Chapter 4
Evaluation of Creative Aesthetics

Harold Cohen, Frieder Nake, David C. Brown, Paul Brown, Philip Galanter, Jon McCormack, and Mark d'Inverno

Abstract This chapter is an edited conversation on the topic of computational evaluation of artistic artefacts. The participants were Harold Cohen, Frieder Nake, David Brown, Jon McCormack, Paul Brown and Philip Galanter. It began at the Dagstuhl seminar on computers and creativity, held in Germany in 2009 and continued over a period of several months via email. The participants discuss their views on the prospects for computational evaluation of both the artistic process and the made artefact.

H. Cohen (✉)
University of California, San Diego, CA, USA
e-mail: hcohen@ucsd.edu

F. Nake
University of Bremen, Bremen, Germany
e-mail: nake@informatik.uni-bremen.de

D.C. Brown
AI Research Group, Computer Science Department, Worcester Polytechnic Institute, Worcester, MA, USA
e-mail: dcb@cs.wpi.edu

P. Brown
Informatics, University of Sussex, Brighton, BN1 9RH, UK
e-mail: paul@paul-brown.com

P. Galanter
Department of Visualization, Texas A&M University, College Station, Texas, USA
e-mail: galanter@viz.tamu.edu

J. McCormack
Centre for Electronic Media Art, Monash University, Caulfield East, Victoria 3145, Australia
e-mail: Jon.McCormack@monash.edu

M. d'Inverno
Department of Computing, Goldsmiths, University of London, London, UK
e-mail: dinverno@gold.ac.uk

J. McCormack, M. d'Inverno (eds.), *Computers and Creativity*,
DOI 10.1007/978-3-642-31727-9_4,

4.1 Introduction

This chapter documents a conversation on the prospects for computational evaluation of art, aesthetics and creativity. The dialogue began in July 2009 at the Dagstuhl seminar (Boden et al. 2009). At the seminar a small group of participants decided to explore the problem of evaluation of creative works. Here "evaluation" included the decision process during artwork production and that following an artwork's completion, including evaluation by others such as audiences and critics.

Following the seminar, the dialogue continued via email over a period of many months (from July to November, 2009) and covered a variety of fascinating issues. What follows here is an edited version of this correspondence, in chronological order. As the reader will understand, the dialogue is relaxed and conversational—points are not always justified, and the unplanned and improvisational nature of the conversation reveals different ideas than would be found in a formal authored chapter. But we hope readers will appreciate the sincerity and openness of all the contributions, the value of candid personal opinions, and the shared sense of trying to explore the complexity of the issues raised.

There is much here that philosophers may be familiar with (and possibly even aghast at). But it does serve as an important historical record, particularly from the perspective of a number of pioneering artists who have been working in this area for decades. Their wisdom and experience brings a compelling perspective to the conversation. The collective insight of these pioneers provides an important point of reference for the next generation of researchers and artists entering the field.

4.2 Background: Evaluation of Artistic Artefacts

Before presenting the edited dialogue, a short background is first provided, in order to establish the context from which these discussions began.[1] The discussion is centred around the idea of computational evaluation of creative artistic artefacts. There are a number of points to be made to flesh out this idea. Firstly, how something is evaluated depends on the evaluator's perspective and role. This role may be as creator or designer, viewer, experiencer, or interactive participant.

This leads to some initial questions:

- What are the main features of human creative and aesthetic evaluation?
- How do these features (and the methods that are used) change according to the evaluator's role in the process?
- What aspects of evaluation can be made computational?
- Is it necessary for computational evaluation to mimic the evaluation methods of humans?
- Does it make sense to automate a task that is so especially human?

[1] Elements of this section are based on the initial Dagstuhl group discussions (Boden et al. 2009).

Answers to these questions have many complex implications, leading to a myriad of further questions, several of which arise in the dialogue that follows. When reading this dialogue it is important to keep in mind that the *role* evaluation plays determines the *kind* of evaluation required. Evaluation of a work as it proceeds leads to changes in that work (and potentially to future works), creating a feedback loop between action, intent, material (physical, musical, virtual) and decision.

Perhaps the most basic evaluation of a work as it proceeds is to know when it is finished. Knowing when a work is "done" arises for artists in almost any medium, working alone or in collaboration. At the opposite end—when beginning a work of art—initial ideas, conditions, moods and decisions can have major impacts on what follows. Cezanne reportedly threw away paintings once an "incorrect" brushstroke was made. Improvising musicians do not have such luxuries: a pianist like Keith Jarrett is acutely aware that what he first plays will shape the rest of the performance.

Computers have the ability to "undo", backtrack, and trial many possible combinations very quickly. But knowing what to undo, how and when to backtrack, and which paths to pursue or abandon requires evaluation *appropriate to the task* if it is to be successful. Evaluation of a work as it proceeds is generally concerned with decision making and prediction, e.g. what are the implications of making this mark, playing this sequence of sounds, or using media in a specific way? Accomplished artists have a seemingly innate intuition about creative decision-making and its impact, developed and fine-tuned over many years of practice. But can such decisions alone lead to the transformational creativity (Boden 1991) we see in the best human artists?

