

# **Generating Mondrian-like Art using Interactive Genetic Algorithms**

**Cameron Burke**

A subthesis submitted in partial fulfillment of the degree of  
Bachelor of Information Technology (Honours) at  
The Research School of Computer Science  
Australian National University

October 2012

© Cameron Burke

Typeset in Palatino by T<sub>E</sub>X and L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub>.

Except where otherwise indicated, this thesis is my own original work.

Cameron Burke

25 October 2012



---

# Abstract

---

Interactive genetic algorithms are a technique that optimises a solution to a problem based on human evaluation. Interactive genetic algorithms have been used to solve a number of problems that require human evaluation or creativity, such as evaluating emotion of images, generating music, and generating art.

This thesis examines the performance of interactive genetic algorithms to generate abstract art in the style of Piet Mondrian. Specifically, this thesis investigates the effect of different representation strategies on the performance of interactive genetic algorithms designed for this problem. Previous research in this field has not examined this effect, and consequently there is little documentation on which representation strategies perform the best.

To determine the effect of different representation strategies, several methods to represent Mondrian-like art in a genetic algorithm are created and described. These representations are then analysed to determine how the different representation schemes contribute to the performance of the genetic algorithms. The results of this analysis shows that, with the correct representation strategy, interactive genetic algorithms can generate Mondrian-like abstract art that is more aesthetically pleasing to the individual users of the algorithm than original Mondrian compositions.



---

# Contents

---

<b>Abstract</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Purpose . . . . .	2
1.2 Organisation . . . . .	2
<b>2 Background</b>	<b>5</b>
2.1 Piet Mondrian . . . . .	5
2.2 Mondrian Compositions . . . . .	6
2.3 Evolutionary Algorithms . . . . .	7
2.3.1 Genetic Algorithms . . . . .	9
2.3.1.1 Interactive Genetic Algorithms . . . . .	9
2.3.2 Evolutionary Programming . . . . .	10
2.3.3 Bacterial Evolutionary Algorithms . . . . .	10
<b>3 Literature Review</b>	<b>11</b>
3.1 Mondrian Art Analysis . . . . .	11
3.2 Mondrian Art Generation . . . . .	12
3.2.1 Schnier & Gero . . . . .	12
3.2.2 MONICA . . . . .	13
3.2.3 Darwindrian . . . . .	14
3.3 Interactive Genetic Algorithms . . . . .	15
3.3.1 GenJam . . . . .	15
3.3.2 Lee & Cho . . . . .	16
3.4 Summary . . . . .	16
<b>4 Research Questions</b>	<b>17</b>
<b>5 Representations</b>	<b>19</b>
5.1 Direct Object Representation . . . . .	19
5.2 Statistical Representation . . . . .	20
5.3 Direct/Statistical Hybrid Representation . . . . .	21
5.4 Darwindrian Representation . . . . .	22
5.5 Summary . . . . .	23

---

<b>6</b>	<b>Representation Performance</b>	<b>25</b>
6.1	Methodology . . . . .	25
6.1.1	Algorithm Implementation . . . . .	25
6.1.2	Experimental Design . . . . .	27
6.1.3	Data Collection . . . . .	28
6.1.4	Statistical Analysis . . . . .	28
6.1.4.1	Statistical Tests . . . . .	29
6.1.4.2	Objective Data Analysis . . . . .	29
6.1.4.3	Subjective Data Analysis . . . . .	30
6.2	Analysis Results . . . . .	31
6.2.1	Objective Performance . . . . .	31
6.2.2	Subjective Performance . . . . .	36
6.3	Qualitative Analysis . . . . .	36
6.4	Discussion . . . . .	38
6.4.1	Objective Performance . . . . .	38
6.4.2	Subjective Performance . . . . .	38
6.4.3	Qualitative Analysis . . . . .	39
<b>7</b>	<b>Future Research</b>	<b>41</b>
7.1	Larger Study . . . . .	41
7.2	Paired Study . . . . .	42
7.3	Followup Study . . . . .	42
7.4	Other Extensions . . . . .	42
<b>8</b>	<b>Conclusion</b>	<b>45</b>
<b>A</b>	<b>Mondrian Compositions</b>	<b>47</b>
<b>B</b>	<b>User Interface</b>	<b>53</b>
<b>C</b>	<b>Questionnaire</b>	<b>55</b>
<b>D</b>	<b>Raw Data</b>	<b>57</b>
	<b>Bibliography</b>	<b>59</b>



---

# List of Figures

---

2.1	Five digital recreations of Mondrian compositions . . . . .	6
2.2	Example of chromosome crossover, at position 4 . . . . .	9
5.1	Chromosome configuration of DOR . . . . .	20
5.2	Chromosome configuration of SR . . . . .	21
5.3	Chromosome configuration of DSHR . . . . .	22
5.4	Chromosome configuration of Darwindrian . . . . .	23
6.1	Diagram of steps performed with each participant in the experiment . .	27
6.2	Highest rated Mondrian-like compositions taken from the final generation, generated by the DSHR representation for participants 11 (left pair) & 14 (right pair) . . . . .	37
6.3	Highest rated Mondrian-like compositions taken from the final (left) & middle (right) generations, generated by the DSHR representation for participants 11 (left pair) & 14 (right pair) . . . . .	37
A.1	Composition with Blue, Yellow, Black, and Red (Piet Mondrian, 1922) . .	48
A.2	Composition II in Red, Blue, and Yellow (Piet Mondrian, 1930) . . . . .	49
A.3	Rhythm of Straight Lines (Piet Mondrian, 1935) . . . . .	50
A.4	Composition with Red, Yellow, and Blue (Piet Mondrian, 1927) . . . . .	51
A.5	Composition with Yellow Patch (Piet Mondrian, 1930) . . . . .	52



---

# List of Tables

---

6.1	Mann-Whitney-Wilcoxon tests (as $U$ statistic and $p$ values) comparing objective ratings of Final and Mondrian compositions per participant (1-28), by algorithm (DWN, DOR, SR, & DSHR), with median ( $M$ ) and interquartile range ( $IQR$ ) per group . . . . .	32
6.2	Participants grouped by whether MWW test indicated significance ( $p < 0.05$ ) between Final and Mondrian ratings, per algorithm (DWN, DOR, SR, & DSHR), with multinomial test indicating significance of the distribution (as $p$ value). . . . .	33
6.3	Pairwise comparisons of distribution of participants in Table 6.2 using Fisher's exact test, given as $p$ value . . . . .	33
6.4	Mann-Whitney-Wilcoxon tests (as $U$ statistic and $p$ values) comparing objective ratings of Final and Middle compositions per participant (1-28), by algorithm (DWN, DOR, SR, & DSHR), with median ( $M$ ) and interquartile range ( $IQR$ ) per group . . . . .	35
6.5	Summary statistics of questionnaire part A responses, median ( $M$ ) & interquartile range ( $IQR$ ) per question (1-4), by algorithm (DWN, DOR, SR, & DSHR) . . . . .	36
6.6	Kruskal-Wallis tests (as $H$ statistic, $df$ , and $p$ values) of questionnaire part A responses by question (1-4) . . . . .	36
D.1	Ratings of each composition per participant (1-28), by type of composition (from final generation, middle generation, or Mondrian composition), grouped by algorithm . . . . .	57
D.2	Responses to part A of the questionnaire per participant (1-28), by question (Q1-Q4), grouped by algorithm . . . . .	58



# Introduction

---

It is a commonly held belief that creativity is fundamentally a human trait, and since art is a product of creativity it is therefore impossible for a computer to truly create art. The dictionary definition of ‘art’ reflects this belief, stating that art is “The expression or application of human creative skill and imagination. . .” [Oxford English Dictionary 2010].

However, artificial intelligence holds promise that computers may one day be able to be as creative as humans. Interactive genetic algorithms, a type of artificial intelligence, are bridging the gap between computers and humans by using human input to perform tasks that computers currently cannot, such as evaluating whether a piece of art is aesthetically pleasing.

This use of human evaluation is not without its own problems. The speed of an algorithm that relies on regular human input is many times slower than an algorithm that can run completely autonomously. In addition, a human subject is usually unwilling to repeat the same task too many times, whereas no such limitation is imposed by a computer. These limitations mean that an algorithm which uses human evaluation must be very efficient in its use of that evaluation, when compared to an algorithm that uses computer evaluation.

Interactive genetic algorithms are a subset of genetic algorithms, which are a type of search algorithm that borrows ideas and techniques from natural evolution, such as reproduction, mutation, and selection. Genetic algorithms differ from traditional search algorithms in that they perform well in searching large problem spaces, or problem spaces that are otherwise hard to navigate using conventional methods.

Each candidate solution to the problem plays the role of an individual in the genetic algorithm, which is analogous to an individual organism in natural evolution. These individuals are evaluated to determine their evolutionary fitness, which is then used to determine which individuals should breed to create new individuals. In an interactive genetic algorithm, this evaluation is performed by human interaction, rather than through computer analysis.

Interactive genetic algorithms have been applied to many problems that traditionally require human creativity, such as generating jazz solos [Biles 1994], image classification based on emotion [Lee and Cho 1999], and generating abstract art [Shen 2007]. It is this last application that will be examined further in this thesis.

Abstract art in the style of Piet Mondrian is a popular choice for studies in computer generated art, for two primary reasons. Firstly, artwork in this style is comprised of simple, well defined elements, which are far easier to generate programmatically compared to more complex art. Secondly, Mondrian claims his placement of each component on the canvas was “with awareness” and “led by high intuition” [Mondrian 1914], indicating that he was very precise in his placement of each element. This leads to the assumption that, while composed of simple elements, there is some difficulty in creating a Mondrian composition that is aesthetically pleasing.

## 1.1 Purpose

This thesis will examine different approaches that have previously been used to generate Mondrian-like art using genetic algorithms, focusing specifically on interactive genetic algorithms.

Three different new approaches are then proposed, and the performance of each approach is compared against the other approaches and previous work. These approaches will be tested in terms of their performance in generating artwork that is more aesthetically pleasing to the user than original Mondrian artwork, as well as subjective performance of each algorithm in engaging the user.

The results of this thesis aim to improve the understanding of how the implementation of an interactive genetic algorithm affects the performance of the algorithm. This understanding will improve the ability of future researchers to create effective interactive genetic algorithms.

## 1.2 Organisation

Chapter 2 introduces the concepts that will be used throughout the rest of the thesis. The artwork of Piet Mondrian is discussed, and a set of rules that this style follows is put forward. Evolutionary algorithms are then introduced, with particular emphasis on genetic algorithms.

Chapter 3 reviews and analyses previous work in this area. Firstly, literature examining the aesthetic properties of Mondrian art is discussed. Then, various evolutionary algorithms that have been created to generate Mondrian-like art are compared

---

and analysed. Finally, different implementations of interactive genetic algorithms in a number of areas are reviewed.

Chapter 4 summarises the current state of the field, and states the specific research questions the rest of the thesis aims to answer. This chapter also provides an overview of the methodology that will be used to answer those research questions.

Chapter 5 puts forward three distinct techniques for implementing interactive genetic algorithms, and describes the differences and similarities between these approaches.

Chapter 6 then describes a study that was used to analyse the performance of each of the techniques in Chapter 5. This analysis examined both the objective performance of each algorithm, as well as the subjective performance of how well the participant felt the algorithm was progressing.

Chapter 7 goes on to propose areas of future research, based on the results of Chapter 6. These areas include a more comprehensive study, and extending the results of the study to other applications of interactive genetic algorithms.

Chapter 8 concludes the thesis, summarises the results of the research, and provides the answers to the research questions posed in Chapter 4.





# Background

---

## 2.1 Piet Mondrian

Piet Mondrian was a Dutch artist who lived from 1872 to 1944. His early art was mostly naturalistic or impressionistic, consisting largely of landscapes, portraits, and other natural subjects. Just after the turn of the century his signature abstract artistic style began to develop, influenced heavily by his budding interest in the theosophical movement. This movement hypothesised that it was possible to obtain a more fundamental knowledge of nature than could be found through empirical means. This idea is reflected in the following quote, taken from a letter Mondrian wrote to H.P. Bremmer when attempting to describe his artistic theory:

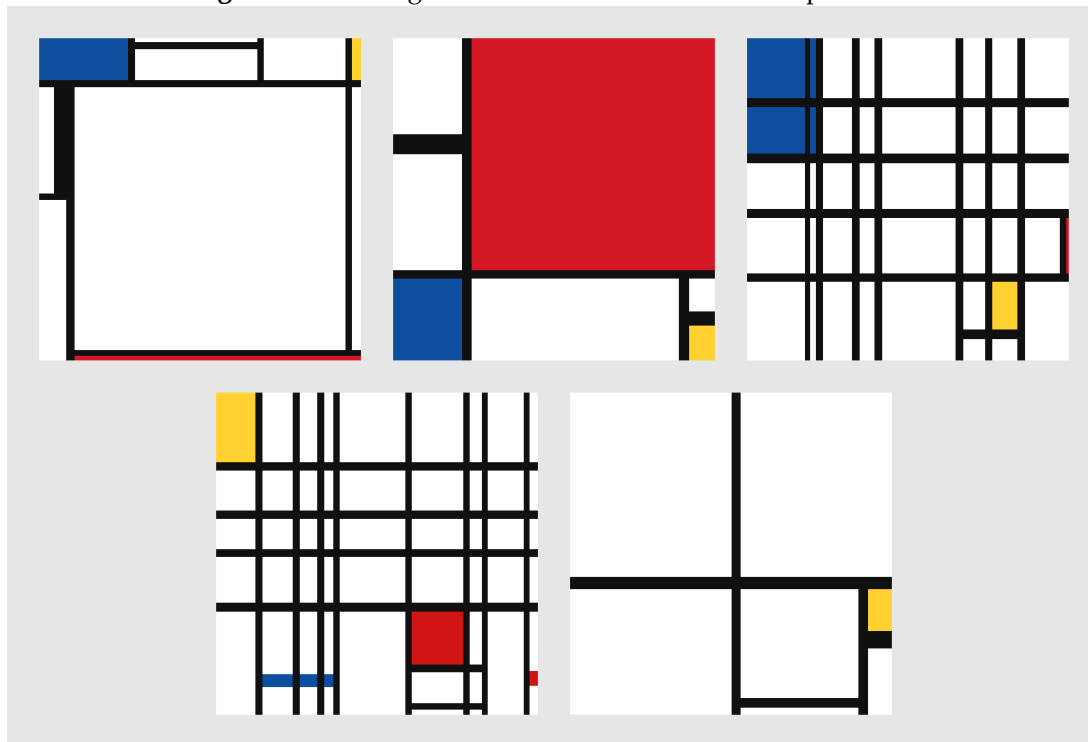
I construct lines and color combinations on a flat surface, in order to express general beauty with the utmost awareness. Nature (or, that which I see) inspires me, puts me, as with any painter, in an emotional state so that an urge comes about to make something, but I want to come as close as possible to the truth and abstract everything from that, until I reach the foundation (still just an external foundation!) of things. . . I believe it is possible that, through horizontal and vertical lines constructed with awareness, but not with calculation, led by high intuition, and brought to harmony and rhythm, these basic forms of beauty, supplemented if necessary by other direct lines or curves, can become a work of art, as strong as it is true. (Mondrian [1914])

By 1920, Mondrian arrived at his signature style – rectangles of primary colours and black lines on a white canvas, each line ending on another line or the edge of the canvas, and each coloured rectangle bordered completely by black lines or the canvas edge. This style manifests itself in the majority of his paintings from this point until his death in 1944. He did not refer to his paintings in this style as artworks, but rather “compositions”.

