The future won’t be pretty: The nature and value of ugly, AI-designed experiments

Michael T. Stuart

Orcid number: 0000-0002-4165-2641

Abstract

Can an ugly experiment be a good experiment? Philosophers have identified many beautiful experiments and explored ways in which their beauty might be connected to their epistemic value. In contrast, the present chapter seeks out (and celebrates) ugly experiments. Among the ugliest are those being designed by AI algorithms. Interestingly, in the contexts where such experiments tend to be deployed, low aesthetic value correlates with high epistemic value. In other words, ugly experiments can be good. Given this, we should conclude that beauty is not generally necessary or sufficient for epistemic value, and increasing beauty will not generally tend to increase epistemic value.

Keywords: aesthetics of science; aesthetics of experiment; internal validity; artificial intelligence; beauty in science

Can an ugly experiment be a good experiment? Philosophers have identified many beautiful experiments and explored ways in which their beauty might be connected to their epistemic value. However in this chapter, I seek out (and celebrate) ugly experiments. Which experiments are ugly? For hardcore theoreticians, perhaps *all* experiments are ugly, and only pure theory can be beautiful. For others, “ugly” experiments will be shorthand for “bad” experiments, in just the same way that “beautiful” can be shorthand for “good” (Todd, 2008). To avoid these ambiguities, this chapter will focus on distinctly *aesthetic* ugliness in scientific experiments to consider what, if anything, the relationship might be between this and epistemic value.

1. Introduction

There are at least two things we should be clear about when evaluating experiments. The first is to specify which *aspect* of an experiment we want to evaluate. Aspects of experiments include the instruments, the results, the objects, the hypothesis being tested, the performance, the interpretation, and the design. Following a number of authors, I will focus on evaluations of the *design* of experiments (Parsons and Rueger, 2000; Murphy, 2020; Ivanova, 2021, 2022). This is because it is the design that defines the experiment. In many cases, we can use the same experimental design to test different hypotheses related to different objects, get different results, interpret the same results differently, and swap in different instruments. But we cannot alter the central features of an experiment’s design without creating a new experiment.

The second thing to be clear about is the sense of goodness at issue. An experimental design can be *ethically* good, e.g., when it does not irresponsibly risk harm or when it aligns with the ethical values of those who stand to profit or be harmed by the outcomes. The design of an experiment can also be *epistemically* good, such that it is likely to produce a high signal-to-noise ratio, eliminates possible confounds, or is easy to replicate. Finally, the design of an experiment can be *aesthetically* good, such that its design is elegant, clear, simple, symmetrical, or beautiful.

These distinctions gloss over a thorny axiological question: are ethical, epistemic and aesthetic values merely different manifestations of a single, maximally general kind of goodness (like the Platonic form of the Good) or are they indeed fundamentally different kinds of goodness? While interesting, we will leave this aside to ask the more applied question concerning how far these different senses of goodness can come apart in the case of experimental design in science. This question is meaningful independently of whether the different kinds of goodness are manifestations of a more general good.

John Horgan (2012) collects some of "science's ugliest experiments," including the Castle Bravo atomic bomb test in 1954, syphilis trials in Guatemala in the mid-1940s, and using electrical brain stimulation to "treat" homosexuality between 1949 and 1980. Some of these experiments were ugly in the sense of being ethically repulsive, while others were epistemically bad, and some were both. However, it seems that an experiment's design can be ethically good while being epistemically bad. Consider members of institutional *ethics* review boards, who do not need to care whether an experiment's design has a high likelihood of producing a result with *epistemic* value, as that is not their job. And going in the other direction, we note that there are experimental designs that would accurately reveal causal dependencies in a system (this is one way to be epistemically good) while causing harm to humans. The Castle Bravo experiment might be an example of this.

Likewise, aesthetically beautiful designs can be ethically bad. For example, imagine a very clear and simple design that tests the effect of some poisonous substance on humans. Going in the other direction, we can also imagine ethically flawless experiments that are convoluted, needlessly complicated, inelegant, and so on.

This brings us to the relation between aesthetic and epistemic goods in experimental design, which is the focus of this chapter. Let's put a few options on the table. It might be that aesthetic values are completely independent of epistemic values, at least when it comes to experimental design, such that having one kind of value tells us absolutely nothing about whether the design has the other kind of value. Alternatively, one kind of value might be *sufficient* for the other, such that possessing one kind of value automatically tells us that an experiment has some value of the other kind. Another possibility is that one of these values is *necessary* for the other. For example, perhaps an experiment's design can't be beautiful if it's not likely to tell us something new about the world, or, a design can't be epistemically good if it's not at least *somewhat* clear, simple, elegant, and so on. Another possibility is that there is no general connection between beauty and epistemic value, but there is/are connections between *certain* aesthetic properties and *certain* epistemic properties. For example, perhaps symmetry is tied to robustness, but not to having a good signal-to-noise ratio. A final possibility is that increasing one kind of value might generally *tend* to increase the other kind of value, without guaranteeing or requiring it.

To conserve space, we will focus only on the one-way inference from aesthetic value to epistemic value. This is the more interesting direction of inference, as it inspires strong and polarizing intuitions. Consider, for example, the statement made by Roger Penrose quoted by James McAllister: “something which looks attractive may have a better chance of being true than something which looks ugly…in fact, it turns out that the more attractive possibility is the true one” (McAllister 1996, 90-1). Some will find this tantalizing while others will find it simply wrong.

We will also restrict ourselves to particular and technical notions of aesthetic and epistemic value. Roughly, by epistemic value I mean something like “internal validity,” that is, the ability of a design to reliably, replicably, and robustly identify the (causal) effect of one variable on (an)other(s) (Campbell, 1957; Guala, 2003, 2005; Hogarth, 2005; Jimenez-Buedo and Miller, 2010). And by aesthetic value, I mean something like a positive, pleasurable feeling resulting from first-hand (intellectual, imaginary, or perceptual) experience (Schellekens, 2022). Other forms of epistemic and aesthetic value (including understanding as a kind of epistemic value and the sorts of aesthetic values more *en vogue* in modern discussions of art) will also be discussed, though these will not be the main focus. Finally, I will mostly be concerned with aesthetic and epistemic value in general rather than the properties that are referred to in justifying aesthetic and epistemic judgments (e.g., simplicity, clarity, signal-to-noise ratio, etc.) since it would be impossible in one chapter to address all the relationships between all the properties. In sum, here are the options.

