Detecting Design by Eliminating Chance:

A Response to Robin Collins

William A. Dembski

419 Pat Neff Hall

Baylor University

Waco, Texas 76798

The Design Inference as a Statistical Inference

In his CSR review of my book *The Design Inference* (henceforth TDI),[1] Robin Collins not only claims to show that my eliminative approach to detecting design is fatally flawed but also offers an alternative likelihood approach to detecting design that, as he puts it, is "much better." Collins's likelihood approach has been around at least 200 years and was certainly familiar to me before I wrote TDI. I found this approach to detecting design inadequate then and I still do. Moreover, I submit that the scientific community regards it as inadequate as well (indeed, design would hardly be as controversial as it is in science were likelihoods the key to its detection). In this response to Collins I shall review my approach to detecting design, indicate why the likelihood approach that Collins proposes is itself highly problematic, and show that even if Collins's likelihood approach could be made to work, it would have to presuppose the very approach to detecting design that I develop in TDI. I admit one small confusion in TDI and one error that admits a quick fix. Other than that, I find Collins's critique overblown.

In TDI I argue that specified events of small probability reliably point to the activity of an intelligent agent. Specified improbability or "specified complexity" as I now prefer to call it (complexity and improbability here being synonymous) is a statistical or logical category. In TDI I argue that it provides a criterion for detecting intelligence or design. The question is, How good a criterion is it for detecting design? Now Collins is correct to observe that the terminology I use in TDI for connecting specified complexity to design is unclear. The problem is that in TDI I use the word "design" both as a logical predicate referring to specified complexity and as a mode of causation referring to intelligent agency. Collins picks up on this when he quotes me as follows: "The design inferred from the design inference does not logically entail an intelligent agent."[2] Indeed, design qua specified complexity cannot entail—in the strict logical sense—design qua

---

[1]William A. Dembski, *The Design Inference: Eliminating Chance Through Small Probabilities* (New York: Cambridge University Press, 1998).

[2]Ibid., 9.

intelligent agency. Nonetheless, design qua specified complexity does serve as a reliable criterion for design qua intelligent agency. I argue this point in section 2.4 of TDI. If there was an equivocation over the word "design," it disappears when these two uses are kept in mind. It also means that Collins is mistaken when he writes, "Dembski does not purport to provide a rigorous method for inferring to intelligent agency." I purport to do just that, using specified complexity as a feature of events, objects, and structures that reliably signals intelligent agency.

The approach to detecting design that I develop in TDI fits squarely within the mainstream statistical literature, and it is helpful to see this. Specified complexity belongs to statistical decision theory. Statistical decision theory sets the ground rules for how to draw inferences about probabilistic claims. The approach I take follows the common statistical practice (popularized by Ronald Fisher) of rejecting a chance hypothesis if a sample appears in a prespecified rejection region. For example, suppose one's chance hypothesis is that a coin is fair. To test whether the coin is biased in favor of heads, and thus not fair, one can set a rejection region of ten heads in a row and then flip the coin ten times. According to Fisher, if the coin lands ten heads in a row, then one is justified rejecting the chance hypothesis that the coin is fair. As a criterion for detecting design, specified complexity extends Fisher's approach to hypothesis testing in two ways: First, it generalizes the types of rejection regions by which chance is eliminated (the generalized rejection regions being what I call "specifications"). Second, it eliminates all relevant chance hypotheses that could characterize an event rather than just a single chance hypothesis (it, as it were, sweeps the field clear of chance hypotheses).

Fisher's approach to hypothesis testing is the one most widely employed in the applied statistics literature and certainly the first one taught in introductory statistics courses. Nevertheless, among analytic philosophers like Robin Collins (much less among working statisticians), Fisher's approach is considered problematic. The problem is this. For a rejection region to warrant rejecting a chance hypothesis, the rejection region must have sufficiently small probability. The question is, How small is small enough? More precisely: Given a chance hypothesis and a rejection region, how small does the probability of the rejection region have to be so that if a sample falls within it, then the chance hypothesis can legitimately be rejected? The problem, then, is to justify what's called a "significance level" (always a positive real number less than one) such that whenever the sample falls within the rejection region and the probability of the rejection region given the chance hypothesis is no greater than the significance level, then the chance hypothesis can be legitimately rejected as explaining the sample. In the applied statistics literature it is common to see significance levels of .05 and .01. The problem to date has been that any such proposed significance levels have seemed arbitrary, lacking what Howson and Urbach call "a rational foundation."[3]

