Proceedings of the National Conference
On Undergraduate Research (NCUR) 2018
University of Central Oklahoma
Edmond, Oklahoma
April 5-7, 2018

A Study on the Perception of Algorithmic Composition Music

Alyssa Caputo
Department of Computer and Information Science
Brooklyn College, City University of New York
2900 Bedford Avenue
Brooklyn, NY 11210 USA

Faculty Advisor: Dr. Devorah Kletenik, Dr. Justin Steinberg

Abstract

This article examines the perception of algorithmic composition – music created through the use of artificial intelligence – in an attempt to answer the question: Is there a common bias against algorithmic (AI) music? This article provides an analysis on the perception of creativity in music and computational creativity, a history of algorithmic composition, and an experiment devised to directly determine whether a bias exists. In the experiment, participants rated two pieces of music, one composed through the use of AI and one from a classical composer, across a variety of metrics. Comparing the ratings of those who knew the identities of the composers and those who did not helped identify bias in ratings. Experimental results showed a trend towards bias against the computer-composed music and a bias towards the human-composed music, indicating that computer-composed music is not perceived as well as human-composed music. However, although there were differences in the way participants viewed algorithmic- vs. human-composed music, the differences were largely not statistically significant, possibly due to the small sample size.

Keywords: Algorithmic Composition, Artificial Intelligence, Bias

1. Introduction

In the recent backfire of Uber's self-driving car that hit and killed a pedestrian on March 18, 2018, I found myself in a discussion about the fears of artificial intelligence taking over the world. My friends, none of whom have a background in computer science or artificial intelligence, were arguing whether we can trust self-driving cars and whether they are truly safer than human drivers. Many people use artificial intelligence every day without knowing how it works and when something unexpected happens, an initial reaction can be to fear the unknown. One pedestrian death caused by an autonomous car led the New York Times to produce headlines titled, "Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam"[1] and "Uber Ordered to Take Its Self-Driving Cars off Arizona Roads."[2] Continuing the discussion, the question, "What can't AI do?" was asked, to which a friend responded, "You can't automate art." And to that, I responded, "You are so wrong." I proceeded to play a song composed using artificial intelligence in the style of Vivaldi, and they were awed.

Automatic computer composition began many decades ago in the 1950s; however, many are still unaware of its existence. Humans pride themselves on art and creativity, believing that they are the chosen ones who can exhibit these abilities. However, now that artificial intelligence can be used to create art, will it still be regarded in high esteem? Does this show that creativity is not essential to art? Or, does it show how advanced the sciences have become? In this work, we conduct an experiment to determine whether there is a common bias against algorithmic (AI) music. We argue that because human-agency and creativity are valued in music, there is a common bias against algorithmic (AI) music. As artificial intelligence creeps more and more into every field of work, it seems only natural that one day most music will be composed using AI. Therefore, if there is a common bias against this music, how will music be regarded in the future? Will it be diminished as an art form and creative outlet? Or will AI act as a tool that will allow for a new branch of creativity, pushing man beyond his limits?

2. Algorithmic Composition

Algorithmic composition is a method of music composition that uses algorithms to compose music, often referring to music that is created from a set of rules "with minimal human intervention."[3] An algorithm is simply "a set of rules for solving a problem in a finite number of steps."[4] In this case, the problem to solve is the creation of music. It may seem, however, that the real problem lies in breaking down the music creation process into a finite number of steps, and more so, how to carry out those steps with minimal human intervention. As strange as "music made with minimal human intervention" sounds, algorithmic composition is a technique that has been used for centuries, with roots tracing back to the complex music theory system of the ancient Greeks. An example of an algorithmic composition technique that was used before the invention of the computer is Wolfgang Amadeus Mozart's musical dice game named Musikalisches Würfelspiel. In this game, pre-composed bars are combined by chance, resulting in a composition that is randomly pieced together.[3] This is done simply by rolling two dice and adding their numbers; the sum decides the bars to be combined and the order of their combination. Here it is clearly seen that the human composer is, at least in part, removed from the decision-making and hands over control to some extra-human process. However, algorithmic composition has been vastly transformed by the invention of the computer and artificial intelligence.

The algorithmic composition (AI) system on which we focus is David Cope's EMI (Experiments in Musical Intelligence) system. This system is rule-based; however it is considered an expert system, which is a program that uses artificial intelligence, because it is able to create its own rules. By evaluating a large corpus of data from a particular composer, EMI is able to determine stylistic tendencies of the composer and produce its own compositions in the style of that composer. David Cope first had the idea of creating EMI in the 1980s when he was commissioned to compose an opera, but was experiencing writer's block.[5] He imagined that a tool that could produce music in his own personal style would help him to overcome writer's block quickly. Cope spent years creating EMI, but once it was running, it was able to produce hundreds of new pieces at the push of a button.[5]

EMI's production of music relies on previous compositions from a specific composer. The heart of EMI's work lies in finding a composer's signature style, through a pattern-matching analysis, and replicating those works, through the use of an augmented transition network (ATN).[6] An in-depth explanation of how EMI works is given in reference [6]. Although there are systems that are newer and further developed than Experiments in Musical Intelligence, we have focused on EMI because it is considered the first successful musical composition system and because there is vast research and literature on EMI.