1. **Independence**: the positive or negative aesthetic value of an experimental design tells us nothing about its epistemic value.
2. **Proportionality**: increasing the aesthetic value of an experiment’s design generally tends to increase its epistemic value.
3. **Necessity**: if an experiment’s design has epistemic value, it has aesthetic value.
4. **Sufficiency**: if an experiment’s design has aesthetic value, then it has epistemic value.
5. **Necessity and sufficiency**: an experimental design has epistemic value if and only if it has aesthetic value.

Which of these is the correct way to understand the relationship between aesthetic and epistemic value? In the next section I will present a case study that takes the last three options off the table, leaving only the first and a specific weakened reading of the second. This may be surprising, since most of the work on the aesthetics of experiments focuses on the beautiful (not the ugly), and it is easy to get the impression from the literature that the connection is quite strong. On the other hand, some readers will already be convinced that there is no (strong) aesthetic-epistemic relationship. In that case, my contribution is to provide new evidence for that intuition.

My argument begins by noting that AI is currently designing experiments. These AI-designed experiments are here to stay, and one day they might represent the majority of experiments in science. (This is important because if AI-designed experiments were just an oddity, we would be tempted to exclude them from any discussion of the aesthetics of experiments). These experiments are ugly, and the nature of the algorithms that produce them helps to explain *both* their ugliness *and* their epistemic value.

2. AI-designed experiments

In the mid-1960s, the success of thinking machines and "analytic engines" in certain computational tasks soon led to the idea of creating a robot scientist. Herbert Simon's thinking machine, the "Logic Theorist," combined with the new Information Processing Language, inspired researchers like Edward Feigenbaum to create discovery programs that were applied in the DENDRAL project, whose purpose was to automate mass spectrometry experiments for the Viking landers, which would test for amino acids on Mars (Lindsay, Buchanan and Feigenbaum, 1993; Feigenbaum, 2007). This idea was scrapped because of the energy and weight constraints of extraterrestrial robotics, but it was taken up elsewhere in chemistry, e.g., by a program called FAHRENHEIT whose purpose was to perform experiments, analyze data, look for errors, determine replicability, and create new theoretical hypotheses, all in the service of (among other things) detecting low-concentration ions (Żytkow, 1987; Żytkow, Zhu and Hussam, 1990).

In the early 2000s, Ross King oversaw the creation of several robot scientists. These were "closed-loop" systems, in the sense that they designed, performed, and interpreted experiments, and then fed what they found back into the system to inspire new experiments, and so on, until a solution was produced. The first of these, Adam (2004-11), worked on functional genomics in yeast, performing knock-out experiments (King *et al.*, 2004). Initially, it found the functions of genes that were already known. Then it discovered genes responsible for the production of important amino acids, which were not known. These discoveries were checked manually, and Adam was correct. King claims this is the first robot scientist to generate completely novel scientific knowledge. His next project, Eve (2008-20), focused on neglected tropical diseases, identifying compounds that might be useful in producing cheap and effective treatments (Williams *et al.*, 2015). Eve was also recently used to test the replicability of studies in cancer research (Roper *et al.*, 2022). A new project, Genesis, is now deployed in the context of systems biology, and it is advertised as being able to perform up to 10 000 closed-loop cycles of experiments at a time.

One admitted shortcoming of these prototype projects is that human intervention is often required, and the experiments are relatively conservative variants of existing designs. However, recent work by Mario Krenn and colleagues has produced an algorithm ("MELVIN"), that designs experiments that no one has ever thought of before. More specifically, Krenn's team could not come up with a viable design for a certain type of experiment, and the algorithm succeeded where the humans couldn't in suggesting a way to arrange laboratory equipment such that certain quantum phenomena could be produced and manipulated in the lab. As Krenn et al. point out, AI is already being used in a number of ways relating to the design of experiments in accelerator physics, plasma physics, nanophotonics, quantum circuits, and superconduction (Krenn, Erhard and Zeilinger, 2020), but using AI to design new experiments from scratch is only now gaining ground.

Krenn's lab wanted to produce a certain type of quantum entangled system. The most common type of entanglement (the "qubit") obtains between two particles and two states (e.g., spin up and spin down). Higher-dimensional entanglement occurs when we have more than two particles or more than two states. This kind of entanglement is potentially very important for quantum computing, as it has greater information-carrying capacity. Trying to produce and manipulate certain types of higher-dimensional entanglement had proven "difficult" (Erhard, Krenn and Zeilinger, 2020): Krenn reports trying with colleagues for several weeks without success to produce a particular form of multiparticle entangled state (Krenn, 2021). The algorithm MELVIN was developed, and it produced an experiment that was "the first genuine high-dimensional multiparticle entanglement" (ibid), using crystals to produce photon pairs. This discovery (Erhard *et al.*, 2018) enabled independent manipulation of the quantum states, invented new tools for quantum experiments, and has now been expanded from photons to non-photonic systems.

Roughly, the algorithm possesses a "toolbox" filled with representations of optical laboratory components, which it combines randomly. The combinations are tested for viability, and those that pass are then tested in more detail. Even if a set-up is not viable, it might produce a new type of tool, which is then added to the toolbox. The final set-up that MELVIN recommended required the use of tools that it produced itself. Experiments are represented mathematically as graphs, with vertices representing photon paths and edges representing amplitude by their weights and quantum properties by their colours. The successful outcome is then simplified by an XAI algorithm that eliminates redundant elements so that humans can understand it well enough to implement it.

AI-designed experiments still make up a minority of all scientific experiments performed, but there are reasons to think that this will change. Frank Wilczek, who won the Nobel prize in physics in 2004, has claimed that in one hundred years the best physicist might be a machine (2016). The Nobel Turing Challenge is a competition to produce a Nobel-winning AI scientist by 2050 (Kitano, 2021). Ross King argues that we should expect the use of AI to expand in science because it is reliable when it comes to memory and mathematics, it can process more data more quickly than humans can, and it can find patterns that are invisible to humans. He points out that AI is very good at anything which can be gamified, and quite a bit of science can be gamified. AI tends to be cheaper, faster, more accurate, more detailed, and it can be scaled up and controlled remotely (King 2021).