A major part of my project in TDI was therefore to provide just such a rational foundation for statistical significance levels. Now it's worth noting one significance level about which both Collins and I agree, namely, a statistical significance of probability zero. Hypotheses that confer zero probability are invariably rejected or passed over in favor of other hypotheses that confer nonzero probability. This is true not only of Fisher's approach to hypothesis testing but also of Collins's likelihood approach. Hypotheses are never confirmed by conferring zero probability. Hypotheses can only be confirmed by conferring nonzero probability. But that raises the question whether hypotheses can also be disconfirmed by conferring nonzero probability. Fisher says yes. Collins and fellow likelihood theorists say no. Collins's approach is essentially comparative, pitting competing chance hypotheses against each other and seeing which confers the greater probability. Consequently, for Collins a disconfirmed hypothesis is simply one that fails to confer as much probability as another hypothesis. Not the absolute value but the relative value of the probabilities conferred determines whether a hypothesis is confirmed in Collins's scheme.

I will discuss Collins's approach to hypothesis testing momentarily. But first I want to indicate how Fisher's approach to hypothesis testing can indeed be placed on a solid rational foundation. In TDI I argue that significance levels cannot be set in isolation, but must always be

---

[3]Colin Howson and Peter Urbach, *Scientific Reasoning: The Bayesian Approach*, 2nd ed. (LaSalle, Ill.: Open Court, 1993), 178.

set in relation to the probabilistic resources relevant to an event's occurrence (i.e., the number of opportunities for an event to occur or be specified). Critics of Fisher's approach to hypothesis testing are therefore correct in claiming that significance levels of .05, .01, and the like that regularly appear in the applied statistics literature are arbitrary. A significance level is supposed to provide an upper bound on the probability of a rejection region and thereby ensure that the rejection region is sufficiently small to justify the elimination of a chance hypothesis (essentially, the idea is to make a target small enough so that an archer can't hit it by chance). Rejection regions eliminate chance hypotheses when events putatively characterized by those hypotheses fall within the rejection regions. The problem with significance levels like .05 and .01 is that typically they are instituted without any reference to the probabilistic resources relevant to controlling for false positives. (Here a false positive is the error of eliminating a chance hypothesis as the explanation of an event when the chance hypothesis actually is operative. Statisticians refer to such false positives as "type I errors.") Given the concept of a probabilistic resource, the choice of significance level is no longer arbitrary but obtains definite meaning. The details for justifying Fisher's approach to hypothesis testing in terms of probabilistic resources can be found in chapter 6 of TDI.[4] Suffice it to say, there exists a non-question-begging account of statistical significance testing. The question remains whether there are more convincing ways to eliminate chance hypotheses.

Design by Comparison

Instead of obtaining design by eliminating chance, Collins prefers a comparative approach to hypothesis testing. In this approach all hypotheses are treated as chance hypotheses in the sense that they confer probabilities on states of affairs. Thus in a competition between a design hypothesis and other hypotheses, one detects design by determining whether the design

---

[4]Collins fails in his review to consider this emendation to Fisher's theory, much less how Fisher's theory challenges his likelihood approach.

hypothesis confers greater probability than the other hypotheses on a given state of affairs. Collins therefore tacitly subscribes to a model of explanation known as *inference to the best explanation*, in which a "best explanation" always presupposes at least two competing explanations. Inference to the best explanation eliminates hypotheses not by eliminating them individually but by setting them against each other and determining which comes out on top.

But why should eliminating a chance hypothesis always require additional chance hypotheses that compete with it? Certainly this can't be a universal requirement for eliminating hypotheses generally. Consider the following hypothesis: "The moon is made of cheese." One doesn't need additional hypotheses (e.g., "The moon is a great ball of nylon") to eliminate the moon-is-made-of-cheese hypothesis. There are plenty of hypotheses that we eliminate in isolation, and for which additional competing hypotheses do nothing to assist in eliminating them. Indeed, often with scientific problems we are fortunate if we can offer even a single hypothesis as a proposed solution (how many alternatives were there to Newtonian mechanics when Newton proposed it?). What's more, a proposed solution may be so poor and unacceptable that it can rightly be eliminated without proposing an alternative (e.g., the moon-is-made-of-cheese hypothesis). It is not a requirement of logic that eliminating a hypothesis means superseding it.