Evaluation of a finished artwork as an "art object" presents a different set of criteria. This may include examination of the emotional response of people experiencing the work under consideration, or an evaluation based on (for example) some aesthetic principles. In this book we are inevitably interested in what aspects of evaluation might be captured in a computational system. One possibility is to employ machine learning techniques where the system is trained on existing art works in order to learn any underlying aesthetic criteria. Certainly, this forms the basis of much current research.

It is also crucial to understand what we are evaluating for: *quality* (artistic, conceptual, aesthetic), *value* (monetary, cultural, critical, emotional), or something else? An important distinction can be made between the evaluation of creativity (appropriate or valuable novelty) and, for example, aesthetics. Something that is aesthetically pleasing may not necessarily be creative (as evidenced by looking at any collection of picturesque wall calendars, for example).

Human evaluation of artistic works typically extends well beyond the artefact itself, encompassing implicit knowledge and cultural norms, such as the intention of the artist who created it, the situation and conditions—social, political, and cultural—under which it is made and presented, the observer's knowledge and experience of similar works, and the dominant social values and norms of the day.[2]

[2]For important considerations of these issues, we refer the reader to the contributions in Part III.

As discussed elsewhere in this book (e.g. Chap. 5), becoming an expert or virtuoso in a particular medium normally takes many years of intense practice and immersion. As expertise and virtuosity mature, so does evaluation: the two appear to go hand-in-hand. Knowledge and experience emerge as decisive factors in producing artefacts of high creative value.

With these statements and questions forming a background, let us now proceed to the discussion.

4.3 A Conversation on Evaluation

The participants are (in order of appearance and identified by their initials), Harold Cohen (HC), Frieder Nake (FN), David Brown (DB), Jon McCormack (JM), Paul Brown (PB) and Philip Galanter (PG).

The conversation begins with a discussion about the aesthetic evaluation of art by people and computers.

Harold Cohen (HC): I sometimes wonder whether Western culture hasn't generated more art evaluation than art over the past few hundred years. How much of it is known outside the art world is another matter. It is worthwhile to make clear that aesthetic evaluation has little to do with conformance to the set of "rules" still being widely taught in art colleges.

As to the evaluation of aesthetics computationally, I confess to paying little attention to what's going on outside my studio, but I'd be very surprised to learn that there's a rich enough history of practical work to fill a book. Why is there so little history? To begin with, AI is still not at a stage where a program could accumulate enough relevant knowledge about an object it didn't make itself to make a non-trivial evaluation, so the discourse is limited, necessarily, to art-making programs, of which there have been relatively few. (I'm unclear about whether the same limitation would apply in other forms: music, for example.)

All of my own sporadic forays until now have been non-starters. But once I relinquish the notion of program autonomy and accept that the program is working with and for me, it becomes clear that it is capable of exercising (my) aesthetic judgement. And it does, to a point. But it's exercised on the work-in-progress, not on the finished work. Thus, it doesn't wait to tell me that an image has too much grey; it evaluates and corrects as it proceeds, provided that I can tell it how much grey is enough. That's a trivial example; one step up I'd need to say how much grey is enough relative to something else. Even if I could find a way of identifying amount-of-grey as an evaluation issue, and say what to do about such issues generally, there is still the problem that they are a moving target. That target moves in the human domain.

Unfortunately, it's a lot easier to say what's wrong with an image than to say what makes it special. I'm looking for the images that are *transcendent;* which means, by definition, that I don't know what it is that makes them special and don't know how

to describe what it is in computational (or any other) terms. The limitation is my own, not the program's.

Evaluation of a work in progress is directed to how to proceed. Evaluation of a finished work is directed to whether it's any good. The procedures required to satisfy the two are likely to be quite different, even when the same aesthetic is informing the procedure in each case.

I think it's very unlikely that "in-line" evaluation[3] can be done algorithmically. The simplest case I can think of would be to determine whether the work is finished. Even that is much harder than one might think. It could only be done algorithmically if one could provide an evaluation function—highly unlikely—which would be, in any case, a shifting target with respect to the many different goals a program (or human artist) may have.

For the general case the problem is much more difficult. The program can't determine how to proceed unless it knows what it has done, and knowing what it has done—the object so far—involves the notoriously troublesome "new term" problem. It knows it has done *a*, *b* and *c*, but can't know that it has introduced a novel and unanticipated relationship between *a* and *c*. Which is exactly what should be the determinant to the next step.

It is true, of course, that many human artists proceed algorithmically—you do this, then you do that, and after you've done all the thises and thats you have an artwork. No evaluation is required; your job is simply to do all the steps well. In human terms this algorithmic approach results in what we call "academic art", which I think has no place in a discussion on creativity.

Post-hoc evaluation is no less troublesome, and I suspect it's likely to be impossible for a program that didn't itself make the artwork. A formal colour evaluation[4] that doesn't take account of the possibility that all the well-balanced colour harmonies may add up to a portrait of an oddly-dressed man making rude hand gestures, for example. It also implies that there are canons for colour distribution and the rest, and evaluation simply measures conformance to those canons. (Impressionism good, German Expressionism bad?)