We therefore have good reason to believe that "experimental design" should not only refer to human-designed experiments, but experiments designed either by humans, by human-AI teams, or by AI themselves.

3. The good, the bad and the ugly

In *Scientific American,* Aephraim Steinberg comments that the experiment designed by MELVIN is "a *gorgeous* first example of the kind of new explorations these thinking machines can take us on" (Ananthaswamy 2021; emphasis added). Steinberg did not work on these experiments, but he uses AI to design experiments of his own. I reached out to Steinberg, and the quotations that follow come from our conversation, which took place over a few weeks in 2021.

So, what did Steinberg mean when he said that the experiment was gorgeous? He recognized that scientists "use anthropomorphic terms like 'beautiful' and even 'sexy' in vague and ill-defined ways all the time," but was this was not what he was doing in this case. "There was something…lip-smacking, entrancing…about this implementation, particularly to experimentalists. Not that it's the most important, the most creative, or the most elegant experiment ever, but…to my eye, at least, there's a 'rhythm' and 'cadence' to the series of interferometers in their experimental setup which indeed looks almost poetic or artistic." He goes on, "I think it was the combination of the experimental prowess and elegance, with the fact of implementing the 'clever' scheme from the AI-invented protocol which pushed me to describe it the way I did…I liked this example because I thought they really did demonstrate that they could do something that as far as I know, no humans had figured out how to do before." In sum, Steinberg's notion of a beautiful experiment involves newness, surprisingness, simplicity of design (i.e., elegance as opposed to arcaneness), looking good (either in the pdf or physically on the tabletop), and having good rhythm or cadence in the performance (whether imagined or physical).

This matches quite closely what philosophers have said about beautiful experiments in the past. Glenn Parsons and Alexander Rueger focus on the aptness of a design, the simplicity of its steps, and the imagination of its designer (2000). Milena Ivanova focuses on the elegance, simplicity, and clarity of a design (Ivanova, 2022). Steinberg agrees with these, and adds rhythm or cadence, which is a nice addition to the set. Grouping together these properties, we have something consistent with what Elisabeth Schellekens calls the "standard conception" of aesthetic value. For Schellekens, the standard conception is the classic, traditional one, which develops from the following two axioms: a) aesthetic experience is grounded in first-hand perception, and b) aesthetic experience is characterized by pleasure (Schellekens 2022, 125). In our case, we should expand the first axiom so that it refers to *experience* rather than perception, since experimental designs are sometimes imagined or contemplated rather than literally perceived (call this expanded version the "standard+" conception of aesthetic value). Then, a beautiful experiment is one whose design is experienced first-hand, where that experience is a pleasurable one. That pleasure will be tied to experiences of particular aspects of the design, like its aptness, clarity, rhythm, etc.

As Schellekens points out, the history of aesthetics is a history of expanding this standard conception to include many different kinds, sources, and media for aesthetic value. But these various kinds of modern and post-modern aesthetic values are not what is at issue in discussions of the aesthetics of science. Scientists and philosophers are not concerned with whether an experiment can have aesthetic value in the same sense as the art of Duchamp, Picasso, Rothko, Pollock, John Cage or Ai Weiwei. Perhaps they should be, but they're not. Therefore, in what follows, by “beauty” we will refer to the most general standard+ aesthetic value, and by “ugly” we will refer the lack of such value.[1]

On the standard+ conception of aesthetic value, is MELVIN’s experiment beautiful? It might seem so, from what Steinberg said above. However, in an interesting turn of conversation that inspired this chapter, Steinberg focused directly on the aesthetic properties of the experiment and in so doing, developed a different view. He went on, “the part of the experiment the AI designed is probably less elegant than many of the simple setups humans have designed for various experiments along the way.” At this point, he begins to separate the contribution of the AI from the contribution of the humans. “When I talk about the ‘experiment,’ I mean this chip the group put together, and…the computer didn’t tell them how to do that.” This is important, as it recognizes a general feature of AI designs. Whether AI is used in designing an experiment, a chair, a car, a building, or something else, many of the features that humans find aesthetically pleasing are those added by the humans who chose and implemented the design.

Further, “if the solution is elegant, it’s not so hard for people to stumble upon, while if it’s arcane, it may be harder for us to discover.” In other words, when the experimental design is “nice and elegant,” we don’t need the AI, because “people would have gotten around to inventing it if they’d had much reason to.” This is important, because it reminds us that the experiments we need AIs to design won’t usually be ones with clear, simple, elegant solutions. They’ll be the arcane, ugly ones.

In general,

> With genetic algorithms, or other optimization techniques, they’ve been able to develop remarkable pulse sequences for controlling quantum systems. And usually, they just look like a godawful mess that happens to work… In some cases, you could squint your eyes and say ‘oh! it’s just trying to do [some technique we already knew about],’ and it would all make sense. And it would turn out that all the ‘mess’ was just ‘noise,’ and if a human had designed it, it would have been ‘prettier’ (simpler, cleaner), but worked just as well. I believe by now there are cases where this isn’t true, and we don’t know of ‘clean’ optimal solutions, so have only the computers.

To illustrate this point, Steinberg referred to some work of his own (fig. 1).

[Fig. 1 here]

[1] We could also characterize ugliness not as a lack of beauty, but, following a tradition that started with Rosenkranz (1853), as its own kind of positive property. Doing so should not affect the following arguments.