But is there something special about chance hypotheses? Collins advocates a likelihood approach to testing chance hypotheses. According to this approach, a hypothesis is confirmed to the degree that it confers probability on a known state of affairs. Unlike Fisher's approach, the likelihood approach has no need for significance levels or small probabilities. What matters is the relative assignments of probability, not their absolute value. Also, unlike Fisher's approach (which is purely eliminative, eliminating a chance hypothesis without accepting another), the likelihood approach focuses on finding a hypothesis that confers maximal probability, thus making the elimination of hypotheses always a by-product of finding and accepting a better hypothesis. But there are problems with the likelihood approach, problems that severely limit its scope and prevent it from becoming the universal instrument for adjudicating between chance hypotheses that Collins intends it. Indeed, I'll argue that the likelihood approach is necessarily parasitic on Fisher's approach, and that properly it can adjudicate only among hypotheses that Fisher's approach has thus far failed to eliminate. For Collins to call the likelihood approach "practically uncontroversial" or "a principle that all sides can agree upon" is therefore off the mark.

One problem, though by no means fatal to the likelihood approach, is that there always exists a chance hypothesis that concentrates all the probability on the state of affairs in question. Within the likelihood approach, what tests a set of chance hypotheses is the probabilities they confer on a given state of affairs. Now it's always possible to stipulate a chance hypothesis that assigns probability 1 (i.e., the maximum probability allowable) to that state of affairs. To be sure, in most contexts such a hypothesis will be highly artificial and thus not up for consideration. But the point is that the likelihood approach depends on a context of inquiry that specifies only certain hypotheses and then adjudicates among them. The likelihood approach thus chooses the best among the hypotheses under consideration, but is silent about how those hypotheses were put up for consideration in the first place. One therefore needs independent grounds for thinking that the hypotheses being considered are worthy of consideration and that adjudicating among them is valuable to our knowledge of the world. Fisher's approach, on the other hand, is not saddled with this problem because it is able to eliminate chance hypotheses individually.

A much more serious problem is that even with independent grounds for thinking one has the right set of hypotheses, the likelihood approach can still lead to wholly unacceptable conclusions. Consider, for instance, the following experimental set up. There are two urns, one with five white balls and five black balls, the other with seven white balls and three black balls. One of these urns will be sampled with replacement a thousand times, but we don't know which. The chance hypothesis characterizing the first urn is that white balls should on average occur the same number of times as black balls, and the chance hypothesis characterizing the second urn is that white balls should on average outnumber black balls by a ratio of seven to three. Suppose now we are told that one of the urns was sampled and that all the balls ended up being white. The probability of this event by sampling from the first urn is roughly 1 in 10^{300} whereas the probability of this event by sampling from the second urn is roughly 1 in 10^{155}.

The second probability is therefore almost 150 orders of magnitude greater than the first, and on the likelihood approach we should vastly prefer the hypothesis that the urn had seven white balls rather than only five. But getting all white balls from the second urn is a specified event of small probability, and on Fisher's approach to hypothesis testing should be eliminated as well (drawing with replacement from this urn 1000 times, we should expect on average around 300 black balls from this urn, and certainly not a complete absence of black balls). This comports with our best probabilistic intuitions: Given these two urns and a thousand white balls in a row, the only sensible conclusion is that neither urn was randomly sampled, and any superiority of the "urn two" hypothesis over the "urn one" hypothesis is utterly insignificant. To be forced to choose between these two hypotheses is like being forced to choose between the moon being made entirely of cheese or the moon being made entirely of nylon. Any superiority of the one hypothesis over the other drowns in a sea of inconsequentiality.

The likelihood principle, being an inherently comparative instrument, has nothing to say about the absolute value of the probability (or probability density) associated with a state of affairs, but only their relative magnitudes. Consequently, the vast improbability of either urn hypothesis in relation to the sample chosen (i.e., 1000 white balls) would on strict likelihood grounds be irrelevant to any doubts about either hypothesis. Indeed, any such doubts could only arise by presupposing specified complexity as a general instrument for eliminating chance (more on this later). Having performed a likelihood analysis and having found that the hypothesis on which this analysis conferred maximal probability is itself dubious, one cannot on purely likelihood grounds add another hypothesis to the mix—there must be good independent reasons for expanding the range of chance hypotheses considered, for without some such constraint one can, as in the previous objection, simply add a hypothesis that confers a probability of 1 (i.e., the maximum allowable probability) on the state of affairs in question. Certainly it is illegitimate to add a "none-of-the-above hypothesis" to the mix. A none-of-the-above hypothesis, which takes the hypotheses in a previous likelihood analysis and then merely asserts that none of these hypotheses accounts for a state of affairs in question, is not properly speaking even a hypothesis capable of conferring probabilities. Indeed, there's no way to assess the probability conferred on the selection of 1000 white balls by the (meta-)hypothesis that it wasn't an urn with five white and five black balls or an urn with seven white and three black balls.

