this situation is equal for all our experimental conditions. Eventually we may find a general difference between simultaneous and successive presentations (see also our discussion of the main effect of SOA). But neither these differences in stimulus quality, nor those between the first and the second stimulus, can explain the eventual differential effect of our experimental conditions with increasing SOA.

 As pre- and post-adaptation fields, and during the empty periods of the SOA, we used a square field of the same size as the photographs, which was illuminated at the same intensity as the backgrounds of the portraits.

 A dim light was arranged in such a position that the S could not see any reflections in the glass-scope. When the S pushed a starting-key the first trial of the series was presented.

Task

In the experiment the S was sitting in front of a display-scope. The visual angle of the faces on the screen was approximately 5 degrees. We did not use a fixation mark. Nevertheless most Ss will mostly more or less fixate the centre of the screen. But sometimes an S might prefer to look at the right or the left side. In the same way the S could try to attend only to the first or the second portrait. It should be noticed, however, that a crucial point of our experimentation is a complete balancing of stimulus-pairs

over all conditions. Together with a specific correction for guessing this has to prevent that any unintended strategy of our Ss can produce a Type I error. For this purpose we are prepared to pay the price of a larger risk of a Type II error, i.e. unjustly accepting the null hypothesis.

 When both faces are in similar positions the S seems to see only one face. When both faces are in different positions the S sees something like a Janus-face. If the S has identified one face, correct identification of the other face is, however, seldom much higher than about chance-level. Therefore we

instructed the S to identify only one model at a trial. This instruction does not offer many problems: the S simply identifies the face he or she has seen most clearly. After the S had signalled a choice to the computer by means of one out of six response-keys, which corresponded to the alphabetically arranged names of the six models, the starting-key could be used again for a new trial. Presentation of a trial started within 20 msec after the starting-key was pushed. Response-errors could be corrected by means of a correction-key, which was on the same response-panel as the starting-key and the response-keys.

Scoring and correction for guessing

To make sensible comparisons between different conditions we should normalise for unequal numbers of trials and correct for blind guessing as well as for guessing on the basis of only seeing the spectacles on the first and/or the second photographs. For this purpose we developed a specific "two-stepped" correction for guessing. From here on we shall simply call the corrected proportion of correct identifications of the first and/or the second photographs in a condition, respectively Pc1, Pc2 and Pc1 + Pc2.

Predictions

Because there will always be some masking as a result of superposition of the two faces, a crucial test will have to compare different conditions with respect to the relative shifts from the identification of the first photograph to the identification of the second photograph. If correct detection of position is indeed primary and essential for identification, there will be stronger initial shifts towards identifying the second photograph as a function of increasing SOA when positions of the two coincide than will be the case

when the two positions are different. Therefore our first prediction is that with increasing SOA there will be a specific interaction: Position X Recency X SOA.

 However, in Calis' (1974) earlier experimentation with a dichoptic stimulus presentation this interaction was actually destroyed because already with short SOAs the preference for the second photograph in case of similarity of position and spectacles increased about as much as it decreased in case of position similarity and spectacle dissimilarity. Thus if the first portrait has the same position as the second one but then belongs to a different set, it may very effectively put the perceiver off the scent. Of course this does not falsify the theory, because it also predicts that an identification process that has already established position as well as the presence or absence of spectacles on the basis of the first photograph, is now "looking" for specific individual features. But precisely these features cannot be found in the second photograph because this belongs to a different set. In this case the identification process would be derailed in such a late phase, that the critical iconic data are already too deficient to make a new round successful. Perhaps this effect was due to the dichoptic stimulus presentation, which might produce less

interference, or speed up the identification process in comparison with the binocular presentation of portraits we used in this experiment. For us it means that finally our second prediction might be more crucial. Once position is correctly established, the shift to the second photograph can be enhanced even more if there is also continuity with respect to the presence or absence of spectacles. This effect would be maximal of course if both photographs are completely identical. However, because we only have Pc1 + Pc2 scores in this case, we cannot directly see such effects in our data. In our analysis of variance,

which for this last reason only concerns the identifications with portraits of two different persons, this crucial test means that we predict a significant higher order interaction: Spectacles X Position X Recency X SOA.



Results and conclusion

The results are presented in fig. 2 and table 1. As can be seen from a comparison of predictions and results, there is significant confirmative evidence for our hypothesis. Hence we believe we have demonstrated some facts concerning the hierarchical-sequential nature of immediate perceptual recognition of familiar faces.