3. Why A Bias?

3.1 Art vs. Science

Today, music is considered one of the five Fine Arts along with painting, sculpting, architecture, and poetry. However, music was not always considered an art. The fine arts as we know them did not even exist as a collective group until the 18th Century. Merriam-Webster defines music as, "the science or art of ordering tones or sounds in succession, in combination, and in temporal relationships."[7] The phrase "science or art" gives evidence to the fluctuation in the categorization of music over the centuries. After Pythagoras discovered that numerical proportions were intrinsic to musical intervals, music theory became likened to the mathematical sciences.[8] Music was then defined as "the science of the harmonic relationships of the universe."[9] To the Ancients, it must be noted, music and dance were considered parts of poetry and "the arts" referred to many activities that would now defined as crafts and sciences. Additionally, music was considered a liberal art along with grammar, rhetoric, dialect, arithmetic, geometry, and astronomy.[8] From the grouping of the liberal arts, it is clear that science and art were not yet separated as clearly as they are in current society.

The Renaissance brought about many changes to the arts and laid the foundation for their grouping. By 1563, the visual arts (painting, sculpture, and architecture) became so popular and distinctive as social and cultural mediums that the artists broke ties with their craftsman guilds in order to establish the first Academy of Art.[8] This is significant in that it caused a shift in their categorization from a craft to an art. Sally Markowitz, a professor of philosophy at Willamette University, writes, "'Art' has positive evaluative connotations that 'craft' lacks."[10] She analyzes the distinction between the two according to aesthetic and semantic criterions. While crafts (such as knitting and pottery) have practical and utilitarian functions, the arts often have only an aesthetic purpose.[10] At the semantic level, crafts are usually not interpreted, whereas arts require an interpretation: "We do not, after all, usually *interpret* finely crafted wooden benches, hand-thrown mugs, woven shawls; we sit on them, drink from them, wear them. In short, we use them."[10]

During the 17th century, many discoveries in the natural sciences had occurred. Realizing the importance of these achievements led to the famous debate *Querelle des Anciens et Moderne* (Quarrel of the Ancients and Moderns), which compared the achievements and authority of the ancients to those of the moderns. In this comparison, it was established that the progress of the moderns over the ancients was most clearly seen in the fields that depended on "mathematical calculation and the accumulation of knowledge."[8] However, in the fields that depended on "individual talent and the taste of the critic," progress could not be clearly seen.[8] This was one of the first times that art and science were distinctly separated. While the ground was set for the present grouping of the fine arts, they were not explicitly grouped until 1746 in Charles Batteux's treatise *Les beaux arts réduits à un même principe* (Fine arts reduced to the same principle). This was the first treatise dedicated entirely to the arts that created a clear system of division. Batteux lists the fine arts, almost in their final form, as those that "have pleasure for their end" and that display an "imitation of beautiful nature:" music, poetry, painting, sculpture, and dance.[8] Music and poetry are imitation "expressed by sounds, or by measured discourse" and are representative of "human vocal utterance, and other forms… of human emotive expression."[11] From his influence in France as well as across Europe, Batteux solidified the concept of fine art, which has persisted until the present. However, the introduction of artificially composed music, especially if a bias against it is present, could diminish the value of music as an art form. Cognitive scientist Douglas Hofstadter writes, "Once some mental function is programmed, people soon cease to consider it as an essential ingredient of 'real thinking.'"[12] This could cause music to return to a closer association with the sciences and even blur the line between art and science once again.

3.2 Three reasons for bias

We hypothesize that a bias against algorithmic (AI) music is present because of three main reasons: first, because computer science is categorized as a science, as opposed to an art, and creativity is more closely associated with the arts; second, because of old programming techniques which did not demonstrate any signs of creativity; and third, because creativity is viewed as a unique human attribute and there is a fear that computers will someday replace mankind. Since science provides knowledge through facts and art is much more subjective, creativity has historically been more closely associated with art. Music is presently considered a fine art and therefore, many view creativity as an essential component of music composition. Furthermore, since computer science is labeled as a science and since computers know only math at their core, computers have not had a reputation of being creative. Computer scientist and mathematician Donald Knuth writes, "Science is what we understand well enough to explain to a computer; art is everything else."[13] The division between science and art makes it easy for many to assume that computers cannot exhibit creativity.

This assumption may be due a lack of knowledge about how computer systems work. Since its invention, the modern computer, previously known as the analytical engine, has known nothing but strings of zeros and ones. A skilled programmer had to tell the computer exactly what to do line by line, which is known as imperative programming. It knew nothing more than what it was explicitly told. This view is still how many believe computers work; however, the computer no longer needs line-by-line instructions. With the development of artificial intelligence and machine learning, computers can now learn to deduce rules from data and make decisions. Cognitive scientist Matthew Elton writes, "The main reason why the whole notion of a creative computer seems so outrageous is due to a crude understanding of what computer systems can do… [They] are generally thought of as consisting of a set of rigid rules."[14] He discusses what he calls implastic and plastic systems, those that require rigid rules, and those that are "'trained' rather than programmed," respectively.[14] A plastic system is self-modifying; it can track its own states and modify its rules of operation through training cycles. However, Elton says, "We too readily slip into the assumption that there is only the implastic view."[14] By learning how computer systems work, humans may become more accepting of the idea of computational creativity. Lastly, humans may show bias against algorithmic (AI) music because we view creativity as a unique human attribute and there is a fear that computers will someday become better than humans in all their abilities. Since creativity is often regarded as a unique human attribute, humans may feel threatened by a computer's ability to act creatively. Colton et al. write, "One has to ask the question, 'Under full disclosure, would people value the artifacts produced by a computer as highly as they would the human produced ones?' … [in] domains such as the visual arts, the answer is very likely to be no."[15]