For most artists, making art is an on-going affair, not a series of isolated (artwork) events. Consequently, the completion of each work provides an extension of the feedback-driven consideration operating in the in-line evaluation. It is similarly concerned with direction, not aesthetics (except to the degree that, for the completed work, the artist must decide at least in part on aesthetic grounds, whether to accept it or reject it).

That is quite different from the aesthetic evaluation by any other agent, who is not engaged in that on-going process. In this case direction is clearly not an issue, acceptance/rejection is not an issue, and the aesthetic principles brought to bear on the work are unlikely to have much correspondence to those of the artist.

[3] By "in-line" Cohen is referring to evaluation of aesthetic decisions as a work proceeds.

[4] At the time of writing this statement, Cohen was very focused on translating his theories of colour and colour harmony into algorithms that AARON could use to colour abstract shapes.

I have some hope for the possibility of post-hoc evaluation by the generating program; no hope at all for evaluation by any other program.

Frieder Nake (FN): Aesthetics is, to a large extent, an evaluative discipline. We would probably not immediately equate evaluation with judgement. But the two are related. "Evaluation" is, quite likely, a more technical approach to aesthetic (or any other) judgement. However, we should be aware of the fundamental difference between value and measure. The temperature in a room can be measured because an instrument has been constructed that shows what physicists have defined as a quantitative expression of the quality of "warmth". The measured temperature is objective insofar as it has nothing to do with any human being present and experiencing the room in the actual situation and context. The human's value may be expressed as hot, warm, cool, or whatever else. Notice these are qualities.

So, in a first approximation, we may relate value with quality (human, subjective), and measure with quantity (instrument, objective).

The value judgement by a human may be influenced by the measured data delivered by an instrument. But the two are definitely and importantly to be kept apart (for intellectual rigour). Even more so in the complex situation of aesthetics.

Aesthetics itself is considered by many as being about our sensual perception of things, processes, and events in the environment. Hence, the subject matter of aesthetics is in itself intrinsically subjective. Those who start from this position cannot accept the claim that there are objective measures that would substantially contribute to human judgement.

HC: However, there have been times when number systems have had special cultural significance, and consequently aesthetics has been bound up with objective measures. For example, the Greek canon of human proportion was quite clear about how big the head should be in relation to the body, and I'm reasonably sure the sculpture critic would have regarded conformity to, or departure from, that canon as an aesthetic issue. There are many other examples.

Objective measures are a component of aesthetics when the measures themselves are important culturally. Today we have no such measures, and attempts to find them in contemporary artworks seem absurd to me, just as Ghikas's[5] attempts to find the golden mean in the art of a culture that knew nothing about incommensurable numbers seems absurd.

FN: Harold, you are absolutely right. By reminding me of some facts of history, you make me aware of a psychological hang-up that I now believe I have created in a dogmatic reaction against Max Bense.[6]

Bense, of course, allowed only objective aesthetic measures. He did so in reaction to German fascism where emotion was the only goal of their grandiose aesthetics

[5]Nikos Hadjikyriakos-Ghikas, a 20th-century Greek artist and academic.

[6]Max Bense was an influential German philosopher and Nake's teacher and mentor in his formative years as an artist exploring the generative possibilities of the computer in the 1960s.

for (against?) the masses. Bense was, at the same time, clear about subjective elements in the building of an aesthetic judgement. But that was outside of scientific investigation and research. It was purely private.

As a young man, I liked and loved this entire attitude. Everything in the world would be rational, mathematical, objective. Everything else was absolutely without interest.

I later adopted the view of aesthetics and sensual perception being tied together. From there it is a short step to my position. Your beautiful hints to some other times carry exactly the message that you summarise above. If some rule or law or proportion or other statement is culturally important, ruling, governing, then—of course—the individual sensual perception is, as always, largely determined by that "objectively" (i.e. culturally) dominating fact.

Having responded to Harold, Frieder now returns to his original discussion on developing algorithms for evaluation of aesthetics.

We seek algorithmic methods of evaluation that might have bearings on individual subjective aesthetic judgement. Yes—some researchers or even critics and artists want to find such measures, to define them, to construct instruments that tell us numbers on a scale. If we want to do this, if we neglect the deeply subjective character of a value judgement, we will try and find or define such measures to replace (or at least approximate) the subjective value. I am afraid, such heroic attempts will not get them very far.

It might be necessary to recall G.D. Birkhoff's definitions of aesthetic measure in the 1920s and 1930s. A lot of psychological work was done afterwards (in the form of empirical measures) with the unceasing intention of scientists to explain complex phenomena in an objective way.

The Birkhoff case is telling. He took up the old and popular idea of "order in complexity" or "unit in complexity" (a clearly subjective value). He put in a formula: $M = O/C$ (to me, this looks beautiful!). Here M is the aesthetic measure, O is the measure for order, C is the measure for complexity.

See how this works? You translate the words with all their connotations into variables. The variables stand for numbers, measured in appropriate units according to a measuring schema. What was a subjective interpretation all of a sudden has become reading scales. Great!

