

In the Eye of the Beholder - Perception of Indeterminate Art

paper1017

Abstract

How do we interpret an object - a scene - a painting? Perception research and art illuminate from different angles how the vast amount of information in our visually perceived environment is processed by the viewer to form a coherent and consistent interpretation of the world. Using drawings and paintings by the artist Robert Pepperell, this work attempts to connect these different world views. Pepperell's paintings at first glance seem to be a baroque fresco, an expressionist still-life, or a cubist collage; taking a closer look, however, this concrete interpretation vanishes and we are left with an indeterminate painting. Using psychophysical experiments and eyetracking measures, in this work we seek to illuminate the visual processing of information in Pepperell's paintings. More specifically, we will investigate how the pattern of fixations - the loci of interest - change as a function of the task ('What is depicted in this scene?' vs. 'Does this image contain people?') and of the image content. The interpretation of the experimental results in the context of perceptual research will give first insights into the perception of (indeterminate) art. Conversely, the results are also relevant for art, as they provide a kind of perceptual, measurable 'validation' of the artist's intentions.

Categories and Subject Descriptors (according to ACM CCS): J.5 [Computer Application]: Arts and HumanitiesFine arts; J.4 [Computer Application]: Social and Behavioural SciencesPsychology;

1. Introduction

From a perception point of view, the interpretation of paintings (and photographs) is different from that of real-world objects, as paintings only deliver a two-dimensional signal. Our brain, however, has learnt to deal with this loss of the third dimension as the projected image on the retina is also only two-dimensional as a result of the perspective projection occurring in the eye. Statistical priors on the distribution of possible object shapes, light sources, and object materials together with compensation mechanisms that allow us to recognize an object from multiple viewpoints, however, make it possible to reliably and robustly interpret two-dimensional images as projections of a three-dimensional world. Monocular depth cues, for example, such as known object dimensions, occlusions, as well as atmospheric effects can be used to infer the depth structure of a given scene. In a way, the inverse problem exists for the artist: the effects of perspective projection, such as foreshortening, for example, have to be learnt and practiced before a realistic image of the world can be produced on canvas.

Art, however, has found different - and decidedly non-(photo-)realistic - ways of expressing aesthetic intentions. Going beyond changes in surface coloring and distortions of shapes, cubism, for example, has resulted in the com-

plete dissolution of perspective laws. Objects lost their fixed position in space and their fixed relations to other objects and individual parts of objects were split apart. Analytic cubism reduced objects to geometric solids (such as cylinders, spheres, and cones) rendering them in monochromatic colors while indicating depth structures through occlusions and sparse shadowing without losing the decidedly two-dimensional impression of the scene. This cubist dissolution of perspective and object coherency has - according to Robert Pepperell [Pep05, Pep06]- resulted in paintings that hide their interpretations and therefore are able to engage the viewer: 'I realized there are very few instances (in our culture at least) where we are deliberately presented with images that withhold their meaning. Among the most brilliant examples in fine art are the paintings by Braque and Picasso made during the so-called 'analytic' phase of cubism around 1910-1914. These paintings confound, tease and frustrate the very act of pictorial recognition.'

In Pepperell's paintings, this ambiguity is taken even further: when in cubist paintings fragments of a guitar or a person can still be recognized, viewers are denied this interpretation possibility in those paintings that succeed according to his intentions [Pep05]: 'Those drawings suggesting the most potential images but fewest actual images I regard

as most successful'. The question for these images is then: what visual information drives this ambiguity? Perceptual theories of scene recognition, for example, show that objects can be better recognized when they are embedded in their context - in Pepperell's indeterminate art, however, this context remains ambiguous [Pep05]: 'The resulting drawing arises from no preconception about what it should 'be', nor does it arrive at any definitive representation of something other than itself [...]. But somehow in the process of making it becomes 'clear' or 'complete', despite in fact being neither clearly 'of' something nor finite in meaning.' Nevertheless, some of the paintings strongly suggest a genre (such as a still-life, a landscape scene, or similar) or even defined objects (a human figure, flowers).