## 2.2 Mondrian Compositions

The abstract style developed by Piet Mondrian has become to be known by his name, and is the focus of this thesis. Specifically, this thesis will focus on Mondrian's abstract compositions in the style that was present in the majority of his artworks between 1920-1942. Samples of these works are shown in Figure 2.1, and a larger version of each composition can be seen in Appendix A.

Figure 2.1: Five digital recreations of Mondrian compositions



While Mondrian claimed that he did not follow explicit rules when creating his compositions [Reynolds 2005, p. 173], a set of rules can still be determined that these compositions follow. The compositions contain only the colours red, blue, yellow, black, and white. They are composed of horizontal and vertical black lines on a white canvas, with coloured rectangles. The lines in the compositions are placed such that there are no right angles - that is, each line ends on either the body of another line, or the edge of the canvas. Each coloured rectangle is completely enclosed by black lines and the edges of the canvas.

It is important to note that while these rules generally hold true for all of Mondrian's compositions in this style, a composition will occasionally break one of these rules. The most common variation from these rules is that a composition will have one or more (usually small) rectangles with at least one side unbounded by a black line. An example of this can be seen in the fourth composition in Figure 2.1.

---

## 2.3 Evolutionary Algorithms

Evolutionary Algorithms (EAs) are a type of optimisation algorithm based on the process of natural evolution in biology, where the fittest individuals survive and pass their characteristics on to the next generation to create an overall fitter population. In an EA, a candidate solution for a problem plays the role of an individual in the algorithm, and how well it performs in solving the problem is taken to be the evolutionary fitness of that individual.

EAs are suited to searching most solution spaces, including large or complicated spaces. This is because the only requirement for implementing an EA is the ability to determine the fitness of each solution in the search space – no assumptions about the shape and size of the fitness landscape are needed. However, as EAs are meta-heuristic optimisation algorithms, there is no guarantee that the optimal solution will be found.

These qualities of an EA make them a suitable candidate for generating Mondrian-like art programatically. Mondrian compositions can be nearly infinitely varied, so the search space is extremely large by conventional measures. In addition, it is likely that the fitness landscape is very complex, with many local optima or local peaks that represent different aesthetics. Finally, it is not required that the search algorithm find the best solution, only that it find a sufficiently good solution. Indeed, it would likely be impossible for a “perfect” solution to exist, due to the subjective nature of art.

The process of an EA can be described by the following steps. Firstly, a population of individuals is created to form the initial generation, and a fitness function is used to calculate the performance of each individual. The fittest individuals are then selected to be used to create a new population of individuals, through recombination. These newly generated individuals form the next generation, which will undergo the same process. This is repeated until a termination condition is satisfied, such as reaching a limit on the number of generations or an individual reaching a satisfactory fitness threshold [Engelbrecht 2007]. This process is described in the following pseudo-code:

```
Generate initial population
Repeat until termination condition:
    Determine fitness
    Select individuals for recombination
    Perform recombination
    Create new population
End
```

While this general structure remains the same between most evolutionary algorithms, many of the components of the algorithm can vary between algorithm implementations. This is because these components often rely on problem-specific knowledge to determine the best implementation.

The function to generate the initial population can be implemented in a variety of different ways. Most commonly, an algorithm will choose to generate the initial population completely randomly. However, some algorithms use individuals generated using a different algorithm, or seeded from a set of known good solutions.

How each individual is evaluated to determine its fitness obviously requires domain knowledge of the problem, and therefore will vary between algorithms. The evolutionary algorithm can be treated as attempting to maximise this fitness function, therefore the more accurate the function is in measuring the performance of an individual the better.

The selection of which individuals are used to create the next generation can also be achieved through a variety of methods. These range from simply selecting the individuals with the highest fitness, the 'roulette wheel' method (which involves assigning each individual a section of a virtual roulette wheel sized in proportion with the individual's fitness, then choosing individuals based on spins of the wheel), to more complex problem-specific methods.

How parent individuals are recombined to form the next generation also varies between algorithm implementations. Recombination is usually achieved through the processes of crossover and mutation. Crossover is performed by choosing a crossover point on each parent, splitting each parent into two pieces at that point, and splicing together the opposite pieces from each parent to form a new individual. Mutation is the process of taking a single genotype, and randomly changing the value of one or more of its genes.

Finally, there is the termination condition that causes the evolutionary algorithm to finish iterating, and return the best result found. This condition also differs between algorithms, as it requires domain-specific knowledge of when a solution can be considered adequate. Common termination conditions include reaching a set number of generations, reaching a target fitness, or finding no further improvements in fitness.

The differences between specific EAs described above vary partly based on the specific problem, and partly due to the type of EA chosen. There are a number of types of evolutionary algorithms, each of which may use different techniques for recombination, fitness, or other features of the algorithm. Each type of EA exists to support slightly different approaches to finding a solution to the problem. The varieties of EA that are relevant to this thesis are described below.

### 2.3.1 Genetic Algorithms

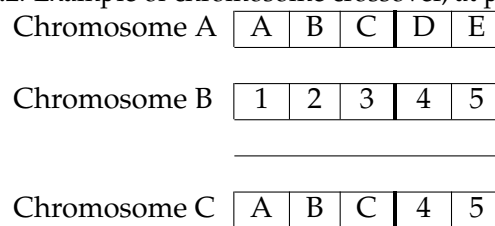
Genetic Algorithms (GAs) are the most popular type of evolutionary algorithm. GAs encode an individual as a sequence of values which represent the features of the individual, much like DNA in a biological system.

An individual is represented in the GA by one or more chromosomes, called a genotype. These chromosomes are composed of a sequence of individual components or ‘building blocks’, called genes. Each gene is a value that describes some aspect of the individual. These aspects or manifested properties of the individual are collectively known as the phenotype. Each location on the chromosome where a gene can reside is referred to as a locus (plural loci). A chromosome usually has a fixed number of loci, however variable-length chromosomes can also be used.

The converse of the genotype is the phenotype. A phenotype is a description of the characteristics of the individual, which in the case of GA is how the individual solves the problem. How this phenotype is represented by the genotype of an individual can affect the performance of the GA.

The most common method of recombination in a GA is a combination of single point crossover and mutation. Single point crossover is a version of crossover which uses a single crossover point on the chromosomes of both parents. However algorithms can implement other crossover variations, such as using more than two parents or using multiple or different crossover points. The process of single point crossover is illustrated in Figure 2.2.

**Figure 2.2:** Example of chromosome crossover, at position 4



#### 2.3.1.1 Interactive Genetic Algorithms

Interactive genetic algorithms (IGAs) are a subset of genetic algorithms that utilise human user feedback to determine the fitness of individuals. This has the advantage of being able to determine the fitness of solutions to subjective problems where the performance cannot easily be measured by a computer, such as design, art, and emotion.

However, IGAs have the disadvantage of being significantly slower than conventional genetic algorithms, as the process of collecting human input is by nature far slower

than the execution of a computer program. In addition, the population size and number of generations must be limited in an IGA, as evaluating many different individuals would impose a large burden on the user [Takagi 2001].

### **2.3.2 Evolutionary Programming**

Evolutionary Programming (EP) is another type of evolutionary algorithm. It differs from a GA in that instead of an individual being represented as a sequence of ‘building blocks’, each individual contains a number of parameters which are then used by a fixed algorithm.

An individual in an EP algorithm is defined as a sequence of parameters such as probabilities, constraints, and multipliers that are used by a fixed algorithm to provide a candidate solution to the problem. This contrasts with a GA, which contains all the information needed to recreate the phenotype within the genotype itself.

Since the information contained within each individual in an EP algorithm is relatively limited when compared to a GA, a mutation-only strategy is often used, instead of a combination of crossover and mutation. In this approach the individuals that are selected for breeding are copied and mutated, and those copies are then added to the existing population. To keep the population size steady, the worst individuals are then removed from the population.

### **2.3.3 Bacterial Evolutionary Algorithms**

Bacterial evolutionary algorithms (BEAs) are a class of evolutionary algorithms that mimic bacterial or asexual reproduction, where a single parent produces offspring. This is contrasted with genetic algorithms, which mimic the more common process of sexual reproduction where two parents produce offspring that are a combination of the parents.

Recombination in a BEA is achieved through first cloning an existing individual in the population, then absorbing certain genes from other individuals in the population to create a new individual. Each gene in a BEA has its own fitness value, which is derived from the fitness of each individual that gene appears in. These gene fitness values are then used during the gene absorption phase of reproduction to determine which genes should be absorbed by the new individual.

# Literature Review

---

## 3.1 Mondrian Art Analysis

Mondrian claimed that his placement of each element on the canvas reflected a fundamental aesthetic of nature [Mondrian 1914]. This has caused many researchers to examine whether this approach resulted in more aesthetically pleasing artwork, and if so, what particular properties caused this increased aesthetic value.

One popular area of research is the effect of line positioning on the aesthetics of a Mondrian composition. McManus, Cheema, and Stoker [1993] presented subjects an original Mondrian composition alongside a composition where the lines were slightly shifted. The results of the study showed that subjects preferred the original Mondrian artwork to the modified composition. The images used in that study were also used by Furnham and Rao [2002], who also found the existence of a significant preference for the original compositions compared to the modified compositions.

Another study looking at line spacing by Wolach and McHale [2005] showed that computer generated Mondrian-like compositions that used line spacings similar to Mondrian's own were preferred over compositions which used near-random line spacings. This suggests that line spacing is one of the key factors which make Mondrian compositions preferred when compared to other compositions in the same style.

The orientation of a Mondrian composition has also shown to be an important factor in the overall aesthetic of the artwork. In a study by Latta, Brain, and Kelly [2000], subjects were shown Mondrian compositions at various angles, through a circular viewport. This study showed that there was a preference for the original orientation of Mondrian compositions compared to rotated orientations. This effect was present in both the compositions created with horizontal/vertical frames, as well as the 'lozenge' compositions created with oblique frames.

The colour of coloured areas in a Mondrian composition has also been shown to have an effect on the perceived "balance" of the composition in a study by Locher, Overbeeke, and Stappers [2005]. The study took a number of Mondrian compositions along

with a number of compositions where the colours of each coloured area had been changed, and participants were asked to state where they felt the balance centre of each image lay, as well as rate the “weight” of each image. The results showed that the colours chosen had a significant effect on the balance and weight of the image, indicating that each colour gives a different weight to the image.

A study by Gedeon [2008] aimed to determine which features of a Mondrian-like composition are most important when describing a subject’s aesthetic preference for some compositions over others. This study used a neural network to model the subject’s aesthetic preferences, using a number of features of the composition such as the number of lines, the distance between lines, the amount of coloured areas in the composition, and the distance between coloured areas. The weights of the neurons in the neural network were then used to determine which features were the most important. The results of this study showed that the most important features were the amount of coloured area on the canvas, the distance between lines, and the distance between coloured areas.

These studies show that the position of each line, the position of each coloured rectangle, and the colour of each rectangle all play a role in making Mondrian compositions aesthetically pleasing to viewers. That is, Mondrian compositions become less aesthetically pleasing when any of these features are modified without care or precision.

## **3.2 Mondrian Art Generation**

Several systems have been created to generate Mondrian-like compositions using evolutionary algorithms. Each of these systems vary in their chosen implementation of evolutionary algorithm, including using different methods of determining fitness, genotype representation, and recombination strategies.

### **3.2.1 Schnier & Gero**

Schnier and Gero [1997] developed a system which attempts to generate Mondrian art that imitates a given example composition as closely as possible, using a novel style of representation.

The genotype representation used in this system was a recursive tree of rectangular divisions, as this can represent many of Mondrian’s paintings. More specifically, each node on the tree contains information defining two areas within the node’s parent area, by defining a single division in that parent area (or the canvas as a whole in the case of the top level node). A node also contains information of the colour of each defined area. Each of these areas may in turn have a node describing two more areas within it, further dividing the composition. At a basic level, this representation is



---

similar to the technique used in genetic programming, where an algorithm or program is built using a tree genotype structure.

The novelty of this system is contained in its additional process of evolving the genes themselves. The system observes how the genotype is being used, and detects any patterns in that utilisation. “Evolved genes” are then created that represent those patterns, and are added to the representation [Schnier and Gero 1995]. These evolved genes take the form of two or more genes that have been linked, to ensure that they remain together through crossover and mutation operations. Each evolved gene has a fitness score, which is derived from the sum of the fitness of all individuals which contain that evolved gene. These evolved genes can also in turn be linked together to form even larger evolved genes.

This technique is further enhanced by extending the evolutionary algorithm to include dominant and recessive genes. This concept is derived from an analogous concept in natural evolution, where two different copies of a gene will be present in a genotype but only the dominant one will be expressed in the phenotype. In the context of artificial evolution, this means that when a crossover operation attempts to merge two segments which have conflicting evolved genes (a pair of evolved genes that contain one or more of the same gene as part of themselves), no information is destroyed.