All that is left to do after this bold step is to "appropriately define" the measuring procedure. When you read Birkhoff's examples, you will be appalled. I was not, when I was young and did this (in the early 1960s). Birkhoff, as his famous example, chose polygons as the class of shapes to measure aesthetically. Complexity was for example the number of edges of the closed polygon. Order was, by and large, the degree of symmetry (plus a few additional features). The square is the best.[7] Wonderful!

When in those days, as a young guy using computers for production of aesthetic objects, I told people, small crowds perhaps, about this great measuring business,

[7] By Birkhoff's formula, the square evaluates to the polygon with the highest aesthetic value.

someone in the audience always reacted by indicating: "young man, what a hapless attempt to put into numbers a complex phenomenon that requires a living and experienced human being to judge".

My reaction then was, oh yes, I see the difficulties, but that's exactly what we *must* do! And we will, I threatened them. I guess, looking back without anger, they shut up and sat down and thought to themselves, let him have his stupid idea. Soon enough he will realise how in vain the attempt is.

He did realise, I am afraid to say.

In the early 1960s, Birkhoff's quotient of order over complexity was taken up again (by Bense, Frank, Gunzenhäuser, Moles, myself). It was given a promising interpretation in information theoretic terms. Helmar Frank, in his PhD thesis of 1959, defined measures of surprise and of conspicuousness (of a sign, like a colour, in an image). All these attempts were bold, strong, promising, radical. But they were really only heroic: the hero always dares a lot, more than anyone else, stupidly much, and always gets defeated and destroyed in the end.

I am sceptical about computer evaluations of aesthetics for many reasons. They are a nice exercise for young people who believe in one-sidedness. Human values are different from instrument measures. When we judge, we are always in a fundamental situation of forces contradicting each other. We should not see this fact as negative. It is part of the human condition.

Harold may be the one who, from his forty years of computational art practice that took him so close to the heroes of AI, would be able to pave the way. But even he is sceptical. "I don't know what it is that makes them (the computer-generated images coming from his program) special", he says. He continues to say he doesn't know how to describe "what it is in computational terms".

If we ever wanted to apply algorithmic methods to aesthetic evaluations, we must first be able to describe what we want to measure. Such a description must be formal and computable. So an explicitly formalised and algorithmic description is what would be needed. And those descriptions would be of works that we are used to calling "art". We all know the situation where five of us are around a small collection of pictures. We discuss them. We describe, bring in comparisons, develop our judgements against the background of our own lives, and of the current situation and discussion. We come up with a judgement in the end that doesn't totally satisfy any participant of the meeting. But all of us feel quite okay. We think we can justify the judgement. Tomorrow it could easily turn out to be different. This is how complex the situation of an evaluation is.

In Toronto in 1968/69, I wrote a program that I proudly called *Generative Aesthetics I*. It accepted as input a series of intervals for information aesthetic measures. They defined boundary conditions that must not be violated. The algorithm then tried to find a solution maximising the aesthetic measure against the boundary conditions. Its result was, of course, only a (probability based) distribution of the colours.

Just see what that program's task was: given a set of numeric (!) criteria, determine a "best" work that satisfies certain given evaluations. Isn't that great? I thought it was. And I was 29 years old.

A second program took this statistical description of an image (really: an infinity of images) and distributed colours into a quadtree structure such that the prescribed (just calculated) frequencies of colours were obeyed. I called the quadtree structure "the topology of the image".

I guess it was one of the most powerful programs ever in computer art, and certainly of its early phase. The program showed how little you achieve this way. As Harold says, you can use such dynamic evaluative measures during the generative process. That's all. Anything beyond this is human value judgement.

Phil Galanter has shared some of the scepticism of others, but says giant leaps are not to be expected. But baby-steps should be tried just to see where they get the baby. Yes, dear Phil, what is there left to do other than doing baby steps. So let us get into those pink knitted tiny shoes that mothers like to put their baby's feet into and move on from there.

David Brown (DB): I think that an analysis of existing methods in order to influence the output of computational systems—via some embedded knowledge (such as rules)—*is* a useful thing to do.

My experience in the design world suggests that you'll find a lot of people who had "techniques looking for a problem"—i.e. the method of evaluation is shaped by their tool.

I think it is better to analyse the problem and then look for techniques. For example, what kinds of evaluations affecting creativity are made during synthesis and what kinds of techniques can make these evaluations? Additionally, what kinds of evaluations can be applied to the descriptions of resulting artefacts, always assuming that all necessary sensing is in place.

For creative evaluation, newness and surprise are key to people judging something as being creative. But judging both of these computationally is tricky, especially during synthesis.

Focusing on learning is putting the cart before the horse. Focusing on a belief that something is "impossible" is not letting either out of the stable: a great way to reduce discovery of, and understanding about, the ingredients that lead to creative artefacts. By taking each challenge and looking at how it might be tackled we can make systematic progress.

Can we get a system to figure out that a blue widget isn't much different from a green widget, even if in some sense it is "new"? How can different types of newness be evaluated? Can a system predict how much a "newer" choice during synthesis will affect the judgement of the creativity of the finished product?