David Cope himself has conducted Turing Tests as a testament to EMI's abilities. A Turing Test is a method of determining a computer's intelligence; if a human being is not able to distinguish between a computer and another human's abilities, the computer has passed the test and is deemed intelligent. Cope has found that, "when [people] assume [EMI's] music is human, they are obviously moved and speak in the same terms as if it had been by Chopin, but when I tell them that there is nothing behind the music but cold hard machinery doing addition and subtraction, they won't admit they were moved."[16] Additionally, in a review of EMI's music, professor of music composition and technology Michael Theodore wrote:

> Does the program create new works that are convincingly in the style of the works in its database? … this is necessarily a subjective question, and it becomes even harder to answer on account of the biases that people invariably bring to the experience of listening to computer-composed music. People listen with a completely different set of ears when told before hand that the music has been composed by computer (by listening for the types of awkward passages they think a computer will create, by assuming that there won't be a perceptible "emotional" core, etc.)[17]

Our goal in this work is to determine whether this suggested bias can be proven.

One of EMI's pieces in the style of Bach was also featured in a Turing Test in 1997, hosted by Professor Hofstadter at the University of Oregon. A pianist performed three pieces in the style of J. S. Bach, which were composed by EMI, Dr. Steve Larson, and Bach himself. The audience selected EMI's piece as the real Bach piece and Dr. Steve Larson's as the computer composition. In response, Dr. Larson stated, "My admiration for [Bach's] music is deep and cosmic. That anyone could be duped by a computer program was very disconcerting."[18] While Professor Hofstadter said, "I find myself baffled and troubled by EMI. To what extent is music composed of 'riffs'? If that's mostly the case, then it would mean that, to my absolute devastation, music is much less than I ever thought it was."[18] Additionally, he stated, "[EMI] has not a trace in it where I think music comes from."[19] This has led Hofstadter to believe that one of the following is true: "music as a whole is not very deep, humans in general are not very deep, or computer programs are much deeper than we ever could imagine."[19] It is clear that EMI's ability to create music at the level of Bach is unsettling because it questions the role of creativity in music. As a result, Professor Hofstadter's view of music was diminished. If the experiment results show that a common bias is present, possible implications are that music will be diminished as an art form and that the clear line between science and art will be blurred.

4. Related Work

In 2008, researchers at Stanford University and California Institute of Technology conducted an experiment to determine whether the price of a bottle of wine has an affect on the perceived quality of the wine's taste.[20] They hypothesized that an increase in price would cause an increase in experienced pleasantness because quality is positively correlated with price.[20] Additionally, they hypothesized that this anticipated pleasure would cause an increase in brain activity in the part of the brain that experiences pleasure, the medial orbitofrontal cortex (mOFC). The participants were twenty healthy subjects who drank red wine on occasion. They were told by researchers that they would be trying five different types of Cabernet, which would be identified by their price. In reality, they tried only three different types of wine. Two wines were presented twice, but with different prices (its real price and a fictitious price).[20] However, the participants stated that they could in fact taste five different wines and found what they perceived as the higher priced wines to be better in taste. Brain scans (fMRI) revealed that activity in the mOFC increased when tasting the perceived higher priced wines as well.[20] This shows that the price of a bottle of wine affects not only the perceived quality of that wine, but also the activity in the part of the brain that experiences pleasure, therefore affecting its actual quality. Although this experiment was conducted specifically for wine, it shows in a general sense how perceived quality can be affected by the reputation of a product's creator. This same effect may be exhibited in our experiment on algorithmic (AI) music. EMI's "reputation" as a computer system, which is not thought of as being creative and artistic, may cause the perceived quality of its music to decrease. On the other hand, Bach's reputation as a genius classical composer could cause the perceived quality of his music to increase, much like the wine.