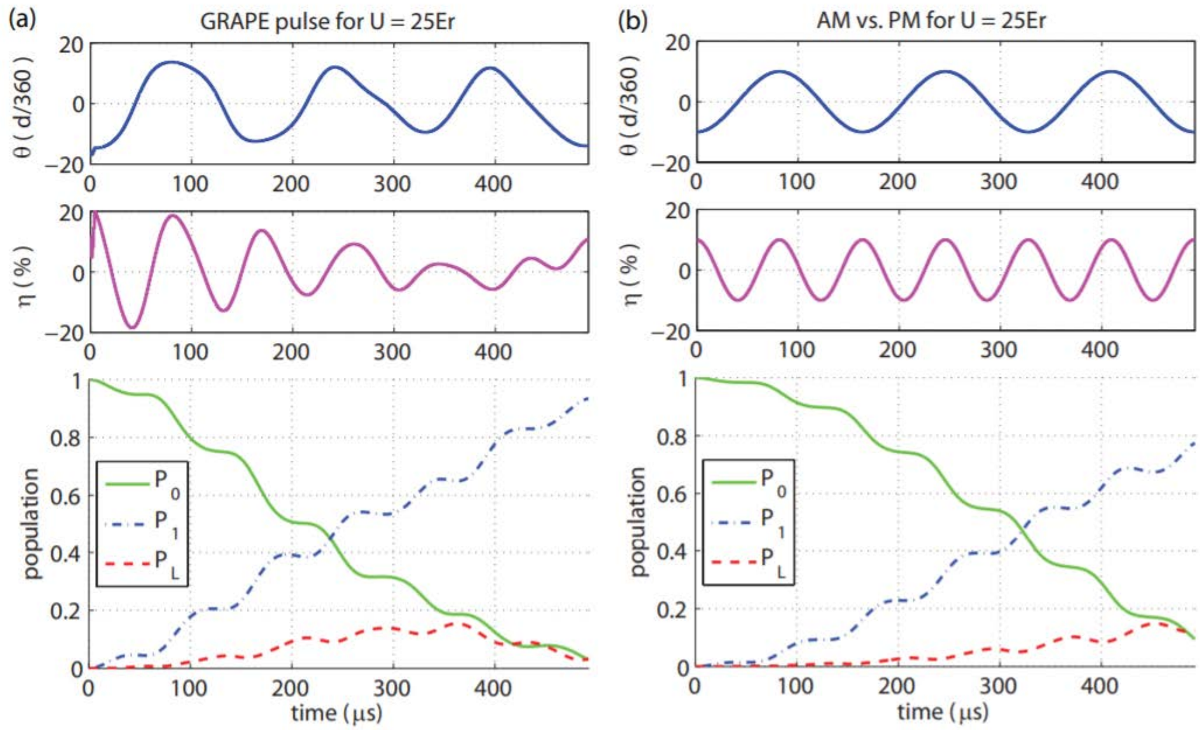

FIG. 6. GRAPE pulse and a similar AM vs. PM pulse. (a) GRAPE pulse and the corresponding simulation results with the lattice Hamiltonian without gravity. (b) A similar AM vs. PM pulse which has the same pulse duration and the variance as the GRAPE pulse and the corresponding simulation results with the lattice Hamiltonian without gravity.

Fig. 1. An example of an "ugly" use of AI in a quantum experiment (from Zhuang et al. 2013).

Steinberg explains that the figure

> Shows two clean sinusoids (in the upper right) which implement a technique that makes physical sense to us, and which we intentionally tried on our atoms; 'GRAPE' (which is a computer optimisation technique) invented its own optimum, shown on the upper left, and you can see that it seems to have more or less discovered the same idea. But if you look at the bottom panels, which show (in simulations) the effectiveness of the techniques (which are meant to bring the line for P_1 as high as possible by the end of the pulse, and the other two [P_0 and P_L] as low as possible), it looks like the GRAPE pulse – at first glance just a 'messier' version of our designed pulse – actually does better. And we don't know anything about why this 'uglier' pulse works better. We also don't know how much you could 'clean it up' and still have it work well.

In the end, experiencing MELVIN's experimental design (on the page or in the imagination) caused pleasure, which earned it the label "gorgeous" on the standard+ conception of aesthetic value. But on closer inspection, when we isolate the aspects of the design that were created by the AI, before it had been cleaned up and implemented by the humans, it is more likely to look a "godawful mess." And this will be true for experiments designed by AI in general. It is not impossible that we could design algorithms that produce beautiful experiments in the standard+ sense. However, the point is that scientists are not motivated to do this. Why would they be? In science we often need a way to control or manipulate systems that have a great deal of variables and/or display some chaotic behaviour, and it is in these contexts that AI will be most valuable. In such contexts, we should not expect AI designs to be beautiful, that is, simple, clear, elegant, and with a satisfying rhythm or cadence. This is because AI algorithms typically employ deep neural networks and machine learning, which consist in brute force exploration of all the options in a pre-selected set (or set of sets) of potential solutions in a problem space, or they alter node weights (i.e., numbers by which to multiply numerical inputs) that were randomly assigned until a certain level of performance is reached, defined in terms of finding a desired local minimum (i.e., "solution"). Their power comes almost entirely from calculation speed. There is nothing in the code that tells the AI to prefer solutions that produce pleasure in humans, and that is a good thing, because it would limit the set of solutions the AI would consider. We use AI precisely because it is not limited in the same ways that humans are. It finds solutions that aren't the ones that we would look for, because its "processing fluency" is completely different from ours. Humans find pleasure in simple, clear, elegant designs, because that is what we like to process, because it is easy to process. AI transcends those limitations. This explains why we are seeing AI used in certain parts of science and not others. Where there are many variables, heaps of data, and we need subtle, complex solutions, scientists will be tempted to turn to AI.

Note that the nature and value of AI designs is similar, both aesthetically and epistemically, to AI proofs in mathematics. Viktor Blåsjö argues that "an ugly proof resorts to computations, algorithms, symbolic manipulation, ad hoc steps, trial-and-error, enumeration of cases, and various other forms of technicalities. The mind can neither predict the course nor grasp the whole; it is forced to cope with extra-cognitive contingencies" (2018). Similarly, Alan Cain writes about computer-assisted proofs that "readers cannot see why this construction or this calculation or that definition is being carried out; they cannot perceive a reason for it that is internal to the proof…the deus is a cataract that their intuition cannot easily navigate" (2010). Cain is inspired by G. H. Hardy's view of mathematical beauty as unexpected inevitability, which a computer-assisted proof does not provide. Ulianov Montano takes this view further by comparing mathematical proofs to narratives, which should have a satisfying plot to be beautiful. Montano claims that "the experience [of studying a computer-assisted proof] has deforming narrative gaps. A computer-assisted proof shall always give us an incomplete experience, something we cannot fully appreciate, despite the fact that the proof is a perfectly acceptable and widespread method. Thus, it is not very plausible that we shall come to regard computer-assisted proofs as beautiful" (2014, 203). These quotations all have in common the idea that the products of digital methods in mathematics are ugly, and for the same reason: their outputs are ad hoc, they use brute force, and they are disfluent to the human mind. But as Montano notes, these methods are "perfectly acceptable" and becoming widespread. In the case of the computer proof of the four-colour theorem, Montano argues that it was never really questioned by mathematicians *as a proof*. "The issue here is not whether the proof is genuine or not, but rather that it is ugly. The ugliness of the proof is thus not explained by mathematicians' epistemic or other technical concerns" (37). The ugliness we experience is similar to watching a movie with a complicated plot that suddenly stops and reports that the main characters lived happily ever after, while refusing to explain anything more. "The assistance of the computer impairs, or rather wrecks, the proof's storytelling" (2014, 37). A similar story can be told for AI-designed experimental designs: what is expected but missing are the rhythm, cadence, narrative arc, underlying coherence, and the feeling that you are in the competent hands of an expert storyteller. All of this contributes to a feeling of ugliness.

