

Parsing eye movement analysis of scanpaths of naïve viewers of art: How do we differentiate art from non-art pictures?

Wolfgang H. Zangemeister *

Claudio Privitera

Neuroophthalmology Unit,
Department of Neurology,
University Hamburg, FRG

Dept. of Physiological Optics
and Optometry, UC Berkeley,
CA 94720, USA

Relating to G. Buswell's early work we posed the questions: How do art-naïve people look at pairs of artful pictures and similarly looking snapshots? Does the analysis of their eye movement recordings reveal a difference in their perception? Parsing eye scanpaths using string editing, similarity coefficients can be sorted out and represented for the two measures 'Sp' (Similarities of position) and 'Ss' (Similarities of sequences). 25 picture pairs were shown 5 times to 7 subjects with no specific task, who were 'art-naïve' to avoid confounding of the results through specific art knowledge of the subjects. A significant difference between scanpaths of artful pictures compared to snapshots was not found in our subjects' repeated viewing sessions. Auto-similarity (same subject viewing the same picture) and cross-similarity (different subjects viewing the same picture) significantly demonstrated this result, for *sequences* of eye fixations (Ss) as well as their *positions* (Sp): In case of global (different subjects and different pairs) sequential similarity Ss we found that about 84 percent of the picture pairs were viewed with very low similarity, in quasi random mode within the range of random values. Only in 4 out of 25 artful-picture snapshot pairs was a high similarity found. A specific restricted set of representative regions in the internal cognitive model of the picture is essential for the brain to perceive and eventually recognize the picture: This representative set is quite similar for different subjects and different picture pairs independently of their art-non art features that were in most cases not recognized by our subjects. Furthermore our study shows that the distinction of art versus non-art has vanished, causing confusion about the ratio of signal and noise in the communication between artists and viewers of art.

Keywords: scanpath, parsing eye movements, art versus non-art pictures, string editing

* **This paper is dedicated to Lawrence W. Stark (1926 – 2004)**

Introduction

Art versus non art

There have been many more or less influential definitions of art throughout the history of art and aesthetics. The distinction between art and non art goes back to Aristotle, who clearly defined it in terms of rhetoric and the role of rhetoric in an argument. It is “a constructed use of factual material”, where the construction is a work of art. However, with the advent of ‘modern art’ at the beginning of the 20th century definitions of art became abundant. Since then there has been no clear consensus about “what is art?” since some artists insisted that “everything is art!” (Vautier, 1972). So it seems obvious to ask with T. Avital: “Is modern art- art at all?” He argues that modern art has thrived on a state of total confusion existing between art and pseudo art and the inability of many to distinguish between these two extremes (Avital, 2007). Art critic A. Danto, however, stresses that “in an age of pluralism in art, when anything might be a work of art (though not everything is), we need a pluralistic critic, willing to see anything as art” (Danto, 1994, 2003).

Postmodernist philosophers (Welsch, 1996) go further by analyzing the postmodern situation from a trans-disciplinary point of view, combining humanities and sciences, as proposed more recently by ourselves (Zangemeister & Stark, 2007). We believe that approaching art this way can lead to a fruitful discussion between “The Two Cultures” (Snow, 1959), resulting in a new way of questioning the very definition of art with respect to aesthetics in its true ambivalent nature (greek: “aesthesis”: the senses; perception).

The present study is the first using the scanpath theory to investigate the underlying eye movement mechanisms of art-naïve subjects’ art perception when viewing pairs of artful pictures and snapshots while not being aware of the pictures background (artfulness or not).

The artist’s painting begins with his model of related objects within a frame, that carry a story either classically in representative art or a story in terms of a particular artistic technology or in terms of certain phases of abstract art, or a combination of these. The process of ‘art-production’ followed by the ‘art- perception’ by a third person is related to Shannon’s Theory of Communication (Shannon & Weaver, 1949) in that it describes – besides the generally high content of information- the inevitable noise created during the phases of communication in this process. Therefore, a suitable signal / noise ratio is need-

ed for the experimental paradigm to be meaningful, i.e. the difference between ‘art’ and ‘non-art’ needs to be beyond noise level.

Questions

Since it would make little sense to attempt to answer questions such as ‘what is art’ in a single study, we came up with some specific questions that can be tackled using the current theories and experimental tools. At first some terms need to be defined: Snapshot, commonly understood as a photographic snapshot, can be generalized as an unselected collection of objects that may not or may be closely (possibly also a selected collection in that case) related. This collection of objects can be modelled with a cognitive schema that is sufficient to drive a scanpath which checks the model to gain more detailed sensory information. Picture originally meant artistic, artful picture, i.e. a painting or a drawing, but it has been generalized to include all levels of picture quality, not negating the artistic component of photography.

Previous findings have shown that artists and art-sophisticated viewers look quite differently at artful pictures than art-naïve viewers (Locher & Nodine, 1987; Locher, 1996; Zangemeister, Sherman & Stark, 1995; Ramachandran & Hirstein, 1999). Similar, but improved methods are used in this paper to search for a cognitive identification of ‘pictorial art/ artful pictures’: Is it the case that some participants of the study indeed distinguish/differentiate paintings (artful pictures) and their assimilations (snapshots) by means of different eye movement patterns?

Given a thematic and geometric similarity (though not in colour) of a pair consisting of a snapshot and an artful picture: Are they able to distinguish between these two, or are both equal contributors to the same spatial cognitive model of these art-naïve viewers? We can divide this question into two sub questions:

- (1) Do naïve subjects perceive a snapshot in a different manner than they perceive an artful picture or is there no difference in perception and thus a high similarity between the spatial and sequential scanpath regions of interests?
- (2) Is the global similarity during scanning of all image-pairs in all subjects low i.e. close to random, or is there a high similarity of scanpaths when artful pictures were viewed, but not in viewing snapshots?

In a second part, we discuss in general the difficulty to distinguish artful pictures from non-artful pictures (snapshots), e.g. signifying or measuring “artfulness”, as since

the advent of the so called modern art there is no compulsory model of art and artfulness any more around – like it existed in earlier times. The lack of an agreed-upon definition of 'art' i.e. the lack of a compulsory cognitive model of art leaves naïve as well as sophisticated art viewers with the problem of arbitrariness. In this situation it might be helpful to apply the theory of information by C. Shannon on to images that may or may not be art – distinguishing information from noise.

Methods

Procedure

Subjects

7 Subjects (4 female and 3 male adults), 28 years on average, with normal eye sight, were tested. 25 different images were used, including terrain photographs, landscapes, and paintings. We also used image modifications of some of these stimuli, such as the embossed effect or binary thresholding. To avoid distortion of the results, *no specific instructions were given*. Subjects were only told that they participated at a pupil recording session. If subjects had known about the aim of the experiment, they might have *consciously* viewed the art pictures in a different way due to pre-existent conceptual knowledge about 'art and non art'. The left right arrangement of the *picture pairs* (see Fig.1a for an example) was randomly combined.



Figures 1a: Example of a snapshot/ picture (Vermeer) pair

All subjects had previously seen each picture at least once since unfamiliarity with the viewed images may affect eye movement patterns and it might correspondingly bias the results for some subjects (Zangemeister, Sherman, & Stark, 1995). Since all observers had some

degree of familiarity with the pictures and since no specific tasks were provided, each observer looked at the pictures using intuitive and natural internal cognitive models. Each subject was asked to repeat the experiments within a few days for a total of five viewing sessions over approximately two weeks. By comparing different viewing sessions, we could study consistency in the way each subject looked at specific visual stimuli. During each experimental run, the complete sequence of images, each time in different order, was displayed to the subjects. After the last session, each subject was asked to describe in which pair out of 25 they believed to recognize an artful picture. As they knew the pictures by then relatively well, they were given four seconds for this decision. On average in 14% they noticed correctly an artful picture. This was only an additional piece of information that we gathered to make sure that our subjects were indeed naïve and unsophisticated viewers of the artful pictures presented to them. We did not ask them for any explanation of the “artfulness” that signified an artful picture though.

In this study, we did not perform a comparison of the pair-similarities through a context free algorithm for defining visual regions-of-interests, although it would have been desirable to measure the similarity between the artful pictures and their snapshot pendants on the basis of ROIs before determining similarities between the eye movements of their viewers: Some of these pairs may be more similar and may lead to more similar eye movements than others. Privitera & Stark (2000) have investigated and developed a methodology that serves to automatically identify a subset of aROIs (algorithmically detected ROIs) using different Image Processing Algorithms, IPAs, and appropriate clustering procedures.

Stimulus presentation and eye movement measurement

Computer controlled experiments presented pictures and carefully measured eye movements using high resolution infrared eye movement devices described in (Stark & Choi, in: Visual Attention and Cognition: Zangemeister, Stiehl & Freksa (eds., 1996.). An infrared source light was projected toward the eyes of the subject, generating a bright Purkinje reflection on the cornea, reflection that was easy to track by a video camera and the eye-tracking server. The subject was instructed to watch the visual stimuli (for 4 seconds, plus a calibration period before and after data acquisition) on a computer screen which was socket-connected to the eye tracking server. The subject was seated in front of the screen with his head

secured onto an optometric chin-rest structure. The viewing distance was approximately 40 cm from the computer screen; stimulus size was an average of 15 cm x 20 cm, yielding a subtended visual angle of approximately 21 to 29 degrees, and the resulting accuracy of the eye-position recording system was of the order of one-half to one degree of visual angle. A fixation analysis algorithm was then applied to the eye movement data to distinguish rapid saccade jumps.

Theoretical Basis

Eye movements are an essential part of human vision because they must carry the fovea and, consequently, the visual attention to each part of an image to be fixated upon and processed with high resolution. An average of three eye fixations per second generally occurs during active looking; these eye fixations are intercalated by rapid eye jumps, called saccades, during which vision is suppressed. Only a small set of eye fixations i.e. human detected regions of interest (hROI), are usually required by the brain to recognize a complex visual input.

The scanpath was defined on the basis of experimental findings. It consists of sequences of alternating saccades and fixations that repeat themselves when a subject is viewing a picture. Only 10 percent of the scanpath duration is taken up by the saccadic eye movements, which thus provide an efficient mechanism for examining the scene or regions of interest. Hence 90 percent of the total viewing consists of intervening fixations or 'foveations' onto human regions of interest (Bahill & Stark, 1979). Through eye movements, i.e. glimpses or fixations, the fovea is moved to place the high resolution fovea on the hROIs. Low resolution peripheral vision completes the mental image. Scanpath sequences appear spontaneously without special instructions to subjects and were discovered to be repetitive. This repetitiveness made Noton and Stark suggest that a top-down internal cognitive model controls perception and active looking of eye movements in a repetitive sequential set of saccades and fixations over features of a scene to check out and confirm the model (Noton & Stark, 1971; Stark & Choi, Zangemeister, et al., 1996). Other evidence comes from studies of eye movements during visual imagery experiments (Brandt & Stark, 1997; Krischer & Zangemeister, 2007; Gbadamosi & Zangemeister, 2001; Zangemeister & Liman, 2007; Liman & Zangemeister, 2012) and ambiguous figures (Ellis & Stark, 1978; Leopold & Logothetis, 1999).

The scanpath theory outlines how a top-down spatial-cognitive model can control active eye movements (EM) and visual perception. The scanpath sequence consists of alternating saccadic EM and fixations

that enable the active looking paradigm. The controlling top-down model can succeed using iconic matching to physical signals arriving at the brain via peripheral nerves and sensory organs. Early experiments by Buswell (1935), Brandt (1940), Yarbus (1967), Noton & Stark (1971a,b) showed the sequential and repetitive character of the scanpath and its idiosyncratic nature with respect to the person viewing and the picture or scene viewed. These experiments suggested the reality of the scanpath EM sequence for several kinds of static pictures like those of Yarbus (1967) which showed evidence of the repetitive sequences now called the *scanpath*. Most scenes are *dynamic*, containing moving objects or as in movies; therefore snapshots depict very often dynamic scenes – "stills".

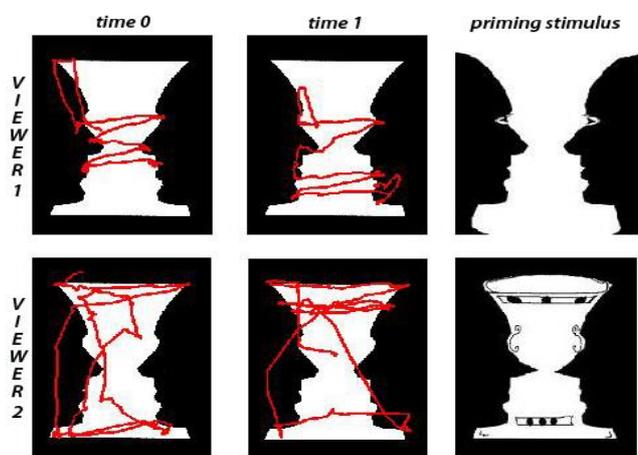


Fig. 1b: 'Two Faces or a Vase' (right hand side: priming stimulus)

Fig. 1c: Two scanpath sequences at time 0 and 1 (left hand side, viewer 1 upper, viewer 2 lower)

Thus a test of the scanpath theory would be to ask whether EM, while looking at such dynamic scenes, could be similarly characterized as a scanpath sequence (Blackmon & Stark 1999; Stark et al. 2001). Ambiguous and fragmented or hidden figures shift in their visual perceptions, so do the scanpaths traced over the constant physical picture (Fig.1c). Thus they appear to be generated from an internal model or schema rather than being controlled by external visual world signals impinging upon the brain. It has been known for some time that the implicit or explicit task setting in which the subject is immersed can strongly modify the scanpath. Thus as a subject continually looks at the scene she may change her point of view, think of different tasks and modify the scanpath. Examples of two such EM scanpaths are shown for the classical ambiguous figure, Two Faces or a Vase

(Fig.1b). Depending upon the TD internal cognitive model, the subject sees one or the other of these two interpretations.