The goal of the system was to create a composition that mimics an example Mondrian composition as closely as possible. To this end, the fitness function used measures the similarity between each individual and the exemplar that was to be copied. This fitness value is composed of a set of values that measure the number of correct and incorrect divisions, the exactness of each division, the correctness of the colour of each division, and the correctness of each line’s thickness.

The results of Schnier and Gero [1997] showed that the technique of using evolved genes, as well as dominant and recessive genes, improved the performance of the system when compared to a system which just used the base representation. However, there were still notable differences between the example composition and the generated composition.

### **3.2.2 MONICA**

Gómez de Silva Garza and Zamora Lores [2011] created the Mondrian-Imitating Computer Artist (MONICA) system, designed to automatically generate Mondrian-like compositions.

MONICA uses an initial population consisting of 60% original Mondrian compositions drawn from a pool of 55 exemplars, and 40% randomly generated compositions.

This ratio of random individuals and exemplars was found to be optimal for increasing the speed of the algorithm while retaining enough variation to sustain a high level of diversity. This is a significant divergence from a standard evolutionary algorithm, in which individuals in the initial population are generated randomly.

Each individual in MONICA is represented by a very literal genotype that directly describes each element of the phenotype. The genotype is composed of 20 sections, each of which corresponds with a component (either a black line or a coloured rectangle) on the canvas. Each section may also be empty, allowing each individual to consist of between 0 and 20 components. Each section is comprised of genes which encode the colour, size, and position of the element it represents. MONICA treats the lines in the composition as thin rectangles, rather than a different kind of component from coloured rectangles.

To determine fitness, MONICA uses a set of 7 manually created evaluation rules. These rules evaluate whether the individual matches the Mondrian style in the categories of colour validity, component location, line thickness, number of vertical and horizontal lines, component adjacency, canvas size, and number of coloured components. The evaluation rules were written such that each of the 55 exemplars included with the algorithm will score 100%.

MONICA uses crossover and mutation for recombination of individuals. 80% of the genotypes for a new generation are created using crossover, and the remaining 20% are generated using mutation. The same crossover point is used for both parents, to preserve the length of the genotype and ensure all offspring genotypes are valid – even though many offspring phenotypes may not conform to the evaluation rules.

The results of Gómez de Silva Garza and Zamora Lores [2011] showed that MONICA created Mondrian-like compositions that, although not obviously distinguishable from Mondrian originals, were picked as being computer generated between 10% and 20% more often than would be expected of an indistinguishable painting. However, it was noted that at least a portion of these results could be attributed to imperfections in the reproduction of the Mondrian compositions that were not present in the MONICA composition.

### **3.2.3 Darwindrian**

Shen [2007] created a system called Darwindrian which uses bacterial evolutionary algorithms to generate Mondrian-like art, using human feedback to rate each composition.

To represent an individual in the algorithm, the system encodes a number of points in a chromosome, each of which will be a point on the canvas where one or more lines will be emitted from. Also included with the location of each point is a set of

---

probabilities of which directions lines will be emitted. After lines are placed on the canvas using this data, areas of the canvas divided by those lines are coloured in using data from another chromosome which contains a set of probabilities of how likely a coloured rectangle is going to be a specific colour.

This generation process ensures that the composition follows the set of rules that define a Mondrian composition. Since the encoding of the individual does not directly describe a composition but rather takes a probabilistic approach, each composition generated using the values of an individual may be similar, but will most likely be different.

Darwindrian uses a human user to evaluate the fitness of each individual. The user is shown a composition generated using each individual, and can choose whether they like the structure of the composition, the colour, both, or neither. This feedback is then used to assign a fitness value to the appropriate portion of the individual. For example, if the user indicates they prefer the colour of the composition, the genes of that individual which control the colour will have their fitness value increased. The value of the individual as a whole is then calculated using the fitness values of the genes it contains.

### **3.3 Interactive Genetic Algorithms**

In addition to the Darwindrian algorithm by Shen [2007] covered in the previous section, interactive genetic algorithms have been used by a number of researchers in an attempt to solve a variety of different problems that require human characteristics.

#### **3.3.1 GenJam**

GenJam was created by Biles [1994] to generate jazz solos using an interactive genetic algorithm. Much like generating art, generating music requires creativity that can currently only be achieved through human interaction. GenJam utilises a human “mentor” to rate the individuals in the genetic algorithm, in this case a phrase or measure of a jazz solo. This is done by playing the mentor a solo constructed from those individuals, and allowing the mentor to increment and decrement the fitness value of each individual as they played.

Biles [1999] noted that different mentors had different styles of giving feedback when using GenJam - some users would give feedback on individuals in time with the beat, others were more reserved and only gave feedback when particularly moved by the music, while others still gave constant feedback at a far more rapid pace. This means that the algorithm should support both frequent and infrequent styles of feedback, and perform well in either situation.

### 3.3.2 Lee & Cho

Lee and Cho [1999] created a human-oriented image retrieval system using interactive genetic algorithms. The goal of the system is to retrieve images that all convey a similar emotion, such as cheerful or gloomy. Interpreting emotion is another area at which humans far outperform computers, and so is a good problem to solve using interactive genetic algorithms. The user interacts with this system by increasing the fitness of images that more closely match the emotion or feeling of the image they wish to retrieve.

The system reduces the amount of user interaction required by clustering groups of individuals, showing the user a representative of each cluster, and extrapolating the fitness of each individual in the cluster based on the fitness assigned to the representative individual. This sparse fitness evaluation reduces the burden on the user by reducing the number of individuals they need to evaluate, while at the same time speeding up the system.

## 3.4 Summary

Analysis of the structure and other qualities of Piet Mondrian's compositions has shown that the position of each element on the canvas has a noticeable effect on the overall quality of the composition. This implies that generating Mondrian-like compositions with computer algorithms is a non-trivial task, since if the position of each element on the canvas is not correct the composition will not be aesthetically pleasing.

Previous research in the field of generating Mondrian-like compositions using computer algorithms has been relatively limited. Additionally, the methods used by previous researchers to implement these algorithms have had mixed results. The MONICA algorithm [Gómez de Silva Garza and Zamora Lores 2011] had reasonable success in generating compositions that were similar to Mondrian's compositions, however these compositions were not tested to gauge whether they were preferred over Mondrian's compositions. The only interactive genetic algorithm approach to generate Mondrian-like compositions has been the Darwindrian algorithm, the performance of which was not tested by Shen [2007].

Interactive genetic algorithms are a promising method for solving problems such as generating aesthetically pleasing artwork. Previous work has shown the success of IGAs in solving various problems that typically require human judgement or intuition, such as generating music and evaluating the emotion of images.

However, little research has been performed on how the performance of an IGA is effected by the implementation of each component of the algorithm. This shows that there is an opportunity for future research in this area to document these effects, leading to the research questions described in the next chapter.

---

## Research Questions

---

This thesis will attempt to build on the work by previous researchers, by looking at the performance of interactive genetic algorithms in generating Mondrian-like art. Specifically, this thesis will study how the performance of an interactive genetic algorithm is affected by the method used to represent an individual's phenotype in the genotype. This phenotype to genotype mapping effectively determines what information is stored in the genotype of the individual and passed down to future generations, and what features are left to random or programmatic generation. This implies that the representation can have a large effect on the performance of the algorithm.

Three research questions that arise are:

- Does the phenotype representation in an interactive genetic algorithm have a significant effect on the performance of the algorithm?
- Can a more effective representation improve on the performance of previous genetic algorithms in generating Mondrian-like art?
- Can an IGA generate art in the style of Piet Mondrian that is more aesthetically pleasing to a subject than Mondrian's own compositions?

The answers to these questions are relevant to the field of generating art using interactive genetic algorithms, but may also be applicable to IGAs in general. Previous researchers have often performed preliminary testing while designing their algorithms, but usually have not covered this in their published results, so the factors which improved representation performance for their algorithm are largely unrecorded. This thesis aims to document techniques which can be used to improve representation of Mondrian art in GAs.

Three distinct representations will be created to answer these questions. Two of these representations will use different representation techniques. The third representation will be a hybrid of the first two approaches, with the goal of combining the strengths of each of each representation. The representation used in the Darwindrian algorithm

by Shen [2007] will be set as a baseline when measuring the performance of each of the representations.

The four representations will be implemented in otherwise identical interactive genetic algorithms. The compositions generated by each of these algorithms will then be compared against digital recreations of original Mondrian compositions, to determine the performance of each algorithm. The hypothesis is that one or more of these representations will perform better than the others, due to the different methods of representation.

---

# Representations

---

In this chapter, three different methods for representing Mondrian-like art in a genetic algorithm are introduced. The first two representations (described in Sections 5.1 & 5.2) are two simple yet very different representation strategies, created to reflect the basic kinds of strategy used in previous works. The third representation (described in Section 5.3) is a hybrid of the first two representations, created to combine the strengths of both representations.

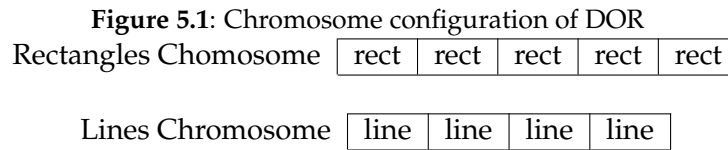
The representation used by the Darwindrian algorithm is also described in this chapter, as this representation will be used as a baseline when evaluating the performance of each algorithm.

## 5.1 Direct Object Representation

This representation technique uses a direct method of phenotype encoding, which will hereafter be referred to as Direct Object Representation (DOR). This representation takes a simple approach, using two chromosomes to encode all of the coloured rectangles and the lines of a composition, as shown in Figure 5.1.

The first chromosome in this representation has five loci, each of which is able to contain information on a single coloured rectangle. The information about every rectangle in the composition is represented in this chromosome. The information stored in each locus includes the rectangle's width, height, position, and colour.

The other chromosome has four loci, each of which may contain information about a line in the composition. The lines that this chromosome represents are lines that, while they appear on the canvas as part of the composition, are not necessary for the composition to be valid. That is, these are lines that do not serve to contain a coloured rectangle. Each locus in this chromosome contains information on whether a line is horizontal or vertical, the starting position, and the width of the line.



The number of loci in these chromosomes restricts the total number of rectangles and lines that can be present in each composition to 5 and 4, respectively.

This information is used to recreate the composition. Firstly, the information from the rectangle chromosome is used to place the coloured rectangles on the canvas. Then, the lines chromosome is used to generate lines on the canvas. Finally, additional lines are emitted from the corner of each coloured rectangle to ensure that each rectangle is fully enclosed by either the edges of the canvas or a line.

This process ensures that each genotype will always generate a valid Mondrian composition. However, due to the generation of lines being partially random, each composition generated from a genotype will be slightly different.

The risk with this representation method is that if certain genes become overly dominant, and therefore become present in most individuals, the majority of the population will be similar to each other. This similarity will be obvious to the user, who may become bored or disinterested with the algorithm due to the lack of variety.

## 5.2 Statistical Representation

This representation technique uses probabilities to represent the individual, and so will be referred to as Statistical Representation (SR). This representation stores a number of different statistical parameters about a composition in a single chromosome, as illustrated in Figure 5.2. These parameters are:

1. Minimum height of rectangles
2. Maximum height of rectangles
3. Minimum width of rectangles
4. Maximum width of rectangles
5. Probability of a rectangle appearing on an edge
6. Line width
7. Probability of a line width varying from normal
8. Amount by which line width varies
9. Probability of lines being vertical/horizontal



10. Probability of lines being generated
11. Number of rectangles
12. Number of lines
13. Colour probability distribution

**Figure 5.2:** Chromosome configuration of SR

Chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13
------------	---	---	---	---	---	---	---	---	---	----	----	----	----

These statistical parameters were chosen based on a study by Gedeon [2008], in which he used a Neural Network to analyse a number of Mondrian-like compositions preferred by a subject to find which features of the compositions best represented the subjects aesthetic preferences.

When generating a composition from an individual using this representation, each of these 13 parameters are taken into account when placing lines and rectangles on the canvas. Firstly, rectangles are generated and placed on the canvas based on parameters 1-5 and 13, until the number of rectangles specified by parameter 11 have been generated. Next, lines are generated and added based on parameters 6-9, up to the number specified by parameter 12. Finally, more lines are emitted from each corner of each rectangle based on the probability from parameter 10.

This process creates a valid composition that fulfils all of the parameters held in the chromosome. However, the choice for each component of the composition is still highly random, meaning that one individual can generate many compositions that differ in appearance.

This representation resembles an Evolutionary Programming (EP) approach, in that the chromosome contains parameters which are used in a fixed algorithm. However, it differs from an EP algorithm in that it still uses the standard GA method of recombination through crossover, whereas an EP algorithm would rely solely on cloning and mutation to create new individuals.

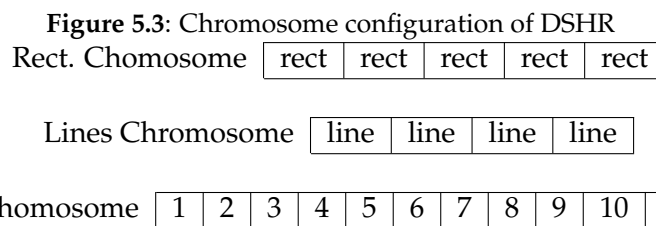
While DOR has the risk of the population becoming too uniform and therefore not interesting to the user, SR has nearly the opposite issue. Since each genotype can generate many different compositions, the specific composition that is generated may not be aesthetically pleasing to the user, even though the genotype represents compositions that the user will usually like. This may cause the user to feel as though the algorithm is not improving, or even that the algorithm is beginning to perform worse.

### 5.3 Direct/Statistical Hybrid Representation

A new representation technique can be created by merging features of both DOR and SR, to create a hybrid representation which will be referred to as Direct/Statistical Hybrid Representation (DSHR).

This representation combines concepts from both DOR and SR to create a genotype that contains information on both specific objects in the composition, as well as general statistics about the composition. This approach is intended to combine the strengths of each representation, by using the information stored in the component chromosomes in unison to create a composition from the genotype. This combined approach is designed to find a middle ground between the opposing issues of DOR and SR, namely that a DOR population risks becoming too uniform while an SR population risks being too random.