It may help to recast the problem here with an analogy. Suppose you are the admissions officer at a prestigious medical school. Lots of people want to get into your medical school; indeed so many that even among qualified applicants you can't admit them all. You feel bad about these qualified applicants who don't make it and wish them well. Nonetheless, you are committed to getting the best students possible, so among qualified applicants you choose only those at the very top. What's more, because your medical school is so prestigious, you are free to choose only from the top. There is, however, another type of student who applies to your medical school, one whose grades are poor, who shows no intellectual spark, and who lacks the requisite premedical training. These are the unqualified students, and you have no compunction about weeding them out immediately, nor do you care what their fate is in graduate school. In this analogy the unqualified students are the hypotheses weeded out by Fisher's approach to hypothesis testing whereas the qualified students are those sifted by the likelihood approach. If, perchance, only unqualified students apply one year, the right thing to do would be to reject all of them rather than to admit the best of a bad lot.

I want next to examine the idea of hypotheses conferring probability, an idea that is mathematically straightforward but that becomes problematic when the likelihood approach gets applied in practice. According to the likelihood approach, chance hypotheses confer probability on states of affairs, and hypotheses that confer maximum probability are selected to the exclusion of others. But what exactly are these hypotheses that confer probability? In practice the likelihood approach is too cavalier about the hypotheses it permits. Urn models as hypotheses are fine and well because they induce well-defined probability distributions. Models for the formation of functional protein assemblages might also induce well-defined probability distributions, though determining the probabilities here will be considerably more difficult. But what about hypotheses like "Natural selection and random mutation together are the principal driving force behind biological evolution" or "God designed living organisms"? Within the likelihood approach any claim can be turned into a chance hypothesis on the basis of which likelihood theorists then assign probabilities. Claims like these, however, don't induce well-defined probability distributions. And since most claims are like this (i.e., they fail to induce well-defined probability distributions), likelihood analyses regularly become exercises in rank subjectivism.

Collins, for instance, offers a likelihood analysis of the "Caputo case" that he claims is superior to mine. Nicholas Caputo, a county clerk from New Jersey was charged with cheating because he gave the preferred ballot line to Democrats over Republicans 40 out of 41 times (the improbability here is that of flipping a fair coin 41 times and getting 40 heads). I'll return to Collins's critique of my analysis later, but for now I want to focus on Collins's own analysis. Here it is: "The actual ballot selection pattern was much more probable under the cheating hypothesis [i.e., the design hypothesis] than under the hypothesis that it occurred by chance. Thus, the ballot pattern confirms that Caputo cheated. Indeed, it is so much more probable under the one hypothesis than under the other that the confirmation turns out to be extremely strong, so strong that virtually any court would conclude cheating was involved."

Actually, the court did not conclude that cheating was involved.[5] Indeed, the probabilistic analysis that Collins finds so convincing is viewed with skepticism by the legal system, and for

---

[5]According to the *New York Times* (23 July 1985, B1): "The court suggested—but did not order—changes in the way Mr. Caputo conducts the drawings to stem 'further loss of public confidence in the integrity of the electoral process.' ... Justice Robert L. Clifford, while concurring with the 6-to-0 ruling, said the guidelines should have been ordered instead of suggested." The court did not conclude that cheating was involved, but merely suggested safeguards so that future drawings would be truly random.

good reason. Nowhere in Collins's likelihood analysis of the Caputo case does he assign precise numbers. The most we see in the likelihood literature are inequalities of the form $P(E|H1) >> P(E|H2)$, signifying that the probability of E (the ballot selection pattern) given H1 (that Caputo was cheating) is "much greater" than the probability of E given H2 (that ballots were selected by chance). What's more, precise numbers are almost never attached to these "probabilities." To say that one probability is "much greater" than another or that one event is, as Collins puts it, "much more probable" than another has no precise mathematical meaning. Collins's appeal to probability theory to make H1 and H2 confer a probability on E is therefore misleading, lending an air or mathematical rigor to what really is just Collins's own subjective assessment of how plausible these hypotheses seem to him. Collins's likelihood analysis is merely an elaborate way of asserting that with respect to the ballot selection pattern, the hypothesis of cheating (or design) is a much more convincing explanation to Collins personally than the hypothesis of pure chance. Nowhere does Collins attempt to show how the design hypotheses that come up in his likelihood analyses issue in well-defined probability distributions. Collins's probabilities are probabilities in name only.

