"VISAGES": A COMPUTER-BASED TEST OF FACE PRECOGNITION

MARIO VARVOGLIS[1] & MICHEL-ANGE AMORIM

LABORATOIRE DE RECHERCHE SUR LES INTERACTIONS PSI

A computer-based psi experiment was conducted to explore whether subjects could precognize the features of a randomly composed face. The experiment was based upon a subset of the "Photo fit" Kit used by police to help identify the facial characteristics of a missing person or a criminal. Forty subjects participated, each contributing a minimum of four runs (16 trials).

Subjects were presented with 4 target packs each containing 16 different instances of a particular facial feature (eyes, nose, mouth and facial-outline with hair). The instances for each element were grouped, so as to suggest different degrees of resemblance between them, and, hence, between the subject's choice and the target.

There were two task-modalities. In the Scanning psi task instances were arranged as a 4 x 4 image array, allowing the subject to consciously choose a particular image using the computer "mouse". In the Timing psi task, the images were presented in a rapidly shifting sequence; here the subject could only choose when to stop the "image roulette" with the mouse. Once the subject had chosen all elements of the face, the program randomly selected an instance for each of the four elements, constructed the target face, and presented it to the subject.

Results were evaluated through goodness-of-fit tests, comparing the obtained distribution of hits, for 5 different levels of scoring, against the expected distribution. The global test yielded a significant chi-square for the experimental condition ($p=.013$), and chance results for a simulation study. Further analyses, examining scoring under the two different task-modalities, yielded a significant chi-square for the Timing task modality alone ($p=.006$).

(i) Main author and experimenter

INTRODUCTION

The possibility of applied parapsychological research has been receiving considerable attention in recent years, both in the U.S. (Agor, 1984; Harary, Targ and White, 1985; Mishlove, 1986; Morris, 1986) and in Europe (Amorim, in press). An application which seems to hold particular promise is the use of psi to help locate missing persons or identify criminals. A number of popular or semi-popular accounts have referred to instances in which psychics helped the police, but little has been done by way of experimental research. One of the few systematic investigations in this area is reported by Reiser et al (1979) who presented 12 psychics with sealed envelopes containing information on two solved and two unsolved crimes. According to the authors, the elicited "psychic impressions" offered little support for the claim that psychics could contribute information necessary for the resolution of crimes. However, in their book "Psychic Criminology", Hibbard & Worring (1982) cite a number of cases resolved with the help of psychics, and criticize the Reiser et al approach as being insensitive to psychological and interpersonal factors. Osis (1984) also cites numerous cases resolved with the help of psychics, and emphasizes the difficulties involved in attempting to address this topic in laboratory contexts.

It is clear that the motivational characteristics of real-life situations cannot be reproduced in the artificiality of laboratory contexts. On the other hand, even if it is impossible to recreate the motivational dynamics of real-life psychic criminology, laboratory experimentation could explore certain facets of this area. One such facet is the identification of an individual. In many crimes, police rely upon eyewitnesses to try to reconstruct the facial characteristics of the criminal. However, witnesses may not be available, or may be unreliable. Can "psychic witnesses" be reliably used to identify the facial characteristics of an unknown person?

The exploration of facial characteristics as psi targets is also interesting in and of itself, independently of any immediate applications. Our perception of the face appears to be a very basic process in human beings; like language, it may constitute an inborn, "hardwired" function, rather than being an acquired capacity. Could the fact that we are "primed" toward face-recognition translate into a special sensitivity toward face -precognition or -clairvoyance? If experimental data were to indicate that faces constitute unusually good psi targets, then this would lend some credence to the idea that psi capacities are tied in to basic neurophysiological and cognitive functions.

The current study, then, was conceived as a preliminary step in exploring the use of faces as psi-targets. Specifically, we explored "face precognition" through a computer-based version of the "Photo-fit" Kit, employed by police to interrogate eyewitnesses, and explored in a number of investigations of face recall and recognition (e.g., Ellis, Sheperd & Davies, 1975; Sergent, 1984). This kit contains a very wide range of noses, eyes, mouths, jaws, etc., drawn on transparencies; it thus allows an interviewer to "mix and match" different instances of facial features, so as to approximate the face recalled by a witness.

In our study, we selected a subset of facial features (face-outline, eyes, nose, mouth) and a small subset of instances for each feature, and passed these into the computer. Then, we created a program which can randomly mix and match these instances, and compose a face. The subject's task was to attempt to choose the facial characteristics which would most approximate the features of the computer-chosen face.