We take questions such as these and look for techniques that might help. For example, could we use the web for assessing newness? Could we take a representation of an artefact that has structural, behavioural and functional components and use that to decide a degrees of newness? Could fuzzy matching techniques be used to detect similarity and therefore newness? And so on...

Jon McCormack (JM): This discussion has made a number of claims as to why objective aesthetic measures seem impossible for an individual or machine. Nevertheless, I do think there is some basis for looking at aesthetic commonality particular

to a specific culture, social group, style or individual. After all, what is taught at art schools? Students learn the basic craft of their medium, they are exposed to many exemplars, they try and fail, try again, receive critique and feedback with a hope of improving with experience. But as has been pointed out by Harold, rule following isn't enough, art is an ongoing dialogue.

A lot of generative art software encodes specific forms of aesthetic judgement. The artist/programmer carefully chooses specific rules so as to create a system that generates pleasing aesthetics for *them* (which in turn may change after being exposed to computer aesthetics or even the aesthetics of the artwork-in-progress). Therefore, in a sense, this software is "evaluating" what it is doing (as it is doing it), but not in the way that a human does. It is an evaluation done for aesthetic purposes. However, the judgement originates with the programmer, not the program, so it becomes a continuous scale of how much is imbued to each.

A program that can adapt can learn, and hence change its judgement. That we know to be possible (using evolutionary algorithms or machine learning for example), but as Frieder points out, the baby may never get out of its tiny pink shoes. Perhaps we need to wait until machines have their own social evolution.

Frieder also raises the point that aesthetics is tied to the phenomenology of sensual perception—how else could we appreciate work like that of the artist James Turrell for example? It is difficult to imagine a machine experiencing such a work and coming to a similar aesthetic understanding, unless that machine had very similar phenomenological perception to us, or had sufficient knowledge about us (our perception, cognition, experience) and physics, to infer what our understanding would be. The same provisos apply to a machine originating such a work.

But while there may be many areas of human aesthetics, cognition and perception that are currently "off limits" to machines, it does not necessarily preclude machines that may be able to originate something that humans find aesthetically valuable. Indeed, a lot of "computer art" has given us very new aesthetics to contemplate.

Paul Brown (PB): I am very aware that writing too briefly opens up the opportunity for misunderstanding (I suspect Darwin said this?). But, to try:

One of the major themes in human development has been the revealing of structure (logic) through the observation and analysis of phenomena. Let me suggest that this point of view, in it's extreme perhaps, believes that all phenomena can be explained in some rational manner. In the history of art this complements the "classical" roots of art and leads directly to the work of Peirce, Saussure, Cezanne, Seurat, etc., and then into the 20th century experiments in constructivism, rational aesthetics, analytical philosophy, cybernetics, conceptualism, systems art, and so on... We could call this approach Modernist but this term is fraught with misunderstanding, especially as it is so (ab)used within the art world.

Another major theme suggests that understanding comes via entering into a relationship with the phenomena that enables the spontaneous emergence of meaning. We use terms like "intuition" and "inspiration". The extreme of this point of view suggests that critical analysis is unnecessary and may actually be counter-productive (and in theological "controlling" manifestations that it should be suppressed). I know of several artists who, after pursuing PhD "practice-based" research, are now

unable to practice since they have lost their spontaneity. Here belief is paramount—the subjective takes precedence over the objective. In the world of art this meme develops in the Romantic tradition. With the same reservations as above we could adopt the term Postmodern to describe this kind of thinking as it developed in the late 20th century.

One important distinctions between these two positions is that the former believes that everything can be explained using rationally/logical methods and the latter does not.

As a member of the former group I believe that the major shortcoming of the latter is that it implicitly invokes the need for a quality of the "unexplainable"—some kind of immaterial "essence" or "soul". However I am also aware that in science we now accept (believe in) dark matter and (even more mysteriously) dark energy—qualities which enable our structural analyses of the universe to make sense but for which we have little or no direct evidence.

Another interesting comment comes from the British biologist/cybernetician Geoff Sommerhoff in his explanation of "freedom of will". He suggests that freedom of will is the response of a simple entity (humans) to an environment (the universe) that seems to be almost infinitely complex. For Sommerhoff freedom of will is no more than a psychological mechanism for helping us maintain our sanity when faced with the actuality of our insignificance and our inability to act independently. Taking this further we can interpret Sommerhoff as suggesting that although everything is knowable, it is not possible for humans to attain all of this knowledge because of our inherent system limitations. This seems to me close to Borges map problem—for a map to be completely accurate it must be—at least—as large (as complex) as the territory it describes. So for us to be able to fully explain the universe we need another universe that is, at least as big, to hold the knowledge.