In this paper, we want to investigate the perception of indeterminate art using eye-tracking techniques in order to take first steps towards characterizing the visual information that might drive interpretation of those paintings. Furthermore, we will examine the different types of visual information that are used to solve two visual tasks when looking at Pepperell's art. The results will help to shed light both on perceptual processing of (ambiguous) scenes as well as on how viewers look at such works of art.

2. Related Work

The first investigation into visual processing of scenes using eye-tracking was done in the early thirties [Bus35]. Participants viewed photographs of various types of fine art, and the resulting scan paths were analyzed. Interestingly, no differences were found between trained and untrained artists (see [NLK93], who found differences between those two groups using a more detailed analysis) - furthermore, no two scan-paths were exactly identical, although global patterns could be derived that defined two overall search strategies of global and local scan paths. Participants also tended to focus on high-contrast regions in the foreground (including faces and people) rather than on the background (see [CBN05] for an interesting study which found that such search strategies depend on the cultural background of the observer). For a review of research done by Locher et al. on eye movements during viewing of works of art, see [Loc06].

In perhaps the most well-known study, Alfred Yarbus in 1967 analyzed the scan paths of an observer viewing I.E. Repin's painting entitled 'They Did Not Expect Him' [Yar67]. Surprisingly, eye movement patterns changed dramatically depending on viewing instructions. For each instruction, the most informative regions of the image were scanned and processed. Similarly, in experiments by Molner [Mol81], participants who were told that they were going to be asked about the aesthetic qualities of a painting made longer fixations than participants who were going to be asked about the contents. Recent experiments on the role of task context on eye-movements have found that fixations fall primarily on task-relevant objects: studies of fixations made during mak-

ing a cup of tea [LMR99], for example, revealed that during pouring of the tea, participants viewed the spout of the teapot, for making a sandwich [HSMP03], participants likewise only focused on relevant objects in the scene such as the plate, the knife, the peanut butter, etc (for a recent overview, see [HB05]).

More rigorous analyses of the visual information contained in the fixations [KRH*00] have found that in comparison to random regions, fixations contain higher spatial variance, i.e., a higher information content. Those regions often correspond to second-order features such as curves, corners, and occlusions rather than first-order changes in intensity such as straight lines and edges. In addition, in the temporal context, fixations have been found to be made in immediate response to the current task demands minimizing local uncertainties rather than minimizing global uncertainty [RVC07].

3. Eye-Tracking Experiment

In the following, we will describe the eye-tracking experiment that was conducted to investigate the perception of indeterminate art. As is known from the early experiments by Yarbus, fixations are critically dependent on the task: How one looks at an image or a painting thus differs for a given task, which means that the pattern of fixations can, for example, be used to characterize the visual information that is necessary for this task. More specifically, in the experiment, we were interested to investigate the difference between local and global processing of Pepperell's paintings. For this we asked participants to view a set of paintings with two different visual tasks: in order to test local information processing strategies we asked whether a person could be found on the painting (person task), whereas global information processing was tested by asking to categorize the painting into one of 7 pre-defined categories (categorization task).

3.1. Methods

A total of 30 indeterminate drawings and paintings were shown to 20 participants. These 15 paintings were split into two sets of 7 and 8 paintings that were used in each task. The first 10 participants saw the first set for the person task and the second set for the categorization task, whereas for the remaining 10 participants we changed the sets for each task in order to counterbalance for possible order effects. Twenty participants, all of them with normal or corrected-to-normal vision, viewed the paintings from a distance of 50cm while resting their head on a chin-rest. Eyetracking data was collected with an 'Iview'-Eyetracker using a master PC for stimulus display with the software package Psychtoolbox [Bra97] and a synchronized control PC for recording of the data. All images were shown centred on a 21" monitor at a resolution of 1024x768 pixels - as the paintings were of different formats, image size varied accordingly