There are three chromosomes in this representation. The first two chromosomes contain the same information as the chromosomes in DOR, and the third chromosome contains the same information as the chromosome in SR. Specifically, one chromosome contains information on specific rectangles in the composition, one chromosome contains information on specific lines in the composition, and the final chromosome contains various statistical parameters used to generate additional components. These chromosomes are shown in Figure 5.3.



The information from all three of these chromosomes are combined to create a composition using an individual with this representation. Firstly, the information from the rectangle and lines chromosomes are used to place specific rectangles and lines on the canvas, as occurs in DR. Then, the information from the statistical chromosome is used to add additional components to the canvas until a valid composition with the desired number of lines and rectangles is generated.

## 5.4 Darwindrian Representation

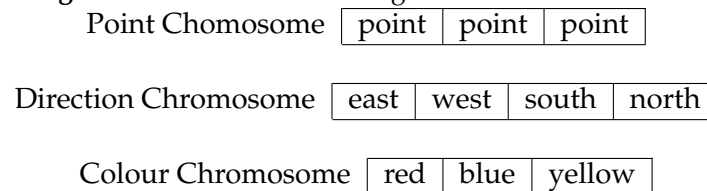
This is the representation used in the Darwindrian interactive bacterial evolutionary algorithm. This representation uses three chromosomes, one to represent points on the canvas where lines will be generated from, another to store the probability distribution for the direction of each generated line, and the final to store the probability distribution for the colour of the coloured rectangles. These three chromosomes are illustrated in Figure 5.4.

The first chromosome contains three loci, each of which contains a point on the canvas. A point is defined by an  $x$  and  $y$  coordinate on the canvas. These points define the locations where lines will be emitted from during line generation.

The second chromosome contains four loci, each of which represents a cardinal direction that a line can be emitted along. Each of these loci contain a series of probabilities for a line to be generated in that direction. A different probability is used from each locus depending on if the line being generated is the first line from a point, the second line, the third, or the fourth. Using a set of distributions allows this chromosome to also encode the probability of lines in each direction being either long or short, as lines generated later have a higher probability of stopping on another line before hitting the edge of the canvas and therefore will be shorter.

The final chromosome contains three loci, corresponding with the possible colours of a coloured rectangle. Each locus contains the probability of its associated colour to be chosen as the colour for each of the coloured rectangles on the canvas.

**Figure 5.4:** Chromosome configuration of Darwindrian



This information is used to create a Mondrian-like composition as follows. Firstly, lines are generated from each point in a round robin manner, using the information from the points and direction chromosomes. Once all lines are generated, a number of rectangular areas created by the lines and the edges of the canvas are filled with colour based on the colour chromosome.

Since this method generates lines and then chooses how to place the coloured rectangles based on the partitions created by those lines, it will always create valid Mondrian compositions. However, because this representation relies on probabilities to generate its lines and rectangles, one genotype can generate many different phenotypes, much like SR.

## 5.5 Summary

The first two representations introduced in this chapter, DOR and SR, intend to embody two popular general approaches used in previous work when designing interactive genetic algorithms for generating art. The third representation, DSHR, attempts to combine the strengths of both these approaches to form a new, more robust approach.

A common problem with direct representations such as DOR is that there is very little degree of randomness or unpredictability. This measure of uncertainty has been found

to be important when generating aesthetically pleasing art [Maiocchi 1991]. However, representations such as SR can have the opposite problem of being too random and providing the user with no sense of improvement. This is the key area where DSHR attempts to strike a balance between DOR and SR.

Now that these four representations have been described, they must be tested to determine which performs best in an interactive genetic algorithm.

---

# Representation Performance

---

It is expected that each of the representations described in Chapter 5 will perform differently when applied as part of an interactive genetic algorithm to generate Mondrian compositions. The performance of each these representations needs to be analysed to determine if this assumption is correct, and if so which representation best supports the generation of aesthetically pleasing compositions. The results of this analysis can then be used to answer the research questions posed in Chapter 4.

The performance of each algorithm will be measured in two ways. The first performance factor is how well the algorithm performs in generating artworks that are as aesthetically pleasing to the participant as Mondrian's compositions. The second factor is how well the algorithm functions in areas such as providing the user with diverse individuals and giving the user a sense of improvement, to keep the user interested in continuing with the algorithm. This second measure is important, since if the user does not feel as though the algorithm is working they are less likely to continue to give accurate feedback or continue to use the algorithm.

Each representation is implemented in a separate interactive genetic algorithm. While each algorithm uses a different representation, all other aspects of the system are kept as similar as possible. The representations implemented in this study are the three representations introduced in Chapter 5, as well as the representation used by the Darwindrian algorithm. For the sake of brevity, each algorithm will be referred to by the representation it uses, with Darwindrian shortened to DWN.

## 6.1 Methodology

### 6.1.1 Algorithm Implementation

Although each algorithm differs in genotype representation, the other components of the algorithms are kept as similar as possible between the algorithms in order to reduce the number of variables that could affect the results. These common components

include the number of individuals and generations, as well as the selection, the breeding strategy, the method for determining individual fitness, and the user interface.

All of these genetic algorithms are interactive, meaning a human user must evaluate every individual. Therefore, the number of individuals and generations should be kept as low as possible, to reduce the total number of individuals the user must evaluate and hence the overall workload on the subject. If the number of generations is too few, a good solution might not have yet been found. If the number of generations is too many, the algorithm will begin to lose genetic diversity and individuals will become very similar to each other. It was decided to have 20 individuals in each generation and have the algorithm run for 15 generations, as this was found in preliminary testing to be an acceptable trade-off between algorithm performance and user fatigue. To test whether this assumption was correct, data is also collected for the best individuals from the middle of the algorithm and compared against the data from the end of the algorithm to test if the algorithm continued to improve in the later generations.

Selection of individuals for breeding is performed using the roulette wheel method. Each individual is assigned a section of a virtual roulette wheel in proportion to its fitness. The roulette wheel is spun twice to choose the two parents who are then used to create a new individual in the next generation. These two parents are recombined using single point crossover on each of the chromosomes. This method is repeated until the next generation is filled.

Due to this small population size and generation count, no mutation strategy is implemented. Additionally, due to the relatively low amount of genetic diversity, a restriction is placed on the creation of a new generation that there may not be any identical individuals in the generation; each individual must be different from every other by at least one gene.

The fitness of an individual is determined by the human user, by rating each individual as either "Disliked", "Neutral", or "Liked". A rating of "Liked" or "Disliked" will increase or decrease the fitness of an individual, respectively. A rating of "Neutral" will slightly modify the fitness to bring it closer to the centre (i.e. decrease the fitness of high fitness individuals and increase the fitness of low fitness individuals).

To retain as much data about the fitness of an individual as possible, the fitness of an individual is initialised to be the average fitness of both of its parents. This is based on the premise that a child of two high fitness individuals is likely to be fitter than a child of two low fitness individuals.

The user interface used by the algorithms, shown in Appendix B, shows the user each individual in the current generation. The user has a set of radio buttons to rate the composition they are currently viewing as either "Liked", "Neutral", or "Disliked". Once the subject has evaluated all 20 individuals in the generation, they tell the algorithm to generate the next generation by clicking on the "Next Generation" button.

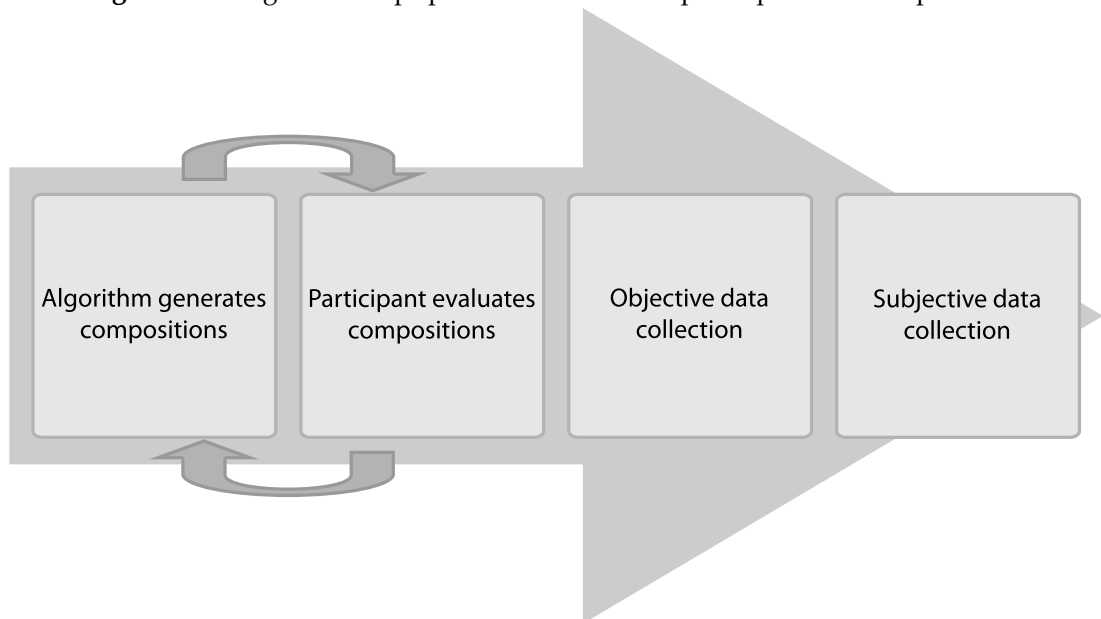
### 6.1.2 Experimental Design

Participants in the study were chosen from a variety of backgrounds and demographics. Each participant was asked a number of questions to determine if they were suitable candidates to participate in the study. These questions determined if the participant was familiar with the works of Piet Mondrian, has formally studied art, or has any trouble using a computer. Only participants that were not familiar with Mondrian's work, had no formal art qualifications, and could navigate a computer interface were selected to participate in the study.

A total of 28 participants were used in this study. To reduce the workload of each participant, each participant only ran a single algorithm. However, this significantly increases the difficulty of analysing the data, as all other possible effects need to be controlled for before combining the results from each participant. Each participant took between 30-50 minutes to complete the experiment.

The steps performed with each participant in the experiment are illustrated in Figure 6.1. First, the algorithm generates a set of compositions for the participant to evaluate. The participant then evaluates each of the 20 compositions in the generation, and then prompts the algorithm to create the next generation. These two steps repeat for the 15 generations of the algorithm, making a total of 300 evaluations performed by each participant. Following that, objective data is collected from the participant regarding their ratings of compositions generated by the algorithm, as well as Mondrian compositions. Finally, the participant is asked a number of questions regarding their subjective experience with the algorithm.

**Figure 6.1:** Diagram of steps performed with each participant in the experiment



### 6.1.3 Data Collection

After completing a full run of the algorithm, each participant was asked to rate a number of compositions generated by the algorithm, as well as complete a questionnaire relating to the performance of the algorithm.

To measure how well the algorithm performed in generating aesthetically pleasing compositions, the participants were shown 15 different compositions. 5 of these compositions are the best individuals from the final generation of the algorithm, 5 are the best individuals from the 7th generation of the algorithm, and 5 are original Mondrian compositions. The participant was then asked to rate how aesthetically pleasing they find each image between 0 and 10, where 0 is extremely disliked and 10 is extremely liked. This is the third step shown in Figure 6.1.

The participants were not informed which group the composition they are viewing is from, and the order in which these compositions are displayed to the participant is randomised to ensure the participant did not know which type of composition they were rating. However, the ratings could be biased toward the generated compositions due to the fact that the ratings were given immediately after the participant has viewed and evaluated a large number of generated compositions. It is possible that this repeated exposure to generated compositions had an effect on the aesthetic preferences of the participant.

After completing the above rating, the participants were presented with the top 5 compositions generated by the algorithm and informed that these are the compositions the algorithm believes they will find the most aesthetically pleasing. To measure how well the algorithm performed in engaging the user the participant was then presented with a questionnaire to fill out, regarding their preferences and overall impression of the performance of the algorithm. This is the fourth step in Figure 6.1. The full questionnaire is given in Appendix C.

The questionnaire contains a number of Likert items regarding how they felt the algorithm performed in generating aesthetically pleasing compositions (Q1), whether the algorithm took into account their aesthetic preferences (Q2), whether the compositions generated by the algorithm were diverse (Q3), and whether the compositions generated by the algorithm improved over time (Q4). The responses to each of these items was a choice of "Strongly Disagree", "Disagree", "Neutral", "Agree", or "Strongly Agree".

### 6.1.4 Statistical Analysis

Throughout this statistical analysis, a statistical test is taken to be significant if its  $p$  value is less than 0.05.



---

Both the objective and subjective data collected from participants is ordinal data. That is to say that, for example, a rating of 5 is better than a ranking of 4, but the ratings 5 & 4 and 4 & 3 are not necessarily different by the same amount. Ordinal data is considered non-parametric, meaning the distribution of the data does not necessarily match a predefined distribution such as a normal distribution. This limits the statistical techniques and tests that can be applied to the data to non-parametric methods. The statistical tests used in this analysis are described below.

#### 6.1.4.1 Statistical Tests

The Mann-Whitney-Wilcoxon (MWW) test is used to compare two groups of ordinal data. MWW is a statistical hypothesis test that determines if one group of values is larger than another group of values. MWW requires that all observations are independent of each other. This test was chosen over a Student's t-test as MWW is a non-parametric test, and so makes no assumptions of the distribution of the data.

The Kruskal-Wallis test is a non-parametric test, equivalent to the parametric ANOVA test, that examines whether data from different groups belong to the same distribution. This test indicates whether at least one of the groups differs from the others, but not which group is different. If this test is positive, a series of MWW tests can then be used to determine between which groups the difference or differences lie.

Fisher's exact test is a significance test which analyses data which is categorised in two different ways, and indicates if there is an association between the two methods of categorisation. The null hypothesis of this test is that there is no correlation between the classification methods.

A multinomial test is a test that determines if a categorical dataset follows a certain multinomial distribution. A multinomial distribution is the set of probabilities of a data point falling in each category. The null hypothesis of this test is that the dataset does follow the specified distribution.