Despite certain superficial similarities, however, this task was not quite analogous to psychic criminology. For one thing, we used "normal" (rather than special) subjects. Also, the psi task was "elementaristic" in nature, insofar as subjects would be focusing upon facial features, rather than attempting to precognize the face as a whole. Some researchers (e.g., Ellis, 1975; Sergent, 1984) suggest that important facets of face perception are holistic, and are not captured by elementaristic approaches. In order to compensate somewhat for this problem, we decided not to provide "piecemeal" feedback following each trial, but give feedback only once the entire face has been composed (i.e., after all four elements have been chosen). Though subjects would still make their selections one feature at a time, at least the moment of feedback would involve a holistic perception; if precognitive information derives from this feedback point, then it would orient the person's psi toward the whole face, rather than an isolated feature.

A more important deviation from psychic criminology, in our approach, was that the experimental context included none of the human elements which lend meaning and significance to the task in real life. Rather, it involved guessing the features of a fictional face, one stripped of any meaningful descriptors or history. To address this, we sought to give the fictional target-face some identity, associating it with a randomly selected name and biography; these were derived from a large pool of possibilities. The relevance of this "meaningfulness" device was to be explored by comparing subject scoring with the biography present vs. absent.

Another factor explored, "psi - task modality", was meant to address the potential problem of response biases. There is little doubt that, to different degrees, we are attracted or repelled by different faces (or facial characteristics). In a psi task in which subjects can freely choose from among all possibilities within a target pack, such aesthetic factors could easily drown out subtle psi information, leading people to choose images they like and avoid those they dislike. As it seemed that this could not be completely avoided, as long as the subject is free to choose among the possibilities, we decided to add a psi-task modality in which the subject could not consciously compose an image.

This second modality was a "timing" psi task, demanding of the the subject only a decision as to when to stop a rapidly changing "image roulette" containing all possibilities. Thus, there were two task-modalities: one based upon the implicit question "when is the target passing by" (the timing task), the other based upon the question "where is the target", and involving the usual scanning of possibilities in order to make a choice (the scanning task).

METHOD

Subjects

The subjects of this study were 35 female and 5 male volunteers, ranging in age from 19 to 59 years old. Thirty four of these participants came to the laboratory following an article in a popular woman's magazine, which presented the laboratory's computer-based psi research. The remaining 6 subjects were either acquaintances, or had heard about the laboratory through acquaintances. Personal and psychological data on all subjects were collected using french versions of the Personal Inventory Form (PIF) and the Myers-Briggs-Type-Inventory (MBTI); these data have not yet been analyzed.

Hardware

The experiment was run using an Amiga 1000 with a color monitor, two disk-drives, a 2-megabyte random-access memory extension, and a "mouse" for subject inputs. The transfer of Photo-fit images into the computer was accomplished using a surveillance camera and an interface which permits the "digitization" of video inputs.

Software

The program controlling the present experiment was based upon a compiler-language named "The Director", similar to BASIC, but explicitly oriented toward graphics- and sound-manipulations.

Pseudo-Random function: The random numbers for the program are generated by the Director language's pseudo-random function, reseeded every cycle by the Amiga clock (read in in micro-seconds). A "Cyclic Redundancy Check" scheme scrambles the clock values and ensures the adequacy of the random distribution. In a personal communication, the creator of the Director language stated that tests of the random function have shown that it yields the expected range and frequency of values. While no detailed assessment of the random function was undertaken by the experimenter, a one-line program was written to at least ensure that the function was reseeded each time. Run immediatly after the "booting" of the computer, this program served to verify that the pseudo-random function was indeed being reseeded, yielding different number sequences each time it was run.

"Visages" program: The Visages precognition test, written by the first author, presents subjects with 4 graphic target packs, each containing 16 distinct instances of a facial element, and, on the basis of the subject's choices, progressively constructs a graphic face. Then, once the subject is satisfied with the face as constructed, the program uses the Amiga's pseudo-random function four times, selecting, for each facial element, one of 16 possible instances. Finally, the program calculates feedback scores (i.e., measures of the proximity between the subject-chosen and the randomly-chosen elements), stores the results, provides feedback (showing the target-face and the score), and offers the subject options to continue or quit.

A slightly modified version of the program serves to collect control or "simulation" trials, in which no subject is present. The program essentially creates two faces, on the basis of two sets of random numbers; the first set substitutes for the subject's guesses, while the second defines the target face as described above.

A more detailed description of the program's operation is given in the Target-preparation and Procedure sections.