Figure 1: All 30 drawings and paintings used in this study

from 500x500 pixels of visual angle to 700x600 pixels - on average images subtended $\approx 20^\circ \times 20^\circ$ of visual angle. Each trial consisted of a fixation cross that was shown for 3 seconds followed by a painting that was shown for 30 seconds. Participants were asked not to direct their gaze outside of the image area during this period. Afterwards, a question for each of the two tasks was shown on the screen and participants had to answer by selecting the appropriate item on a list. For the person task, the question was 'Was there a person in this image?' whereas for the categorization task the question was 'To which of the following categories does this image belong?' - here the choice consisted of 7 categories ('Biblical scenes', 'Landscapes with person', 'Landscape without a person', 'Portrait', 'Stilllife', 'Battle scene', and 'None of the above' - these categories were determined in prior pilot experiments). For this part there was no time limit imposed. A keypress started the next trial.

Before each experiment, a calibration of the eyetracker had to be done in order to align the camera of the 'Iview'-System to each participant. For this, we used a standard 9-point calibration sheet, which consisted of nine points with fixed pixel positions that were distributed on the screen and which participants were asked to fixate in a fixed order.

3.2. Analysis of Behavioral Data

Responses to the categorization task were collected and the genre which received the most responses was chosen as the winning category. In addition, we calculated a consistency rating for each image, which sums up the differences between the number of responses for the winning category and all other categories (without the 'none-of-the-above' option) and divides by the total number of responses - a value of 0.0 would mean totally inconsistent category responses, whereas a value of 1.0 would mean maximum consistency. Similarly, we also collected the number of participants which detected a person in each image. The results are summarized in Fig. 1.

In general, as the data shows, participants were able to answer consistently for some images - given the ambiguous nature of the paintings, an average consistency of 0.49 shows that these images, indeed, correspond nicely to Pepperell's intention of indeterminate art. Accordingly, participants used the 'none-of-the-above' category for 26% of all responses again showing that some paintings defy easy categorization.

3.3. Analysis of Eyetracking Data

For the analysis, we examined the behavioral data (consisting of the answer as well as the reaction time) and the eye-tracking data that consist of the individual fixations for each image. The raw data of the eyetracker, however, still contain saccades and eye blinks which need to be removed for the final analysis. For this, we used a toolbox for Matlab called

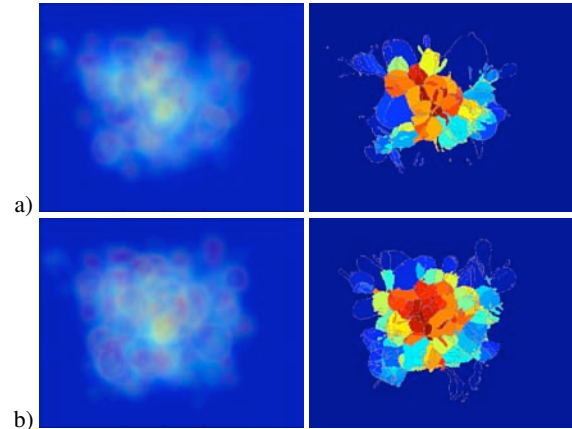


Figure 2: Averaged fixation map and fixation clusters for the a) categorization and the b) person task

'Ilab' [Git02] which processes the raw eyetracking data to filter out the fixations. This requires the specification of two parameters that determine the minimum duration t_{fix} ms for which the eye gaze stays within a given region of diameter r_{fix} pixels. Here, we chose $t_{fix} = 100$ ms and $r_{fix} = 5$ pixels, where the latter corresponds to a fixation region of $\approx 1^\circ$ of visual angle.

The fixation analysis is visualized in this paper using the following procedure: First, for each painting and each task, we place at all fixation locations for a single participant a Gaussian 'blob' of diameter $\approx 2^\circ$ visual angle which represents a rough approximation of the foveated area. These individual maps are then averaged over all 10 participants. Finally, in order to aid interpretation of the fixated areas, each map is alpha-blended with the original painting.