#### 6.1.4.2 Objective Data Analysis

Since each participant only used one algorithm, the analysis of the ratings data is quite complicated. This is because other possible effects on the ratings need to be eliminated prior to combining the results from each participant. Some possible effects on the ratings could include the participants overall preference for the style of artwork, the general voting pattern of the participant, and any cultural biases of the participant. To control for these effects, the data for each participant is first tested for statistical significance, and the results of those tests will then be used to measure and compare the performance of each algorithm.

Firstly, the performance of an algorithm for a specific participant must be determined. To do this, the ratings given for the final compositions are compared against the ratings for the Mondrian compositions. A Mann-Whitney-Wilcoxon test can be used to determine if either group has statistically significantly higher ratings than the other group, thereby indicating whether the algorithm performed well for that participant.

The performance of each algorithm can then be determined by grouping the participants by the results of these MWW tests, per algorithm. A multinomial test can be used to determine if the distribution of participants in these groups is statistically significantly different from what would be expected by random chance, indicating whether the algorithm generates compositions that are more or less aesthetically pleasing than Mondrian's compositions.

The results of the above MWW tests can also be used to determine the relative performance of each algorithm compared to the others. Fisher's exact test can be used to determine if the distribution of participants where the algorithm generated compositions were preferred by the participant can be explained by random chance, between any two algorithms. If this test indicates that the distribution is unlikely to be through random chance, then it is evidence that one algorithm performs statistically significantly better than the other.

Finally, the ratings given to compositions from the final generation of the algorithm can also be compared to the ratings given to compositions from the middle of the algorithm to determine if the number of generations the program ran for was necessary. This comparison is achieved through the same method as was used to compare the final compositions to the Mondrian compositions. Specifically, an MWW test can be used to test the hypothesis that the individuals from the end of the algorithm performed statistically significantly better or worse than individuals from the middle of the algorithm. A multinomial test can then be used to determine if there was a statistically significant number of participants for which the final compositions were either better or worse than the middle compositions.

#### **6.1.4.3 Subjective Data Analysis**

The data collected from part A of the questionnaire can be analysed to determine the participants subjective experience with the algorithm, as well as determine how engaging the algorithm was. These questions were on a Likert scale, so each response can be represented numerically as a rating from 1-5.

Descriptive statistics can give an insight into the typical answers to each question, per algorithm. The average response is given by the median of the set of responses, and the interquartile range of responses gives the range within which the middle 50% of responses fall. This gives an indication of the subjective performance of each algorithm.

The relative subjective performance of each algorithm compared to the others can also be examined. A Kruskal-Wallis test can be used to discover if there is a statistically significant difference between the answers to a question by algorithm. This test determines if there is a statistically significant difference between the distribution of responses between at least two of the questions, however it cannot indicate between which questions this difference lies, or the direction of the difference.

The algorithms can then be compared against each other to determine which algorithm attracted the best results for each question, assuming the Kruskal-Wallis test indicated that there was a difference between at least two of the algorithms for that question. An MWW test can be used for this comparison, testing if the set of responses for one algorithm is significantly different to the responses for another algorithm. A series of these MWW tests will be needed, to test every possible pair of algorithms against each other.

## 6.2 Analysis Results

### 6.2.1 Objective Performance

The objective rating data for each participant was analysed, using an MWW test to compare the distribution of ratings between the best compositions from the final generation and Mondrian compositions. The results of these MWW tests ( $U$  &  $p$ ), along with the median ( $M$ ) and interquartile range ( $IQR$ ) for each group, are shown in Table 6.1. The raw rating data is given in Appendix D.

The results of these MWW tests are summarised in Table 6.2 for each algorithm, by the number of participants passing the significance threshold ( $p < 0.05$ ) in either direction. Also included in this table is a multinomial significance test, with the null hypothesis that the results will fall in the "Same" category 95% of the time (due to our significance threshold). The null hypothesis will be true if the number of participants where the MWW test indicated a significant difference can be explained by random chance.

The significance results in Table 6.2 show that DOR, SR, and DSHR all generated compositions that were rated statistically significantly higher than Mondrian's compositions (significance threshold  $p < 0.05$ ). DWN, however, generated compositions that were not rated significantly differently from Mondrian's compositions ( $p = 0.698$ ).

The groupings shown in Table 6.2 can also be compared between algorithms, using a series of 2x3 Fisher's exact tests. Each of these tests indicates whether the association between the categories and algorithms are statistically significant. The p-value from these tests are shown in Table 6.3. Only the difference between the algorithms DWN & DOR is statistically significant according to this test.

**Table 6.1:** Mann-Whitney-Wilcoxon tests (as  $U$  statistic and  $p$  values) comparing objective ratings of Final and Mondrian compositions per participant (1-28), by algorithm (DWN, DOR, SR, & DSHR), with median ( $M$ ) and interquartile range ( $IQR$ ) per group

Alg.	Subject	Final		Mondrian		$U$	$p$
		$M$	$IQR$	$M$	$IQR$		
DWN	1	1	1.0 - 1.0	0	0.0 - 1.0	5	0.151
	2	7	2.0 - 10	2	0.5 - 3.5	5	0.151
	3	4	3.0 - 7.0	5	4.0 - 6.5	11	0.841
	4	3	2.0 - 6.0	2	2.0 - 4.5	9	0.548
	6	5	3.5 - 6.5	3	2.5 - 6.0	8.5	0.421
	26	6	1.5 - 8.0	3	1.5 - 6.5	10	0.690
	28	4	2.5 - 5.5	5	0.0 - 6.5	12	1.000
DOR	5	5	1.0 - 6.5	8	3.0 - 10	6.5	0.222
	7	7	6.0 - 8.0	2	1.5 - 3.5	0	0.008
	8	5	4.5 - 7.0	3	2.0 - 5.5	5	0.151
	9	6	5.0 - 7.0	2	1.0 - 2.0	0	0.008
	10	7	5.5 - 7.5	3	1.5 - 4.5	1.5	0.016
	23	8	5.0 - 9.0	3	0.5 - 5.5	2.5	0.032
	24	8	5.5 - 9.0	4	3.0 - 5.0	2.5	0.032
SR	12	7	5.0 - 7.5	4	3.0 - 6.0	4	0.095
	13	3	1.5 - 5.5	1	0.0 - 1.5	2.5	0.032
	15	8	5.0 - 8.0	2	0.0 - 4.5	2.5	0.032
	17	6	4.5 - 8.0	7	5.5 - 7.5	11	0.841
	18	6	4.5 - 6.5	0	0.0 - 2.0	0.5	0.008
	20	9	3.0 - 9.5	4	0.0 - 8.0	5	0.151
	22	5	3.5 - 6.0	1	0.0 - 2.5	0.5	0.008
DSHR	11	3	2.5 - 4.0	1	0.5 - 1.0	0	0.008
	14	6	5.5 - 6.5	3	1.0 - 6.0	5	0.151
	16	5	4.0 - 6.5	3	2.0 - 4.5	3.5	0.056
	19	3	2.5 - 4.0	1	0.0 - 1.5	0.5	0.008
	21	5	3.5 - 6.5	2	0.5 - 5.5	5	0.151
	25	7	3.5 - 8.0	1	1.0 - 4.5	4	0.095
	27	8	7.0 - 9.0	3	0.0 - 5.5	1	0.016

**Table 6.2:** Participants grouped by whether MWW test indicated significance ( $p < 0.05$ ) between Final and Mondrian ratings, per algorithm (DWN, DOR, SR, & DSHR), with multinomial test indicating significance of the distribution (as  $p$  value).

Alg.	Lower <sup>a</sup>	Same <sup>b</sup>	Higher <sup>c</sup>	$p$
DWN	0	7	0	0.698
DOR	0	2	5	<0.001
SR	0	3	4	<0.001
DSHR	0	4	3	0.004

<sup>a</sup> Number of participants where the MWW test found the Final ratings were significantly below the Mondrian ratings

<sup>b</sup> Number of participants where the MWW test found the Final ratings were not significantly different from the Mondrian ratings

<sup>c</sup> Number of participants where the MWW test found the Final ratings were significantly above the Mondrian ratings

**Table 6.3:** Pairwise comparisons of distribution of participants in Table 6.2 using Fisher's exact test, given as  $p$  value

Pair	$p$
DWN & DOR	0.021
DWN & SR	0.070
DWN & DSHR	0.192
DOR & SR	1.000
DOR & DSHR	0.592
SR & DSHR	1.000

The ratings of the best compositions taken from the middle of the algorithm and the best compositions from the end of the algorithm can also be compared using the same method of a series of MWW tests, as used above. The results of these comparisons are shown in Table 6.4.

The significance results in Table 6.4 show that only one participant (subject 7) has a statistically significant ( $p < 0.05$ ) difference in ratings between the final and middle compositions. However, this distribution is consistent with the null hypothesis that there is no difference between the two groups (multinomial test,  $p = 0.17$ ).

**Table 6.4:** Mann-Whitney-Wilcoxon tests (as  $U$  statistic and  $p$  values) comparing objective ratings of Final and Middle compositions per participant (1-28), by algorithm (DWN, DOR, SR, & DSHR), with median ( $M$ ) and interquartile range ( $IQR$ ) per group

Alg.	Subject	Final		Middle		$U$	$p$
		$M$	$IQR$	$M$	$IQR$		
DWN	1	1	1.0 - 1.0	1	0.0 - 4.0	12.5	1.000
	2	7	2.0 - 10.0	8	4.5 - 9.5	11.5	0.841
	3	4	3.0 - 7.0	4	3.5 - 5.0	11	0.841
	4	3	2.0 - 6.0	4	4.0 - 5.5	9	0.548
	6	5	3.5 - 6.5	7	4.5 - 7.5	6.5	0.222
	26	6	1.5 - 8.0	7	6.5 - 7.5	8	0.421
	28	4	2.5 - 5.5	5	3.5 - 6	8.5	0.421
DOR	5	5	1.0 - 6.5	5	4.5 - 8.0	8	0.421
	7	7	6.0 - 8.0	4	2.0 - 5.5	2.5	0.034
	8	5	4.5 - 7.0	5	3.5 - 7.5	11	0.841
	9	6	5.0 - 7.0	5	3.0 - 7.5	10	0.690
	10	7	5.5 - 7.5	7	6.5 - 8.0	10	0.690
	23	8	5.0 - 9.0	4	2.5 - 5.5	3.5	0.056
	24	8	5.5 - 9.0	7	5.5 - 8.0	9	0.548
SR	12	7	5.0 - 7.5	7	5.0 - 7.5	12.5	1.000
	13	3	1.5 - 5.5	3	1.5 - 4.5	11.5	0.841
	15	8	5.0 - 8.0	5	0.0 - 7.0	5	0.151
	17	6	4.5 - 8.0	7	6.0 - 8.0	9	0.548
	18	6	4.5 - 6.5	3	1.5 - 5.0	3.5	0.056
	20	9	3.0 - 9.5	6	2.5 - 9.5	11	0.841
	22	5	3.5 - 6.0	7	3.0 - 7.5	9.5	0.548
4	11	3	2.5 - 4.0	3	2.0 - 3.5	9	0.548
	14	6	5.5 - 6.5	5	4.0 - 6.0	6	0.222
	16	5	4.0 - 6.5	5	3.5 - 8.0	11.5	0.841
	19	3	2.5 - 4.0	3	2.0 - 7.0	11.5	0.841
	21	5	3.5 - 6.5	6	4.5 - 7.5	7.5	0.310
	25	7	3.5 - 8.0	7	3.5 - 8.5	11.5	0.841
	27	8	7.0 - 9.0	7	4.0 - 8.0	6	0.222

### 6.2.2 Subjective Performance

The raw data of responses to part A of the questionnaire can be found in Appendix D. This data is summarised in Table 6.5 by the median ( $M$ ) and interquartile range ( $IQR$ ) for each algorithm and question.

**Table 6.5:** Summary statistics of questionnaire part A responses, median ( $M$ ) & interquartile range ( $IQR$ ) per question (1-4), by algorithm (DWN, DOR, SR, & DSHR)

Alg.	Question 1		Question 2		Question 3		Question 4	
	$M$	$IQR$	$M$	$IQR$	$M$	$IQR$	$M$	$IQR$
DWN	3	2 - 4	3	3 - 4	2	2 - 4	4	3 - 5
DOR	4	3 - 5	4	3 - 4	4	3 - 4	3	2 - 4
SR	4	4 - 4	4	3 - 4	4	2 - 4	4	3 - 5
DSHR	4	4 - 4	4	4 - 4	4	2 - 4	3	3 - 4

A Kruskal-Wallis test can be performed on the response data, to determine if any algorithm had statistically significantly different responses to the others. The null hypothesis for this test is that the responses to the question does not vary between algorithms. This test is run on the responses for each of the questions, and the results can be seen in Table 6.6, which shows the test statistic  $H$ , the degrees of freedom  $df$ , and the  $p$ -value approximation.

**Table 6.6:** Kruskal-Wallis tests (as  $H$  statistic,  $df$ , and  $p$  values) of questionnaire part A responses by question (1-4)

Question	$H$	$df$	$p$
Q1	2.35	3	0.503
Q2	3.22	3	0.359
Q3	0.66	3	0.883
Q4	3.94	3	0.268

These results show that the null hypothesis, that the response to a question is not dependent on the algorithm, cannot be rejected at the  $p > 0.05$  significance level for any of the 4 questions.

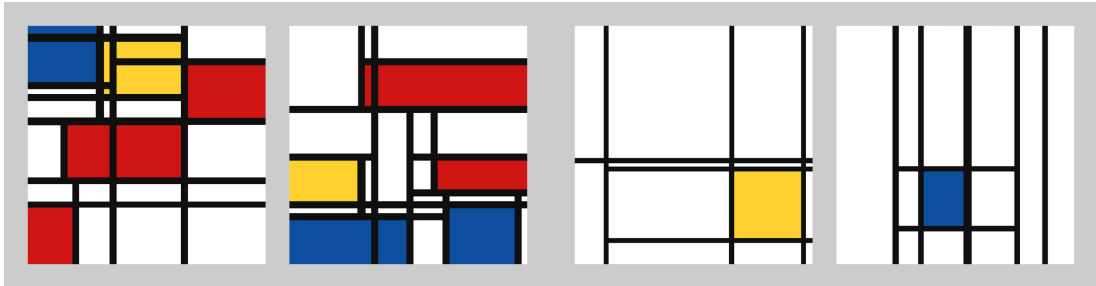
## 6.3 Qualitative Analysis

The performance of each algorithm is also evident in the compositions they generate. By looking at the compositions generated for two different participants using the same algorithm, it becomes clear that the algorithm is taking into account the user's preferences when generating the compositions, and that these preferences can



be quite varied between users. Figure 6.2 shows the two compositions from the final generation that the participant rated highest, for participants 11 and 14.