The final problem with Collins's likelihood approach that I want to consider is his treatment of design hypotheses as chance hypotheses. For Collins any hypothesis can be treated as a chance hypothesis in the sense that it confers probability on a state of affairs. As we've just seen, there's a problem here because Collins's probabilities float free of well-defined probability distributions. But even if we bracket this problem, there's a problem treating design hypotheses as chance hypotheses, using design hypotheses to confer probability (now conceived in a loose, subjective sense) on states of affairs. To be sure, designing agents can do things that follow well-defined probability distributions. For instance, actuaries, marketing analysts, and criminologists are all interested in probability distributions connected with the actions of intelligent agents (e.g., murder rates). Such probability distributions ride, as it were, epiphenomenally on design hypotheses.

Collins, however, is much more interested in assessing probabilities that bear directly on a design hypothesis than in characterizing chance events that ride epiphenomenally on it. With intelligent design in biology, for instance, Collins wants to know what sorts of biological systems should be expected from an intelligent designer, and not what sorts of random epiphenomena might be associated with such a designer. Yet to demand this sort of knowledge seems ill-conceived. We infer design regularly and reliably without being able to assess what a given designer is likely to do on the basis of past experience or even on theoretical grounds. For instance, the Smithsonian Institute has a collection of obviously designed artifacts for which no one has a clue what their function is, much less who designed them.[6] Collins's likelihood approach puts designers in the same boat as natural laws, locating their explanatory power in an extrapolation from past experience. To be sure, designers, like natural laws, can behave predictably. Yet unlike natural laws, which are universal and uniform, designers are also innovators. Innovation, the emergence to true novelty, undermines expectations and therefore a likelihood analysis. A likelihood analysis generates expectations about the future by conforming the present to the past and extrapolating therefrom. Consequently, design, by involving innovation and invention, cannot be subsumed under a likelihood framework.

But the problem goes deeper. Not only can't Collins's likelihood approach account for the unpredictability inherent in design, but also it can't account for how we recognize design in the first place. Given his likelihood approach, Collins can regard the intelligent design hypothesis as fruitful for biology only if it confers sufficient probability on biologically interesting

---

[6]See Del Ratzsch's article "Design, Chance, and Theistic Evolution," in *Mere Creation*, ed. W. A. Dembski (Downers Grove, Ill.: InterVarsity, 1998), 294. Elliott Sober in critiquing my work admits this as well: "To infer watchmaker from watch, you needn't know exactly what the watchmaker had in mind; indeed, you don't even have to know that the watch is a device for measuring time. Archaeologists sometimes unearth tools of unknown function, but still reasonably draw the inference that these things are, in fact, *tools*." Elliott Sober, "Testability," *Proceedings and Addresses of the American Philosophical Association* 73(2) (1999): 73, n. 20.

propositions. But take a different example, say from archeology, in which a design hypothesis about certain aborigines confers a large probability on certain artifacts, say, arrowheads. Such a design hypothesis would on Collins's likelihood approach be rationally preferable to the chance hypothesis (i.e., that these rocks are merely the result of natural random processes). But what sort of archeological background knowledge had to go into that design hypothesis for Collins's likelihood analysis to be successful? At the very least, we would have had to have past experience with arrowheads. But how did we recognize that the arrowheads in our past experience were designed? Did we see humans actually manufacture arrowheads? If so, how did we recognize that these humans were acting deliberately as designing agents and not just randomly chipping away at random chunks of rock (carpentry and sculpting entail design; but whittling and chipping, though performed by intelligent agents, do not)? As is evident from this regress, any likelihood analysis capable of recognizing design must ultimately appeal to a characterization of design that transcends the likelihood approach.[7] Our ability to recognize design must therefore arise independently of Collins's likelihood framework.

The direction of Collins's logic is from design hypothesis to designed object, with the design hypothesis generating predictions or expectations about the designed object. Yet in practice we start with objects that initially we may not know to be designed. Then by identifying general features of those objects that reliably signal design, we infer to a designing intelligence

---

[7]Thomas Reid argued as much over 200 years ago: "No man ever saw wisdom, and if he does not [infer wisdom] from the marks of it, he can form no conclusions respecting anything of his fellow creatures.... But says Hume, unless you know it by experience, you know nothing of it. If this is the case, I never could know it at all. Hence it appears that whoever maintains that there is no force in the [general rule that from marks of intelligence and wisdom in effects a wise and intelligent cause may be inferred], denies the existence of any intelligent being but himself." Thomas Reid, *Lectures on Natural Theology*, eds. E. Duncan and W. R. Eakin (1780; reprinted Washington, D.C.: University Press of America, 1981), 56.

responsible for those objects. Still further downstream in the logic is an investigation into the specific causal factors involved in the construction and continued operation of those objects (e.g., How was the object constructed? How could it have been constructed? What is its function? What effect have natural causes had on the original design? Is the original design recoverable? How much has the original design been perturbed? How much perturbation can the object allow and still remain functional?). But what are the general features of designed objects that sets the design inference in motion and reliably signals design? The answer I give in TDI is *specification* and *complexity*. As I argue there, specified complexity is a reliable empirical marker of intelligent design.