In 2012, researchers from KTH Royal Institute of Technology in Sweden conducted a listening study in which fourteen computer compositions were compared to twenty-one human compositions.[21] They created a computer program that algorithmically composed popular music. This program was built on a statistical analysis of popular music, which served as the model for the generation of new music.[21] The user could control settings such as tempo and rhythm pattern through the interface, and the composition process would then begin. Although the program created rhythm, harmony, and phrase structuring, the melody was the focus of the composition. Five groups of songs were tested, each of which contained seven songs. There were two groups of computer compositions, one whose songs were selected by the author and one whose songs were randomly selected. The third group contained human compositions, which were the winners of a Norwegian music competition called the "Melodi Grand Prix" from the year 1976 to 1982. The two remaining groups were composed by the author Elowsson; they contained songs with the same melody, but with a good and bad Global Joint Accent Structure (GJAS), respectively. GJAS refers to the alignment of melody and rhythmic accents.[21] The songs were played in a random order to eighteen participants who were between the ages of twenty and thirty. Their self-rated experience as musicians averaged to 4.2 out of 7. The participants were asked to judge the following attributes on a seven-point scale: good, groove, human, and stressed. Their results showed a slightly lower rating for the computer compositions, but the difference was statistically insignificant. This study was blind and attempted to analyze the real differences in ratings of music; it did not take into effect the perception of quality caused by knowledge of the composer, which is the subject of our experiment.

Another experiment, which was conducted at the University of Rochester, compared chorales by Johann Sebastian Bach to those produced by David Cope's EMI system.[22] Although this experiment was not a listening study, it compared entropy levels in 200 songs composed by Bach and EMI in order to evaluate EMI's ability to create music that is truly in the style of Bach. Entropy measures the unpredictability of information; in terms of music, it measures "the frequency of musical events."[22] Specifically, entropy window (the entropy of localized, equal-length segments) and entropy profile (the entropy of increasing length segments) were evaluated on both the pitch and duration level. The results showed a stark difference in overall entropy levels between EMI's and Bach's music; EMI's music had higher entropy levels in general, but especially at the pitch level. This indicates that EMI's music is less predictable and less repetitive.[22] High entropy levels are not necessarily desirable or undesirable in music; however, in this examination, the closer EMI's entropy levels are to Bach's, the better EMI is said to imitate Bach's style. They showed that EMI's pieces begin at a higher entropy level and continue to increase, whereas Bach's pieces begin at a lower entropy level and fluctuate much more. Researcher Iris Yuping Ren says, "The unnaturalness [of EMI's music] is showing itself in the growing smooth curves."[22] This higher entropy profile is due to the lack of repeated pattern in EMI's music. When listening to a piece of music, we form expectations about the patterns that we will hear; therefore, EMI's high entropy value suggests that its patterns do not correspond to the listener's expectations as well as Bach's.[22]

5. Experimental Design

The experiment that we conducted sought to answer the question: is there a common bias against algorithmically composed (AI) music? This study was conducted in the form of an online questionnaire, which was made through Google Forms. Each questionnaire contained two songs: one composed by a human (Johann Sebastian Bach) and one composed algorithmically, using artificial intelligence (EMI). The participants were randomly assigned into control and experimental groups. Participants in both groups were asked to listen to one song and answer the questions that followed. Then, they were asked to listen to the second song and answer the same questions. The control group did not know the composer of each song; this group acted as the baseline rating for each song. The experimental group was given the composer identities, which tested whether the ratings of EMI's or Bach's music changed given the knowledge that a computer system or human composed it. When providing the composer identities, the questionnaire stated, "This song was composed by an Artificial Intelligence system named EMI" and "This song was composed by Johann Sebastian Bach."

Music that was composed by David Cope's EMI system was used as the algorithmically composed music because EMI was one of the first successful music composition programs. Additionally, EMI's music imitated that of J. S. Bach's particularly well. Therefore, two songs were selected that were composed by EMI in the style of Bach: Bach-2 and Bach-3. These pieces are chorales, which are hymns written in a four-part harmony.[23] They were available to download from David Cope's website.[24] Cope has released five thousand chorales in the style of Bach, which gives evidence to the power of EMI as a compositional tool. Since there was a wide range of chorales to choose from, the song selection was limited to two pieces from the first ten in the list. Because those pieces by EMI were chorales composed in the style of Bach, chorales composed by Bach himself were used as the human-composed music. This ensured that the difference in judgment between human- and computer-composed music would not be due to a difference in style or form of the music, which decreases the participants' subjectivity as a factor. The selected songs that were composed by Bach were Freuet euch, ihr Christen alle (Rejoice, you Christians all) and Ich dank dir, lieber Herre (I thank you, dear Lord).

Two songs were selected from each composer to ensure the validity of the experimental results. Choosing more than two per composer would have exponentially increased the experimental size, necessitating even more participants to obtain valid results. We controlled for variables that could affect the difference in ratings, including length of the songs, combination of songs, and ordering. All four songs are approximately one minute long and we created a set of four questionnaires that features every possible combination of the two Bach and two EMI songs. The songs were evenly ordered among the questionnaires, with two beginning with a Bach song and two beginning with an EMI song. Finally, the division between the control and experimental groups was created by duplicating each of the four questionnaires so that one set did not name the identity of the composer (the control group) and the other set did (the experimental group); this gives a total of eight questionnaires.