So for me this objective/subjective question can be expressed:

1. I implicitly believe that everything is rationally explainable (there is no essence or soul);
2. I acknowledge, however, that there are many things that may never be explained;
3. Nevertheless I do not believe that this acknowledgement of limitation should prevent us from seeking explanations—however hard the problems we address may be;
4. I believe that the rational analysis and synthesis of aesthetics (and other perceptual, cognitive, conceptual and creative processes) is one of the key issues for humanity to address in the 21st century—we must now apply our systematic methodologies to our own internal mechanisms (and I'm here using the word "mechanism" deliberately);
5. If we do not then we are in danger of handing our world over to the priests, fascists and other bigots whose only wish is to enslave us.

In response to this on-going discussion, Philip Galanter responds in order to draw out some of the underlying assumptions.

Philip Galanter (PG): In terms of epistemology the (undefended here) subsuming view is that there really are intrinsic unknowns "out there" even though "out

there" is a noumenal world that is mechanical, rational and logical. Meaningful, objective and verifiable general explanation is possible. However such explanation is, as a matter of principle, incomplete and statistical. Specific past events may elude explanation, and future events may be unpredictable as a matter of principle even though they are not irrational.

FN: I think I have mentioned before, how much my admired teacher in philosophy, Max Bense, was motivated in all his thinking and writing by his experience as a thinking individual in Nazi Germany.

Nobody should allow him- or herself to let any emotions, anything non-rational creep into their aesthetic (or other) judgement. Rationalism was the weapon in thinking against fascism and other totalitarian movements.

As young students we loved him for these messages. Radically I tried to follow his traces. An exercise that helped me for a long time and occupied my thinking in the most beautiful and satisfying way.

Why then did I later start deviating from this line? And why do I today no longer believe that aesthetic judgement rationalism will get me very far?

It seems to me that, at this moment, I cannot pin down a specific event or insight or influence that caused me to change in the way indicated. In very simple terms, my position is: of course, we try to analyse a painting, a piece of music, a novel, etc. in rationalist concepts and in a rationalist method; such an approach will give us a lot of insight and a way to discuss and criticise without attacking us personally, but only in issues of the subject matter; often, and for many, this is enough and nothing more needs to be done; for others, however, the final judgement remains to be a personal statement based on acquired feelings.

It has happened to me more than once that I enter a gallery room, take a look around, and immediately (and unmediated) react in a positive, excited, interested, attracted way to one of the paintings there. I move closer, study it carefully, think, compare, visit the other paintings in the room, build up a judgement. Often, the immediate impression survives a more careful consideration, and is enforced. Not always though. At times, closer investigation leads to a revision of the first and immediate impression.

I do know that everything I have learned and experienced about Artificial Intelligence, everything I have read from Hubert Dreyfus, Joe Weizenbaum, the Scandinavians, David Noble, from Herbert Simon, Allen Newell, … all the heroes of AI—all that built up in me, and reinforced again and again, a deep rejection of anything that seems close to the separation of mind and body.

Cartesianism has had a great time, and has led to exciting results. But it has had its time. The belief in "progress" has disappeared from me. Change, yes. Permanent change.

Hannah Arendt refers to Kant as having said that aesthetic judgement relies on examples, not on general concepts. This I believe. I say "believe", not more.

After several weeks of silence, the discussion continues, this time initiated by a report from Harold on his progress with AARON in creating new images for a forthcoming exhibition…

HC: A report from the front. A couple of weeks ago I decided I wanted to see more saturated colour in AARON's output. I gave the program what I thought would be a suitable colour profile (controlling the frequency with which it would choose one of the nine possible combinations of lightness and saturation) and then watched in increasing frustration as the program generated several hundred rotten images.

Yesterday I bowed to what I've always known to be the unyielding dominance of value—lightness—over saturation, and substituted a different colour profile that generated colours from very light to very dark. And this morning I selected forty stunning images: my "aesthetic evaluation"? from more than two hundred mostly excellent images.

What was I looking for when I made the selection?

A sense of place. All the images make use of the same set of form generators; I chose those images that transcended mere arrangement of forms, those that generated the sense that they represented something external to themselves, those that seemed to carry the authenticity of the thing seen.

What contributes to this sense of place?

There are relatively few variables in the program that exercise critical control over the nature and reading of the output. One is the choice of colour profile. Others are the scale of forms relative to the size of the image; the proportions of the image; the background colour (hue, lightness and saturation) relative to what builds in the foreground; the proportion of space allocated to background and foreground; the mode of distribution of the forms.

You'll see that these are all quantifiable. (There are several possible distribution modes, each of which is controlled by quantifiables.)

Is the nature and quality of the output—the sense of place—then quantifiable?

I am aware that there are no intrinsically good or bad values for the variables that control the output. The sense of place—and everything else—results from the combination of all the variable values. That's a multidimensional space with perhaps fifteen or twenty dimensions that I know about; way beyond my own mathematical capabilities if I thought that was a good way to go. But notice that the same set of values generated more than two hundred images, of which I judged only forty to have an adequate sense of place. Evidently there are other elements involved beyond the variable settings; specifically, I suspect, the "clustering" of forms which emerges from distribution and scale and population and all the rest.

Is this emergent property—clustering—quantifiable? I doubt it.