Design by Elimination

The defects in Collins's likelihood approach are, I believe, so grave that it cannot provide an adequate account of how design hypotheses are confirmed. The question remains, however, whether specified complexity can provide an adequate account for how design hypotheses are confirmed. The worry here centers on the move from specified complexity to design. Specified complexity is a statistical notion. Design, as generally understood, is a causal notion. How, then, do the two connect? In *The Design Inference*, and more explicitly in *Intelligent Design*,[8] I argue the connection as follows. First I offer an inductive argument, showing that in all cases where we know the causal history and specified complexity was involved, that an intelligence was involved as well. The inductive generalization that follows is that all cases of specified complexity involve intelligence. Next I argue that choice is the defining feature of intelligence and that specified complexity is how in fact we identify choice.

Although I regard these two arguments as utterly convincing, Collins regards them as less so. The problem according to Collins is that specified complexity detects design purely by

---

[8]William A. Dembski, *Intelligent Design: The Bridge Between Science and Theology* (Downers Grove, Ill.: InterVarsity, 1999), ch. 5.

elimination, telling us nothing positive about how an intelligent designer might have produced an object we observe. Collins regards this as a defect. I regard it as a virtue. I'll come back to why I regard it as a virtue, but for the moment let's consider this criticism on its own terms. Take, for instance, a biological system, one that exhibits specified complexity, but for which we have no clue how an intelligent designer might have produced it. To employ specified complexity as a marker of design here seems to tell us nothing except that the object is designed. Indeed, when we examine the logic of detecting design via specified complexity, at first blush it looks purely eliminative. The "complexity" in "specified complexity" is a measure of improbability. Now probabilities are always assigned in relation to chance hypotheses. Thus, to establish specified complexity requires defeating a set of chance hypotheses. Specified complexity therefore seems at best to tell us what's not the case, not what is the case.

In response to this criticism, note first that even though specified complexity is established via an eliminative argument, it is not fair to say that it is established via a *purely* eliminative argument. If the argument were purely eliminative, one might be justified in saying that the move from specified complexity to a designing intelligence is an argument from ignorance. But unlike Fisher's approach to hypothesis testing, in which individual chance hypotheses get eliminated without reference to the entire set of relevant chance hypotheses that might explain a phenomenon, specified complexity presupposes that the entire set of relevant chance hypotheses has first been identified. This takes considerable background knowledge. What's more, it takes considerable background knowledge to come up with the right patterns (specifications) for eliminating all those chance hypotheses and thus for inferring design. Design inferences that infer design by identifying specified complexity are therefore not purely eliminative. They do not merely exclude, but they exclude from an exhaustive set in which design is all that remains once the inference has done its work. Design inferences, by identifying specified complexity, exclude everything that might in turn exclude design.

The question remains, however, What is the connection between design as a statistical notion (i.e., specified complexity) and design as a causal notion (i.e., the action of a designing intelligence)? Now it's true that simply knowing that an object is complex and specified tells us nothing positive about its causal history. To be sure, it tells us something negative about its causal history, namely, that the phenomenon in question could not have been produced solely by undirected natural causes. Even so, specified complexity by itself provides no causal details and thus from a likelihood perspective no explanation. Collins regards this as a defect of the concept. Yet it might equally well be regarded as a virtue for enabling us neatly to separate whether something is designed from how it was produced. Once specified complexity tells us that something is designed, there's nothing to stop us from inquiring into its production. A design inference therefore does not avoid the problem of how a designing intelligence might have produced an object. It simply makes it a separate question.

The claim that design inferences are purely eliminative is false, and the claim that they provide no (positive) causal story is true but hardly relevant—causal stories must always be assessed on a case-by-case basis independently of general statistical considerations. So where is the problem in connecting design as a statistical notion (i.e., specified complexity) to design as a causal notion (i.e., the action of a designing intelligence), especially given the close parallels between specified complexity and choice, and also the absence of counterexamples for generating specified complexity apart from intelligence?