Some control over the current interpretation can be induced by priming the subject with a non ambiguous distortion of the ambiguous figure, Fig. 1b, rightmost column. Subjects are not able to “see” both variants of the ambiguous figure simultaneously. However, there are electrophysiological transitions (Leopold & Logothetis, 1999).

Scanpath recordings present a range from verbal descriptions to qualitative comparisons to quantitative measures available to statistical analysis. How can two scanpaths be quantitatively compared as to their loci of fixations? This has been done using a position similarity index (Sp), a Euclidean distance or a binary measure dependent upon a typical clustering of fixations about a ROI. To compare sequencing of these fixations with respect to how similar strings of fixations are, a *sequential similarity index* (Ss) can be used. To measure the Ss a string editing algorithm is used, where three basic cases are possible as depicted in figure 2 (Fig.2): Similarity index of 1 with identical strings (Sp=Ss); Similarity index of 0 (Sp=0; Ss=0) completely different strings; similarity index between one and zero (Sp=1, Ss=0 or Sp=0, Ss=1).

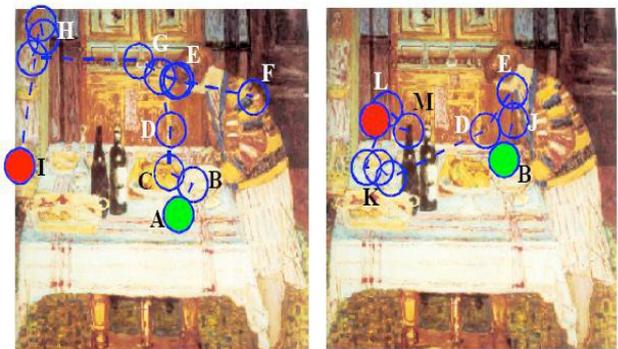


Figure 3a: Examples of two scanpaths and their ROIs.

A sequence of eye fixations can be represented by a string of letters (Figure 3a) with each letter corresponding to a different region of interest. In the two examples reported here, we have on the left a total of thirteen fixations that can be condensed and represented by the string ABCDEFEGHI and, on the right, a total of ten fixations represented by the string BJEDKLML. Fixation strings are then used by two indices Sp and Ss for comparing the spatial distribution and ordering of different viewing session scanpaths. Modified from: Bonnard’s representation of the perception of substance (Privitera, Stark & Zangemeister, 2007).

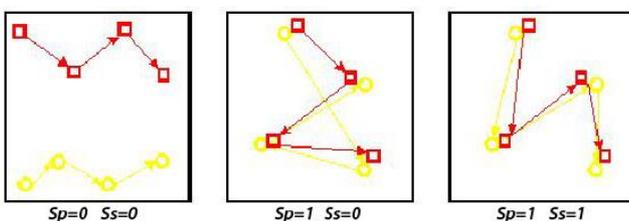


Figure 2: The three basic possibilities of similarity as given by Sp and Ss.

Random similarities are used so as to enable statistical tests of the results. Methodological results are detailed as to the various aspects of the scanpath recordings from EM data acquisition, to fixation identification, to analysis of results and presentation as parsing diagrams.

A generalization of the Levenshtein distance (Damerau-Levenshtein distance, Wagner & Fischer, 1974) allows the transposition of two characters as an operation. It is often used in applications that need to determine how similar, or different, two strings are, such as spell checkers.

For example, the Levenshtein distance between "kitten" and "sitting" is 3, since the following three edits change one into the other:

- kitten → sitten (substitution of 's' for 'k')
- sitten → sittin (substitution of 'i' for 'e')
- sittin → sitting (insert 'g' at the end)

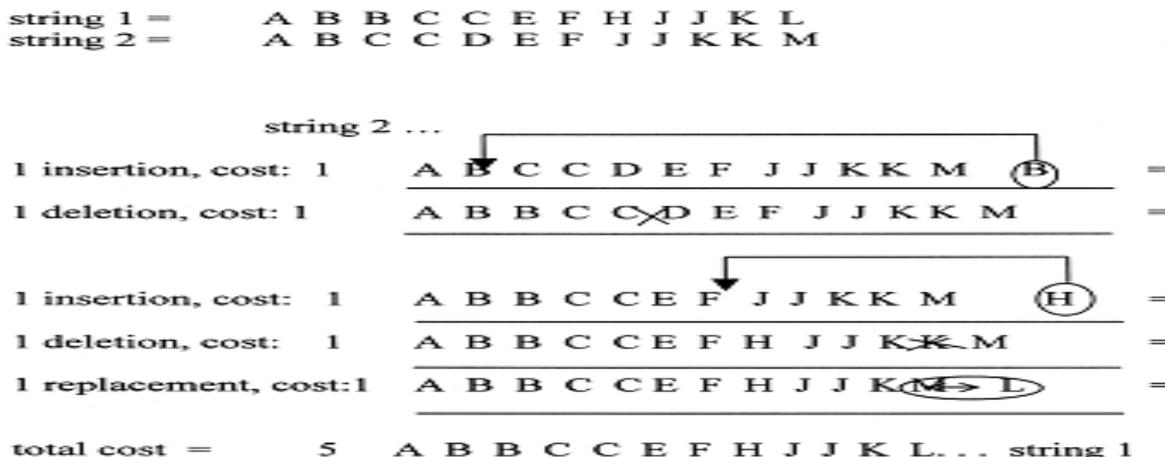


Fig. 3b: Minimum necessary changes (total cost) to change string 2 into string 1.

For the distance measure to be meaningful, the minimum number of edits counts: the distance is three, as the following three edits are *necessary* to change one string into the other. Of course, it is always possible to obtain the same result with more edits, but this would be arbitrary.

*Analyses of the Vectors of Looking:
String Editing and Parsing Diagrams of sensory cognitive motor operations*

Parsing a *sensory cognitive motor* operation like the eye scanpath into components, means characterizing the parallel and the serial nature of this flow; understanding what each process ultimately contributes to the specific response, are fundamental questions in cognitive neuroscience. EM during active looking is a very complex mechanism controlled by an internal cognitive, top-down representation which can be in general exemplified by a Bayesian inference framework (Privitera & Stark 2000, 2003, 2005). High cognitive representations depend on visual particularities to support the overall visual perception process, confirming and correcting the cognitive spatial model. Thus, degree of visual information, top-down, particularities and bottom-up, are intimately interconnected.

The string editing operations of the recorded scanpaths follow Levenshtein (1966). In information theory and computer science, the Levenshtein distance (LD) is a metric for measuring the amount of difference between two sequences (the so called edit distance). The LD between two strings is given by the minimum number

of operations needed to transform one string into the other. Possible operations of transformation are insertion, deletion, or substitution of a single character.

a)

$$S_{SE}^{ab} = 1 - \frac{D_{SE}^{ab}}{D_{SE, \max}^{ab}}$$

b)

$$D_{SE, \max}^{ab} = n^a \chi + (n^b - n^a) \delta$$

Fig 4: Editing operations for calculating the similarity index (SE).

4a: Applying this formula we get a measure of similarity.
4b: The different editing operations will be weighted in different ways, like pay expense. So, for insertion or deleting one label you have to pay 2, for changing a label you pay 1. The maximum distance of 2 strings n^a , respectively n^b labels results with normalization by the maximum distance between two strings in a similarity of range from 0 to 1 as shown in the formula; χ represents the cost of changing and δ stands for the cost of deleting or inserting a label.

String-Editing Algorithm

The string-editing algorithm is a discrete dynamic programming method. Using the operations: insertion, *In*, deletion, *De*, and replacement, *Re*, the algorithm of (Wagner & Fischer, 1974) finds the minimum distance or

cost to convert from string2 to string1; this defines the matrix (Tab.1) Insertions result in horizontal shifts, deletions in vertical ones, and replacements produce shifts along the diagonal. Each operation may add to the cost; the coefficients of the matrix are the hypothetical costs to reach that cell. The editing costs used here are arbitrary examples.

Our use of string editing in matching loci and sequences in images is a bit unusual. However, once we have established a finite state automaton and equivalently, a Markov model the sequences are inherently in a form appropriate for application of the string-editing algorithm. The widest use of string-editing algorithms is perhaps in spellcheck programs. The use in matching of double-stranded chromosomes and sequences of nucleic acids within them, is an important current application.

The Y-matrix (Tab.1)

Sequences of experimental ROIs (regions of interest) can be compared pairwise and then averaged for different conditions. The simplified matrix shows all possible pairwise comparisons for two subjects, two pictures and two random algorithms, with the subjects viewing each picture twice. A total of four experimental ROIs for each subject (in this example) results in 28 possible comparisons for the two subjects. Small bold letters indicate how comparisons are averaged together and reported in the parsing diagram:

For all the subjects who participated in the art-nonart eye movement experiments and for all images these numbers represent the averages of these similarities in a more collected and intuitive fashion. R for Repetitive scanpaths, same subject looking at the same picture at different times; L for Local, different subjects, same picture; I for Idiosyncratic, same subject, different pictures; G for Global, different subjects and different pictures.

Table 1: Y-matrix

Sp similarity coefficients	Subject 1				Subject 2				Algorithm		
	Picture 1		Picture 2		Picture 1		Picture 2		Picture 1	Picture 2	
	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2			
Sub.1 Pic.1	Time1	1	R 0.63	I 0.35	I 0.3	L 0.55	L 0.5	G 0.29	G 0.31	Lah 0.53	Gah 0.24
	Time2		1	I 0.25	I 0.31	L 0.52	L 0.48	G 0.2	G 0.27	Lah 0.5	Gah 0.25
Sub.1 Pic.2	Time1			1	R 0.6	I 0.24	G 0.31	L 0.55	L 0.5	Gah 0.24	Lah 0.45
	Time2				1	G 0.3	G 0.22	L 0.5	L 0.58	Gah 0.21	Lah 0.48
Sub.2 Pic.1	Time1					1	R 0.64	I 0.24	I 0.2	Lah 0.5	Gah 0.2
	Time2						1	I 0.22	I 0.25	Lah 0.45	Gah 0.22
Sub.2 Pic.2	Time1							1	R 0.65	Gah 0.28	Lah 0.49
	Time2								1	Gah 0.23	Lah 0.5
Alg. Pic.1	Time1									1	Gah 0.23
	Time2										1

Data Analysis

Y-Matrices, Parsing Tables (Choi & Stark 1996 ; Privitera & Stark 2005).

Similarity coefficients can be sorted and represented for the two measures Sp and Ss and explicitly displayed in a table, named the Y-matrix (Stark et al. 2001; Privitera & Stark 2005), having as many rows and columns as the number of the different sequences ROIs considered. Pair wise comparisons of all scanpaths were averaged and assembled in Y-matrices (Tab.1).

Parsing tables refer to all images and subjects. They show the average values of similarity coefficients, i.e. the cross correlations between the subjects' eye fixations collected from the arrays of the Y-matrices, and are a compact and intuitive alternative way to look at the data: *R* for *Repetitive* scanpaths, same subject looking at the same picture at different times; *L* for *Local*, different subjects, same picture; *I* for *Idiosyncratic*, same subject, different pictures; *G* for *Global*, different subjects and different pictures. For our experiments, the truncated Y-matrix, represents only a small part of the entire set of comparisons and refers to only two images and two subjects. This Y-matrix, however, is sufficient to illustrate how Y-matrices are translated into a parsing diagram. For example, the Y-matrix diagonal represents the *auto similarity coefficients* (labelled R) of each subject looking at the same picture over different times; these coefficients then generate a unique averaged coefficient reported in the Repetitive box of the parsing diagrams (Table 1). The same ordered collection of the coefficients of the Y-matrix arrays is applied for the other types of comparisons: Local, Idiosyncratic, and Global. The most important distinction is that between repetitive similarity, R, which is usually high, and Global similarity, G usually low, close to random.

Statistical Data Analysis (ANOVA)

Our question was, are the three different treatments (T) i.e. scanpaths of all subjects over pictures of either art (T1) or nonart (T2) compared to randomly generated scanpaths (T3), quite similar, or are they statistically significantly different from each other with $p < 0.01$; i.e. are they different enough (compared to the variability within the individual treatments) for us to conclude that they correspond to three different populations: Could we conclude that, based on those means, the same statistical differences generated in our experiments hold for the hypothetical infinite population of all images and

viewers? The ANOVA (Fisher, 1925; 1978) is finally applied to further validate whether or not the different experimental treatment means, are different enough (compared to the variability within the individual treatments) for us to conclude that they correspond to three different populations.