The implication seems to be that a program might be able to pick out the good ones, but couldn't pick out the exceptional ones; which are, of course, the ones I'm interested in. But even this might be going too far, partly because it may not be possible to identify the original variable values from the output, partly because in doing so it would only have identified this particular work as belonging to a particular group and would reject any work that didn't belong to this or another successful group. Clearly, that's not the way to go. The transcendent images that don't belong to any group are precisely the ones I want.

The more important point to make, however, since we appear to be talking about aesthetic evaluation, is that I've not said a word to suggest that beauty is an issue

for me. In fact, I don't think I've ever met an artist who did think that beauty was an issue. Beauty is emergent, apparently, from the relentless pursuit of the individual's holy grail, whatever that might be, bearing in mind that my grail and yours are unlikely to have the same shape. That does not necessarily mean that a purely formal evaluation of the work itself, without regard to how it got to be that way—harmony, balance, golden mean and whatnot—are non-starters, but I have yet to see one finish.

And, yes, you certainly do run into cultural issues. Impressionism has been the epitome of "beautiful" painting for a long while now; but the Impressionists were accused of shooting the paint on to the canvas with a pistol. Not good. Though today we'd probably think of that as a great idea; after all, Pollock didn't go in for brushes, either.

FN: I, as one occasional participant in this dialogue, love in particular your comments and deep insight, the insight of a life in art and science, Harold. By necessity our discussion must get closer and closer, as it continues, to the fundamental philosophical question of objective vs. subjective. This discussion would then have to ask what the "thing" would be, what the "work" would be, and much more…

We all know to some extent that these issues cannot be solved (as a mathematical equation may be solved), but that they remain the eternal discourse of philosophy. It produces the question itself in new forms, and therefore also with new answers.

Our question here is, of course, much more pragmatic and mundane. I guess a few statements could be made in this regard. Like, perhaps:

The making of art is subjective. The appreciation of art is subjective. The making of art relies on certain general and specific objective conditions. So does the appreciation.

Humans, as cultural groups or as individuals, like to emphasise how nice it would be to have objectivity. But there is only a little objectivity when humans are involved. There is, however, also little subjectivity if "subjective" is what pertains to this individual, here and now. If the striving for objectivity is taken as an attempt to enter discourses with others (individuals, groups, living or dead), and conduct such discourse with passion and patience, decidedly and forgiving, ready to accept a position, ready to give in and not to win but to convince—if factors like those determine the process then those involved will eventually agree that there is little objectivity, little subjectivity, but lots of historic and societal impact.

Judgement is different from evaluation. The absolute pre-condition for programming (and thus for using computers) is formalisation and computability. This is so even in the most interactive and sensor-prone situation.

The concreteness in your argument, dear Harold, is marvellous, it is telling, it is itself artistic. You know that—if I understand my own thinking well enough—I totally agree with your sentiments. You summarised them beautifully by saying: "at the lowest level of machine evaluation, I can see that the program might be able to tell me which images *not* to print". More, I also think, is not possible. The others say: "we are just at the beginning, give us enough money and time". Birkhoff and all those of the 1930s debate failed. Bense and all those of the 1960s debate (including Nake) failed.

It is perfectly legitimate to use computational methods for some first and preliminary evaluations, as we use the thermometer, the speedometer, the yardstick. When a distance is measured as five meters, some of us say, "oh, I can long-jump this easily". Others will never make it. But all try very hard.

When the temperature in a room is measured as 22 degrees Celsius, some react with "too hot for me", others with "rather cool after a while". Measure, value; evaluation, judgement.

And let us not forget, how you, Harold, continue after your optimistic remark about what the machine might be capable of. You say that you would still take a look before, upon the program's evaluation, you delete the file...

PG: I think that this is the kind of discussion that can always be paused but never ended. For now I'd be happy just to clarify what the differences are.

If it turns out that non-trivial computational aesthetic evaluation is impossible, that in itself would be worth better understanding. It seems to me such a statement might come in two forms. There might be some kind of formal sense, or there might be an engineering analysis leading to absurdly expensive, or quantitatively impossible, practical requirements.

Frieder seems to lean towards the former by saying that aesthetic evaluation would have to be formally computable, but is not. But this leads to (in my mind) an even more interesting question. How is it that the mind is capable of "computing" the uncomputable? Is the mind more than the result of a mechanistic brain?

And if the objection is more practical and in the realm of engineering a similar question is raised. What aspect of the mechanistic brain can we know to be beyond the reach of human engineering? How is it that nature has brought the costs and quantities within reach in a way we will never be able to duplicate?

The strongest objection, to me, would also be the one that claims the least, i.e. that computational evaluation as an engineering challenge is impossible *for the time being*. Maybe it will be within reach in ... 10 years? 50 years? 100 years?

But if the operative objection is this last one it changes the entire conversation. Because then computational aesthetic evaluation is possible in principle and merely contingent. All discussions of creativity should allow for it in principle.

Frieder also mentions that, "Judgement is different from evaluation". In our Dagstuhl discussion Margaret Boden rejected such a notion out of hand. Perhaps they are referring to two different kinds of judgement, or two different kinds of evaluation, or both. In any case this confirms in my mind that the language involved will need more precision than everyday speech, and technical definitions are probably called for. For example, when a human takes a given work of art and merely classifies it to an art movement, can that be called "evaluation" or should some other word be used?