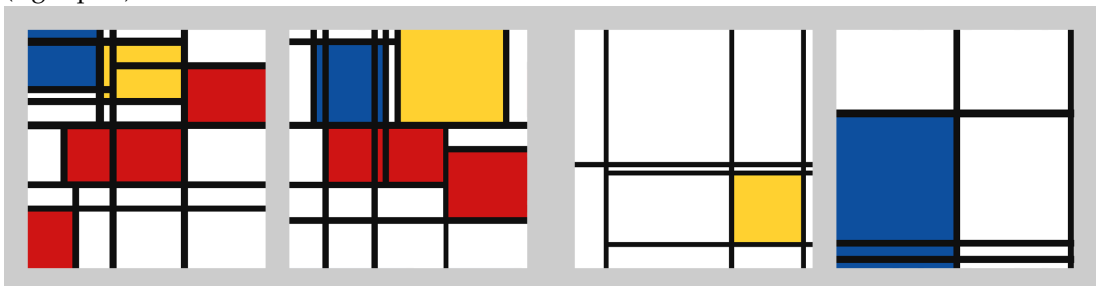
**Figure 6.2:** Highest rated Mondrian-like compositions taken from the final generation, generated by the DSHR representation for participants 11 (left pair) & 14 (right pair)



The compositions in Figure 6.2 show very different aesthetic styles. The compositions generated for participant 11 are quite busy, with many coloured rectangles, and very few unnecessary lines. This contrasts with participant 14, whose compositions contain only one fairly small coloured rectangle, and at least 2 unneeded lines. This supports the hypothesis that the algorithm can capture the aesthetic desired by the user.

This same comparison can also be performed between the highest rated composition from the final generation and the highest rated composition from the middle generation. Figure 6.3 shows this comparison for participants 11 and 14.

**Figure 6.3:** Highest rated Mondrian-like compositions taken from the final (left) & middle (right) generations, generated by the DSHR representation for participants 11 (left pair) & 14 (right pair)



The comparisons in Figure 6.3 the compositions from the end of the algorithm are reasonably similar to the compositions from the middle of the algorithm. This supports the results of the statistical analysis, which showed no significant difference between the ratings of these two groups.

## 6.4 Discussion

The statistical analysis of this data was difficult for a number of reasons. The relatively small sample size decreased the power of the statistical tests, as well as amplified the affect of any outliers. The data showed a fairly large range of results for each algorithm, making it hard for any of the statistical tests to detect significant trends in the data. Finally, the need to analyse the results from each participant separately before aggregating the results for an algorithm removed the ability to use the underlying distribution of data to improve the power of the tests.

### 6.4.1 Objective Performance

The analysis of the objective performance of each algorithm in Table 6.2 showed that the three algorithms that used the DOR, SR, and DSHR representations generated compositions that were, on average, preferred over the example Mondrian compositions. The algorithm that used the Darwindrian representation generated compositions that were not rated significantly better or worse than the Mondrian compositions.

However, the comparison between algorithms in table Table 6.3 showed only the difference between DOR and Darwindrian to be statistically significant, every other pair of algorithms produced inconclusive results. This means there is not enough data to rank the algorithms, or determine which algorithm performed best out of the four. A larger sample size may be able to determine the relative performance of each algorithm and thereby discover which algorithm performs best, however no such conclusions can be made from the current data.

The comparison between compositions taken from the middle of the algorithms and the end of the algorithms in Table 6.4 showed that there was no statistically significant difference in ratings between those two groups. This indicates that the compositions taken from the middle of the algorithms would perform similarly to the compositions taken from the end. This suggests that the length of the algorithm (15 generations) was too long, given that the algorithm performed similarly at half as many generations. However, the compositions from the final generation did not perform significantly worse, indicating that there was no large detrimental effect from running for the full number of generations, as would be the case if there was a large loss of genetic diversity.

### 6.4.2 Subjective Performance

The summary statistics of the subjective data in Table 6.5 give some insight into the average response to each question, per algorithm. The medians show that the average

---

response to every question for every algorithm was either “Agree” or “Neutral”, with the exception of Question 3 (regarding whether the generated compositions were diverse) for the Darwindrian representation, which was “Disagree”. This indicates that, overall, the subjective experience of using the algorithms was relatively positive.

However, the statistical tests applied in Table 6.6 could not determine any statistically significant difference between the responses for each of the algorithms, for any of the questions. This means that no conclusions can be drawn in regards to the relative subjective performance of the algorithms.

The high degree of similarity in responses between algorithms could either be due to there being no significant difference in the subjective performance of each algorithm, or the sample size being too small to adequately detect the presence of such an effect. In either case, a larger sample size would be able to draw more conclusions from this data.

### 6.4.3 Qualitative Analysis

The qualitative analysis in this chapter examined the compositions generated by the DSHR algorithm for two participants, 11 and 14. The goal of analysis was to determine whether the results of the statistical analysis were reflected in the appearance of the generated compositions.

The compositions shown in Figure 6.2 showed that the compositions generated for a particular participant do share many similarities in aesthetics with each other. Additionally, the compositions generated for one participant are quite different to the compositions generated for another participant, indicating that the aesthetic preferences can vary greatly between participants, and the algorithm adapts to those preferences.

The comparisons in Figure 6.3 between the highest rated composition from the final generation of the algorithm and the highest rated composition from the middle generation show that there is little visual difference between compositions taken from the middle of the algorithm and the end of the algorithm. This is consistent with the results of the statistical analysis which compared these two groups, which indicated that there was no significant difference between the ratings of the two groups.



---

## Future Research

---

The results in the previous chapter showed that while the three representations DOR, SR, and DSHR all created compositions that were, on average, preferred over the example Mondrian compositions, there was not enough data to draw statistically significant conclusions with regards to which of the representations performed best. In addition, there were no statistically significant differences in the responses to the questionnaire designed to measure the subjective performance of each algorithm.

These mixed results highlight a number of areas where future research could be useful in this field. This future research can focus both on expanding on the study presented in this thesis in an attempt to improve the significance of the results, as well as implementing complimentary studies to extend the results of this research to other areas.

### 7.1 Larger Study

The study presented in this thesis was comprised of 28 participants spread over 4 algorithms, equaling 7 participants per algorithm. Each participant took around 30-45 minutes to complete the study, making the process of data collection quite slow. Given more resources, the study could be expanded to cover a larger number of participants. This would reduce any affect from outliers in any of the algorithms, as well as improve the ability to draw statistically significant results from the data.

A larger sample size would aide the objective analysis of which algorithm produced compositions which were on average rated higher than Mondrian compositions. This analysis would then have a greater chance of being able to conclusively determine which of the four representations performed the best. Once the best representation technique was determined, a number of additional variations based on that approach could be made to discover if any further improvements can be achieved.

Additionally, a larger study would improve the likelihood of finding some statistically significant differences in the answers to the subjective questions in the questionnaire, regarding how well the participant felt the algorithm performed.

## 7.2 Paired Study

Alternatively, a new study could be implemented, based on the results of the previous study. The algorithms in this study could use fewer generations (7 as opposed to 15), as indicated by the results of the study in this thesis that showed no significant difference in ratings between compositions from the 7th generation and the 15th generation.

Fewer generations would allow each participant to complete two algorithms rather than one, without increasing the overall time required from the participant. Due to the paired nature of the resulting data, this study would be able to use more powerful statistical tests and comparisons, increasing the likelihood of drawing statistically significant conclusions from the data.

## 7.3 Followup Study

One possible effect on the results of the study presented in this thesis is a bias toward the generated compositions due to the structure of the experiment. The participants' aesthetic preferences may have been affected by viewing and evaluating a large number of generated compositions directly before giving their ratings for the final generated compositions and Mondrian compositions.

If this is the case, and the preference for the generated compositions over the Mondrian compositions is due to this effect, a followup study would reveal this and also give a more accurate measure of the actual aesthetic preferences of each participant. This followup study would consist of asking the participants of the previous study to again rate the same 15 compositions they originally rated.

This new rating data could then be analysed in much the same way as was done for the previous study, comparing the number of participants for which the algorithm generated compositions that were rated higher than the example Mondrian compositions, on average. This analysis will show if the effect found in the previous study was accurate, or if it was due to an anomaly in testing. The data from this analysis could then be used to confirm or reject the results of the previous study, by comparing the results of the two sets of analysis.

## 7.4 Other Extensions

If the results of a larger study find a statistically significant difference in performance between the three representations proposed in this thesis, further research could be

---

performed in examining whether those effects can be extended to other implementations of genetic algorithms.

This further research could take a similar approach to the research presented in this thesis, by creating a number of representations using opposing or diverse strategies and comparing the performance of each representation against the others. By studying the performance of direct and indirect representations over a number of genetic algorithm implementations, the existence of a general trend on the relative performance of these different representation strategies could be determined.

If such a trend was found to exist, it would greatly improve the ability of researchers to create more efficient genetic algorithms, especially in the area of interactive genetic algorithms for problems requiring human creativity, such as generating art or music.





## Conclusion

---

The research in this thesis has shown that interactive genetic algorithms can be used to generate artwork in the style of Piet Mondrian. The compositions generated by the algorithms in this research were liked by participants more than Mondrian's own compositions. This is a significant result that has not been previously achieved using interactive genetic algorithms.

The study did not conclusively determine several things which it was hoped it would. The analysis of the data could not draw statistically significant conclusions regarding which representation strategy performed best, nor which representation provided the best subjective performance. These issues stem from a combination of a relatively small sample size of only 7 participants per algorithm making 28 total participants, as well as a fairly noisy dataset making the real effect of the representations hard to detect.

Nevertheless, the study did show how each representation performed when compared to sample Mondrian compositions. It was shown that the three representation techniques proposed in this thesis generated compositions that were, on average, rated higher than the sample Mondrian's compositions. The representation technique used by the previous interactive genetic algorithm in this field, Darwindrian, was not rated significantly differently from the Mondrian compositions.

Previous research in this field has been relatively sparse. Interactive genetic algorithms designed to generate Mondrian art have never before been tested against original Mondrian compositions to determine if the algorithm can create more aesthetically pleasing compositions. The results in this thesis are therefore showing the feasibility of this technique, and documenting the factors to consider when designing IGAs for this purpose.

Future research in this field can use the results in this thesis as a foundation to further document the effect of factors such as phenotype representation on interactive genetic algorithms designed to generate Mondrian-like art. Additionally, research could be performed on whether the information in this thesis can be extended to other areas

that involve capturing human qualities, such as in generating music or evaluating emotion.

Interactive genetic algorithms hold great promise in connecting humans and computers to find solutions to problems that have never before been able to be solved by computer algorithms. This field has many varied applications, of which generating art is but one. Ultimately, it is hoped that this thesis has contributed to the knowledge of how interactive genetic algorithms function, and that this knowledge can be used to further the quality of IGA implementations.