3.4. Overall Results

In the following, we will first discuss the averaged fixation patterns for all paintings for each of the two tasks. Fig. 2 shows the averaged fixation maps for each task on the left and a processed version showing *clusters of fixations* on the right. For the processed versions, we applied a watershed transformation in Matlab to the fixation maps and visualized the resulting cluster membership using a 'jet' false-color map. The watershed transformation can be imagined as if the image intensities were treated as three-dimensional hills and valleys - water is then poured into the image and depending on the water level, different parts of the image become submerged with the visible parts forming connected clusters.

One of the differences between the two tasks might be the scale with which information needs to be processed: scene categorization will presumably rely on global information requiring only few fixations as opposed to person detection

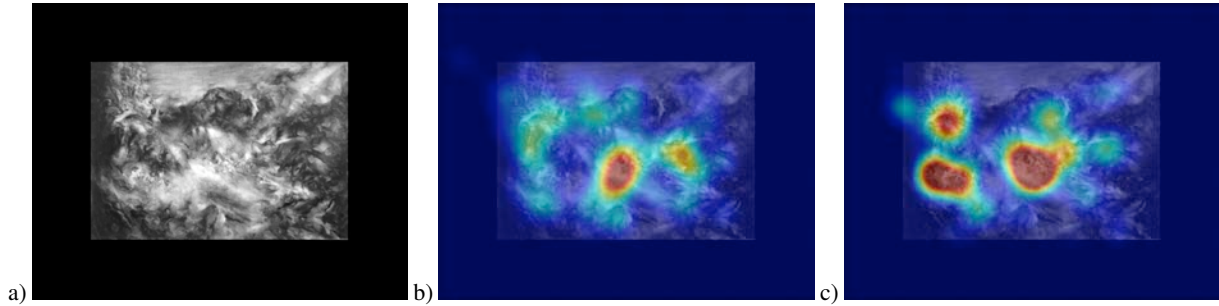


Figure 3: a) Image 16, distribution of fixations for b) categorization c) person task

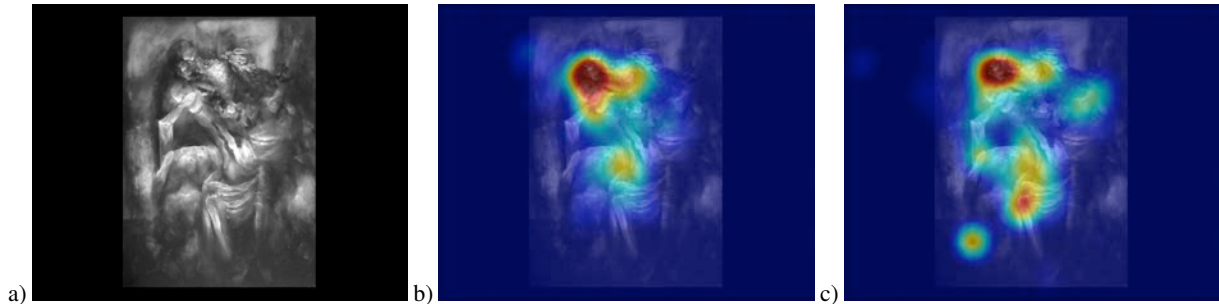


Figure 4: a) Image 18, distribution of fixations for b) categorization c) person task

for which the whole image and specific locations need to be scanned. Indeed, this pattern was confirmed in the experiments: Fig. 2a) shows a much more centered distribution of fixations, whereas Fig. 2b) shows a broader, more localized spread. The processed versions showing the fixation clusters demonstrate this effect even clearer. The average number of fixation clusters in the categorization task is **288**, whereas for the person detection task it is much higher at **413** unique clusters.

Note also, that although the two tasks differ in terms of their demands and search strategies they also share some commonalities - the search for a person, for example, seems also required if a painting is to be classified as a 'Biblical scene', a 'Landscape with person', or a 'Portrait'. Indeed, if we repeat the cluster analysis just for those paintings that were grouped into these categories, the number of clusters in the categorization task rises to **340** - it still does not, however, reach the granularity of the person task showing that clear differences remain between the two tasks.