Target-preparation

The Target pool was based upon a portion of the Penry Photo-fit Kit, kindly provided by the central police department of Paris (Ministere de l'interieur), in photocopy form. The kit involves transparencies showing different male facial elements (eyes, noses, mouths, etc.); these can be freely combined and mixed, and so as to produce a very wide range of possible male facial types.

Four facial elements were used for this study: eyes, nose, mouth, and facial outline (showing hair, forehead, and jaw). To select from among the many instances provided, we used our subjective judgement and several criteria; for example, selection of as wide a range of characteristics as possible, for each facial element and avoidance of facial characteristics which are too striking or weird. We then passed this subset of photo-fit images into the computer through a "digitization" process, and each digitized image was treated with diverse computer graphic tools, so as to maximize definition and clarity. Then, for each element, we selected 16 different instances (i.e., sixteen noses, sixteen mouths, etc.), and arranged these images into 4 computer bit-map screens or "pages", which would serve as target packs (Two of these pages are illustrated in Figures 1 and 2).

The 16 instances of each page were arranged in a 4 x 4 array, images being grouped according to different levels of resemblance between them. Taking Figure 1 as an example, we see that the top two rows are distinguishable from the bottom two ("little hair" vs. "lots of hair"). Then, the 4 instances of a facial element in a [illegible] distinct from those [illegible] row (near-bald, short hair,

Figure 1. Target pack for face-outline

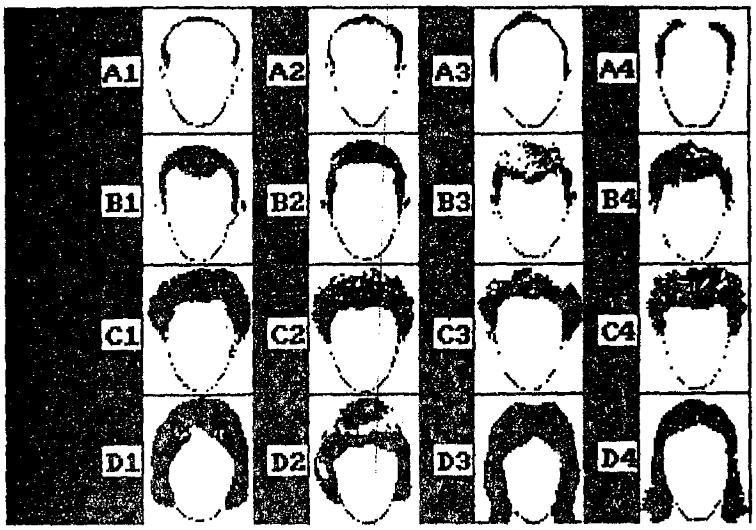


Figure 2. Target pack for Lips

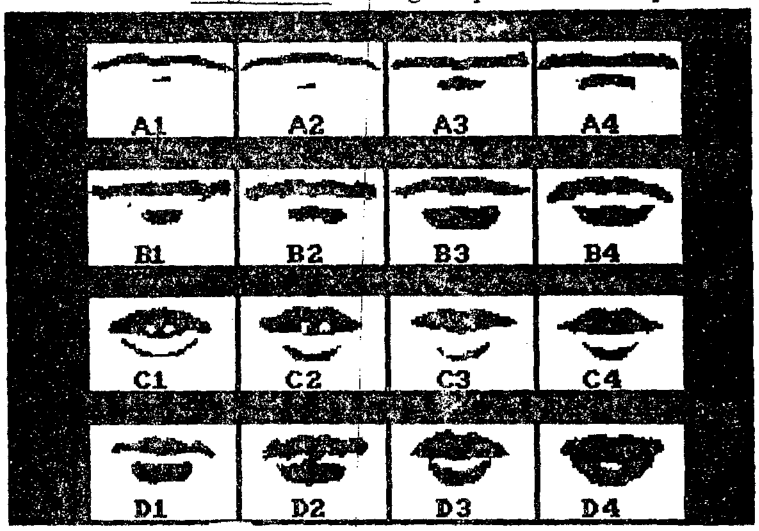


Figure 3. Face with three elements selected

full hair and long-hair). Finally, within each row, 2 groups are distinguishable (e.g., in row D, D1 / D2 and D3 / D4).