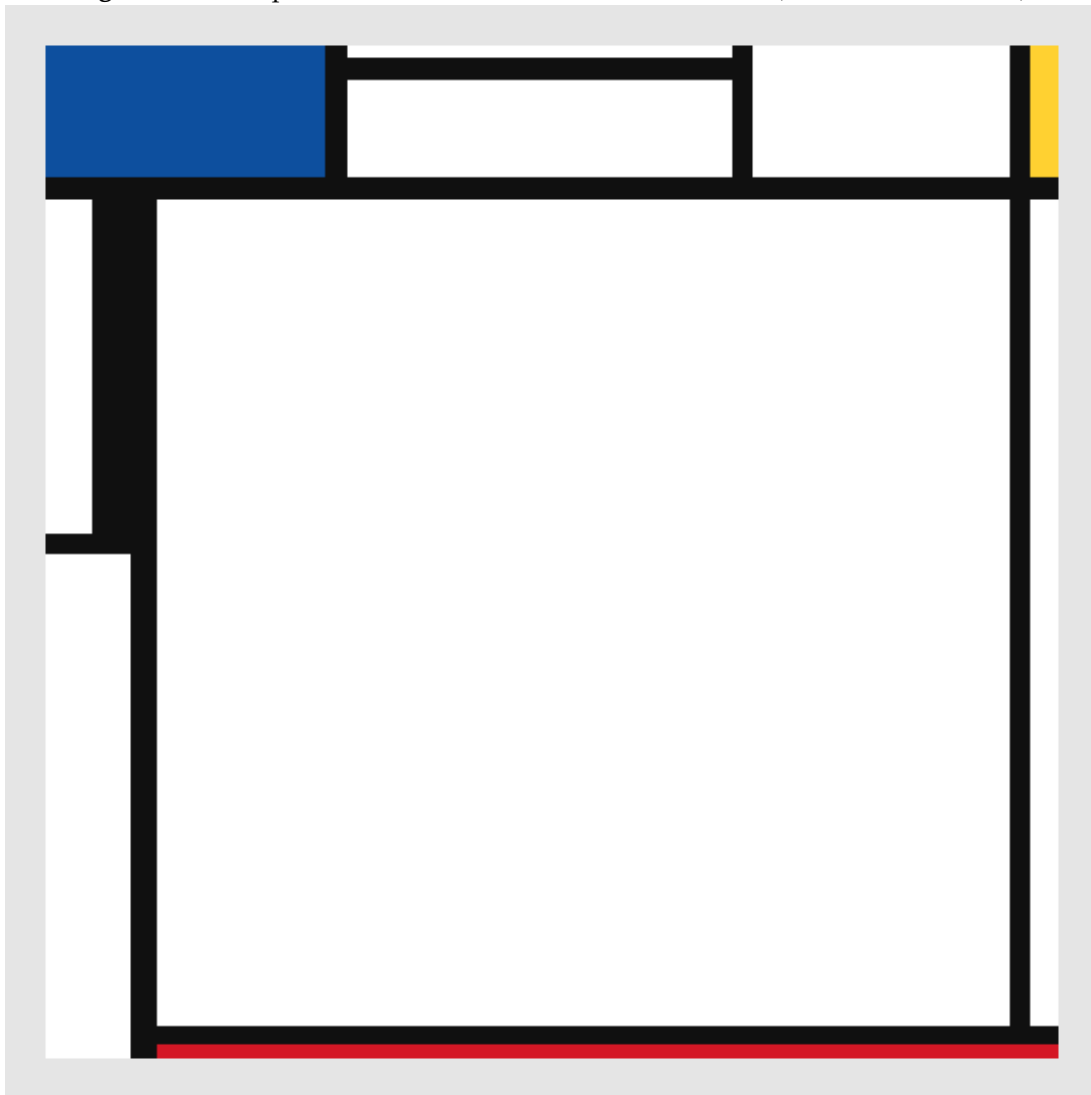
---

# Mondrian Compositions

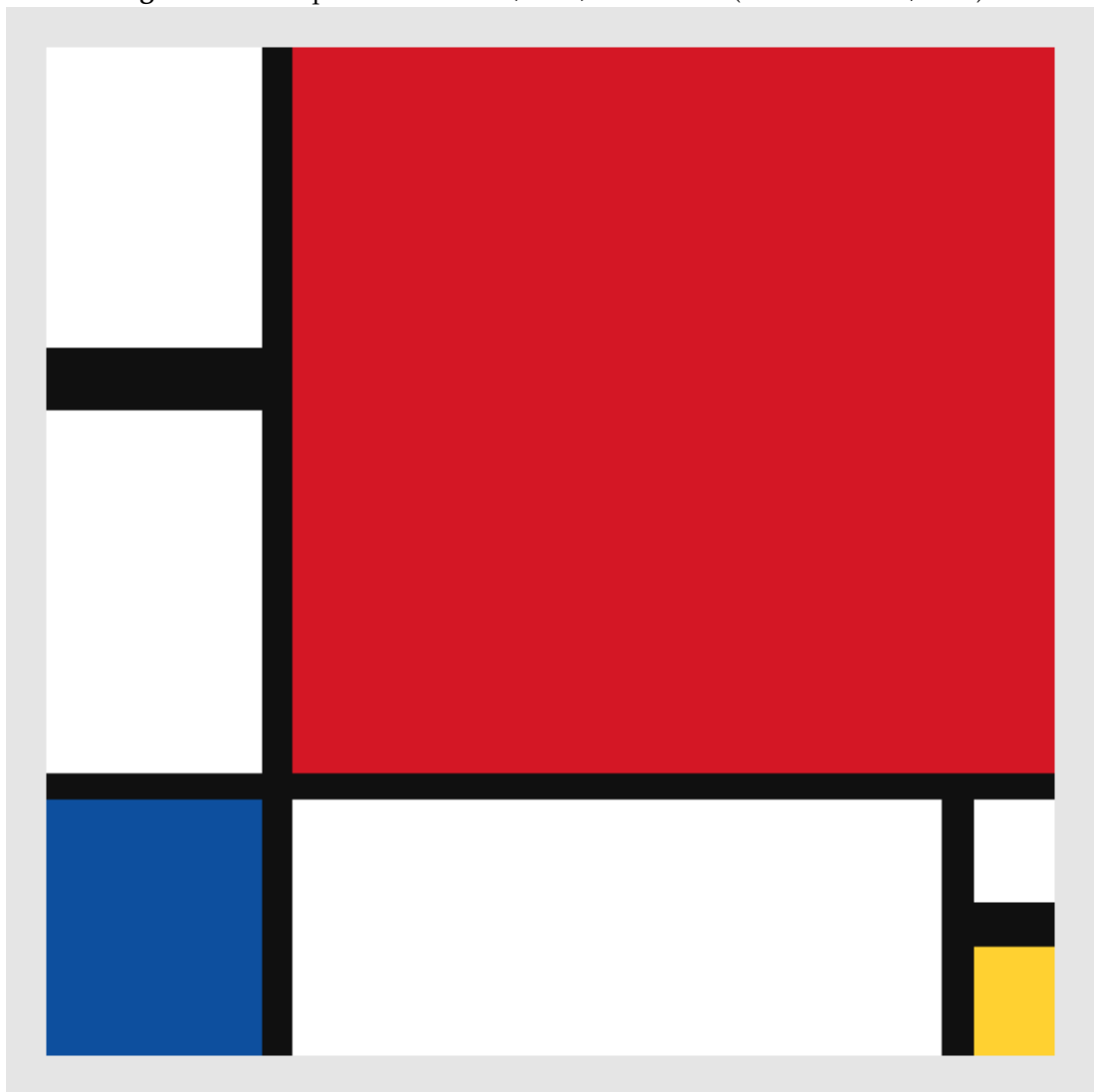
---

All Mondrian compositions in this thesis are digital recreations of the original artworks to ensure consistency with computer generated compositions. The compositions are displayed on a light grey background, to differentiate the white of the image from the white of the page.

**Figure A.1:** Composition with Blue, Yellow, Black, and Red (Piet Mondrian, 1922)



**Figure A.2:** Composition II in Red, Blue, and Yellow (Piet Mondrian, 1930)



**Figure A.3:** Rhythm of Straight Lines (Piet Mondrian, 1935)

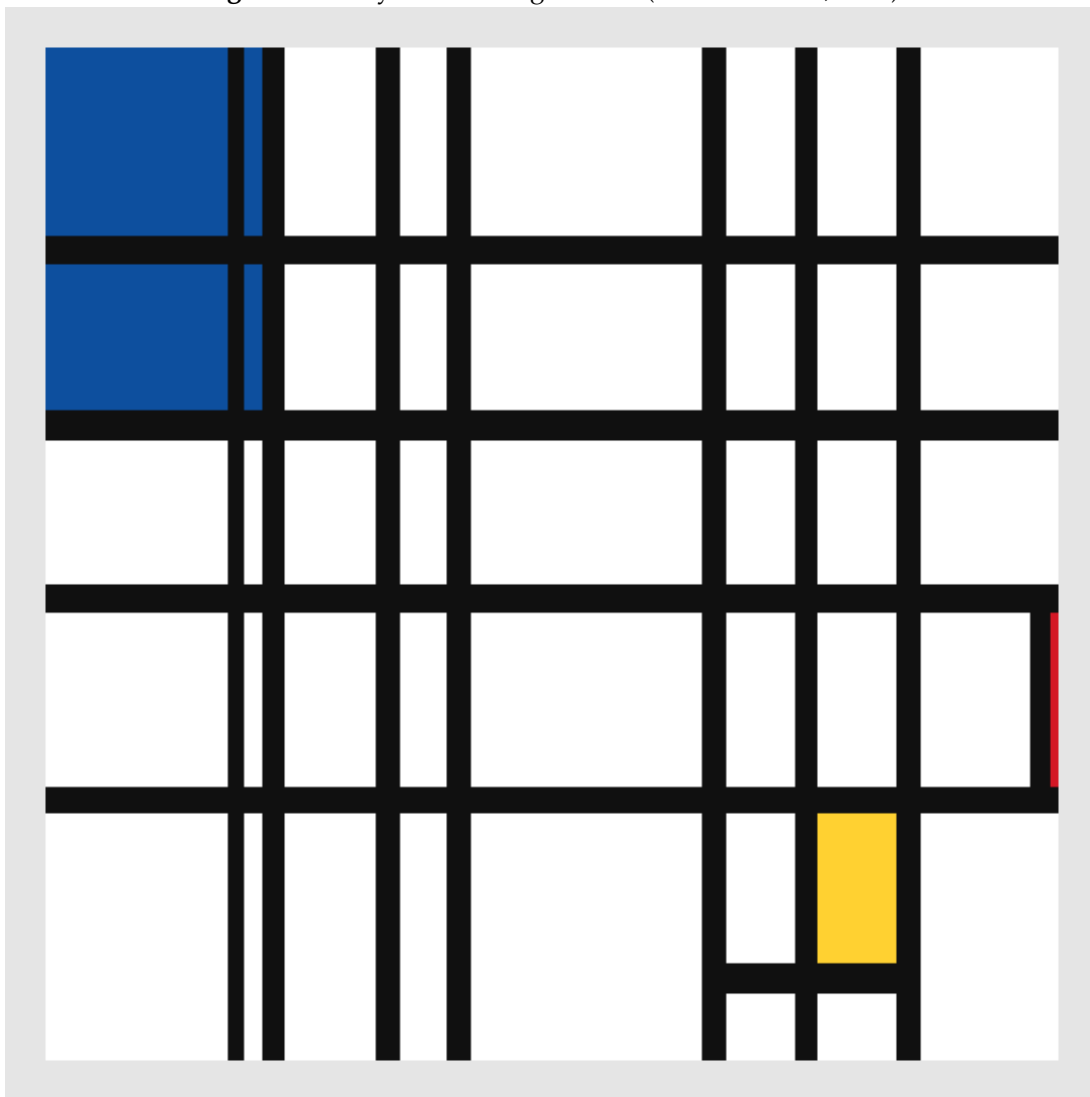
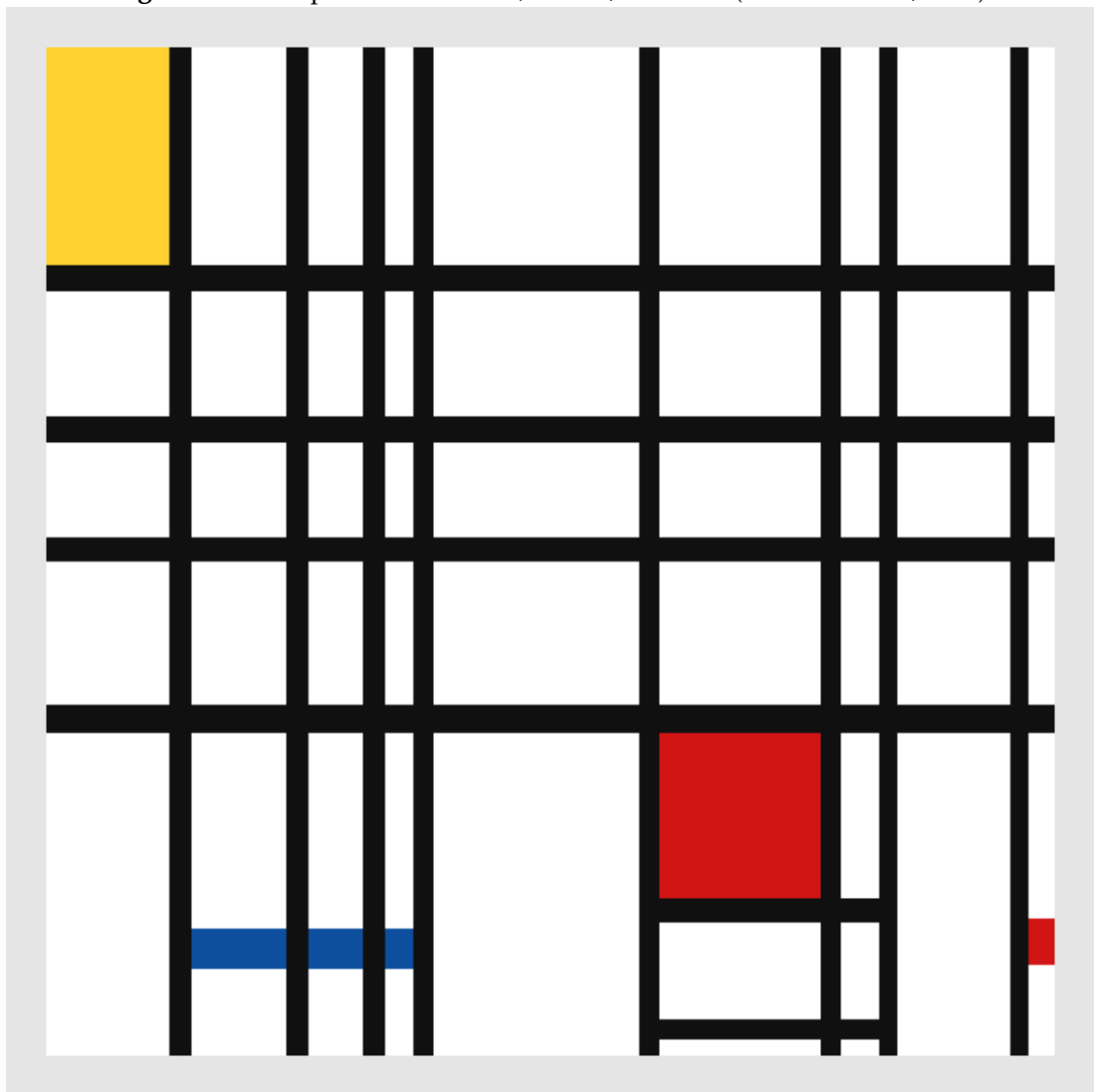
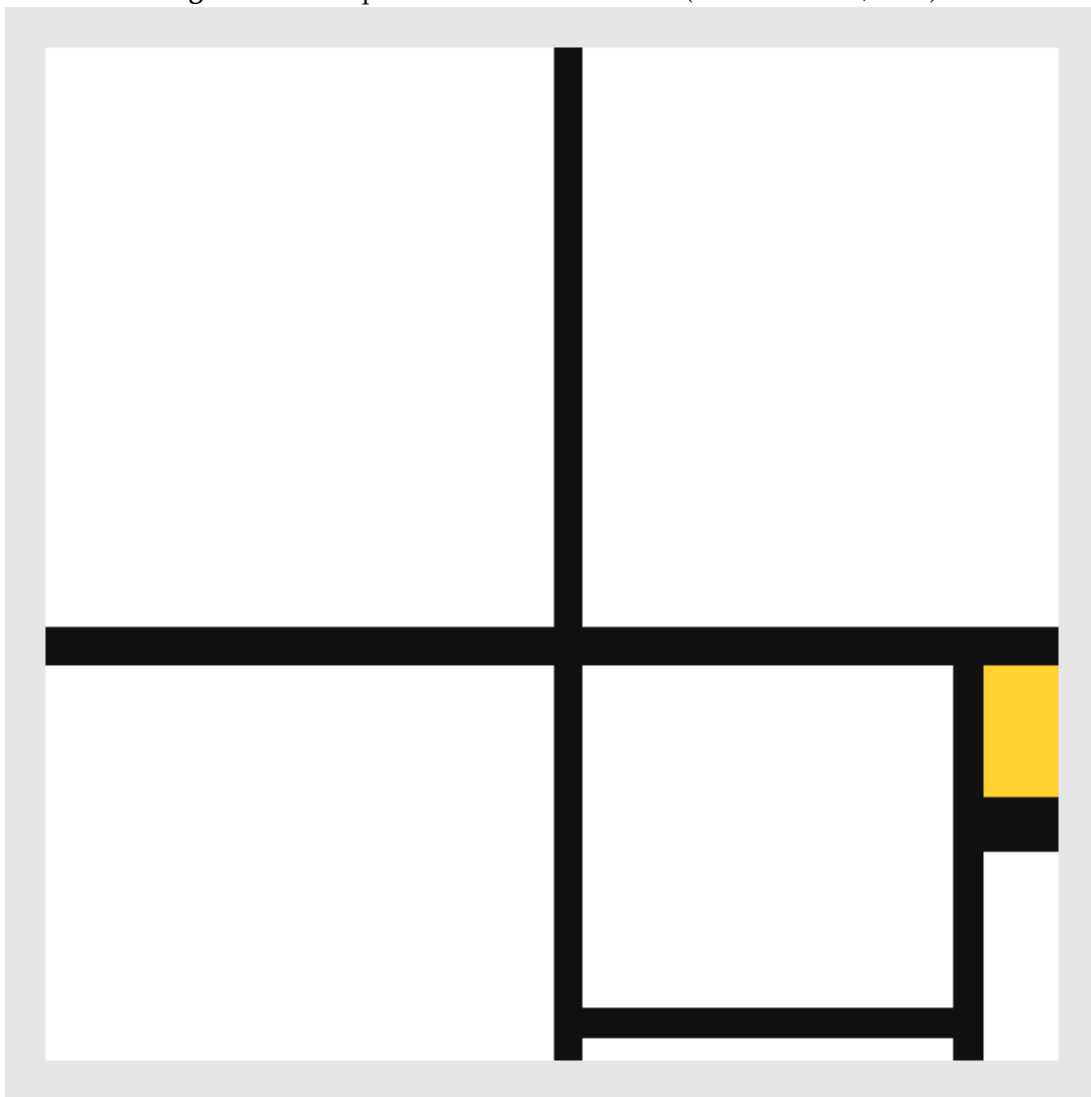