In summary, the averaged fixation maps and the cluster analysis have shown that that global information seems to be more important for the categorization task than local information. This finding is in accordance with current approaches to scene categorization - psychophysical studies conducted by Vogel et al. [VSWB06], for example, show that scene categorization relies critically on global information: 'Categories with many different local semantic concepts presented in an image [...] require global context information for categorization.' Global information is defined 'as

the overall 'context' of a scene generated through the presence of large spatial structures (e.g. horizon lines) and the spatial arrangement of lighter and darker blobs in an image'.

3.5. Image Analysis

In the following, we will take a closer look at a few selected paintings that are representative of the general trends we observed in the data and interpret the fixation patterns and behavioral results for each of them.

3.5.1. Image 16 - Fig. 3

Categorization task: This image was classified as a landscape (with or without a person) in the categorization task. The fixation pattern clearly reflects the search strategy discussed above: participants during categorization scanned nearly all of the area of the painting compared to the person task which has a much more localized distribution.

Person task: 7 out of 10 participants recognized a person in this painting. As this painting was unanimously classified as a landscape scene, we assume that participants looked for persons in the image that were 'at a distance', i.e., objects that occupied only a small image area. The center of the image, which gathered the largest amount of fixations in this task, does indeed contain a few possible items of interest. This area is especially salient as its immediate neighborhood only contains rather low-contrast areas. Fixated areas also do not contain high differences in contrast - taking into account, however, the context of this landscape scene, persons in the

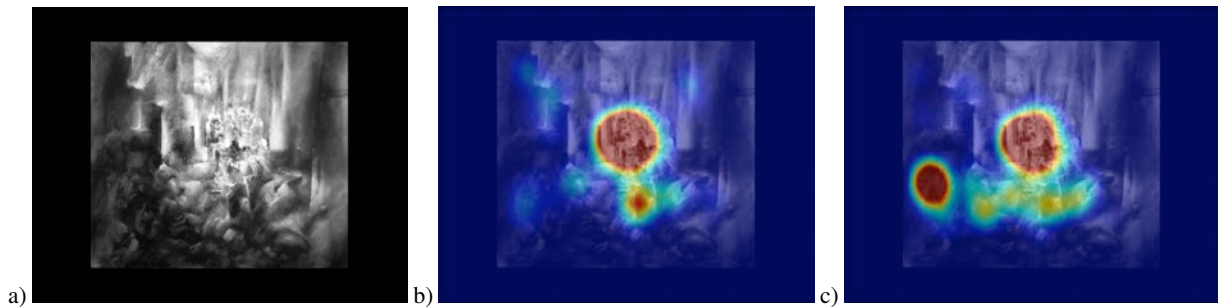


Figure 5: a) Image 22, distribution of fixations for b) categorization c) person task

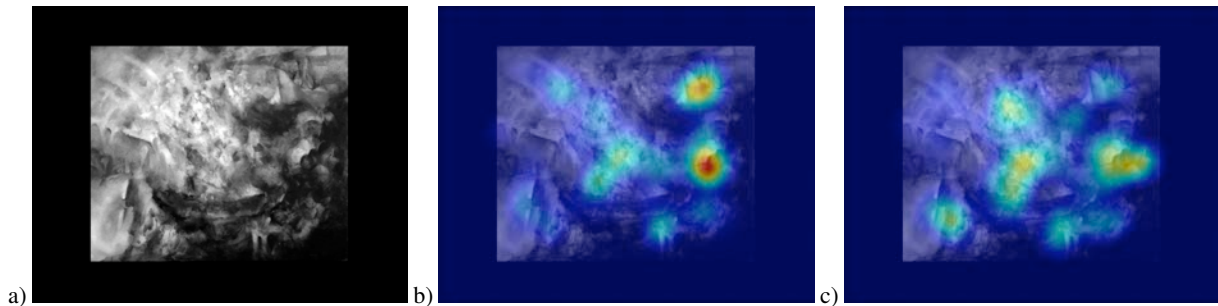


Figure 6: a) Image 19, distribution of fixations for b) categorization c) person task

image might also be expected to be rendered in low contrast due to distance to the virtual observer.