The analysis was carried out with standardized computer programs (SPSS 16). As the number of subjects permitted no meaningful conclusions on the normality of the data distribution, non-parametric statistical tests were applied throughout. In the visual imagery evaluation for comparing the similarities of the computed strings (viewing, imagery) Kruskal -Wallis ANOVA on Ranks were used. We calculated the viewing/ imagery scanpath similarities using string editing comparison methods mentioned above. The evaluation of basic saccade parameters was performed using Mann - Whitney rank sum test. For more than 2 groups we used Kruskal and Wallis Test with post hoc Dunn's test for pair wise group comparison. A p- value less than 0.05 was considered to be a statistically significant difference.- In addition we computed the post hoc power using G*Power 3.1.6 (Erdfelder, Faul, & Buchner, 1996) – given alpha (0.01), sample size (35), and effect size (0.7): The power was $(1 - \beta \text{ err prob}) = 0.939$.

Results

General Results

EM comparisons of Sp with Ss show that the R repetitive values, 0.837 and 0.458, are significantly different from random, Ra, and from global, G, the two bottom values. It is notable that while Sp-Local has a relatively high value, indicating that different subjects selected similar ROIs, the Ss-Local value is lower, suggesting that different subjects used different sequences for the same picture and similar loci across subjects. The Ss-Ra values are much lower than the Sp-Ra values, since there are many ways to establish sequences among similar loci.

The global similarity value, G, represents any invariant components of eye movements, e.g. the use of some global eye movement strategy controls the tendency to start at the centre of the image and then scan circularly around the periphery. Indeed, reading eye movements have a high G value since all English readers start at the upper left and proceed horizontally across each line and descend vertically line by line, (Stark & Choi, 1996).

High similarity values in the global condition is actually the antithesis to our basic theory, since it would prove that a general and invariant motor program rather than an idiosyncratic internal motor model based on image-specific modelled regions of interest controls eye movements. However, our and later findings (Privitera, Fujita, Chernyak & Stark, 2005) showed that this component is usually very low, close to random. Consequently, global similarity is considered to be a bottom value for our scale of comparisons. The random similarity value, Ra, would be more intuitive than global similarity and it is usually

considered to be a more important bottom value for our comparisons: It represents the similarity of randomly generated scanpaths. This value for Sp is 0.11, which is equivalent to the similarity value between randomly generated scanpaths and human hROIs.

Table 2: Overall parsing diagram for Sp (position) and Ss (sequence).

SD – standard deviation (in brackets), * $p < 0.05$, ** $p < 0.01$, ns - not significant

Same Subject	Different Subject		Same Subject	Different Subject
Repetitive	Local	Same Picture	Repetitive	Local
0.837 (0.085) **	0.466 (0.114) *		0.458 (0.092) *	0.352 (0.115) *
Idiosyncratic	Global	Different Pictures	Idiosyncratic	Global
0.34 (0.150) ns	0.28 (0.151) ns		0.176 (0.150) ns	0.161 (0.150) ns
	Sp Random			Ss Random
	0.11 (0.012) ns			0.0016 (0.005) ns

Even without any specific task instruction for general viewing conditions when different subjects look at the same picture, they are fairly consistent in identifying regions of interest as indicated in this study by the high local (L) and repetitive (R) values. The strong scanpath consistency reported in human experiments when no specific objective is given to the subjects means that only a specific restricted set of representative regions in the internal cognitive model of the picture is essential for the brain to perceive and eventually recognize the picture. This representative set is quite similar for different - naïve- subjects and different picture pairs independently of their art – non art features.

In the case of global (different subjects and different pairs) sequential similarity Ss we found that about 84 percent of the picture pairs where viewed in very dissimi-

lar modes, meaning that their similarity indices lie within the range of the random value or slightly higher. Only in 4 out of the 25 artful-picture snapshot pairs was a high (non-significant) similarity found.

The graphical depiction of this result is shown in figure 4 with examples of picture pairs on the right hand side from Ss global. For 21 pairs of paintings and snapshots subjects either did not show any viewing similarities, such that their EM scanpaths demonstrated global similarity indices that approached random, i.e. close to 0.11; or, they did show some higher viewing similarities between snapshots and artful pictures, such that their EM scanpaths demonstrated similarity indices that somewhat differed from random; however this was only true for 4 out of the 25 artful-picture snapshot pairs. This could have represented differences between scanpath sequences

due to the art-non art selection of pair-combinations. We did not, however, perform a comparison of the pair-similarities, although it would have been desirable to measure the similarity between the artful pictures and their snapshot pendants -- e.g. on the basis of ROIs -- before determining similarities between the eye movements of their viewers: Some of these pairs may be more

similar and may lead to more similar eye movements than others. Privitera and Stark (2000) have investigated and developed a methodology that serves to automatically identify a subset of aROIs (algorithmically detected ROIs) using different Image Processing Algorithms, IPAs, and appropriate clustering procedures.

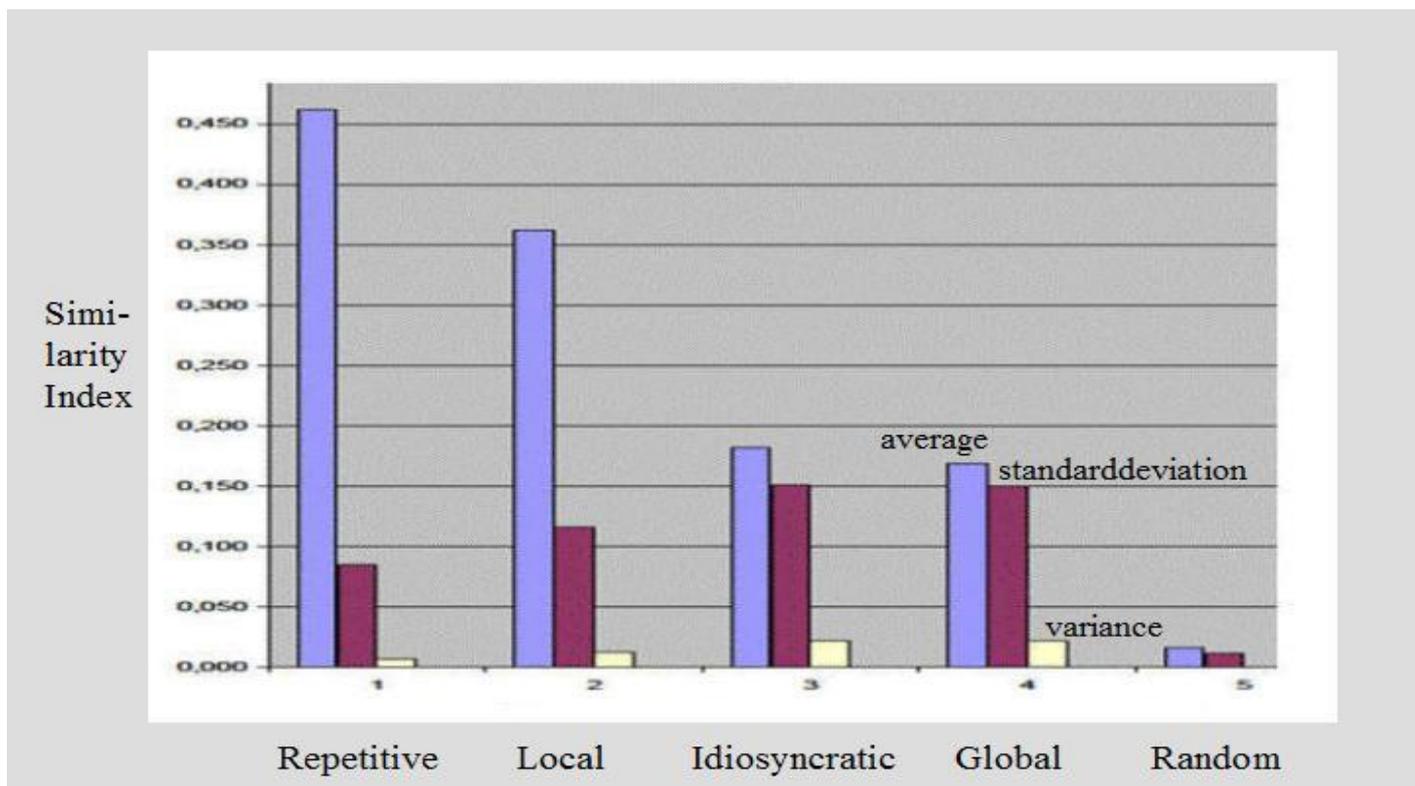


Fig.5a: Overview of picture pairs for sequential (Ss) averaged global responses of 25 pairs. For 21 pairs of paintings and snapshots subjects either did not show any viewing similarities, i.e. similarity index close to 0.11; or, they did show some higher viewing similarities between snapshots and artful pictures, such that their EM scanpaths demonstrated similarity indices that somewhat differed from random (non significantly). For four pairs of paintings and snapshots subjects did show relatively higher viewing similarities between snapshots and artful pictures, such that their EM scanpaths demonstrated similarity indices that differed clearly from random, i.e. higher than 0.1.

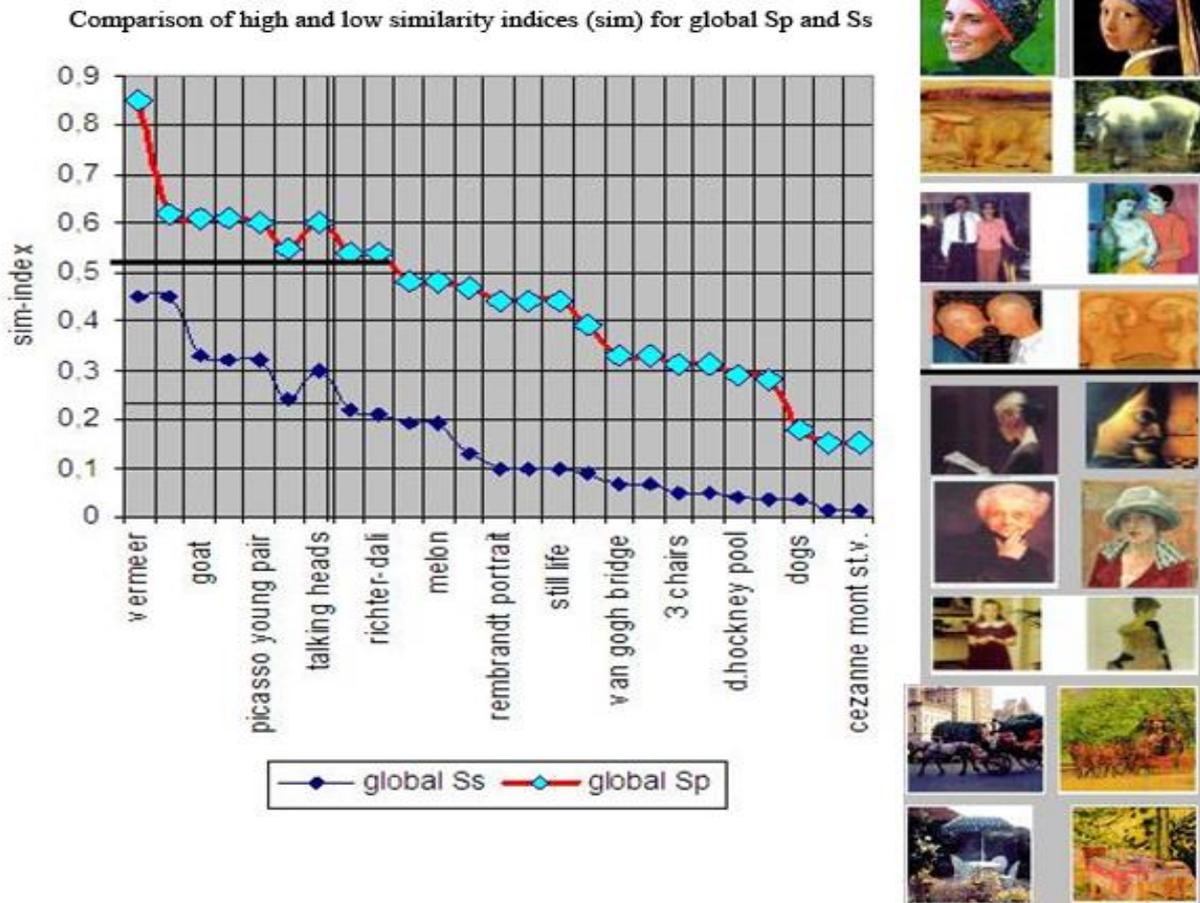


Fig.5b: Parsing Diagram of similarity of sequence (Ss) of repetitive (R), local (L), idiosyncratic (I) and global (G) conditions. Note the the high similarity index for R and L as expected from previous findings; also note the low similarity index for I and R that is in the range of the standard deviation; blue column: average sim value; red column: standard deviation; yellow column: variance.