The questionnaire contains twenty questions, eighteen of which were Likert-scale responses on a scale of one to ten, to enable a quantitative analysis of the responses. The questions took the form of: In your opinion, how "x" did the song sound? The attributes that the participants were asked to judge were: *good, groovy, human-like, stressful, mechanical, robotic, emotional, expressive, fluid, natural, creative, dynamic, emotive, original, repetitive, rhythmic, lifeless,* and *rich*. Four of these attributes (*good, human-like, stressful*, and *groovy*) are based on the attributes in [21]; *repetitive* was based on the study in [22] on the difference in entropy levels between human- and algorithmically-composed music. Some of the attributes listed are synonymous with each other, such as *mechanical* and *robotic*, or opposites, such as *human-like* and *lifeless*. This was done not only because the judgment of music is highly subjective, but also because two words, no matter how similar in definition, never have the same exact meaning or connotation. By listing opposing attributes, the participants explicitly judge a piece on that attribute, rather than implicitly from judging the opposite attribute. In addition to the scaled questions, the questionnaire also features two open-ended questions. The first asks, "If there was a specific emotion that you felt while listening, please list." The second is the last question of the study; it asks, "Do you consider this music a work of art?" The responses to this question are particularly important in that they provide a glimpse into how music may be perceived in the future, if algorithmic (AI) composition is to become a pillar of the music industry.

6. Experiment Results

A total of 86 students from Brooklyn College participated in the study, 59.3% of whom were either majoring or minoring in computer science or multimedia computing. 45 participants were assigned to the experimental group and were given the composer identities, and the remaining 41 were assigned to the control group and were not given the identities. The analysis was conducted using one-way ANOVA, which stands for analysis of variance. This analysis determines whether there is a statistical difference between two groups that differ by one variable. Because we were actually measuring two variables, two-way ANOVA could have been used. However, two-way ANOVA requires equally sized samples and because of the random assignments to groups, our control and experimental groups were not balanced. Rather than discard some data to balance the sizes of our groups, we used one-way ANOVA twice.

The analysis was done once for each of two variables: 1) participant knowledge of composer and 2) identity of composer (Bach vs. EMI). Therefore, four different groups were analyzed: named EMI and unnamed EMI, named Bach and unnamed Bach, named EMI and named Bach, and unnamed EMI and unnamed Bach. The hypothesis (called the *alternate hypothesis*) for each set of analyses was that there would be a difference between the groups; the *null hypothesis* for each of the analyses was that there would be no difference between the groups. The null hypothesis should be rejected if there is sufficient evidence in favor of the alternate hypothesis. A p-value, or probability, of 0.05 was used to test for significance, which represents the probability of correctly rejecting the null hypothesis. A low p-value indicates that the null hypothesis should be discarded because it is inconsistent with the evidence. This implies a high probability that the alternate hypothesis is correct and that there is a statistically significant difference between the two groups

Our hypothesis was that the results would show a common bias against EMI's music, which is indicative of a lower perception of quality that is not backed by actual feelings. This hypothesis can be broken down into two sub-hypotheses: 1) Unnamed EMI would be given higher ratings than named EMI and 2) named Bach would be given higher ratings than named EMI but that this difference would not be seen in the unnamed Bach vs. unnamed EMI comparisons. Overall, the results showed a strong trend towards consistency with the hypotheses, though the differences were largely not statistically significant (possibly due to a small sample size). The analysis was first done with a linear function of the attributes, in which each response was given a score: the sum of each of the attributes multiplied by a constant coefficient. The "positive" attributes, for which a higher score was considered a better rating (e.g. *good, creative*), were given a coefficient of +1. The "negative" attributes, for which a lower score was considered a better rating (*stressful, mechanical, robotic, repetitive*, and *lifeless*), were each given a coefficient of -1. The attribute *work of art* was given a numeric value (0 if it was not considered a work of art and 1 if it was) and a coefficient of +10 in order to balance it with the other attributes. Then a sum of the attributes was calculated to give one combined score from each participant. For example, if the participants responses were good: 5, groovy: 4, human-like: 3, emotional: 2, expressive: 4, fluid: 7, natural: 5, creative: 3, dynamic: 3, emotive: 2, original: 7, rhythmic: 9, stressful: 2, mechanical: 8, robotic: 8, repetitive: 6, lifeless: 2, work of art: 1 (for *yes*). The responses are transformed into the function $5 + 4 + 3 + 2 + 4 + 7 + 5 + 3 + 3 + 2 + 7 + 9 + 4 – 2 – 8 – 8 – 6 – 2 + 10 = 42$.

42 then becomes the score for that particular participant. Note that the response to the question "If there was a specific emotion that you felt while listening, please list" was omitted because it is not quantitative at all. Overall, the named Bach group received higher scores than the named EMI group, with average scores of 59.73 and 48.93, respectively. This difference was marginally significant (p-value of 0.053).

The analysis was then repeated for each individual attribute, some of whose ratings were significantly different between the groups. The following diagrams depict the trends in the data for the questions in which a higher rating (closer to 10) was considered better. The questions in which a lower score was considered a better rating were omitted from these diagrams. The open-ended question "Do you consider this music a work of art?" was given a numeric value (0 for *no* and 1 for *yes*) for the purposes of the analysis.

Figure 1 is a graph of the average ratings between the group that knew EMI was the composer and the group that did not. In almost every case, the value of EMI's music decreased when the composer identity was provided. The two attributes with significant differences between these groups were *stressful* and *robotic*, which are not included in Figure 1. The average rating for *stressful* when EMI was not known as the composer was 2.5, but the average rose to 3.56 when EMI was named as the composer. This had a p-value of 0.007. Similarly, the average rating for *robotic* was 4.25 when the composer was not known, but 5.28 when EMI's name was provided, having a p-value of 0.05. This suggests that participants think of computer compositions as robotic and associate them with stress.