I conclude that AI-designed experiments can be both ugly and epistemically good (again, where "epistemically good" refers to something like internal validity). This ugliness is characteristic of AI-use in science. Given that AI plays ever greater roles in science, we can expect more and more ugly experimental designs in science. Of course, what makes AI-designed experiments ugly is nothing new: many experiments in science can be found that will be arcane and without a satisfying rhythm in very similar ways. But the case with AI experiments is especially clear, so I will continue to draw on that in what follows.

4. The uglier the better

As a reminder, here are the options we outlined at the beginning for possible relations between aesthetic and epistemic value.

1. **Independence**: the positive or negative aesthetic value of an experimental design tells us nothing about its epistemic value.
2. **Proportionality**: increasing the aesthetic value of an experiment's design generally tends to increase its epistemic value.
3. **Necessity**: if an experiment's design has epistemic value, it has aesthetic value.
4. **Sufficiency**: if an experiment's design has aesthetic value, then it has epistemic value.
5. **Necessity and sufficiency**: an experimental design has epistemic value if and only if it has aesthetic value.

I don't think we need to say much about sufficiency because no one claims that simplicity or clarity (etc.) on their own guarantee epistemic value. But there are good reasons to think that beauty is *necessary* for epistemic value. I will consider two arguments for this claim. First, we might think that beauty is necessary for epistemic value because without beauty, scientists would have no motivation to do science, and without science, we would lose all the epistemic value that comes with science. Indeed, we know that scientists *do* find beauty motivating, especially in their search for fundamental theories, even if this sometimes holds science back (Hossenfelder, 2018; Vaidyanathan and Jacobi, 2022). A version of this idea is expressed by Poincaré's famous quotation: "The scientist does not study nature because it is useful to do so. He studies it because he takes pleasure in it, and he takes pleasure in it because it is beautiful. If nature were not beautiful it would not be worth knowing, and life would not be worth living " (Poincaré, 1914/2003, 22). Poincaré has a particular "intellectual" notion of beauty in mind, and few would agree that the lives of those without such beauty are not worth living. A more moderate form of this argument would be that some beauty is required for science to be the kind of thing that humans pursue. Such a view is perhaps suggested by Schellekens (2022), Ivanova (2020) and Breitenbach (2020), though I do not want to claim that any of these authors really supports a strong necessitarian like this.

There are several ways we might respond. First, beauty is motivating, but it is not the only kind of value that motivates. The Work and Well-Being in Science project finds that while 62% of scientists were motivated to become a scientist due to considerations of beauty, this leaves 38% who were not (Vaidyanathan and Jacobi, 2022). Besides beauty, scientists are motivated by the ethical good that they can achieve by doing science, as well as epistemic concerns like finding truth, knowledge and understanding. So beauty is not necessary for scientific motivation. Another reply would be to consider closed-loop AI scientists, mentioned above, which design, perform, and interpret their own experiments. In this case, the experimental designs need not have any amount of beauty in the standard+ sense for the AI to remain "motivated." All it requires is a power outlet. Thus, beauty is not necessary for epistemic value, either for humans or machines.

A second argument for the necessity claim would be that some amount of beauty is necessary for the experiment's design to be understandable, and this is one kind of epistemic value that experimental designs can have. This line of thinking is inspired by arguments made by Elgin (2020) and Ivanova (2020). For Catherine Elgin, aesthetic value can act as a gatekeeper, telling us what counts as acceptable in science. For Ivanova, our preference for beautiful things is "deeply ingrained," such that "taking aesthetic values in science as conditions of our cognitive makeup reflecting our intellectual interest and capacities explains why aesthetic values persist even when the best theories do not seem to quite fit our aesthetic requirements" (2020). Again, neither author states that they are arguing for the necessity of aesthetic value for epistemic value, but we can draw inspiration from them to construct such an argument.

Again, there are several ways to respond. One is to point out that understandability is only necessary if humans are always needed to implement experimental designs. If we are talking about a closed-loop AI scientist, the kind of simplicity and clarity required for human understandability or epistemic gatekeeping is not necessary. Thus, even if some beauty is required for *human* understandability, human understandability is not required for all experimental designs to have epistemic value, because those designs that are not understandable to humans can still be understood by machines. Another reply would deny that beauty is required for human understandability. After all, we can understand complicated, ugly things: we simply don't enjoy the process very much. Finally, we might also argue that human understandability isn't the main epistemic good that scientists want experimental designs to have. In this chapter, I have focused on internal validity, which is a measure of how well an experiment identifies the nature and size of the causal effects of one variable on (an)other(s). Signs of internal validity might include a high signal-to-noise ratio, robustness to perturbation, proper isolation of the experimental system, and an established theory of the instruments used. A scientist might be able to establish that all of these are possessed by a certain experimental design, even for an experiment that is "more like a Rube Goldberg machine" (Elgin, 2020).

At this point, we can reject options 3-5 from the above list. Let us then turn to the second option. The claim here is that increasing aesthetic value generally *tends* to increase epistemic value. Some of the arguments presented above for the necessity claim can be weakened to support this position. For example, increasing aesthetic value generally tends to increase the motivation of scientists, which tends to increase epistemic value in science. Or, increasing aesthetic value tends to increase human understandability, which tends to increase epistemic value. In reply to each, it is important to keep in mind that the claim should not be about tending to increase epistemic value in general, but about tending to increase the epistemic value of a particular experiment's design. For example, suppose it was part of an experiment's design to play music in the laboratory while performing the experiment. We might find that this made scientists more motivated, and that might lead to more epistemic value. But it would not increase the epistemic value – at least in the sense of internal validity – of that particular experiment's design. Likewise, it might be the case that a simpler design is easier for a scientist to understand, and therefore, it might pass the gate of acceptability and scientists can easily grasp how the experiment works. But these on their own don't say much about the likeliness of that experiment to correctly identify the causal structure of a system.