Discussion

Using high resolution infrared eye recordings in 7 young naïve subjects we recorded their scanpaths of 4 sec viewing 50 pairs of 25 artful pictures and 25 snapshot photographs on 5 different days. The pictures were selected with respect to similarity of size and scene between the snapshots and the artful pictures. After string editing and

parsing analysis we compared the repetitive and the global parsing indices for Sp and Ss. Sp (position) and Ss (sequence) showed in principal similar values, although Ss values were in general some 40% lower than Sp due to the greater variance within sequences of viewing. Our general results were not unexpected inasmuch they confirmed the general statement that follows from the methodical consideration explained in Table 3:

Table 3: Number of strings that contribute to each index in the parsing table (25 picture-pairs = 50 pictures, times 5 presentations, times 7 subjects), which means that different subjects (7) looking at the same picture(1) should show less variant strings than different subjects (7) looking at many different pictures (50). This is due to the low numbers of strings that contribute to that measure; Similarity of strings in decreasing order: **R**epetitive, **I**diosyncratic, **L**ocal, **G**lobal. Ss parsing tables show that the probability of similarities decreases to about one half of Sp, due to the high number of vector directions compared to the lower and constant number of possible fixation positions.

	Same Subject	$p \times s \times n$	Range	Diff. Subject	$p \times s \times n$	Range
Same pic.	R	$1 \times 1 \times 5 = 5$	10^0	L	$1 \times 7 \times 5 = 35$	10^1
Diff. pic.	I	$50 \times 1 \times 5 = 250$	10^2	G	$50 \times 7 \times 5 = 1750$	10^3

Zero Hypothesis

We asked, do naïve subjects perceive a snapshot in a different manner than they perceive an artful picture or is there no difference in perception and thus a high similarity between the spatial and sequential scanpath regions of interests; and: Is the global similarity during scanning of all image-pairs in all subjects low i.e. close to random, or is there a high similarity of scanpaths when artful pictures were viewed, but not in viewing snapshots? Since Sp Repetitive showed for all situations the highest sim-index, it follows that the zero hypothesis is correct and that there was no basic difference between viewing artful pictures or snapshots within our somewhat naïve group, as far as viewing ‘art’ was concerned.

Even without any specific task instruction for general viewing conditions when different subjects look at the same picture, they are fairly consistent in identifying regions of interest as indicated in this study by the high local (L) and repetitive (R) values. The strong scanpath consistency reported in human experiments when no specific objective is given to the subjects means that only a specific restricted set of representative regions in the internal cognitive model of the picture is essential for the brain to perceive and eventually recognize the picture. This representative set is quite similar for different, in our case art-naïve subjects and different picture pairs independently of their art – non art features.

In the case of global (different subjects and different pairs) sequential similarity Ss we found that about 84 percent of the picture pairs were viewed in very dissimilar modes, meaning that their similarity indices lie within the range of the random value or slightly higher. Only in 4 out of the 25 artful-picture snapshot pairs (16%) was a

higher (non-significant) similarity found. This indicates that only a comparatively small proportion of our subjects may have been aware of the artfulness of some pictures; This was corroborated by our post-test question to quickly select possible artful pictures (14% on average) within pairs.

Global-invariant components of eye movements - the use of some global eye movement strategy control- is highest in reading, i.e. ‘start at the upper left, proceed horizontally and downwards’. ‘Reading artful pictures’ (Gombrich, 1969 & 1984) is a similarly difficult and only with long term training achievable skill. Therefore we might expect high Sp and even higher Ss only in subjects trained in “reading art”, but not in art-naïve subjects, which has been demonstrated previously (Zangemeister, Sherman & Stark, 1995).

Does the skilled artist control our eye movements?

We started with the hypothesis: Since the artful picture of each pair has a more tightly interrelated set of objects, and thus, regions of interest, it could be that the Sp Local for art pictures are higher than the Sp Local for snapshots. This could suggest that the artist simplifies, emphasizes or somehow *controls* (this is what most artists claim) each component’s degree of information in the top-down cognitive model (Zangemeister & Stark, 2007). Note that in this case the correlation of saliency with degree of information is avoided. If the above hypothesis were true, then sequential effects of presentations of the artistic pictures in the snapshot pictures would occur. These effects we tried to exclude through randomising presentations of the art-non art sequences. The results

demonstrate that these hypotheses were not verified.

'Artfulness': Bottom Up and Top Down

A strong Top-Down component generates repetitive *Intra-* and *Inter-*subject scanpath sequences. This is the high end of the similarity-index. Inter-subject similarity is often high only for Sp, but not for Ss. Bottom-up particularity in similarly arranged picture pairs gives rise to very high Sp sim-indices, independent from the mode of fabrication of a particular picture.

Global-invariant components of eye movements may generally be small. However, with skillful internal models as in *reading* sentences or artful pictures we might expect significantly different Sp and Ss Global sim-indices from the repetitive and local conditions. Degree of information (top- down), and particularity (Bottom-Up) in our paradigm were shown to be intimately interconnected. Thus, differential viewing of art pictures compared to snapshots rarely showed up in our naïve subjects. Obviously, the *knowing* viewer must apply a pre-existent top down sophisticated model of „artfulness” in a particular picture, in order to differentiate it from a simple snapshot that looks similar. Overall, one could divide viewers into three subgroups with respect to the capability of distinguishing art from naïve (almost no differentiation) to sophisticated (capable to differentiate art - non art in *traditional aesthetics*), to the modern and postmodern professional viewer (capable to distinguish any piece of work, including a ‘ready- made’ as art from non-art).

Definitions and distinctions of artful paintings, snapshots and mental images of artful pictures – how do our results compare to Buswell’s and his followers findings?

In Buswell’s early study (1935) photographic records were made of eye movements of 200 subjects while they were looking at reproductions of paintings (coloured and uncoloured), of vases and dishes, of furniture and design, of statuary and museum pieces, of tapestries, buildings, posters, outlines, and geometric figures. Records were made both of direction of the subjects’ eye movements and their fixation durations. The results showed that colour had little effect on eye movements, which, however, were influenced by the *instructions given the subject*, by training in art, and by the length of time that the picture was inspected (overall viewing time). In terms of numbers, Buswell’s and the present study are quite different: subjects recorded, pictures and scenes viewed, differ-

ences of items viewed, duration of viewing time; also the present study is quite different in that subjects were naïve, i.e. with no art training. In other words, this study deliberately went for a carefully controlled repetitive trial in only few subjects that were chosen for art naïveté and were left free to view the presented pictures as they chose within a precisely limited time of 4 sec which accounts on average for a sum of maximally twelve eye movements (fixations) altogether.

The main result relates to Buswell’s in that instructions as well as art training appear to influence subjects’ EM strongly. Amongst others, this has been shown later on by Yarbus (1967), Locher and Nodine (1987, 1996), Zangemeister, Sherman & Stark (1995), Ramachandran & Hirstein (1999). In fact, these well known results have been one of the reasons to initiate the study presented here: To ask how *without* specific instructions, *without* art training, viewing only for a short time but repetitively, subjects may be able to differentiate artful pictures from snapshots. Our results demonstrate that –in this paradigm- it is highly unlikely that this differentiation happens since subjects’ were unable to perform it consistently. Evidently there is no consistent *intuition* or *feeling* for the *artfulness* of the given pictures, even when presented repeatedly.

The artistic process and Shannon’s Theory of Communication

What is a Picture? A counterexample is a large field-of-regard, as in a complex *scene*, with many unrelated objects located in it. A *picture*, then, is some localized sub region of a broad scene, with a frame of peripheral content, and perhaps with related objects, that can be modelled in a top-down fashion and checked with the high-resolution fovea. The movement one’s gaze direction does may be considered as framing a *sub segment of a broad vista* in such a way as to produce a ‘picture’. A more focused and framed picture could be an *art picture*. It is a constructed use of factual material, wherein the construction is a work of art.

This is relevant to the artist attempting to *communicate through a painting* or as an example of an art picture. The artist’s painting begins with his model of related objects within a frame, that carry a story either classically in representative art, or a story in terms of artistic technology, or in terms of certain phases of abstract art. Other phases, of course, combine these. The artistic process in light of Shannon’s Theory of Communication (Shannon & Weaver, 1949) is depicted in the following figure (modified from: Zangemeister & Stark 2007). When we view a painting, our eye focuses on curves, angles, line crossings, shadows, colours. In the BU bottom up

scheme, the viewer follows the given basic stimuli and cues like the drop of a line or the red of a peaked angle. In the TD top down scheme, the viewer uses from a set of pre-existent models, which one could optimally match the picture one is looking at.

While the matching process is going on between the viewer's internal model and the picture, there is a continuous exchange between TD and BU. In both cases we are looking at the painting or any object over some period of time. Thus we can fixate on objects or regions of interest (ROIs) only in sequence, such that duration of these fixations becomes an important parameter, since during longer lasting fixations of small areas most of the picture's detailed information is taken in. In figure 5, this process

is depicted in terms of his theory. A picture of art may be the complex message – perhaps confounded by unfocused ideas and unclear representations of the artist – that the artist-sender has imagined using sets of imagery scanpaths, and encoded with his technique and skill. In addition, faded colour, poor restoration, loss of picture pertinent information due to changing historical times may contribute to some *noise* in this information channel. So the receiver (viewer), while he is actively looking i.e. generating a set of scanpaths in order to decode the artist's message, may have difficulties to form a good image in his mind's eye, because of artistic insufficiencies, noise, and a gap in understanding due to time and culture gaps.



Fig. 6: The artistic process and Shannon's Theory of Communication

As Duchamp points out, the creative act is not performed by the artist alone since “the spectator brings the work in contact with the external world by deciphering and interpreting its inner qualifications and thus adds his contribution to the creative act. This becomes even more obvious when posterity gives its final verdict and sometimes rehabilitates forgotten artists” (Duchamp, Ready-mades 1921). Creative communication (artist-sender) and evaluating communication (viewer-receiver) are separate phases of a process that produces art. In reality they are superimposed on each other all the time, since every creative act in the sense of producing innovations is made up of partial creations, interspersed with judgments, acts of acceptance or dismissal: Evaluative aesthetics is relevant with respect to the interpreter (Max Bense, Introduction to information-theory and aesthetics, 1965).

Scanpath theory and Schrödinger's Cat

Erwin Schrödinger imagined a thought experiment, known as ‘Schrödinger's Cat’, in which a cat is placed in a sealed box, its life depending on a radioactive atom whose probability of decay for a certain period is known. The atom is connected to a mechanism that poisons the cat in case of the atom's decay. However, unless one

opens the box, one cannot know the cat's actual state of being dead or alive. Thus, the cat lives in a ‘*superposition*’, i.e. the actual state of being dead or alive in the light of the probabilistic interpretation of quantum physics is occupying two states simultaneously.

A similar concept can be extracted from the scanpath theory: A ‘scene’, i.e. the whole of a visual perception at one instant can be decomposed into a number of designed, abstracted or composed separate pictures. The term “*passing scene*” may be better to illustrate a likely non-mentally created picture as an extension of the plain word “scene” that has been used in D.Chernyak's paper on scene analysis (Chernyak & Stark, 2001). The word ‘passing’ is added to the word ‘scene’ to indicate a set of visual signals not yet included to the mental image by the brain. George Berkeley (1685-1753) was the first to suggest that one *cannot* see non-internally-imagined pictures (Berkeley, 1960) which was more recently described by Chatterjee (Chatterjee et al. 1995). In the instant of ‘seeing’ one constructs a hypothesized schema (*mental image hypothesis*) with widespread representations in the brain attributing to it the qualities which it shares with other images, especially art pictures. Now, assuming this to be the case: When perceiving a scene, how does the internally generated image

and the model of the perceived external, physical world relate to each other in one's brain? Does the external-world-perception interfere with the mental image, cancelling it out? Or do these two images exist at the same time in our brain just as 'Schrödinger's Cat' does in the box? (And even more intriguing: what happens when we are able to 'open the box', i.e. decoding the relevant brain mechanisms?). This is best exemplified through our scanpaths when viewing *ambiguous pictures* as in *Two Faces or a Vase* (cf. Fig. 1b,c). Depending upon the TD internal cognitive model, the subject *sees* only one the two interpretations.

Thus, subjects are not able to 'see' *both* variants of the ambiguous figure *simultaneously*; in primates, however, there are electrophysiologically recordable transitions (Leopold & Logothetis, 1999) that show that to some, although minor extent *transitional overlaps do occur*.

Relating to this aspect of ambiguity in picture viewing, Semir Zeki (2001) has distinguished between three modes of artistic ambiguity that may evoke aesthetic pleasure: *Metastable works*, in which the recessional plane of one border shifts continuously and in an obligate manner with the recessional planes of the two abutting borders, implying inhibitory interactions and instability in responses of cortical cells. *Determinate ambiguity* (Fig.1b,c: vase/face) as also found in the Necker cube or works of artists as Salvador Dali, where objects can take on one of two forms. *Open ambiguity*, that is a characteristic of some completed works of art, for example those of Johannes Vermeer, as well as unfinished works, such as those of Michelangelo, who left three fifths of his sculptures unfinished (e.g. the *Pietà Rondanini*).