The idea behind this arrangment was to create a psi task which could allow for different degrees of psi -accuracy or -resolution - from vague feelings to detailed information. The scoring scheme, accordingly, was meant to reflect different degrees of resemblance between subjects' choices and the target image. For example, let us assume that the target for facial-outline were D2. Selection of any instance within row C - the other row of the same half-page - implies having correctly identified that the target-face generally has "lots of hair"; this would be a "half-page" hit. Selecting D3 or D4 - the other pair on the same row, or a "row" hit - implies having identified the target face as having specifically long hair. Selecting D1, the other member of the pair, would be a "pair" hit - whereby the subject has found the instance which most resembles the actual target. Selecting D2, of course, is a direct hit.

As mentioned in the Introduction, the target face was accompanied by a name and, in half the trials, a biography. The names were drawn from a file containing 80 names commonly found in France. The biography was drawn from a second file, containing 200 statements, organized into 10 theme-related groups (sports and leisure, living quarters, childhood and education, mood and temperament, social life, paranormal experiences, reactions to world events, beliefs and philosophy, favorite sayings, health).

Procedure

Upon arrival at the laboratory and preliminary exchanges, the subject was placed in front of the Amiga, and instructed on the utilisation of the mouse. The subject then took computer-based (French) versions of the PRL Personal Inventory Form (PIF) and the Myers-Briggs-Type-Inventory (MBTI). Following feedback on the MBTI, the subject was switched to the Apple-based computer-RNG test "Volition". Then, after a minimum of two Volition runs, the subject was brought back to the Amiga, for the Visages precognition test; the experimenter remained present throughout the Visages session.

The subject was told that, unlike Volition, the Visages test was geared toward receptive psi. It was explained that the computer would create a face, randomly selecting instances for the four facial elements; the person was asked to use their intuition to guess which instances of each element would be selected by the computer. It was emphasized that the computer would not select those instances on the basis of any aesthetic criteria, but on the basis of random decisions.

The run, consisting of four trials (one for each facial element), begins with the presentation of a Menu on the monitor screen, naming the four elements as "Hair", "Eyes", "Nose", "Lips". The program awaits the subject's selection of one of these, using the mouse. (For the first run, the experimenter encouraged the subject to start with face-outline, and progressively fill in the other elements of the face). Once an element is selected, the computer presents the subject with the target pack, i.e., the 16 instances of that element.

Depending on the psi-task modality, the target-pack is presented in one of two different ways. In the scanning condition, all 16 possibilities are present on the screen simultaneously, arranged in the 4x4 array described above; the person uses the mouse to place the cursor over one of these 16 instances and then "clicks" to select it. In the timing condition, only one of the 16 instances is visible on the screeen at any moment; the images succeed each other very rapidly in a random sequence (giving the impression of a nose changing shape, a mouth talking, etc.), and selection is made by clicking on the mouse and stopping the "image roulette" at some particular image. The image actually selected, however, is not the one last seen by the subject, but rather one which is randomly generated just after mouse input; irrespective of how fast their reaction time might be, subjects cannot consciously select a particular target.

The order of task presentation, fixed across subjects, was based upon a predetermined schedule allowing for different permutations of the biograhpy and task-modality variables. The first four runs were scanning/biography, scanning/no biography, timing/biography, timing/no biography.

In both scanning and timing modes, the specific instance chosen by the person is immediately added to those previously selected. Thus, as subjects proceed through the four facial elements and select a particular face-outline, set of eyes, nose, and mouth, they see the face being constructed. (Figure 3 illustrates a face with three features already chosen and lips not yet selected). The process of face construction is automatic: placement of the feature chosen on the face depends not upon the subject, but upon predefined coordinates.

Following the subject's selection of all four elements, and thus the completion of the face, the individual is presented with options 5:"Review Face", and 6:"See target". Option 5 allows subjects to review the face constructed, in case they've changed their mind about a particular selection (in which case, they can re-initiate the selection process by clicking on the corresponding number in the Menu).

Option number 6, once clicked, launches the construction of the target face. The program generates four random numbers, between 1 and 16, each corresponding to a particular instance of the four features. The program also randomly selects a name out of the name-file, and, in the "biography" condition, constructs a biography by randomly selecting 6 statements from the 20 categories of statements. The program then stores all results on a disk file, and finally projects the target face on the screen, along with a name, a graphic "button" for re-viewing the subject-chosen face, and another button for reading the biography (*).

The screen with the subject-chosen face allows for comparisons with the target-face; it also shows the scores obtained for each of the four elements. These scores give subjects a numerical estimate of the proximity of their choices to the target-instances. For each element, the possible scores are 0 (no relation between target and choice), 2 (half-page success), 4 (row success), 8 (pair success) and 16 (direct hit). Thus, the total score for the run could range from 0 to a very unlikely 64 (direct hits on every trial).