3.5.2. Image 18 - Fig. 4

Categorization task: This image was categorized as a portrait. Here, the pattern of fixations shows one of the few deviations from the general trend: the categorization task results in a less global scan pattern than the person task. The observed, local pattern of fixations, however, can be explained by the unambiguous classification as a portrait resulting in a strong focus on the 'face' of the person. The category decision therefore had been possible already at a very early stage making an exhaustive search in the image unnecessary.

Person task: 8 out of 10 participants detected a person in this painting. The main focus of fixations lies in a contrast-rich region of the painting, which contains very salient features such as curves and edges. The interplay of these image features, indeed, bear resemblance to a human skull (eye sockets and nose). A second focus area is put on the lower corner of the painting: this area is located at a location where one would expect a human hip given the scale of the scene. The full pattern of fixations can be seen as describing the outline of a person - elbows and shoulders are further loci of attention showing that observers tended to follow a full outline of the imagined figure in the image.

3.5.3. Image 22 - Fig. 5

Categorization task: This painting was classified as a biblical scene. The main focus lies in the centre of the paint-



Figure 7: Profile of a person found in the image.

ing, which is an area with high contrast surrounded by border regions that are more 'blurred'. Surprisingly, the high-contrast 'columns' in the image do not draw much attention which shows that the categorization task could be solved using high-resolution information of the centre alone.

Person task: 6 participants recognized a person in this image. Again, the main focus of fixations for this task falls in the centre of the painting which contains the highest density of 'people-like' features. Interestingly, another main focus is on an area in the dark shadows at the left border of the image - closer examination of this area reveals that, indeed, this could be interpreted as the profile of a person (Fig. 7). This small but very consistent effect again demonstrates the crucial differences between the types of information needed for both tasks.

3.5.4. Image 19 compared to Image 8 - Figs. 6,8

In image 19, no participant was able to spot a person - in addition, the painting was categorized as 'none-of-the-above'

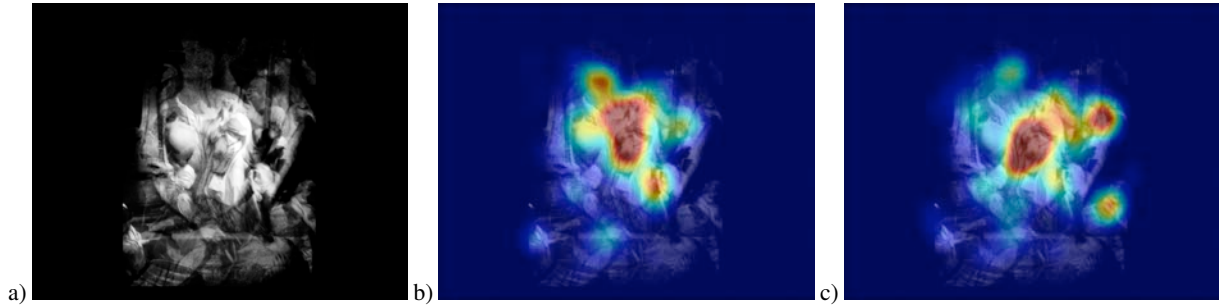


Figure 8: a) Image 8, distribution of fixations for b) categorization c) person task

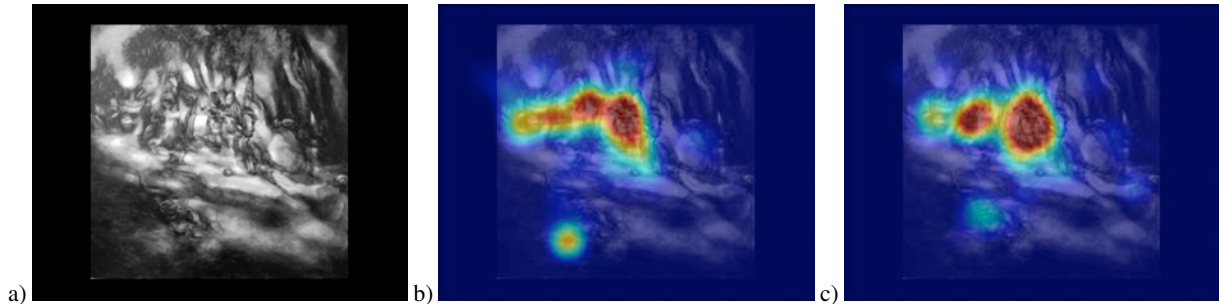


Figure 9: a) Image 3, distribution of fixations for b) categorization c) person task