Views and definitions of art: Classical and modern views

Aristotle was the first to introduce the theory that art imitates nature (*mimesis*) and attributed the origin of art to the human affinity for imitation. Based on *mimesis* he distinguished three classes of art: 1st, difference in the means of imitation: rhythm, language, harmony relating to music, poetry, dance and drama; 2nd the examination of the object being represented; 3rd the manner in which the object is presented. Hence, art is a productive science: It is found within the object produced, not within the mind of the artist, and this determines the quality of the art. The viewer (evaluator) of the piece of art does not need to consider the message or intent of the artist or the history or circumstances behind the work when evaluating it critically: Even if the message of the artist may

be absent or unclear, the object itself may be a perfect imitation and therefore a perfect piece of art. Aristotle's theory of art as imitation in this way provides a basis for classification of art forms. His theory appeals to human nature – especially in view of the more recent findings on mirror neurons in the human cortex and their variant functions (Rizzolatti, 2004) - but it lacks more refined ideas about the creativity of the artist, about the viewer's response and about abstract art forms.

Immanuel Kant and G.F.W.Hegel ascribed far greater importance to natural rather than to artistic beauty, so far as there were grounds for distinguishing them: for them the assurance of a deep intended harmony between the world and us. "*Natural* beauty is perhaps always external, unless we see the world itself as a work of art, and its meaning the symbol of its goodness." (Danto, 1994; Zaidel, 2005).

Marcel Duchamp has changed this view radically: "Art is a drug: Art has absolutely no existence as veracity, as truth. The onlooker is important as the artist." (Duchamp, 1921). According to Duchamp, in the creative act, the artist goes from intention to realization through a chain of totally subjective reactions. His struggle toward the realization is a series of efforts, pains, satisfactions, refusals, decisions, which also cannot and must not be fully self-conscious, at least on the aesthetic plane. The result of this struggle is a difference between the intention and its realization, a difference which the artist is not aware of: It is like an arithmetical relation between the unexpressed but intended and the unintentionally expressed. The creative act takes another aspect when the spectator experiences the phenomenon of transmutation: through the change from inert matter into a work of art in the mind's eye of the viewer who determines the weight of the work of art on an aesthetic scale. The spectator brings the work in contact with the external world by deciphering and interpreting its inner qualifications and thus adds his contribution to the creative act. As an active, top down process, vision and higher order cognitive influences such as memory retrieval and expectation, attention, perceptual task as well as motor signals are fed into the sensory apparatus (Gilbert and Li, 2013). Duchamp's view has had a major influence on art of the 20th century in many respects that are beyond the scope of this paper.

The other major influence and also departure from traditional art is represented by the work of Andy Warhol and his lasting influence on Pop Art and its followers. Alluding to the pure and perfect surface of things he said: "There I am. There's nothing behind it. I see everything that way, the surface of things, a kind of mental Braille. I just pass my hands over the surface of things. The reason I'm painting this way is that I want to be a machine, and I feel that whatever I do and do ma-

chine-like is what I want to do. I like boring things. I like things to be exactly the same over and over again. (Warhol, 1975).

The aspect of postmodern, post-pop art hybrids was put in to the cover image of Danto's book „*Beyond the Brillo Box*“, as a citation of Rembrandt's famous „*Anatomy Lesson*“ wherein Andy Warhol's 'Brillo Box' was transported back to a much earlier time and replaced the cadaver in Rembrandt's picture – as if Dr.Tulp's eager 17th century auditors were listening to a discourse on mid-20th-century American art.

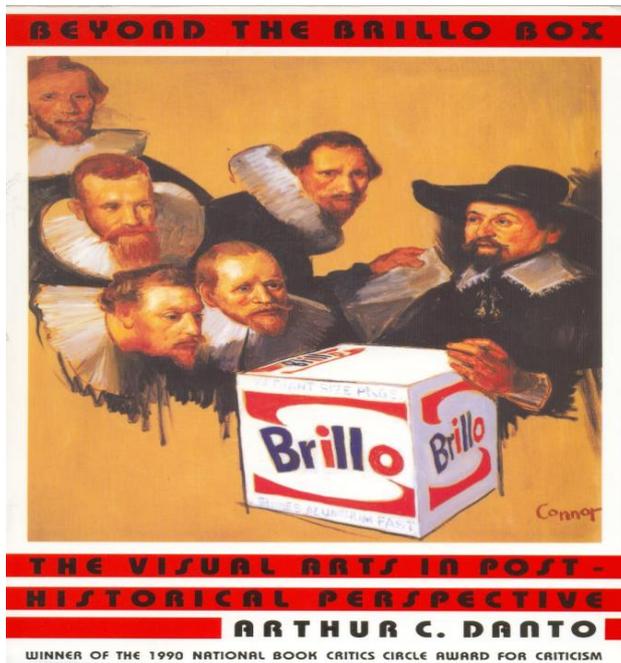


Fig.7: Cover image [by Russell Conner] for Danto's book „*Beyond the Brillo Box*“.

Conclusion

Definitions and distinctions of artful paintings, snapshots and mental images of artful pictures relate to Buswell's original findings through understanding of the continual exchange between top down and bottom up in viewing pictures and snapshots. The *knowing* viewer must apply a pre-existent top down sophisticated model of „artfulness“ in a particular picture, in order to differentiate it from a simple snapshot that looks similar. As an active top down process, vision and higher order cognitive influences such as memory retrieval and expectation, attention, perceptual task as well as motor signals are fed into the sensory

apparatus while viewing pictures. The artist simplifies, emphasizes or somehow *controls* -this is what most artists claim- each component's degree of information in the top-down cognitive model (Zangemeister & Stark, 2007). George Berkeley was the first to suggest that one *cannot* see non-internally-imagined pictures. In the instant of 'seeing' one constructs a hypothesized schema (*mental image hypothesis*) with wide-spread representations in the brain attributing to it the qualities which it shares with other images, especially art pictures.

Artists attempt to *communicate* through a work of art like a painting. The artist begins with his model of related objects within a frame, that carry a story either classically in representative art, or a story in terms of artistic technology, or in terms of certain phases of abstract art. This leads to the view of the artistic process as a process of communication. When we look at a painting, our eye focuses on different features of the picture. While the matching process during the viewing is going on over some period of time between the viewer's internal model and the picture, there is a continuous exchange between TD (the viewer's pre-existent models) and BU (basic stimuli within a picture). Thus a picture of art is a complex message extending in time – perhaps confounded by unfocused ideas and unclear representations of the artist, encoded with his technique and skill: I.e. dense artistic information content accompanied by noise. The receiver –viewer-, while she is actively looking i.e. generating a set of scanpaths in order to decode the artist's message, may have difficulties to form a good image in her mind's eye, because of artistic insufficiencies, noise, lack of attention, and a gap in understanding due to time and culture gaps. Creative communication (artist-sender) and evaluative communication (viewer-receiver) are separate phases of a process that produces art.

Aristotle was the first to introduce the theory that art imitates nature (*mimesis*) and attributed the origin of art to the human affinity for imitation. Aristotle's theory of art as imitation in this way provides a basis for classification of art forms. His theory lacks more refined ideas about the creativity of the artist, about the viewer's response and about abstract art forms. Immanuel Kant and G.F.W.Hegel ascribed far greater importance to natural rather than to artistic beauty. For them this was the assurance of a deep intended harmony between the world and us. Marcel Duchamp has changed this view radically in claiming that the onlooker is as important as the artist: "A work of art is like an arithmetical relation between the unexpressed but intended and the unintentionally expressed". The spectator then brings the work in contact with the external world by deciphering and interpreting its inner qualifications and thus adds his contribution to

the creative act, determining the weight of the work of art on an aesthetic scale, such that everything could be art. Without giving subjects a clear definition of what has to be seen as art and what not, it is impossible for them to distinguish art from non-art as shown above in naïve, unsophisticated viewers of art. Only if subjects are given a clear clue about 'what is art' they might successfully be able to make that distinction. This could only be done by using a restricted range of art pictures that are clearly seen as art. Nowadays this distinction of earlier times has vanished. This has been causing some confusion about the ratio of signal and noise in the communication between artists and viewers of art.

Of course, it would be highly desirable to provide an outlook on a research strategy that could be pursued by scientists who want to follow up on this work using sophisticated analyzing methods. Instead to try to answer the question of how to determine artfulness – which is likely to lead to a never ending discussion - it would be preferable to try to answer the question: Do (untrained, trained, experienced) viewers look at artistically assembled visual material (e.g. paintings whose creators claim to convey messages that are hidden beneath the surface structure of the images) differently than at more accidentally assembled material with comparable visual surface structure. That there is a significant difference between naïve, sophisticated and professional viewers viewing the same pictures, has been shown by us previously (Zangemeister, Sherman & Stark 1995). This would allow for more differentiated answers, as 'art' seems to be such an ill-defined notion, as we have pointed out in this paper.

References

- Avital T. (2003). *Art versus nonart: Art out of mind*. Cambridge: Cambridge University Press.
- Bahill T.A. & Stark L.W. (1979). Trajectories of Saccadic Eye Movements, *Scientific American*, 240, 84-93.
- Vautier, B. (1972). *Everything is art*. Kassel: Documenta 5.
- Brandt, H.F. (1940). Ocular patterns and their psychological implications. *American Journal of Psychology*, 53, 260-268.
- Brandt, S.A., & Stark, L.W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience*, 9(1) 27-38.
- Buswell, G. T. (1935). *How people look at pictures: a study of the psychology and perception in art*. Chicago: University of Chicago Press.
- Bense M. (1965). *Aesthetica – Einführung in die Aesthetik (Introduction to aesthetics)*. Baden-Baden: Agis Verlag.
- Berkeley, G. (1960). *A new theory of vision and other writings* (Reprint). New York: Dutton
- Blackmon T., Yeuk F., Chernyak D.A., Azzariti M., Stark L.W. (1999). Dynamic scanpaths: eye movement analysis methods. *Proceedings SPIE 3644, Human Vision and Electronic Imaging IV*, 511 (May 19), doi:10.1117/12.348471
- Chatterjee, A., Southwood, M.H. (1995). Cortical blindness and visual imagery. *Neurology* 45: 2189-95
- Chernyak, D.A. & Stark, L.W. (2001). Top-down guided eye movements. *IEEE Transactions SMC, B* 31(4), 514-522
- Danto, A.C. (2003). *The abuse of beauty*. Chicago: Open Court Publications.
- Danto, A.C. (1994). *Embodied meanings: critical essays and aesthetic meditations*. New York: Farrar, Straus, Giroux.
- Duchamp, M. (1921). *Apropos of "Readymades"*. In: Tomkins, C. (Ed.), *Duchamp: A biography*. New York: Henry Holt.
- Ellis, S. R., & Stark, L. (1978). Eye movements during the viewing of Necker cubes. *Perception*, 7(5), 575-581.
- Erdfelder, E., Faul, F. & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, 28, 1-11.
- Fisher L. (1978). *Fixed effects analysis of variance*. New York: Academic Press
- Fisher, R.A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.
- Gbadamosi, J., & Zangemeister, W.H. (2001). Visual imagery in hemianopic patients, *Journal of Cognitive Neuroscience*, 13, 855 – 866