In fact, Collins disputes the absence of counterexamples. I propose therefore next to run through Collins's presumed counterexamples. Let me start with potentially the most damaging one. This counterexample rightly identifies an inaccuracy in TDI. Although I'm grateful to Collins for identifying this inaccuracy, the problem admits a quick fix in terms of the apparatus laid out in TDI. The inaccuracy turns up in my formulation of the tractability condition, a condition that helps determine whether a pattern is a specification. According to this condition, to have a specification requires having an item of background knowledge that enables one to construct a pattern to which an event conforms. The problem is that in formalizing this ability to construct patterns on the basis of background information, I left open the possibility of unrestricted pattern generation. The number of patterns generated needs to be factored into the specificational resources (a type of probabilistic resource) that goes into assessing whether a significance level is small enough to warrant the elimination of chance. Once this is done, my analysis carries through.

Collins disputes this as well. For instance, he argues that my analysis of the Caputo case will now have to be fundamentally altered. Central to his argument is calling in computers as pattern generators. Once they are called in, the number of specificational resources will increase drastically, and as a consequence it will no longer be possible to eliminate chance in the Caputo case. Collins sees this as extremely damaging to my project, but I do not see it that way at all. Why, after all, should computers be allowed to inflate specificational resources in the Caputo case? As I point out in TDI, probabilistic resources are typically chosen according to pragmatic considerations that balance the need to eliminate chance when chance is not operating with the need to avoid eliminating chance when chance is operating. In civil trials, for instance, we allow far bigger probabilities to convict someone than in criminal trials (cf. "preponderance of evidence" vs. "to a moral certainty and beyond reasonable doubt") and thus require far fewer probabilistic resources than in criminal trials because mistakes in finding against someone are presumably less severe in civil trials than in criminal trials. And even in criminal trials one is not going to want to let probabilistic resources get too large for otherwise one will never convict anybody (the greater the number of probabilistic resources, the smaller the probability to eliminate chance—which here means conviction). Thus in bringing people like Caputo to justice for committing "probabilistic fraud," a court will be uninclined to increase probabilistic resources by using a computer to generate patterns.

It's important here to note that this contextual approach to setting probabilistic resources does not run into the same problems of subjectivism that Collins's likelihood approach faces. Although subjective factors often influence the setting of probabilistic resources, I argue in TDI that it is also possible to set them objectively by calculating a universal probability bound.[9] Such a universal probability bound takes into account all the specificational resources that might ever be encountered in the known physical universe. Universal probability bounds come up regularly in cryptographic research where they are used to assess the security of cryptosystems against brute force attacks in which the entire universe is treated as a (non-quantum) computer.[10] In TDI I compute a universal probability of 1 in 10^{150} and argue that it is beyond the capacity of the observable universe to generate sufficiently many specifications so that a specified event of that improbability could wrongly be attributed to design. And since improbabilities this small regularly arise in biology, specified complexity based on this universal probability bound is objectively given and directly applicable to biology.

Collins also appeals to physics to undermine specified complexity as a reliable indicator of design. He notes that the state function of a quantum mechanical systems can take continuous values and thus assume infinitely many states. Collins then claims, "This means that in Dembski's scheme one could only absolutely eliminate chance for events of zero probability!"

---

[9]A universal probability bound is a level of improbability that precludes specified events below that level from occurring by chance in the observable universe. Emile Borel proposed 10^{-50} as a bound below which probabilities could be neglected universally (i.e., neglected across the entire observable universe). See Emile Borel, *Probabilities and Life*, trans. M. Baudin (New York: Dover, 1962), 28 and Eberhard Knobloch, "Emile Borel as a Probabilist," in *The Probabilistic Revolution*, vol. 1, eds. L. Krüger, L. J. Daston, and M. Heidelberger, 215-233 (Cambridge, Mass.: MIT Press, 1987), 228. In TDI I justify a more stringent universal probability bound of 10^{-150} based on the number of elementary particles in the observable universe, the duration of the observable universe, and the Planck time. See Dembski, *Design Inference*, sec. 6.5.

[10]See Kenneth W. Dam and Herbert S. Lin, eds., *Cryptography's Role in Securing the Information Society* (Washington, D.C.: National Academy Press, 1996), 380, n. 17, where a universal probability bound of 10^{-95} is computed.

Presumably Collins thinks that because quantum systems can produce infinitely many possible events, this means that infinitely many probabilistic resources would be needed to eliminate chance in their explanation. And since infinitely many probabilistic resources coincide with a significance level of zero, my scheme could therefore only eliminate chance for events of probability zero. The problem here is that Collins fails to distinguish between the range of possible events that might occur and the opportunities for a given event to occur or be specified. Probabilistic resources always refer exclusively to the latter. The range of possible events might well be infinite. But this has no bearing on the probabilistic resources associated with a given event in that range.