On the other hand, the value of Bach's music increased in every instance when the composer's identity was provided, which is shown in Figure 2. This indicates that the perception of music's value is affected by its composer, similar to the findings in Stanford's experiment on the perception of wine.[20] Since Bach is a famous classical composer and known to be one of the greatest composers of all time, the perceived value of his music is increased when his name is provided; however, since EMI is an artificial intelligence system, the perceived value of its music is decreased when its named was provided. Additionally, the difference in ratings for the attribute *original* was statistically different in this comparison, with a p-value of 0.022. This is possibly due to the association of originality with a human or with a genius, which many consider Bach to be.

The control group, which did not know the composer identities, was meant to provide the baseline, true rating of the songs. The hypothesis for this group was that it would show the same or very similar ratings for each song, since EMI's music was composed in the style of Bach. The results were surprising in that the perceived value of EMI's music, although similar to that of Bach's, was almost always higher than the perceived value of Bach's music, which is shown in Figure 3. The one characteristic for which this is not true is *emotional*, which means that the listeners may have been able to detect the lack of emotions in EMI's music. However, the attributes *emotive* and *expressive* have higher averages for EMI's music. It is important to note that this was a within-subject analysis, meaning the same participants were in both the unnamed EMI and unnamed Bach groups. Therefore, a difference in ratings cannot be due to different people's opinions between groups. In contrast, the previous two analyses (named EMI vs. unnamed EMI and named Bach vs. unnamed Bach) were between-subjects, meaning the named groups contained different participants from the unnamed groups. The differences in the ratings from those analyses could therefore be due to different people's opinions. We note that there were no statistically significant differences between these two groups (unnamed EMI and unnamed Bach), showing that the quality of these two pieces seemed to have been viewed similarly.