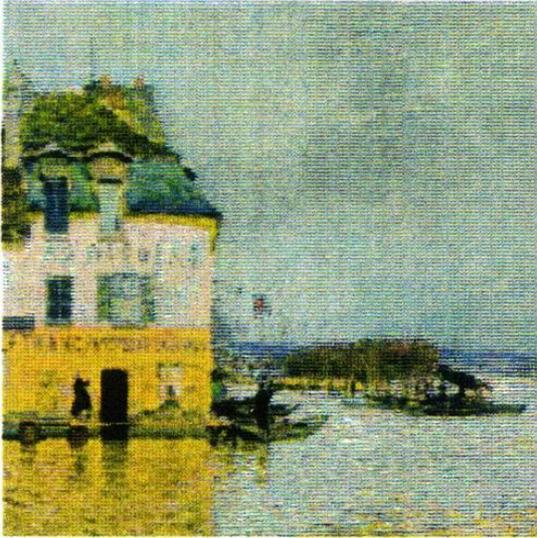
- Gilbert, C.D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience* 14(5), 350-363, doi:10.1038/nrn3476
- Krischer, C., & Zangemeister, W.H. (2007). Scanpaths in reading and picture viewing: Computer-assisted optimization of display conditions. *Computers in Biology and Medicine*, 37, 947 – 956.
- Leopold, D.A., & Logothetis, N.K. (1999). Multistable phenomena: changing views in perception. *Trends in Cognitive Science*, 7, 234-264.
- Levenshtein, V.I. (1965). В.И. Левенштейн (1965) Двоичные коды с исправлением выпадений, вставок и замещений символов. *Доклады Академии Наук СССР* 163(4), 845–848. Appeared in English as: V. I. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals (1966). *Soviet Physics Doklady* 10, 707–710.
- Liman, T. G. & Zangemeister, W. H. (2012). Scanpath eye movements during visual mental imagery in a simulated hemianopia paradigm. *Journal of Eye Movement Research* 5(1):2, 1-11.
- Locher, P.J., & Nodine, C.F. (1987). Symmetry catches the eye. In J.K. O'Reagan & A. Levy-Schoen (Eds.) *Eye movements: From physiology to cognition* (pp. 353-361). Amsterdam: Elsevier Science.
- Locher, P.J. (1996). The contribution of eye-movement research to an understanding of the nature of pictorial balance perception: A review of the literature. *Journal of the International Association of Empirical Studies* 14(2), 143-163.
- Noton, D., & Stark, L.W. (1971). Scanpaths in eye movements during pattern perception. *Science*, 171, 308-311.
- Noton, D., & Stark, L.W. (1971). Eye movements and visual perception. *Scientific American*, 224, 34-43.
- Privitera, C.M., & Stark, L.W. (2000). Algorithms for defining visual regions-of-interest: Comparison with Eye Fixations, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22 (9), 970 – 982.
- Privitera, C.M., Stark, L.W. (2003). Human-vision-based selection of image processing algorithms for planetary exploration. *IEEE Transactions on Image Processing*, 12(8), 917-923.
- Privitera, C.M., Fujita, T., Chernyak, D., & Stark, L.W. (2005). On the discriminability of hROIs, human visually selected regions-of-interest. *Biological Cybernetics*, 93(2), 141-152.
- Privitera, C. M., Stark, L. W., & Zangemeister, W. H. (2007). Bonnard's representation of the perception of substance. *Journal of Eye Movement Research*, 1(1):3, 1-6
- Ramachandran, V.S., & Hirstein, W. (1999). A theory of human artistic experience and the neural mechanisms that mediate it. *Journal of Consciousness Studies*, 6(6-7), 15-51.
- Rizzolatti G. & Craighero L. (2004). The mirror neuron system. *Annual Review of Neuroscience* 27, 169-192
- Shannon, C., & Weaver, W. (1949). *A mathematical theory of communication*. Urbana Illinois: University of Illinois Press.
- Snow, C.P. (1959). *The two cultures*. Cambridge, UK: Cambridge University Press.
- Stark, L.W., Privitera, C.M., Yang, H., Azzariti, M., Ho, Y.F., Blackmon, T., & Chernyak, D. (2001). Representation of human vision in the brain: How does human perception recognize images? *Journal of Electronic Imaging* 10(1), 123-151.
- Wagner, R., & Fischer, M (1974). The string-to-string correction problem (Damerau). *Journal of the Association of Computing Machinery* 21(1), 168–173.
- Warhol, A. (1975). *The philosophy of Andy Warhol, From A to B and back again*. New York: Barnes & Noble.
- Welsch, W. (1996). *Grenzgaenge der Aesthetik (Borderline courses in aesthetics)*. Stuttgart: Reclam.
- Yarbus, A. (1967). *Eye movements and vision*, New York: Plenum.
- Zaidel, D.W. (2005). *Neuropsychology of art: Neurological, cognitive, and evolutionary perspectives*. Los Angeles: Psychology Press.
- Zangemeister, W.H, Sherman, K., & Stark, L.W. (1995). Evidence for a global scanpath strategy in viewing abstract compared with realistic images. *Neuropsychologia* 33(8), 1009-1025.

- Zangemeister, W. H., Oechsner, U., & Freksa, C. (1995). Short-term adaptation of eye movements in patients with visual hemifield defects indicates high level control of human scanpath. *Optometry & Vision Science* 72(7), 467-477.
- Zangemeister, W.H., Stiehl, S. & Freksa, C. (1996). *Visual attention and cognition*. Amsterdam: Elsevier.
- Zangemeister, W.H., & Stark, L.W. (2007). *The artistic brain beyond the eyes*. London & Bloomington: Authorhouse.
- Zangemeister, W. H., & Liman, T. (2007). Foveal versus parafoveal scanpaths of visual imagery in virtual hemianopic subjects. *Computers in Biology and Medicine* 37, 975–982.
- Zeki, S. (1999). *Inner vision: An exploration of art and the brain*. Oxford: Oxford University Press.
- Zeki, S. (2001). *Artistic creativity and the brain*. *Science*, 293, 51-52.

Appendix:

All picture-snapshot pairs used in the experiment

Sisley:



Picture 1

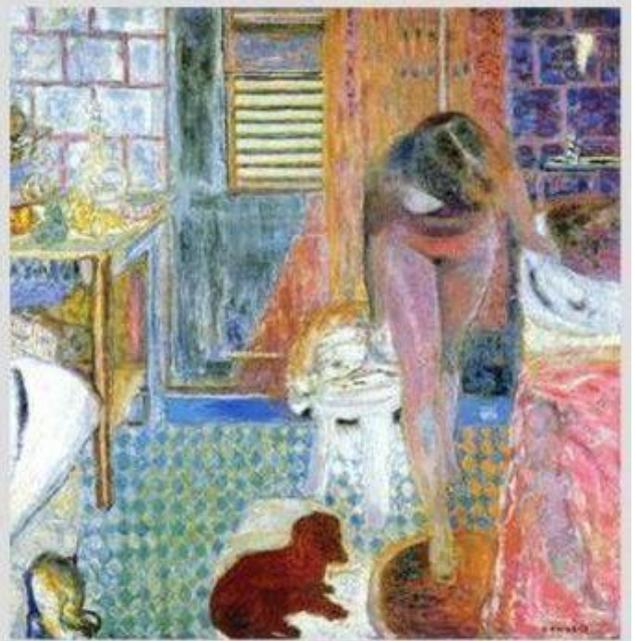


Picture 3

Bonnard 1:

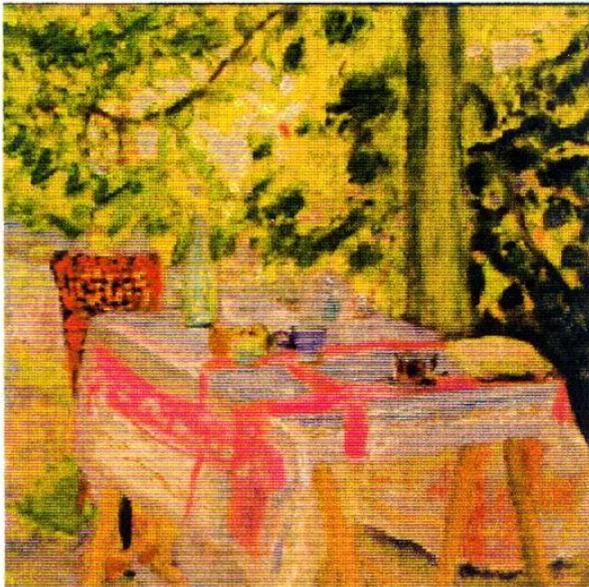


Picture 2



Picture 45

Bonnard 2:

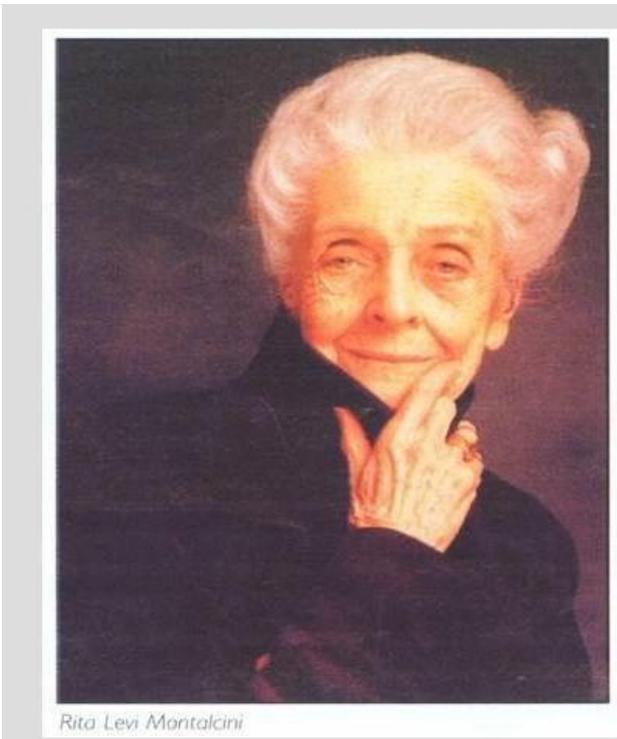


Picture 26



Picture 41

Bonnard 3:

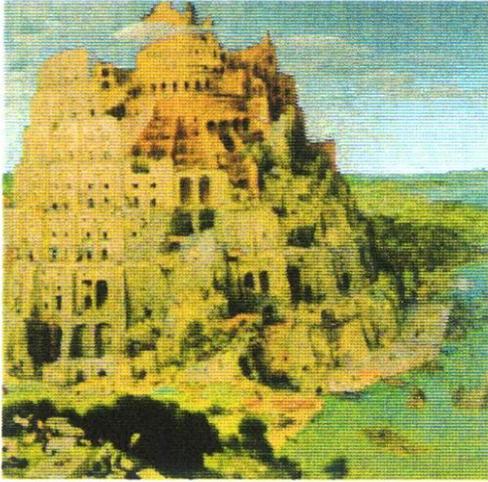


Picture 11



Picture 16

Breughel:



Picture 7

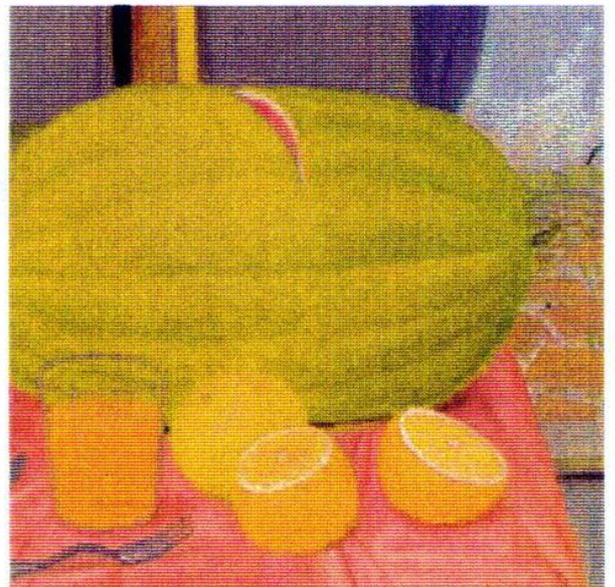


Picture 49

Cezanne:

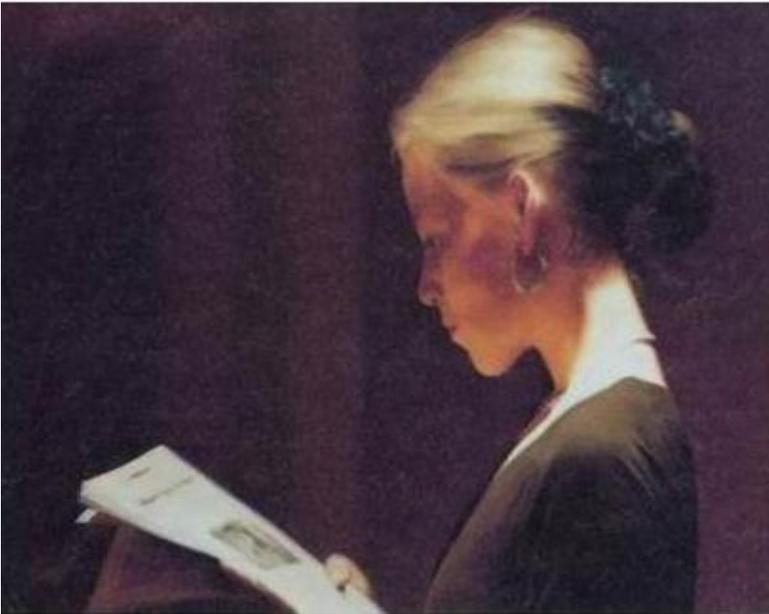


Picture 15

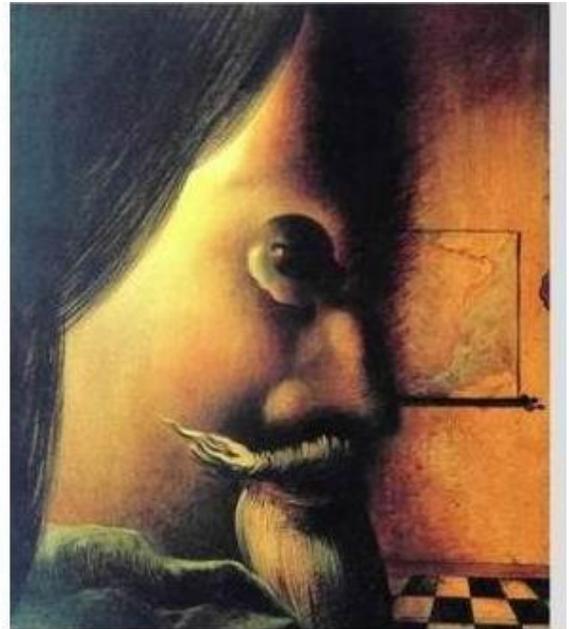


Picture 23

Dali:



Picture 43



Picture 46

de Stael:

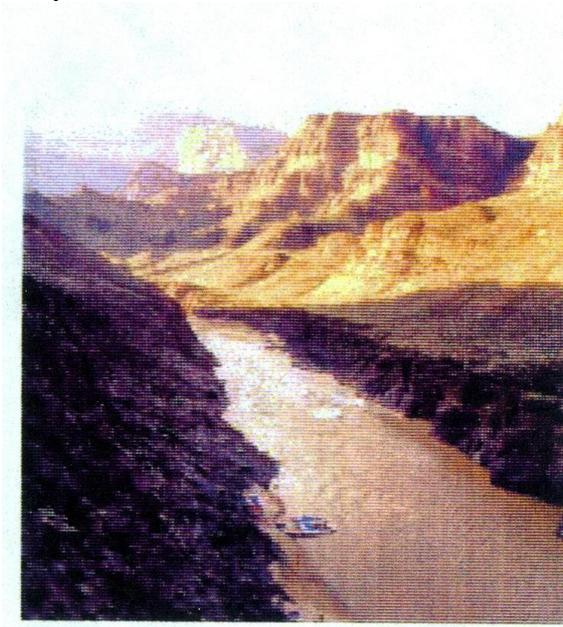


Picture 17

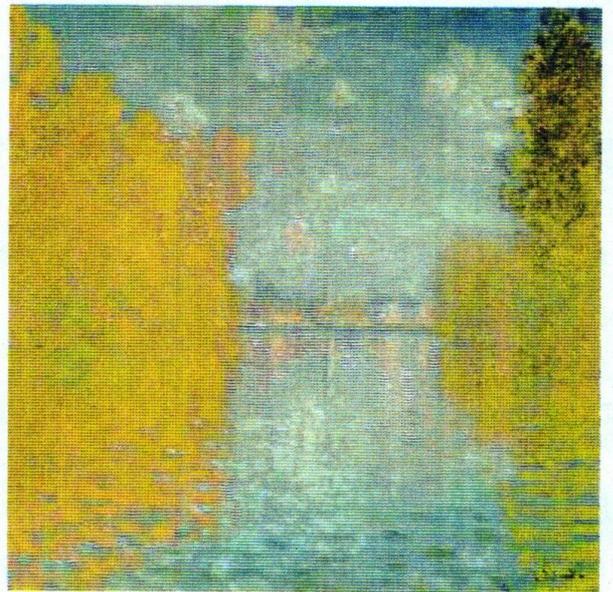


Picture 35

Dufy:

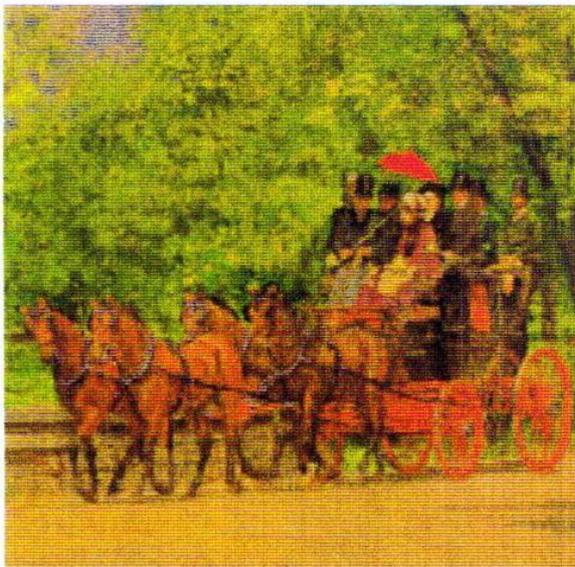


Picture 12

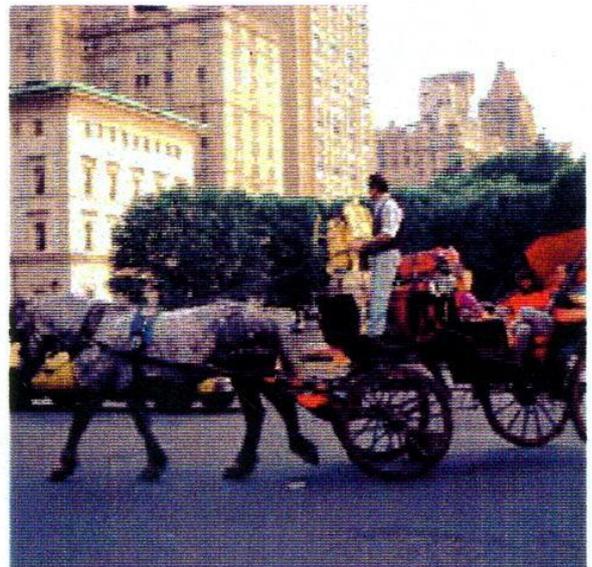


Picture 31

Eakins:



Picture 14



Picture 36

Feininger:

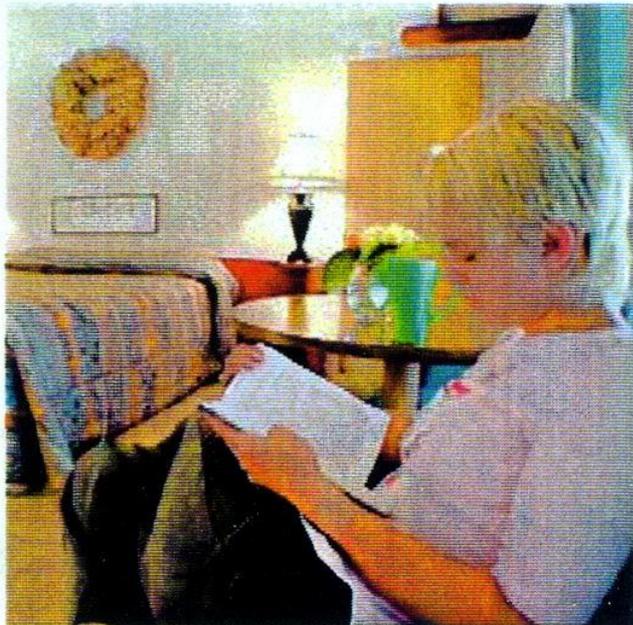


Picture 37



Picture 40

Fragonard:

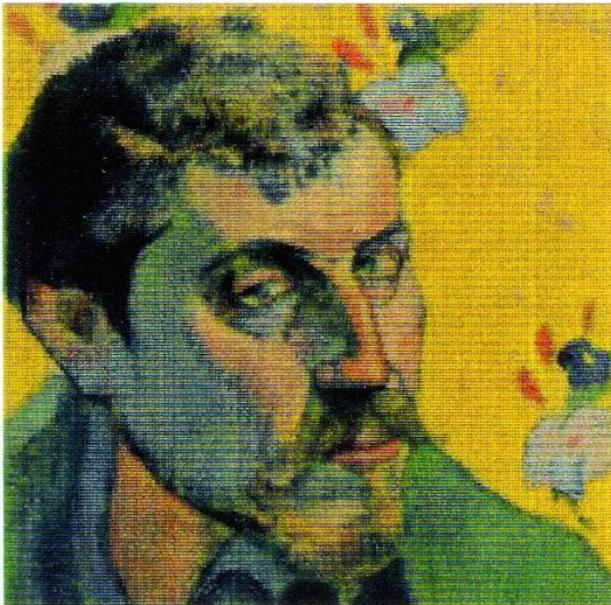


Picture 21

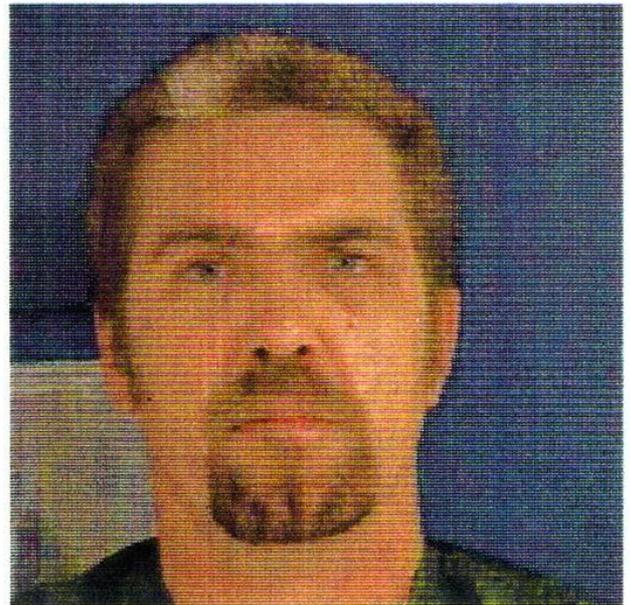


Picture 47

Gauguin:

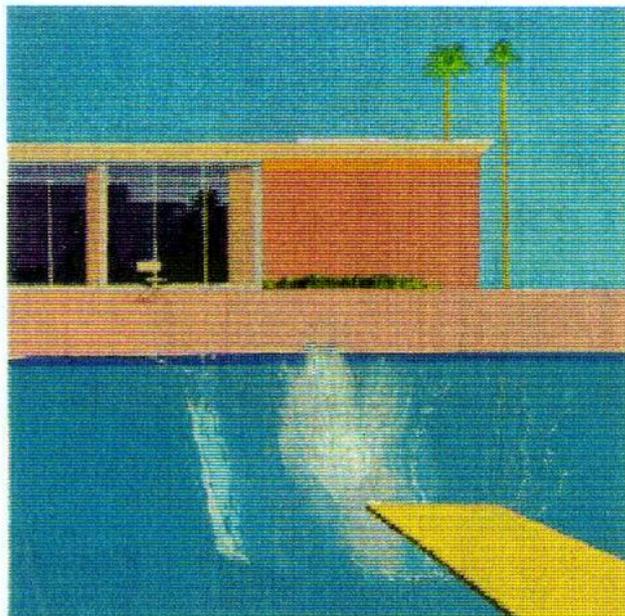


Picture 27

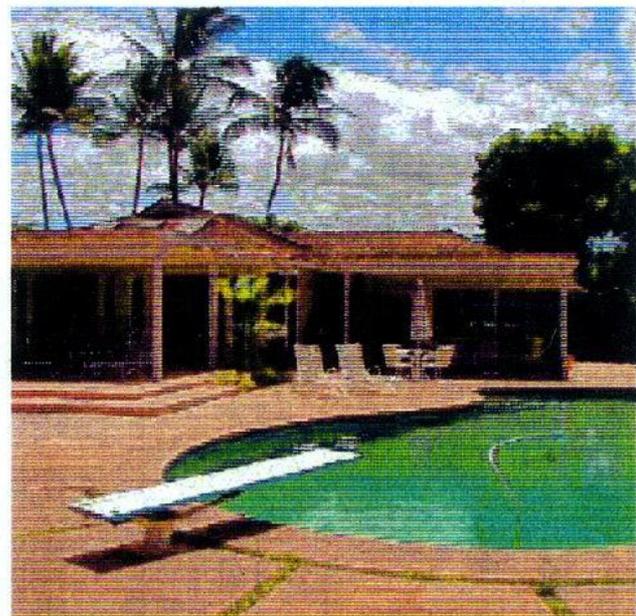


Picture 42

Hockney 1:



Picture 38



Picture 48

Hockney 2:

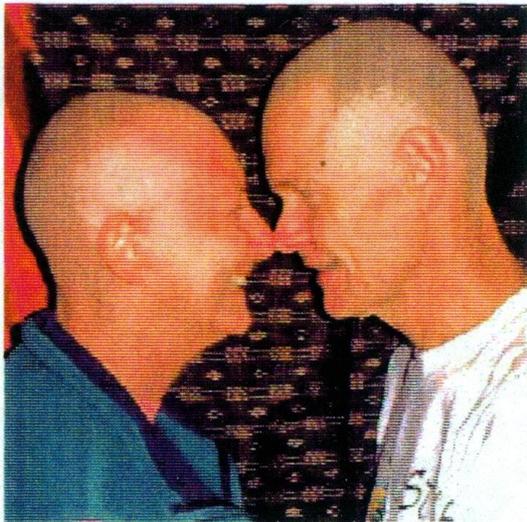


Picture 28

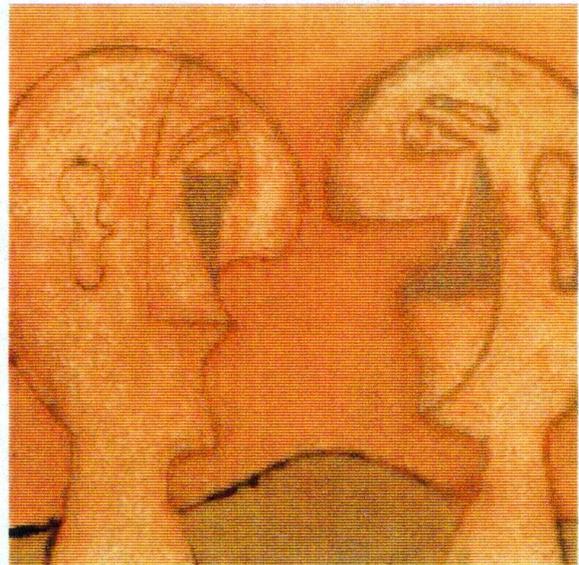


Picture 33

Klee:

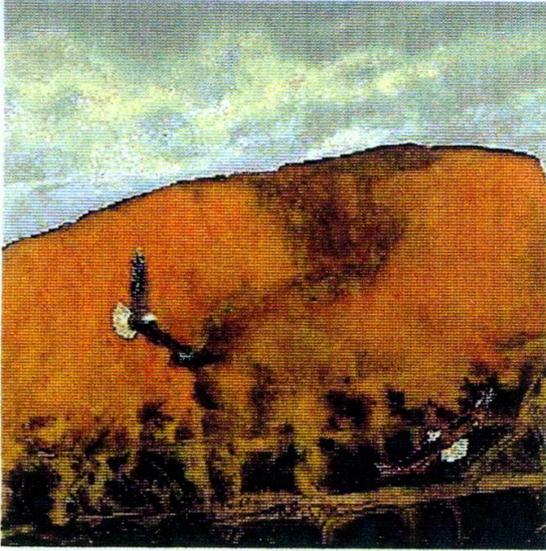


Picture 13



Picture 24

Leland:

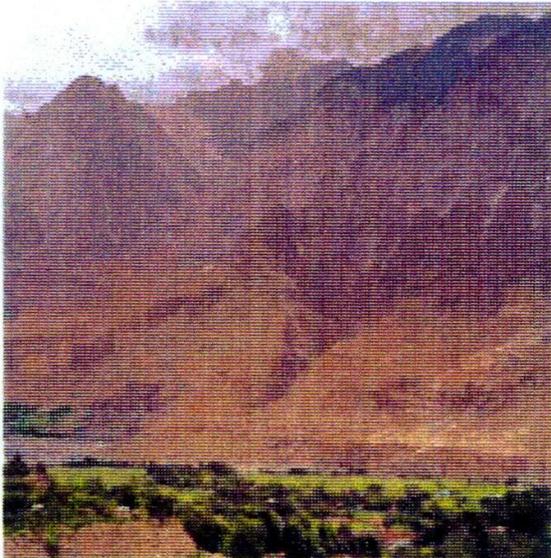


Picture 11

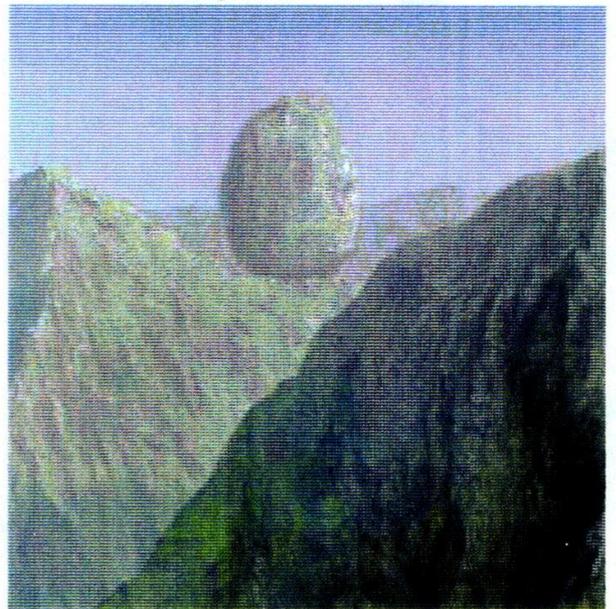


Picture 16

Magritte:



Picture 10



Picture 20

Matisse:



Picture 19

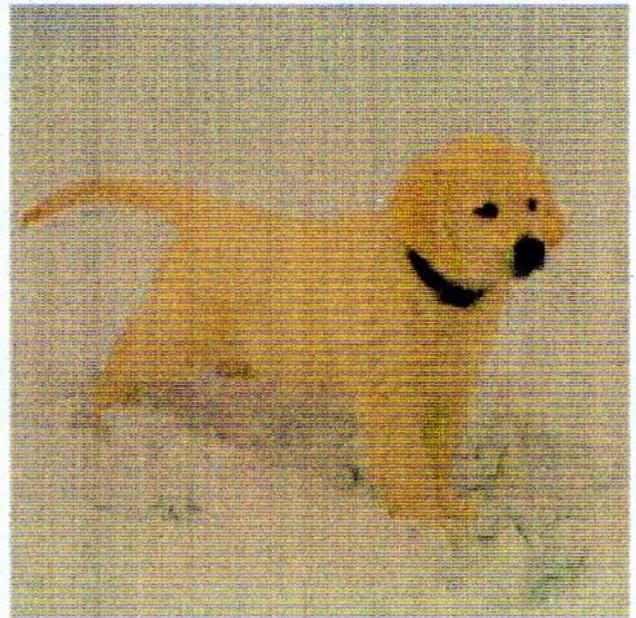


Picture 25

Marc:



Picture 30



Picture 39

Monet 1:

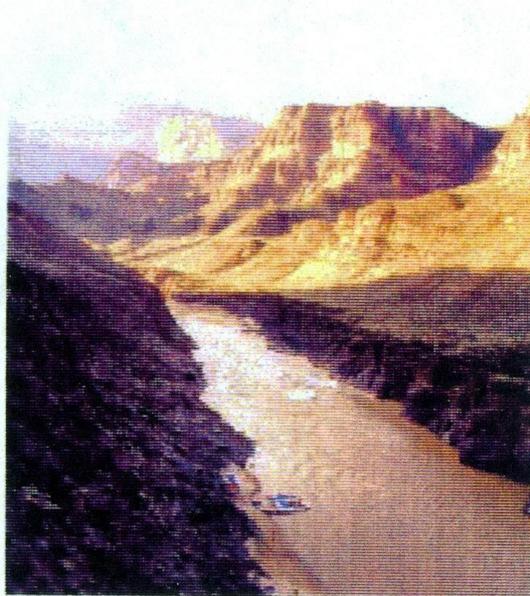


Picture 5

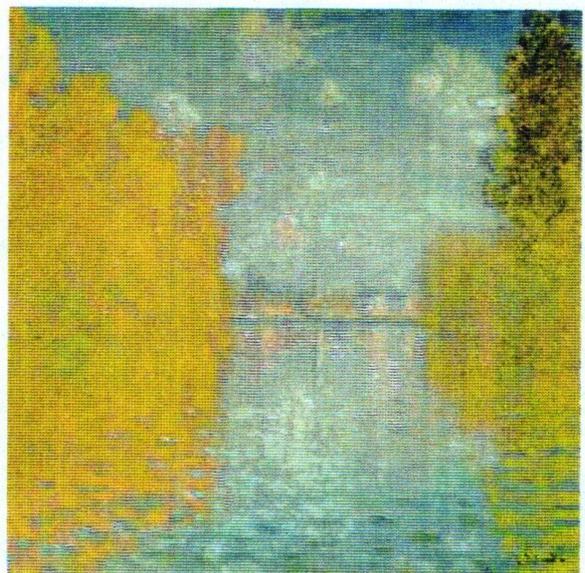


Picture 49

Monet 2:

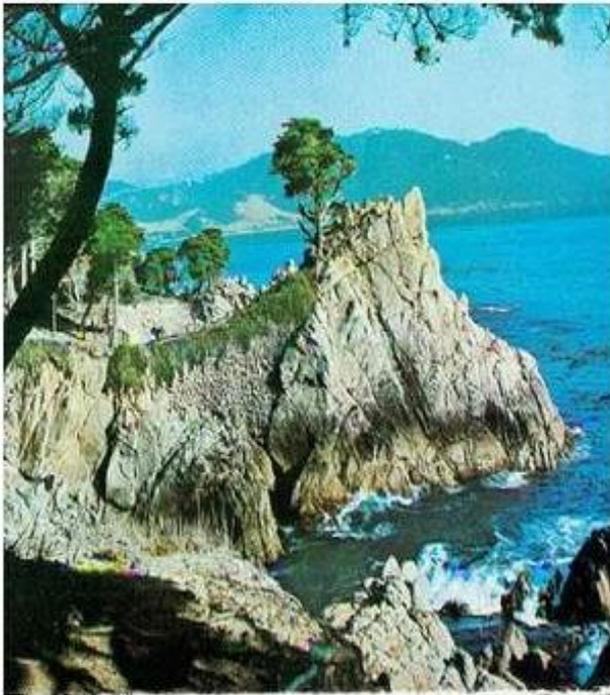


Picture 12

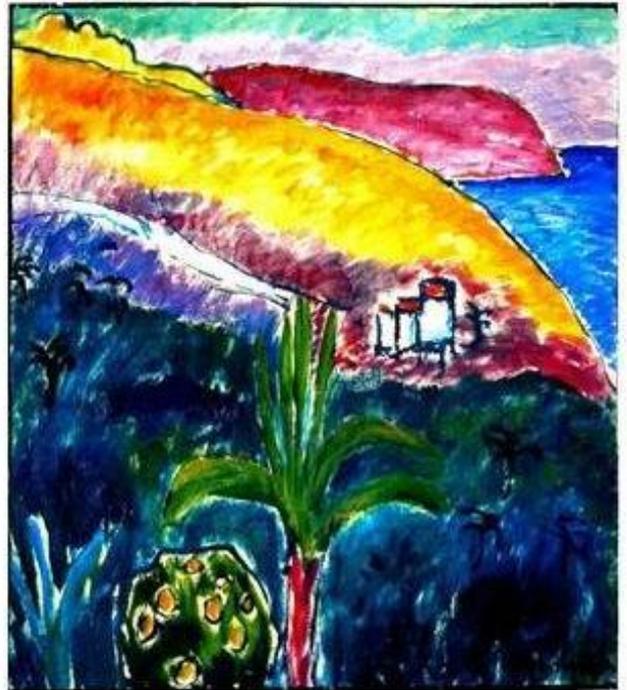


Picture 31

Pechstein:

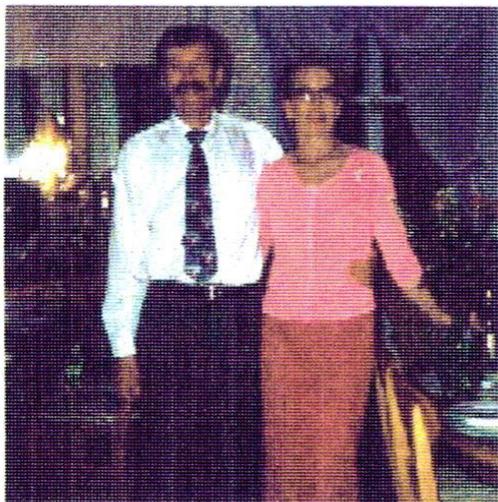


Picture 4

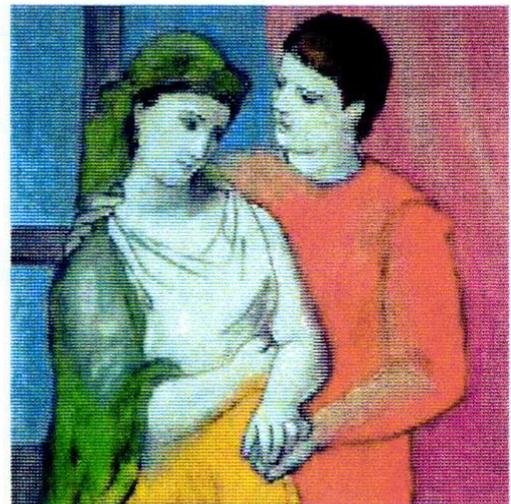


Picture 29

Picasso:



Picture 6



Picture 32

Schiele:

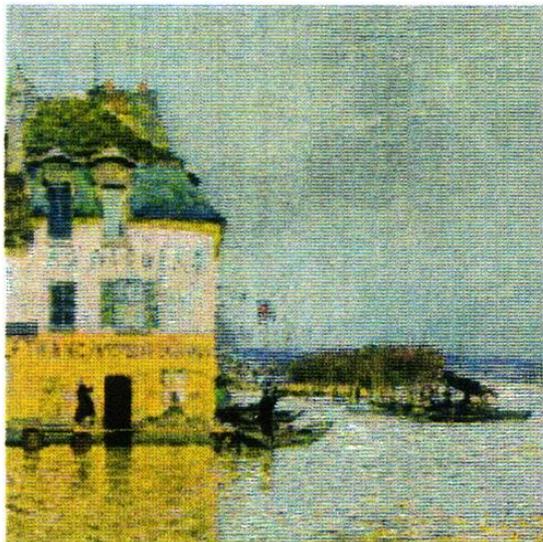


Picture 4



Picture 34

Sisley:



Picture 1



Picture 3