Figure A.4: Composition with Red, Yellow, and Blue (Piet Mondrian, 1927)

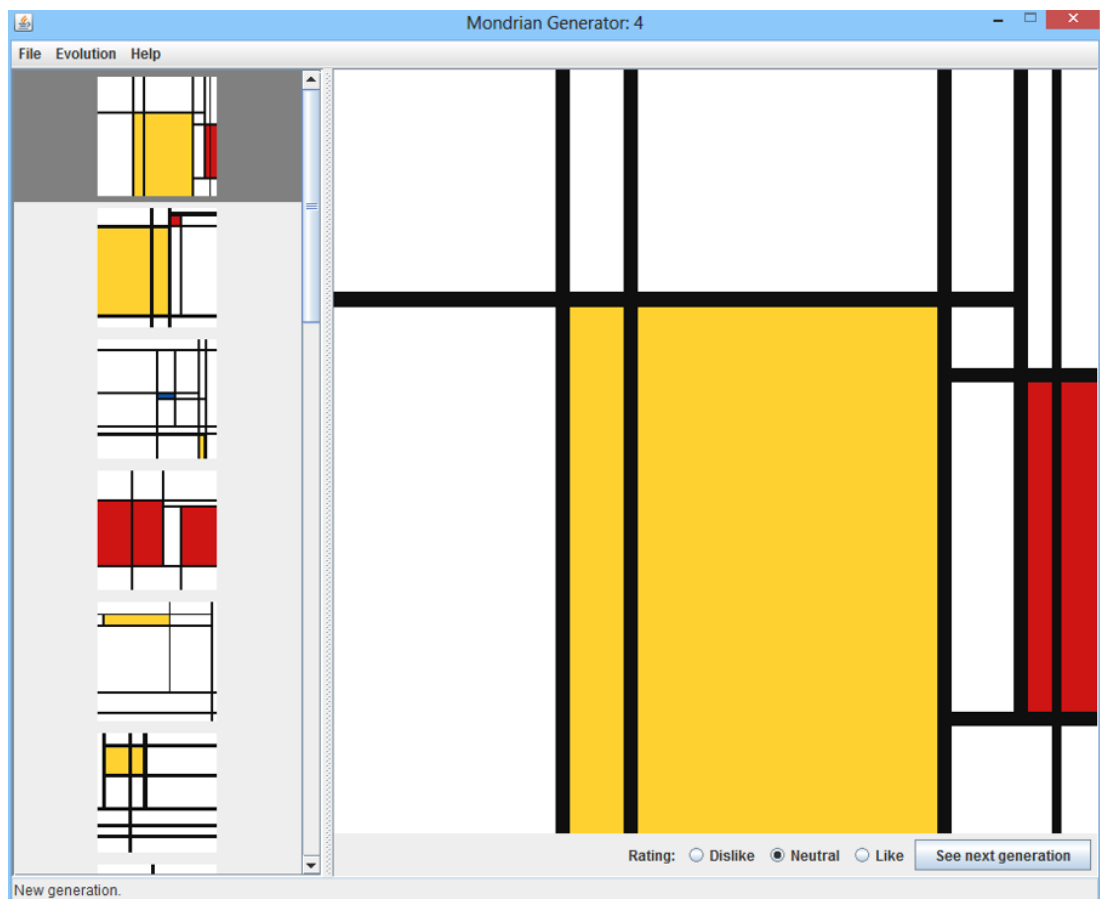


**Figure A.5:** Composition with Yellow Patch (Piet Mondrian, 1930)





# User Interface





---

# Questionnaire

---

This is this questionnaire that was given to participants once they completed an algorithm. Part A was designed to gauge the participants subjective experience with the algorithm. Part B was designed to be used to determine the cause of any abnormal data in the case of an obvious outlier.

## Generating Abstract Art using Human Interaction Questionnaire

### Part A

**1. Overall, I was happy with the final artworks that the program generated**

Strongly Disagree      Disagree      Neutral      Agree      Strongly Agree

**2. I felt as though the program took into account my preferences when generating the final artworks**

Strongly Disagree      Disagree      Neutral      Agree      Strongly Agree

**3. I felt as though the final artworks were diverse, or differed from each other**

Strongly Disagree      Disagree      Neutral      Agree      Strongly Agree

**4. I felt as though the quality of the artwork improved over the course of the program**

Strongly Disagree      Disagree      Neutral      Agree      Strongly Agree

### Part B

For pieces of artwork you liked, what features of the artwork made you like them?

---

---

For pieces of artwork you disliked, what features of the artwork made you dislike them?

---

---

Other comments or feedback:

---

---

---

---

# Raw Data

**Table D.1:** Ratings of each composition per participant (1-28), by type of composition (from final generation, middle generation, or Mondrian composition), grouped by algorithm

Alg.	Subject	Final Gen.					Middle Gen.					Mondrian				
DWN	1	1	1	1	1	1	3	0	1	0	5	0	2	0	0	0
	2	2	2	10	7	10	10	8	9	2	7	2	3	1	4	0
	3	7	2	7	4	4	5	4	5	3	4	6	7	3	5	5
	4	6	1	3	3	6	4	4	4	5	6	2	2	2	6	3
	6	7	5	5	6	2	8	7	3	6	7	8	3	2	4	3
	26	6	7	9	3	0	7	8	6	7	7	4	9	1	2	3
	28	3	2	6	4	5	5	2	6	6	5	7	6	0	0	5
DOR	5	7	1	1	6	5	5	5	8	8	4	5	10	1	10	8
	7	6	6	7	8	8	1	7	4	4	3	1	2	2	4	3
	8	6	5	4	8	5	4	3	8	5	7	2	6	5	2	3
	9	7	5	6	7	5	3	3	7	8	5	2	1	2	2	1
	10	7	7	8	7	4	6	8	7	7	8	5	4	2	1	3
	23	9	3	7	8	9	5	2	3	4	6	1	6	0	3	5
	24	9	4	9	8	7	5	7	7	6	9	3	4	3	5	5
SR	12	7	8	7	6	4	8	7	7	5	5	4	6	4	6	2
	13	3	5	6	2	1	6	3	2	3	1	0	0	1	1	2
	15	8	8	8	2	8	8	6	0	5	0	6	0	3	2	0
	17	4	5	8	8	6	8	6	7	6	8	7	8	6	7	5
	18	4	6	7	6	5	1	3	6	4	2	0	0	0	0	4
	20	1	9	9	5	10	6	9	1	4	10	4	8	0	8	0
	22	7	4	5	3	5	8	7	4	7	2	1	3	0	0	2
DSHR	11	4	4	3	2	3	3	4	3	2	2	1	0	1	1	1
	14	6	5	6	6	7	5	3	7	5	5	5	2	7	0	3
	16	4	5	7	4	6	8	3	5	8	4	6	3	3	2	2
	19	3	3	3	2	5	8	3	6	2	2	0	1	2	1	0
	21	4	5	3	5	8	6	6	9	6	3	0	1	8	2	3
	25	2	5	8	8	7	8	9	4	7	3	1	1	1	8	1
	27	7	9	7	8	9	7	7	3	5	9	0	0	7	3	4

**Table D.2:** Responses to part A of the questionnaire per participant (1-28), by question (Q1-Q4), grouped by algorithm

Alg.	Subject	Q1	Q2	Q3	Q4
DWN	1	3	3	2	2
	2	3	4	1	5
	3	4	3	4	4
	4	2	2	2	4
	6	4	4	4	3
	26	4	4	5	5
	28	2	3	2	3
DOR	5	3	3	4	3
	7	2	2	3	2
	8	4	4	4	3
	9	3	3	3	2
	10	4	4	4	4
	23	5	5	1	4
	24	5	4	4	4
SR	12	4	4	2	3
	13	2	3	4	3
	15	4	4	3	3
	17	5	5	5	5
	18	4	4	2	4
	20	4	3	4	5
	22	4	4	4	4
DSHR	11	2	4	2	3
	14	4	4	4	5
	16	4	4	4	2
	19	4	4	4	3
	21	4	4	2	3
	25	4	4	4	3
	27	4	4	3	4

---

# Bibliography

---

- BILES, J. 1994. GenJam: A genetic algorithm for generating jazz solos. In *International Computer Music Conference* (1994), pp. 131–137. (pp.2, 15)
- BILES, J. 1999. Life with GenJam: Interacting with a musical IGA. In *Systems, Man, and Cybernetics* (1999), pp. 652–656. (p.15)
- ENGELBRECHT, A. P. 2007. *Computational Intelligence: An Introduction* (2nd ed.). John Wiley & Sons. (p.7)
- FURNHAM, A. AND RAO, S. 2002. Personality and the aesthetics of composition: A study of Mondrian & Hirst. *North American Journal of Psychology* 4, 233–242. (p.11)
- GEDEON, T. D. 2008. Neural network for modeling esthetic selection. In *Neural Information Processing, ICONIP 2007, Revised Selected Papers, Part II*, Volume 4985 of *Lecture Notes in Computer Science* (2008). (pp.12, 21)
- GÓMEZ DE SILVA GARZA, A. AND ZAMORA LORES, A. 2005. Case-based art. In *International Conference on Case-based Reasoning* (2005), pp. 237–251.
- GÓMEZ DE SILVA GARZA, A. AND ZAMORA LORES, A. 2011. Evaluating an evolutionary method of design style imitation. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 25, 1–13. (pp.13, 14, 16)
- LATTO, R., BRAIN, D., AND KELLY, B. 2000. An oblique effect in aesthetics: Homage to Mondrian. *Perception* 29, 981–987. (p.11)
- LEE, J.-Y. AND CHO, S.-B. 1999. Sparse fitness evaluation for reducing user burden in interactive genetic algorithm. In *IEEE International Fuzzy Systems Conference Proceedings*, Volume II (1999), pp. 998–1003. (pp.2, 16)
- LOCHER, P., OVERBEEKE, K., AND STAPPERS, P. J. 2005. Spatial balance of color triads in the abstract art of Piet Mondrian. *Perception* 34, 169–189. (p.11)
- MAIOCCHI, R. 1991. Can you make a computer understand and produce art? *AI & society* 5, 3, 183–201. (p.24)
- MCMANUS, I. C., CHEEMA, B., AND STOKER, J. 1993. The aesthetics of composition: A study of Mondrian. *Empirical Studies of the Arts* 11(2), 83–94. (p.11)
- MONDRIAN, P. 1914. Letter to H.P. Bremmer. (pp.2, 5, 11)
- OXFORD ENGLISH DICTIONARY. 2010. “art”. In *Oxford English Dictionary*. Oxford University Press. (p.1)
- REYNOLDS, D. 2005. *Symbolist Aesthetics and Early Abstract Art: Sites of Imaginary Space*. Cambridge University Press. (p.6)

- SCHNIER, T. AND GERO, J. 1995. Learning representations for evolutionary computation. In *Australian Joint Conference on Artificial Intelligence* (1995), pp. 387–394. (p. 13)
- SCHNIER, T. AND GERO, J. 1997. Dominant and recessive genes in evolutionary systems applied to spatial reasoning. *Lecture Notes in Computer Science* 1342, 127–136. (pp. 12, 13)
- SHEN, J. Y. 2007. Cyber-genetic neo-plasticism – an AI program creating Mondrian-like paintings by using interactive bacterial evolution. (pp. 2, 14, 15, 16, 18)
- TAKAGI, H. 2001. Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation. *Proceedings of the IEEE* 89, 1275–1296. (p. 10)
- WOLACH, A. H. AND MCHALE, M. A. 2005. Line spacing in Mondrian paintings and computer-generated modifications. *The Journal of General Psychology* 132:3, 281–291. (p. 11)