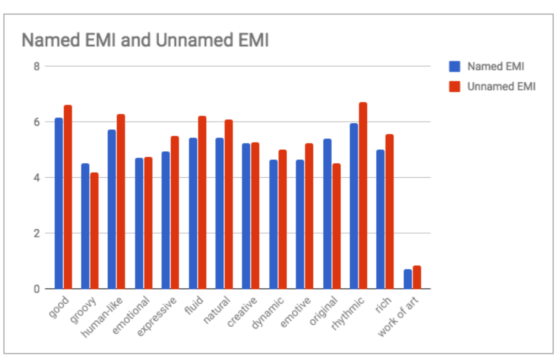

Figure 1. Average ratings for the named EMI and unnamed EMI groups

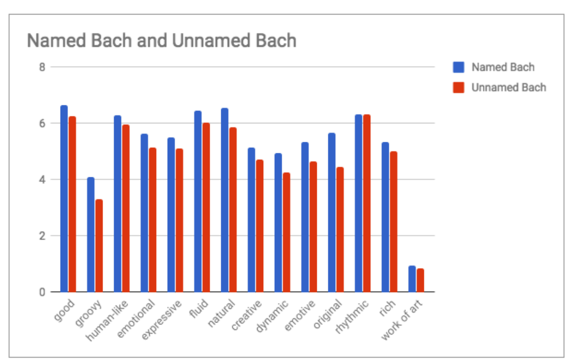

Figure 2. Average ratings for the named Bach and unnamed Bach groups

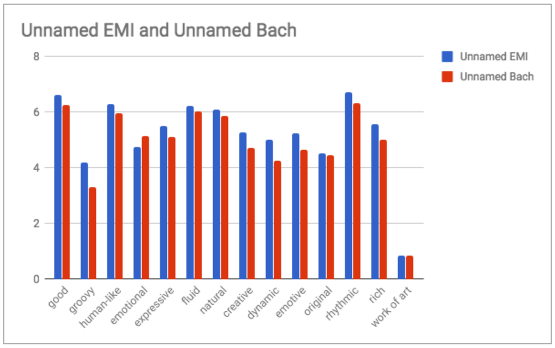

Figure 3. Average ratings for the unnamed EMI and unnamed Bach groups

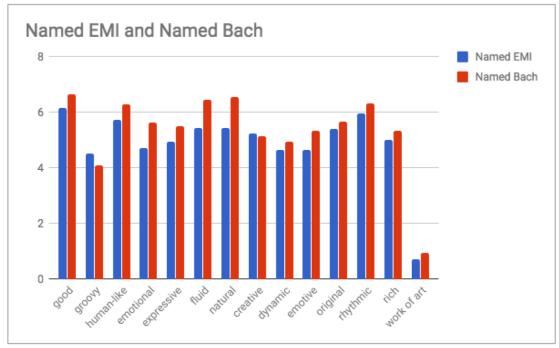

Figure 4. Average ratings for the named EMI and named Bach groups

The final analysis was conducted on the named EMI and named Bach groups. The number of significantly different ratings in attributes was highest between these groups. This is important because this group was also within-subjects. Therefore, as in the previous analysis, the statistically significant differences in this analysis were truly due to the participants' valuing Bach's music over EMI's. This is shown in Figure 4. The attributes for which there was a significant difference were *natural, stressful,* and *fluid,* which had *p*-values of 0.016, 0.016, and 0.023, respectively. The average ratings for *natural* and *fluid* were at least an entire point lower for EMI's music, and the average rating for *stressful* was an entire point higher for EMI's music. Additionally, *emotional* and *robotic* were close to the 0.05 threshold with *p*-values of 0.070 and 0.073, respectively. Surprisingly, the participants were not able to distinguish the different levels of entropy between EMI's and Bach's music. In the named groups, the average ratings for *repetitive* were almost identical, with 6.86 for Bach's piece and 6.81 for EMI's. In the unnamed groups, the ratings were very close again with an average of 6.17 for Bach's piece and 5.97 for EMI's.

Most importantly, the attribute *work of art* was shown to be statistically significant (*p*-value of 0.002), with an average value over 0.2 points (20%) higher for Bach. The *work of art* attribute was asked through the open-ended question, "Do you consider this music a work of art?" and some comments were insightful. For those who were told the composer names, answers for EMI's music included: "Yes but a rather stale work of art," "No, it didn't really have layers. It sounds like a piece that someone who was learning how to play the piano would use to practice with," and "no, Its [sic] seems very robotic or computer generated." Even more interesting were these comments from participants who were *not* told that the song was computer-generated: "Sounds like AI generated music which I actually fully support. Answer would be yes, contingent of better production of midi" and "It can be considered a work of art however is lacking in emotions." When asked, "If there was a specific emotion that you felt while listening, please list," one participant responded, "Nothing at all. Felt like it was made by a computer," even though the composer identity was not provided. Since 59.3% of participants were studying computer science or multimedia computing, it is possible that they have studied or heard AI music before; however, these comments show that EMI's music is not completely on par with Bach's.

One limitation of the study was the number of participants. Although a large number of students did participate (86), there were eight different forms; therefore, an average of about 11 participants answered each form, which is not a large enough number to conduct a true statistical analysis. Because the results were consistent and in accordance with the hypothesis, it is likely that that if this experiment were repeated with more participants, stronger results would show a bias against EMI's music, as well as a bias in favor of Bach's music. Additionally, as stated earlier, the majority of students who participated in this study were computer science or multimedia majors/minors. In Section 3, I argued that one reason why people are hesitant to believe that computers can act creatively is because they only have the 'implastic' view of computers. If this is the case, a computer science student may exhibit less of a bias because he/she does understand how these systems work.

Although this experiment did not reveal an overall statistically significant bias against EMI's music, it did show consistent results in line with the hypothesis. We leave as future work to repeat this study with a larger number of and variety of participants. We also leave as future work to include in this study songs composed by more advanced and autonomous computer systems, such as Iamus, which creates music without any human intervention in a way analogous to evolution.[25]

7. Conclusion

In this work, we set out to answer: Is there a common bias against algorithmic (AI) music? Although the results of the experiment did not statistically show a common bias against this music for each attribute, the trends consistently showed a lower average rating for EMI and a higher average rating for Bach when their identities as the composers were provided. Furthermore, although the experiment results did not show a statistical bias in a general sense, they did show that music composed using artificial intelligence is less likely to be considered a work of art. What does this mean for the future of music? Music could revert to being associated more with the sciences than the arts. This brings into question whether the other fine arts, such as painting and poetry, would also be less likely to be considered an art. If that would happen, the gap between art and science may be bridged until there is no longer a clear distinction between them.

In fact, this bridge could be extremely beneficial to society. Combining natural human abilities such as language processing, image recognition, and creativity with the computational power of computers could open a world of endless possibilities. Viewed properly, humans and computers are not competitors, but collaborators working together; computers lend humans the processing power necessary to augment humans' abilities and enable them to overcome physical and mental limitations. Skype and Facebook enable us to overcome geographic and temporal constraints to help us connect with people around the world. Google organizes the world's information and gives us access to knowledge at the push of a button. EMI naturally extends this sort of collaboration to the world of music, allowing composers to experiment faster and more freely than they have in the past. EMI produced 5,000 chorales in the style of Bach alone, which is more than any one composer could produce in a lifetime. This high-speed composition process could help composers to test more ideas, and more variation on those ideas, than they would normally be able. Therefore, algorithmic composition does not replace composers, but acts as a tool for their use. This can lead composers to produce more creative art and possibly even a new branch of creativity. As David Cope said, "We could do amazing things if we'd just give up a little bit of our ego and attempt to not pit us against our own creations – our own computers – but embrace the two of us together and in one way or another continue to grow."[26]

Acknowledgments

The author wishes to express their appreciation to Professor Devorah Kletenik, Professor Justin Steinberg, and Professor Jacob Sachs-Mishalanie at Brooklyn College for helping to develop this experiment and paper, as well as Stephanie Montalti and the rest of the thesis class for their feedback on this paper.