Still, it is true that as designs get uglier, humans will have tend to have less understanding of how the experiments work, in all the main senses of understanding (see, e.g., Stuart, 2018). In some cases, XAI algorithms will be able to explain how an experiment works, but this is not always possible, as Steinberg points out. And even when this is possible, it is likely that the grasp a scientist has via such an explanation is less than it would be for a more beautiful experiment. And while an ugly experiment can greatly enhance the abilities (or "practical understanding") that a scientist has with respect to a target system, ugly designs can make it harder for a scientist to develop the relevant abilities to use that experimental design in fruitful ways. Given this connection between beauty and human understanding, ugly designs will tend to be less understandable. However, if scientists are consequentialists about their tools (Stuart, 2022), we should expect scientists to be comfortable trading lessened understandability for the other epistemic benefits of AI designs. AI designs are becoming more common in science precisely because of their overall epistemic advantages. So, it would be incorrect to claim that changes in aesthetic value generally tend to correlate with equal changes in epistemic value, or in other words, that aesthetic and epistemic value are proportional to one another.

To put the point another way, ugly AI-designed experiments tend to have more epistemic value *overall* than human-designed experiments aimed to test the same hypotheses. This is why they are used. Thus, insofar as AI-designed experiments are not outlying oddities but are here to stay, we cannot claim that aesthetic and epistemic value are *generally* proportional. The best way to defend such a general (science-wide) proportionality claim in the face of the above arguments would be to show that AI ugliness is somehow a special case, and that AI will not become increasingly common in experimental design. But it seems to me that the sort of ugliness we see in AI designs is simply a more pronounced version of the same kind of ugliness we sometimes see in human-made designs. The difference is one of degree, not kind. And given that science has a long future ahead of us, it does not seem prudent to bet against the increasing use of AI methods.

Thus, we are left with the first option (independence) and a significantly weakened version of the second option (proportionality). The second option is weakened in the sense that it would hold only for those experiments implemented by humans, and only in terms of the understandability of the design, not internal validity. With respect to the first option, things are still unclear. The reason that AI-designed experiments are ugly *and* epistemically good is partially explainable thanks to a common cause: the nature of the algorithms involved. In this sense, aesthetic and epistemic value are not independent. But an experiment can be ugly and good, ugly and bad, beautiful and good, or beautiful and bad. In this sense, the two different kinds of values are independent because the presence of one does not (on its own) tell us anything about the presence of the other.

5. Objections

Someone might object that AI can't "design" experiments because AI can't *do* anything, in the properly intentional sense required of human action. This might be because AI is not conscious, or because it is not connected in the right way to the objects of its "cognition." To such an objection, we can simply remove all references to AI as an active designer, and allow that AI is merely a tool, like a hammer or paintbrush. On this presentation, we claim that when *humans* use AI to design experiments, the designs are ugly. The rest of the arguments go through as before, since we have identified an important set of experiments that are both ugly and epistemically good.[2]

Another kind of objection is really a lament: If we are headed into a future in which science has only (or mostly) ugly experiments, the world would be less beautiful, and it is our duty to prevent this. To argue this way accepts the conclusion of this chapter but rejects it as repugnant. I can sympathize: a future in which all scientific experimentation was done by AI and AI alone, perhaps in big steel factories in the desert, seems somehow deficient. But in reply, we can point out that such a future, however unlikely, would trade one source of beauty for others, and it is not clear that aesthetic value would decrease overall. Scientists who find aspects of experimental design work beautiful would have to move on to other pursuits, perhaps in applying what the Science Machine revealed to us. However, a future like that might be one in which diseases are cured, climate change is reversed, energy and food needs are met, and so on, such that our time could be freed up for longer, happier lives full of the production and appreciation of beauty.

Another objection would refer to human aesthetic adaptation, a feature of history so fruitfully employed by McAllister (1996). The idea is that human aesthetic values have evolved over time, and they will continue to evolve, such that the more useful AI-designed experiments come to be, the more beautiful scientists will come to find them. This accepts the premises of the argument but denies the conclusion. That is, it allows that AI-designed experiments are *currently* ugly, but claims that we will eventually emerge from this transitory phase, aesthetically acclimatized. After all, there was more than a century of baroque art and architecture, why deny that we might bring baroque values back? While human aesthetic values will certainly adapt to a world increasingly suffused with AI, I think it is unlikely that future humans could change (or indeed *reverse*) their deeply psychologically ingrained preferences for simplicity, symmetry, elegance, clarity, and so on, in the way required to appreciate AI designs as beautiful in the standard+ sense. Even the most baroque art is still beautiful in that standard+ sense. Such art demands more

[2] Of course, hammers and paintbrushes change over time, and some are adorned and beautiful. But the point is that the basic functions of these tools hasn't changed. Equally, user-interfaces for AI algorithms in science will become easier to work with, and perhaps one day they will be beautiful. But unless the fundamental structure of AI changes, the processes and output of AI will remain ugly.

attention to be paid to the details, and this can be challenging on first exposure. But that attention reveals details that are themselves elegant, symmetrical, and so on, and the overall result is clearly pleasurable. This is not the case for designs that make no attempt whatsoever to produce something understandable, clear, simple, symmetrical, or united by a satisfying narrative structure. It seems unlikely that humans could adapt in the foreseeable future to find designs with such properties pleasurable.

A counterargument would be that human aesthetic values have already moved on far beyond the baroque to embrace all the mad diversity of contemporary art, much of which already resembles AI-outputs, and which has aesthetic value. This must be admitted. But the point of the chapter is to challenge the claim that AI does or will design experiments that have aesthetic value *in the standard+ sense*. That is the sense with which philosophers and scientists are generally concerned, and it is not the sense with which artists, art critics and afficionados are concerned. Perhaps philosophers and scientists need an update. But that is another topic.