Subjects were asked to complete at least four runs (sixteen trials), but were allowed to contribute additional runs, if so desired. Thus, following feedback they could either click on a Replay button, to initiate a new run, or, if they had completed 4 runs, click on a Stop button to close the Visages program and end the session.

Simulation Runs: In order to ensure that the RND function of the Amiga operates correctly, and that there were no problems in the program's logic, we conducted a simulation study, based upon a slightly modified version of the Visages program. In this progam, the subject's scanning or timing guesses for each element were replaced by the generation of random numbers between 1-16. Thus, the program would construct a face on the basis of 4 random numbers, and then a second, target-face on the basis of 4 more random numbers.

Once launched, the simulation program ran automatically, until it completed 9 runs; it was then re-launched by the experimenter. This process continued until the number of runs accumulated equalled the total of experimental runs.

----

* The screen with the biography text was intended to examine the meaningfulness factor mentioned in the Introduction. From the first few sessions, subjects appeared to be confused as to the role and purpose of the statements; the biography seemed incongruent with the stated nature of the task-precognizing a randomly constructed face. Following repeated negative comments by several subjects, the experimenter realized that the biography was not appropriate for assessing meaningfulness, and decided to drop assessment of this factor from the study. From that point on, he no longer directed subjects to click on the biography button, and practically no one did.

RESULTS

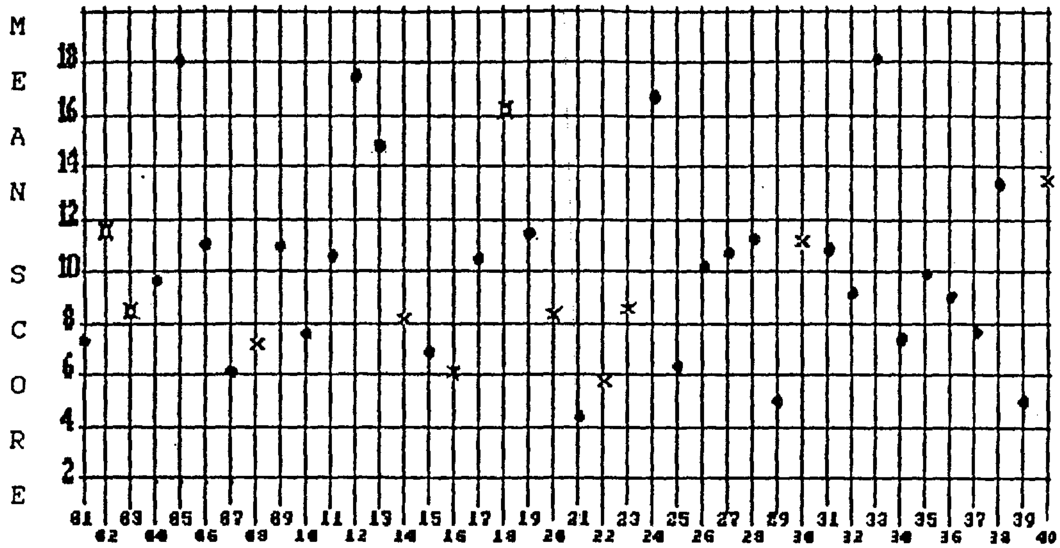
Collectivly, the 40 participants contributed a total of 212 experimental runs (848 trials). Individuals' contribution to this database was quite uneven: 28 of the 40 participants completed just the minimum of 4 runs each, while the remaining 12 contributed between 5-14 runs. Using subjects' mean feedback score as an index of individual performance, we find that the average score for the group contributing 4 runs is 10.16, while for the group contributing more runs it is 9.18. A t-test for independent means shows no difference between the two groups (t=.752, 38 df, ns). Figure 4, depicting mean feedback scores for all subjects, also shows that there are no consistent trends distinguishing the scores of the 28 subjects who contributed exactly 4 runs, from the 9 contributing 5-9 runs, and the 3 contributing 10-14 runs.