References

1. D. Wakabayashi, “Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam,” *New York Times*, Mar. 19, 1028. [Online]. Available: https://nyti.ms/2u3QDYx. [Accessed Mar. 28, 2018].
2. D. Wakabayashi, “Uber Ordered to Take Its Self-Driving Cars off Arizona Roads,” *New York Times*, Mar. 26, 2018. [Online]. Available: https://nyti.ms/2pGUPZE. [Accessed Mar. 28, 2018].
3. A. Alpern, “Techniques for Algorithmic Composition of Music,” Nov. 1995. [Online], Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.23.9364&rep=rep1&type=pdf [Accessed Oct. 5, 2017].
4. Dictionary.com "algorithm," in *Dictionary.com Unabridged*. Source location: Random House, Inc. http://www.dictionary.com/browse/algorithm. Available: http://www.dictionary.com/. [Accessed Oct. 13, 2017].
5. T. Adams, “David Cope: ‘You pushed the button and out came hundreds and thousands of sonatas,’” *The Observer*, July 2010. [Online], Available: https://www.theguardian.com/technology/2010/jul/11/david-cope-computer-composer. [Accessed November 21, 2017].
6. D. Cope, “Computer Modeling of Musical Intelligence in EMI,” *Computer Music Journal*, Vol. 16, No. 2, pp. 69-83, 1992.
7. Merriam-Webster.com, “music,” Merriam-Webster. [Online]. Available: https://www.merriam-webster.com/dictionary/music. [Accessed Feb. 12, 2018].
8. P. O. Kristeller, “The Modern System of the Arts: A Study in the History of Aesthetics Part I,” *Journal of the History of Ideas*, Vol. 12, No. 4, pp. 496-527, Oct. 1951.
9. F. d’Olivet, J. Godwin, “The Ideas of the Ancients on Music,” in *Music Explained as Science and Art*. Rochester, Vermont: Inner Traditions International, Ltd., 1987, Ch. 1, pp. 39-50.
10. S. J. Markowitz, “The Distinction between Art and Craft,” *The Journal of Aesthetic Education*, Vol. 28, No. 1, pp. 55-70, 1994.
11. P. Kivy, “Is Music An Art?” *The Journal of Philosophy*, Vol. 88, No. 10, pp. 544-554, Oct. 1991.
12. D. Hofstadter, *Gödel, Escher, Bach*. New York: Basic Books, 1999.
13. D. E. Knuth, Foreword in *A*=B. New York: A.K. Peters Ltd., 1996, pp. vii.
14. M. Elton, ‘Artificial Creativity: Enculturing Computers,” *Leonardo*, Vol. 28, No. 3, pp. 207-213, 1995.
15. S. Colton, R. López de Mántaras, O. Stock, “Computational Creativity: Coming of Age,” *AI Magazine*, no. 30.3, pp. 11+, 2009.
16. G. Johnson, “Undiscovered Bach? No, a Computer Wrote It,” *New York Times*, Nov. 11, 1997. [Online]. Available: https://nyti.ms/2sFHvDI. [Accessed Apr. 28, 2018].
17. M. Theodore, Review: “Virtual Music by David Cope,” *Computer Music Journal*, Vol. 26, No. 4, Languages and Environments for Computer Music, pp. 92-95, 2002.
18. C. Garcia, “Algorithmic Music – David Cope and EMI,” *Computer History Museum.* Apr. 2015. [Online]. Available: http://www.computerhistory.org/atchm/algorithmic-music-david-cope-and-emi/. [Accessed Nov. 21, 2017].
19. D. Cope, “Facing the Music: Perspectives on Machine Composed Music,” *Leonardo Music Journal*, Vol. 9, pp.79-87, 1999.
20. H. Plassmann, J. O’Doherty, B. Shiv, A. Rangel, “Marketing Actions Can Modulate Neural Representations of Experienced Pleasantness,” *Proceedings of the National Academy of Scienes of the United States of America*, Dec. 3, 2007. [Online]. Available: http://www.pnas.org/content/105/3/1050. [Accessed Mar. 14, 2018].
21. A. Elowsson, A. Friberg, “Algorithmic Composition of Popular Music" in *Modeling Music: Studies of Music Transcription, Music Perception and Music Production,* 2018, paper C2, pp. 34-38.
22. I. Y. Ren, “Comparing Artificial Intelligence Music and Human Music: A Case Study of Prof. David Cope’s Emmy System,” University of Rochester.
23. Dictionary.onmusic.org, “chorale,” [Online]. Available: https://dictionary.onmusic.org/terms/688-chorale. [Accessed Mar. 23, 2018].
24. D. Cope, “5000 Chorales” [Online]. Available: http://artsites.ucsc.edu/faculty/cope/5000.html. [Accessed Nov. 14, 2017].
25. C. Sanchez Quintana, F. Moreno Arcas, D. Albarracin Molina, J. D. Fernandez Rodriguez, F. J. Vico, “Melomics: A Case-Study of AI in Spain,” *AI Magazine, 34*(3), pp. 99-103, 2013.
26. R. Moss, “Creative AI: Computer Composers are changing how music is made,” *New Atlas*, Jan. 2015. [Online]. Available: https://newatlas.com/creative-artificial-intelligence-computer-algorithmic-music/35764/. [Accessed Nov. 3, 2017].