Finally, Collins points to the sharp increase in probabilistic resources that necessarily follows from an inflationary cosmology, stating that I "simply dismiss inflationary cosmology claiming that the only evidence in its favor is 'its ability to render chance plausible'." Collins here is referring to section 6.6 of TDI titled "The Inflationary Fallacy." He implies that I fail there to take inflationary cosmology seriously. I do take it seriously, and in that section I argue that if one assumes an inflationary cosmology, specified complexity is still how we detect design. I offer there an argument that anticipates Collins's objection, but that Collins in his review fails to engage. I argue that inflating probabilistic resources destroys induction and therefore human rationality. It's this argument I would have liked to see Collins engage.

Although death by counterexample would certainly be a legitimate way for specified complexity to fail as a reliable indicator of intelligence, Collins suggests that there is still another way for it to fail. According to Collins this criterion fails as a rational reconstruction of how we detect design in common life. In TDI I argue that the reconstruction works, especially when Fisher's concept of statistical significance testing is supplemented with the notion of a probabilistic resource. What's more, I've argued in this response to Collins that the likelihood approach fails to provide an adequate account of design. Still more problematic for Collins, however, is that the likelihood approach—even if it could be made to work—can make sense of design only by already presupposing specified complexity.

To see this, take an event that is the product of intelligent design, but for which we haven't yet seen the relevant pattern that makes its design clear to us (take a Search for Extra-Terrestrial Intelligence example in which a long sequence of prime numbers, say, reaches us from outer space, but suppose we haven't yet seen that it is a sequence of prime numbers). Without that pattern we won't be able to distinguish between the probability that this event takes the form it does given that it is the result of chance, and the probability that it takes the form it does given that it is the result of design. Consequently, we won't be able to infer design for this event. Only once we see the pattern will we, on a likelihood analysis, be able to see that the latter probability is greater than the former. But what are the right sorts of patterns that allow us to see that? Not all patterns signal design. What's more, the pattern needs to delimit an event of sufficient improbability (i.e., complexity) for otherwise it can readily be referred to chance. We are back, then, to needing some account of complexity and specification. Thus a likelihood analysis that pits competing design and chance hypotheses against each other must itself presuppose the legitimacy of specified complexity as a reliable indicator of intelligence.

But the problem goes even deeper. Collins suggests that the likelihood approach might be suited for detecting design in the natural sciences in cases where my Fisherian approach to specified complexity breaks down. Thus he mentions the irreducible complexity of biological systems as a state of affairs upon which a design hypothesis confers greater probability than a naturalistic competitor (e.g., the neo-Darwinian synthesis). Irreducible complexity is biochemist Michael Behe's notion. According to Behe, a system is irreducibly complex if it is "composed of several well-matched, interacting parts that contribute to the basic function, wherein the removal of any one of the parts causes the system to effectively cease functioning."[11]

Collins is therefore looking for some property of biological systems upon which the design hypothesis confers greater probability than its naturalistic competitors. This sounds reasonable until one considers such properties more carefully. For Collins specified complexity is

---

[11]Michael Behe, *Darwin's Black Box* (New York: Free Press, 1996), 39.

disallowed because it is a statistical property that depends on Fisher's approach to hypothesis testing, and Collins does not regard Fisher's approach as rationally justified (which, as I've argued, it is once one introduces the idea of a probabilistic resource). What Collins apparently fails to realize, however, is that any property of biological systems upon which a design hypothesis confers greater probability than a naturalistic competitor must itself presuppose specified complexity.

Ultimately, what makes irreducible complexity work is that it is a special case of specified complexity. Behe admits as much in his public lectures whenever points to my work in TDI as providing the theoretical underpinnings for his own work on irreducible complexity. The connection between irreducible complexity and specified complexity is easily seen. The irreducibly complex systems Behe considers require numerous components specifically adapted to each other and each necessary for function. On any formal complexity-theoretic analysis, they are complex. Moreover, in virtue of their function, these systems embody independently given patterns that can be identified without recourse to actual living systems. Hence these systems are also specified. Irreducible complexity is thus a special case of specified complexity. More generally, any property of biological systems that confirms a design hypothesis over non-design alternatives will have to be a special case of specified complexity—if not, such systems would readily be referred to chance. In fine, all the evidence to date suggests that specified complexity—and not a likelihood analysis—is the key to detecting design.