Finally, one might argue that if an AI were advanced enough, it might come to prefer certain strategies, or develop little traditions, and experience a sense of rhythm and momentum which it could celebrate in its work. In other words, if it were a "live creature" in John Dewey's sense, then AI-designed experiments might be beautiful to the AIs themselves. We are now firmly in the realm of science fiction, but if such a future obtained, we could accept the premises but reject the conclusion of this chapter by allowing that AI-designed experiments are ugly for humans, while insisting that they have aesthetic value for *someone*, and that this aesthetic value might be connected to their epistemic value. In reply, note that AI is so useful because we can program the means to identify solution states without including aesthetic preferences. And this is one important reason that AI is able to find solutions that appear so "alien" to us (see Halina, 2021). If we want AI to continue to produce useful alien solutions, either we must prevent AI from developing aesthetic values, or we must continuously build new AIs that have entirely different sets of aesthetic values then their predecessors. The second option is costly and comes with no obvious advantage, so scientists might be motivated to continue with the status quo. And in that case, AI designs will continue to be ugly to everyone who is capable of aesthetic displeasure.

6. Conclusion

The design of an experiment can have at least three kinds of value: ethical, epistemic, and aesthetic. These kinds of value can come apart. This chapter concerned the strength of the relationship between aesthetic and epistemic value in experimental designs: are these kinds of values independent, or does the first generally tend to increase the second, or is the first necessary or sufficient (or both) for the second? To investigate, we examined a case study of an AI-designed experiment, demonstrating that AI-designed experiments can be ugly and epistemically valuable at the same time. Whether AI-designed experiments are ugly in the same sense as ugly human-designed experiments or not, we can expect the use of AI in experimental design to continue to expand, and therefore this kind of ugliness should not be dismissed. Its existence puts pressure on any strong relationship postulated between aesthetic and epistemic value, since ugliness is not only acceptable in science, it can be epistemically desirable. Specifically, I conclude that beauty is neither necessary nor sufficient for epistemic value, and increasing beauty does not generally tend to increase epistemic value. Whether aesthetic and epistemic values are wholly independent is a question left open.

In terms of experimental design, the future might not be pretty. But given the consequentialist nature of science, we can expect this to help, not hinder, the achievement of science's main epistemic goals.

Acknowledgments

I would like to thank Milena Ivanova and Alice Murphy for their kind invitation to contribute to this volume, for many stimulating conversations before and after that invitation, as well as for extremely helpful comments on a draft of the paper. I would also like to thank Milena for organizing two workshops leading up to this volume, sponsored by the British Society of Aesthetics. Thanks also to audiences at those workshops, audiences at the University of York and Aarhus University, and finally, thanks to Aephraim Steinberg for the exciting conversation that planted the seed for this chapter.

References

Ananthaswamy, A. (2021) 'AI Designs Quantum Physics Experiments beyond What Any Human Has Conceived', *Scientific American*. Available at: https://www.scientificamerican.com/article/ai-designs-quantum-physics-experiments-beyond-what-any-human-has-conceived/ (Accessed: 28 November 2022).

Blåsjö, V. (2018) 'Mathematicians Versus Philosophers in Recent Work on Mathematical Beauty', *Journal of Humanistic Mathematics*, 8(1), pp. 414–431. Available at: https://doi.org/10.5642/jhummath.201801.20.

Breitenbach, A. (2020) 'One Imagination in Experiences of Beauty and Achievements of Understanding', *The British Journal of Aesthetics*, 60(1), pp. 71–88. Available at: https://doi.org/10.1093/aesthj/ayz048.

Cain, A.J. (2010) 'Deus ex Machina and the Aesthetics of Proof', *The Mathematical Intelligencer*, 32(3), pp. 7–11. Available at: https://doi.org/10.1007/s00283-010-9141-z.

Campbell, D.T. (1957) 'Factors relevant to the validity of experiments in social settings', *Psychological bulletin*, 54(4). Available at: https://doi.org/10.1037/h0040950.

Elgin, C.Z. (2020) 'Epistemic Gatekeepers: The Role of Aesthetic Factors in Science', in *The Aesthetics of Science*. Routledge.

Erhard, M. *et al.* (2018) 'Experimental Greenberger–Horne–Zeilinger entanglement beyond qubits', *Nature Photonics*, 12(12), pp. 759–764. Available at: https://doi.org/10.1038/s41566-018-0257-6.

Erhard, M., Krenn, M. and Zeilinger, A. (2020) 'Advances in high-dimensional quantum entanglement', *Nature Reviews Physics*, 2(7), pp. 365–381. Available at: https://doi.org/10.1038/s42254-020-0193-5.

Feigenbaum, E. (2007) ‘Oral History of Edward Feigenbaum’. Available at: http://archive.computerhistory.org/resources/access/text/2013/05/102702002-05-01-acc.pdf.

Guala, F. (2003) ‘Experimental Localism and External Validity’, *Philosophy of Science*, 70(5), pp. 1195–1205. Available at: https://doi.org/10.1086/377400.

Guala, F. (ed.) (2005) ‘External Validity’, in *The Methodology of Experimental Economics*. Cambridge: Cambridge University Press, pp. 141–160. Available at: https://doi.org/10.1017/CBO9780511614651.008.

Halina, M. (2021) ‘Insightful artificial intelligence’, *Mind & Language*, 36(2), pp. 315–329. Available at: https://doi.org/10.1111/mila.12321.

Hogarth, R.M. (2005) ‘The challenge of representative design in psychology and economics’, *Journal of Economic Methodology*, 12(2), pp. 253–263. Available at: https://doi.org/10.1080/13501780500086172.

Horgan, J. (2012) ‘What Are Science’s Ugliest Experiments?’, *Scientific American Blog Network*. Available at: https://blogs.scientificamerican.com/cross-check/what-are-sciences-ugliest-experiments/ (Accessed: 28 November 2022).

Hossenfelder, S. (2018) *Lost in Math: How Beauty Leads Physics Astray*. New York: Basic Books. Available at: https://www.basicbooks.com/titles/sabine-hossenfelder/lost-in-math/9781541646766/ (Accessed: 28 November 2022).

Ivanova, M. (2020) ‘Beauty, Truth and Understanding’, in *The Aesthetics of Science: Beauty, Imagination and Understanding*. London: Routledge. Available at: https://www.routledge.com/The-Aesthetics-of-Science-Beauty-Imagination-and-Understanding/Ivanova-French/p/book/9781032337180 (Accessed: 28 November 2022).

Ivanova, M. (2021) ‘The aesthetics of scientific experiments’, *Philosophy Compass*, 16(3), p. e12730. Available at: https://doi.org/10.1111/phc3.12730.