Finally there is a bit of a paradox worth pointing out here. Most attempts to define creativity I heard at the Dagstuhl workshop included a provision that the innovation must not only be new but it must also be of value. Now if computational aesthetic evaluation is more or less impossible does this mean computational creativity is impossible? Or does this mean a computer can be creative without being able to measure the value of its own output?

If so, then turn this back on human creativity. If a creative computer need not "understand" the value of its own creations, does that mean a human can be deemed creative even though they are incapable of knowing whether their creations are valuable?

To me it seems odd to demand that creativity result in value but allow that the creator may not know that it does. It would be similar to crediting someone as being "ethical" even though they cannot discriminate between right and wrong.

My response to these problems is implicit in the chapter I present. I think it will ultimately be more fruitful to disconnect definitions of creativity from questions of value.[8] Just as it's a mistake to connect the definition of art to the definition of good art, I believe it's a mistake to connect the definition of creativity to the definition of valuable creativity.

I see creativity as being more related to issues around complexity and the behaviour of complex systems. For me creativity is simply what complex adaptive systems "do", nothing more and nothing less. From this point of view the value of a given creative act is relative to the (possibly co-evolutionary) situation at hand and the contribution it makes towards adaptation by the creative entity. In this case humans, computers, and all manner of things/processes are capable of some degree of creativity.

PB: Thanks for this good summary of the situation. It seems to me to hit several of the important issues head on. If aesthetic evaluation is uncomputable then how does the mind/brain do it? As you comment, an interesting question in itself. As I briefly mentioned previously, it seems to me that the only way beyond this point is to posit the existence of a metaphysical (super-mechanical) entity which is unacceptable to me. Therefore I assume it has to be computable.

You infer the work of Gödel and Turing and we know that within any finite axiom system there will exist propositions that cannot be resolved. However this doesn't answer the problem since again we must ask: then how does the mind/brain (a finite system) resolve aesthetic evaluation?

I return also to my earlier mention of Sommerhoff's description of freedom of will. He implies that things like creativity and aesthetic evaluation may not be computable until the computing engine is at least as complex (or can reflect the same degree of variety—to use Ross Ashby's term) as the human brain. As suggested in this discussion, this is a long way off.

Nevertheless we have to start somewhere and it seems to me that starting with the assumption that computational aesthetic evaluation is not possible is counter productive—we *must* begin from the belief that it can be achieved.

My glass is half full!

[8] This view is also shared by Dorin and Korb in Chap. 13.

4.4 Conclusion

As you might expect from a topic as complex as computational evaluation of art, there is no real consensus or closure from this discussion, nor could this be realistically expected. Yet it is interesting to examine the different perspectives participants consider to be useful or practical in approaching computational evaluation. As Paul Brown's concluding remarks emphasise, unless you think there is something fundamentally uncomputable and ineffable in what humans do, then computational modelling of human evaluation is at least a possibility. But just because something is possible doesn't make it easy, or even practical. It is tantalising to think that future computational models will shed a different light on evaluation of art (and more generally on human behaviour), complementing and informing other discourses such as critical and cultural theory, or philosophical aesthetics. However, computational models of this kind are still very much in their infancy.

It is also interesting to consider the mirror question to the one that is the main topic of this chapter. Namely, can art made by an individual computer program (or social network of autonomous computer agents) ever be fully understood and evaluated by humans? Such considerations raised in this chapter, and many others running through the entire volume, raise many crucial questions to investigating creativity through computing, a number of which are listed in the final Chap. 16 of this book.

Evaluation remains a difficult and vexed issue for understanding creativity from a computational perspective. No doubt it is something that artists and musicians are involved with at almost every moment of their creative practice, but so far attempts to mimic this process in a machine fall short of what any human being can easily do. Interestingly, the two artists with perhaps the longest experience in this field (Nake and Cohen) see little merit in pursuing the idea of developing creative or aesthetic measures, precisely because they have tried to use them in their own art practices and found them to be creative dead-ends. This should at least give us cause for reflection. While understanding exactly what evaluation is and how it is performed by humans remains an open problem, anyone wanting to make serious inroads into developing machine creativity cannot afford to ignore it.

Acknowledgements We acknowledge the contribution from all the participants, including the original Dagstuhl discussion group on this topic, which consisted of Harold Cohen, Margaret Boden, David Brown, Paul Brown, Oliver Deussen and Philip Galanter. The discussion group notes can be found at http://drops.dagstuhl.de/opus/volltexte/2009/2212/. The interview in this chapter was edited by Jon McCormack.

References

Boden, M. A. (1991). *The creative mind: myths & mechanisms*. New York: Basic Books.

Boden, M., d'Inverno, M., & McCormack, J. (Eds.) (2009). *Computational creativity: an interdisciplinary approach. Dagstuhl seminar proceedings: Vol. 09291*. LZI. http://drops.dagstuhl.de/portals/index.php?semnr=09291.