not showing any clear tendencies towards a consistent classification. Accordingly, very few consistent fixation regions were found in the analysis for both tasks. It seems that regions rich in local contrast were fixated more than other regions - none of the shapes, curves, or edges, however, seems to deliver enough information to recognize a person or to unambiguously classify the genre.

In contrast to image 19, 9 out of 10 participants detected a person in image 8. Similar to the previous images, the main focus lies in the centre of the painting on an object, which could be seen as a veiled person - a bedouin (see Fig. 3.5.4a). An additional fixation spot on a round shape in the right area of the painting could also be interpreted as a potential face. Interestingly, in the categorization task (the image was classified as a biblical scene), another region of the image was found to be of importance that was not fixated in the person detection task: a fixation spot in the left part of the painting that might resemble a 'vase' (see Fig. 3.5.4b). This clearly shows the difference in attentional localization depending on the context of the image: person detection again focuses only on candidate regions that are already rather similar to human features whereas categorization focuses also on other parts ('objects') of the scene.

3.5.5. Image 3: composition - Fig. 9

The distribution of fixation spots in this painting (which was classified as a landscape with or without a person) exemplifies just how much image composition can actually in-

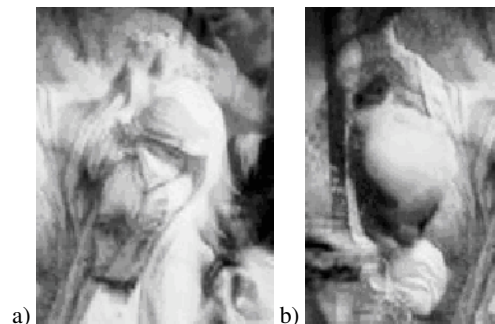


Figure 10: a) Detail from Image 8: the bedouin, b) Detail from Image 8: the vase

fluence eye gaze pattern. Fixations in this painting clearly fall along the border between the 'ground' and the 'forest'. Another consistent fixation pattern that can be observed in both tasks is a detail in the otherwise smooth 'ground' that seems to draw attention - this is possibly due to it being a high-contrast region in an otherwise low-contrast neighborhood. The difference between the two tasks lies mainly in the stronger focus in the person task on the middle of the scene where one might expect a person to appear.

3.6. Questionnaires

After the eye-tracking experiment, we asked participants to fill out a questionnaire with questions related to their expertise in art, the difficulty of the task, and most importantly to

the strategies they used to solve the two different tasks. On a scale of 1 (low level of expertise) to 5 (high level of expertise), on average participants rated themselves as an average 3.11 (± 0.88). Analysis of the last question revealed that participants were able to give quite detailed accounts of those strategies - both in terms of the features they were looking for, as well as of the differences between the two tasks. Search strategies that were mentioned in general, were, for example, the search for 'salient points' in the image to try and 'compose those small features into larger objects'. Participants also mentioned that they were looking for 'sharp contrasts' and 'known shape-features' to infer the context of a scene from those. In addition, the 'overall impression' of the painting as well as 'associations to well-known works of art' were indicated as important. For the categorization task, these were also rated as the most important types of information: the 'first impression' as well as the overall 'look' or 'impression of the scene' - these introspective ratings correspond very well to the global pattern of fixations that we found for this task. Conversely, local information was given more importance for the person task: unanimously, participants were looking for faces or parts of faces and bodies.

4. Conclusions

In these experiments, we have tried to take first steps towards characterizing the visual information in the processing of modern art for two specific tasks. Using the behavioral data, we have been able to validate Pepperell's artistic intentions of visual indeterminacy to a certain degree: participants' consistency in determining a genre for a painting or drawing was on average neither high nor low - in addition, participants made use of the 'none-of-the-above' option for a quarter of all responses. For a more objective validation, of course, one will need to compare these responses to a set of representational and fully abstract images in order to get a proper baseline - these experiments are currently underway.