The evaluation of overall results, utilizing the trial as unit, was based upon two goodness-of-fit tests - one for experimental and one for simulation data. These analyses examine whether the observed distribution of hits, for all scoring levels, conforms to the binomial expectation (the probability corresponding to each scoring level multiplied by the number of trials). The probabilities used to estimate expectation for each scoring level represent the likelihood of obtaining exactly (rather than "at least") a pair hit, a row hit, etc.; they thus allow each scoring level to be treated independently. The probabilities corresponding to each level of hitting are direct hit, 1/16; pair hit, 1/16; row hit, 1/8; half-page hit, 1/4; and miss, 1/2. (For example, in the facial-outline example cited earlier, with D2 as target, there is exactly 1 way to obtain a direct hit, 1 way to obtain specifically a pair hit (D1), 2 ways to obtain a row hit (D3, D4), 4 possibilities for a half-page hit (all of row C) and 8 ways to obtain a miss (rows A and B)).

Table 1 summarizes the results of the goodness-of-fit tests. The first row represents the expected number of hits for each scoring-level, given a total of 848 trials. The second and third rows show the obtained number of hits for simulation and experimental trials (respectively). As can be seen from this table, simulation trials conformed quite closely to expectation. In contrast, the distribution of scores in experimental trials departs significantly from expectation (chi-sq [4 df] = 12.632; p=.013). This latter result is associated with an effect size of .076 (obtained by converting the p-value to a one-tailed z-score, and dividing the latter by the square root of N, i.e., of 848).

The significant effect for the experimental trials was mainly due to a shift from the expected number of hits in the three partial-hit levels (pair, row and half-page). Post-hoc chi-square analyses, comparing each of the five hitting levels with the other four, suggest that the main effect was due to a shortage of hits at the pair-hits level (chi-sq [1 df]=5.[illegible], p=.0[illegible]) and an excess of hits at the half-page level (chi-sq [1 df]=4.25, p=.039). Neither of

FIGURE 4: MEAN FEEDBACK SCORES FOR 40 SUBJECTS



MEAN SCORE BASED UPON: 4 RUNS: ● / 5-9 RUNS: ✕ / 10-14 RUNS: ⌘

Table 1: Frequency of hits for 5 scoring levels for Experimental and Simulated trials

	DIR	PAIR	ROW	H.PGE	MISS	CHI-SQ (4 DF)
EXPECTED	53	53	106	212	424	
SIMULATION	50	56	112	213	417	.799
EXPERIMENTAL	60	36	88	238	426	12.632

Table 2: Frequency of hits for 5 scoring levels for Scanning and Timing task modalities

	DIR	PAIR	ROW	H.PGE	MISS	CHI-SQ (4 DF)
EXPECTED	26.5	26.5	53	106	212	
SCANNING	33	20	47	104	220	4.207
TIMING	27	16	41	134	206	14.453

these values remain significant when corrected for multiple analysis (i.e., by multiplying each p-value by 5).

Table 2 examines the experimental results for scanning vs. timing psi tasks separately (424 trials each). For the scanning task, the chi-square was non-significant (chi-sq °4 df§ = 4.207, n.s.) For the timing task, the result is significant (chi-sq [4 df] =14.453; p=.006).

Post-hoc chi-square analyses, comparing each of the five hitting levels with the other four, suggest that the effect in the timing-task condition was largely due to an excess of hits at the half-page level (chi-sq [1 df]=9.861, p=.0017). This value remains significant even when corrected for multiple analysis.

DISCUSSION

As indicated in the Results, whereas the chi-square for the simulation trials conformed to expectation, the chi-square for experimental trials was significant. The overall chi-square analysis thus suggests a relationship between subjects' guesses, and the targets which were randomly selected following their guesses.

As mentioned, subjects had the option to stop after a minimum of four runs, or continue. This option had been introduced because pilot sessions had suggested that some subjects tired quickly of Visages, whereas others liked it. As it turned out, only 12 of the 40 subjects contributed more than the required 4 runs. It might therefore be suspected that it was the few subjects who scored well that kept on going; this, of course, would detract from the generalizability of the results. However, as shown earlier, the mean scores for those who stopped after four runs was not lower than those who continued; if anything, they were slightly higher. Overall results cannot be attributed to the scoring of a few subjects who contributed large amounts of data.

What does seem clear is that the overall significant results were largely due to the timing condition runs. When the data were broken down in terms of psi-task modality, we found that the distribution of scores in the scanning condition did not depart significantly from chance, whereas the result for the timing task was significant.

The effect observed in this study thus appears to be associated with the relatively effortless and game-like task-modality rather than with the one obliging subjects to consciously choose the elements of the face. The lack of results in the scanning condition may well reflect the operation of response biases, and subjects' frustration in having to fight their feelings during the task. During the scanning condition, participants repeatedly complained about difficulties in discriminating between their intuition and [illegible] reaction to a particular instance. Indeed, though instructed to try to avoid choosing [illegible]s of