Ivanova, M. (2022) ‘What is a Beautiful Experiment?’, *Erkenntnis* [Preprint]. Available at: https://doi.org/10.1007/s10670-021-00509-3.

Jimenez-Buedo, M. and Miller, L.M. (2010) ‘Why a Trade-Off? The Relationship between the External and Internal Validity of Experiments’, *THEORIA*, 25(3), pp. 301–321. Available at: https://doi.org/10.1387/theoria.779.

King, R. (2021) ‘The Automation of Science’. *Workshop on Artificial Scientific Discovery 2021*, Max Plank Institute for the Science of Light. Available at: https://mpl.mpg.de/divisions/marquardt-division/workshops/2021-artificial-scientific-discovery/ (Accessed: 28 November 2022).

King, R.D. *et al.* (2004) ‘Functional genomic hypothesis generation and experimentation by a robot scientist’, *Nature*, 427(6971), pp. 247–252. Available at: https://doi.org/10.1038/nature02236.

Kitano, H. (2021) ‘Nobel Turing Challenge: creating the engine for scientific discovery’, *npj Systems Biology and Applications*, 7(1), pp. 1–12. Available at: https://doi.org/10.1038/s41540-021-00189-3.

Krenn, M. (2021) ‘A journey from computer-designed quantum experiments to computer-inspired scientific understanding’. *Workshop on Artificial Scientific Discovery 2021*, Max Plank Institute for the Science of Light. Available at: https://mpl.mpg.de/divisions/marquardt-division/workshops/2021-artificial-scientific-discovery/ (Accessed: 28 November 2022).

Krenn, M., Erhard, M. and Zeilinger, A. (2020) ‘Computer-inspired quantum experiments’, *Nature Reviews Physics*, 2(11), pp. 649–661. Available at: https://doi.org/10.1038/s42254-020-0230-4.

Lindsay, R.K., Buchanan, B.G. and Feigenbaum, E.A. (1993) ‘DENDRAL: a case study of the first expert system for scientific hypothesis formation’, *Artificial Intelligence*, 61(2), pp. 209–261. Available at: https://doi.org/10.1016/0004-3702(93)90068-M.

McAllister, J.W. (1996) *Beauty and Revolution in Science*. Cornell University Press. Available at: https://www.jstor.org/stable/10.7591/j.ctv5rf5v0 (Accessed: 28 November 2022).

Montano, U. (2014) *Explaining Beauty in Mathematics: An Aesthetic Theory of Mathematics*. London: Springer (Synthese library). Available at: https://www.academia.edu/25566436/Introduction_to_a_Naturalistic_Aesthetic_Theory (Accessed: 29 November 2022).

Murphy, A. (2020) ‘The Aesthetic and Literary Qualities of Scientific Thought Experiments’, in M. Ivanova and S. French (eds) *The Aesthetics of Science: Beauty, Imagination and Understanding*.

Parsons, G. and Rueger, A. (2000) ‘The epistemic significance of appreciating experiments aesthetically’, *The British Journal of Aesthetics*, 40(4), pp. 407–423. Available at: https://doi.org/10.1093/bjaesthetics/40.4.407.

Poincaré, H. (1914) *Science and Method*. London: Dover.

Roper, K. *et al.* (2022) ‘Testing the reproducibility and robustness of the cancer biology literature by robot’, *Journal of The Royal Society Interface*, 19(189), p. 20210821. Available at: https://doi.org/10.1098/rsif.2021.0821.

Rosenkranz, K. (1853) *Aesthetics of Ugliness*. Translated by A. Pop and M. Widrich. Bloomsbury. Available at: https://www.bloomsbury.com/us/aesthetics-of-ugliness-9781472568878/ (Accessed: 30 November 2022).

Schellekens, E. (2022) ‘Aesthetic Experience and Intellectual Pursuits’, *Aristotelian Society Supplementary Volume*, 96(1), pp. 123–146.

Stuart, M.T. (2018) ‘How Thought Experiments Increase Understanding’, in *The Routledge Companion to Thought Experiments*. Routledge. Available at: https://doi.org/10.4324/9781315175027.ch30.

Stuart, M.T. (2022) ‘Sharpening the tools of imagination’, *Synthese*, 200(6), p. 451. Available at: https://doi.org/10.1007/s11229-022-03939-w.

Todd, C.S. (2008) ‘Unmasking the Truth Beneath the Beauty: Why the Supposed Aesthetic Judgements Made in Science May Not Be Aesthetic at All’, *International Studies in the Philosophy of Science*, 22(1), pp. 61–79. Available at: https://doi.org/10.1080/02698590802280910.

Vaidyanathan, B. and Jacobi, C. (2022) *The Role of Aesthetics in Science: Survery Findings from Work and Well-Being in Science: An International Study*. Department of Sociology, The Catholic University of America, p. 10. Available at: https://workandwellbeingstudy.com/wp-content/uploads/2022/05/vaid-aesthetic-fact-sheet.pdf (Accessed: 30 November 2022).

Wilczek, F. (2016) ‘Physics in 100 years’, *Physics Today*, 69(4), pp. 32–39. Available at: https://doi.org/10.1063/PT.3.3137.

Williams, K. *et al.* (2015) ‘Cheaper faster drug development validated by the repositioning of drugs against neglected tropical diseases’, *Journal of The Royal Society Interface*, 12(104), p. 20141289. Available at: https://doi.org/10.1098/rsif.2014.1289.

Zhuang, C. *et al.* (2013) ‘Coherent Control of Population Transfer between Vibrational States in an Optical Lattice via Two-Path Quantum Interference’, *Physical Review Letters*, 111(23), p. 233002. Available at: https://doi.org/10.1103/PhysRevLett.111.233002.

Żytkow, J.M. (1987) ‘Combining many searches in the FAHRENHEIT discovery system’, in P. Langley (ed.) *Proceedings of the Fourth International Workshop on MACHINE LEARNING*. Morgan Kaufmann, pp. 281–287. Available at: https://doi.org/10.1016/B978-0-934613-41-5.50032-5.

Żytkow, J.M., Zhu, J. and Hussam, A. (1990) ‘Automated discovery in a chemistry laboratory’, in *Proceedings of the eighth National conference on Artificial intelligence - Volume 2*. Boston, Massachusetts: AAAI Press (AAAI’90), pp. 889–894.