Analysis of the eye movements has revealed clear differences between the two tasks: a preference for a more global search strategy in the categorization task compared to the person detection task was found, which is in accordance with current theories of scene processing. In addition, we have confirmed earlier findings on selection of fixation regions: on average - and perhaps not surprisingly - high-contrast regions are preferred over low-contrast regions; in addition, it seems that only immediately relevant parts of the scene are fixated. Interestingly, we could not replicate the general search pattern identified, for example, by [Loc06] of a coarse, global scan path followed by more small-scale local fixations. This was most probably due to the task constraints which may have biased participants towards very specific global and local strategies, respectively. Again, in order to make a more broad argument about the visual information, which also will analyze the low-level image properties of the fixated regions, we are currently working on follow-up ex-

periments with more paintings and drawings from Pepperell compared to both representational and fully abstract works of art.

References

- [Bra97] BRAINARD D. H.: The Psychophysics Toolbox. *Spatial Vision* 10, 4 (1997), 433–436.
- [Bus35] BUSWELL G.: *How people look at pictures: A study of the psychology of perception in art*. Univ. Chicago Press, Chicago, 1935.
- [CBN05] CHUA H. F., BOLAND J. E., NISBETT R. E.: Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of the Sciences USA* 102, 35 (Aug 2005), 12629–12633. Comparative Study.
- [Git02] GITELMAN D. R.: ILAB: a program for postexperimental eye movement analysis. *Behavioral Research Methods Instruments Computers* 34, 4 (Nov 2002), 605–612.
- [HB05] HAYHOE M. M., BALLARD D.: Eye movements in natural behavior. *Trends Cogn Sci* 9, 4 (Apr 2005), 188–194.
- [HSMP03] HAYHOE M. M., SHRIVASTAVA A., MRUCZEK R., PELZ J. B.: Visual memory and motor planning in a natural task. *J. Vis.* 3, 1 (2 2003), 49–63.
- [KRH*00] KRIEGER G., RENTSCHLER I., HAUSKE G., SCHILL K., ZETZSCHE C.: Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics. *Spat Vis* 13, 2-3 (2000), 201–214. Comparative Study.
- [LMR99] LAND M., MENNIE N., RUSTED J.: The roles of vision and eye movements in the control of activities of daily living. *Perception* 28, 11 (1999), 1311–1328.
- [Loc06] LOCHER P. J.: The usefulness of eye movement recordings to subject an aesthetic episode with visual art to empirical scrutiny. *Psychology Science* 48 (2006).
- [Mol81] MOLNAR F.: *Advances in Intrinsic Motivation and Aesthetics*. Plenum Press, New York, 1981, ch. About the role of visual exploration in aesthetics, pp. 385–413.
- [NLK93] NODINE C., LOCHER P., KRUPINSKI E.: The role of formal art training on the perception and aesthetic judgment of art compositions. *Leonardo* 26 (1993), 219–227.
- [Pep05] PEPPERELL R.: Visual indeterminacy and the paradoxes of consciousness. In *Proceedings of Altered States Conference* (2005).
- [Pep06] PEPPERELL R.: Seeing without objects: Visual indeterminacy and art. *Leonardo* 39 (2006).
- [RVC07] RENNINGER L., VERGHESE P., COUGHLAN J.: Where to look next? Eye movements reduce local uncertainty. *J. Vis.* 7, 3 (2 2007), 1–17.
- [VSWB06] VOGEL J., SCHWANINGER A., WALLRAVEN C., BÜLTHOFF H. H.: Categorization of natural scenes: local vs. global information. In *APGV '06: Proceedings of the 3rd symposium on Applied perception in graphics and visualization* (New York, NY, USA, 2006), ACM Press, pp. 33–40.
- [Yar67] YARBUS A.: *Eye movements and vision*. Plenum Press, New York, 1967.