OPEN ACCESS

BIO-complexity

Research Article

A Unified Model of Complex Specified Information

George D. Montañez*
Department of Computer Science, Harvey Mudd College, Claremont, CA, USA

Abstract

A mathematical theory of complex specified information is introduced which unifies several prior methods of computing specified complexity. Similar to how the exponential family of probability distributions have dissimilar surface forms yet share a common underlying mathematical identity, we define a model that allows us to cast Dembski's semiotic specified complexity, Ewert et al.'s algorithmic specified complexity, Hazen et al.'s functional information, and Behe's irreducible complexity into a common mathematical form. Adding additional constraints, we introduce canonical specified complexity models, for which one-sided conservation bounds are given, showing that large specified complexity values are unlikely under any given continuous or discrete distribution and that canonical models can be used to form statistical hypothesis tests, by bounding tail probabilities for arbitrary distributions.

Cite As: Montañez GD (2018) A Unified Model of Complex Specified Information. BIO-Complexity 2018 (4):1-26. **doi:**10.5048/BIO-C.2018.4.

Editor: Robert J. Marks II

Received: June 29, 2018; **Accepted:** November 30, 2018; **Published:** December 14, 2018

Notes: A *Critique* of this paper, when available, will be assigned **doi:**10.5048/BIO-C.2018.4.c.

*Email: gmontane@alumni.cmu.edu

"Life must depend on a higher level of complexity, structure without predictable repetition, he argued."
- James Gleick, *"The Information: A History, a Theory, a Flood"*, ch. 9

INTRODUCTION

Specified complexity, the property of an object being both unlikely and structurally organized, has been proposed as a signifier of design [1–3]. Objects exhibiting specified complexity must be both complex (e.g., unlikely under the relevant probability distribution) and specified (e.g., conform to an independent or *detached* specification). While mathematical methods for measuring specified complexity have changed over the decades [1, 4–6], the underlying components of specification and complexity have remained constant. Possible applications of specified complexity include differentiating malicious network behavior from non-intentional anomalies, identifying genetically-modified material in agricultural products (such as artificially engineered gene sequences), identifying intelligently engineered extra-terrestrial signals, and testing for artificial general intelligence [7]. At its core, it has been proposed as a method of sorting artificial or teleological structures from naturally occurring unintentional ones. Specified complexity has been severely criticized, both for its underlying logic and its various formulations [8–11].

Recent work in specified complexity has proposed using algorithmic information theory as a way of computationally grounding specified complexity, under the name of *algorithmic specified complexity (ASC)* [5, 6]. Ewert and collaborators have shown that under any proposed discrete distribution, sequences exhibiting high levels of ASC must be unlikely [6]. We extend this body of work by generalizing the result of Ewert et al. [6], showing that all specified complexity models of a certain class, which we call *canonical models*, share this property. We re-derive the existing result of Ewert et al. [6] as a corollary to our general theorem, and prove a number of previously unknown results, including a conservation result for a variant of *semiotic specified complexity*, introduced in Dembski (2005) [4]. These results add to a body of work generally known as *conservation of information* theorems, which prove that in many contexts it takes as much information to find an object as you might gain by using that object, rendering explanations that rely on such specialized objects as cases of special pleading. Viewed as a conservation theorem, if one characterizes specified complexity as a form of information signal measurable in bits, the probability under the pro-

posed distribution of encountering a structure exhibiting b or more bits of complex specified information is no greater than 2^{-b}, thus incurring an information cost of at least b bits.

More importantly, this work formally reduces specified complexity testing to a form of statistical hypothesis testing, a view inspired by Dembski (2005) [4] (and preceded by Milosavljević [12]). Accordingly, we demonstrate the connection for canonical specified complexity models, where the model core (called a *kardis*) plays the same role as a p-value under a null distribution. We prove that for such models the null-distribution probability mass of values as extreme as the one observed is upper bounded by the significance level α, which is the crucial property for p-value significance tests. Thus, we formally establish the canonical model as a particular statistical hypothesis test statistic, one which would prove useful in cases where only a likelihood and a specification value are available and the behavior of the distribution outside of the given observation is otherwise unknown.

We begin by reviewing the general structure of specified complexity models and defining their common and canonical model forms, along with providing a general recipe for constructing new specified complexity models. We then review statistical hypothesis testing and the level-α property, formally reducing p-value testing to a form of specified complexity model hypothesis testing. Conservation and level-α test properties of canonical specified complexity models are proven, and instructions for conducting specified complexity model hypothesis tests are also given. Algorithmic specified complexity, semiotic specified complexity, and functional information are reduced to common form, canonical variants are derived, and conservation theorems are proven for each as corollaries to our results for the general canonical specified complexity models. While criticisms of specified complexity [8–11] (and their rebuttals [2, 13–16]) are duly acknowledged, we do not attempt to address them here, as the purpose of this manuscript is to prove the mathematical commonality of several complexity metrics, not to defend them individually.

1. SPECIFIED COMPLEXITY MODELS

A *specified complexity model* is a function of an observation X taking values in some space $\mathcal{X}$, where X is distributed according to probability distribution $p(x)$ (denoted as $X \sim p$). It contains two components, one measuring the likelihood of the observation under p (referred to as the *complexity* term for historical reasons), and another component that measures the *specificity* of the observation, according to a *specification function* $\nu(x)$. This function should be computable or capable of being estimated given x and should be definable independently of the particular observation x, to avoid cherry-picking and fabrications. This can be done by defining the function prior to observing x, but need not be, as long as the function itself is not conditioned on x when computing the value for x. In other words, specification functions of the form $\nu_x(x)$ or $\nu(x|x)$ are invalid. This danger of fabrication is a general vulnerability of statistical hypothesis testing, since one could inappropriately choose the significance level α (acting as the specification function) in response to the observed test statistic value, or even define a test statistic $T_x(X) = c \cdot \mathbb{1}\{X = x\}$ (where $\mathbb{1}\{\cdot\}$ denotes the indicator function, which equals 1 when the argument evaluates to true, and 0 otherwise) that always assigns a large value c to the particular x observed and zero to all other possible observations, leading to the smallest possible critical region, a singleton set containing only the observation itself. If the danger of fabrications forces us to reject specified complexity hypothesis testing then we must also reject many other statistical hypothesis tests, which all suffer from the same vulnerability.

1.1 Common Form and Canonical Models

We begin with a few definitions.

Definition 1 ($\nu(\mathcal{X})$)**.** For any integrable nonnegative specification function $\nu : \mathcal{X} \to \mathbb{R}_{\geq 0}$, define $\nu(\mathcal{X})$ as follows:

$$\nu(\mathcal{X}) := \begin{cases} \int_{\mathcal{X}} \nu(x)dx & \text{if continuous,} \\ \sum_{x \in \mathcal{X}} \nu(x) & \text{if discrete,} \\ \int_{\mathcal{X}} d\nu(x) & \text{in general.} \end{cases} \tag{1}$$

Definition 2 (Common Form and Kardis)**.** For any probability distribution $p(x)$ on space $\mathcal{X}$, any strictly positive scaling constant $r \in \mathbb{R}_{>0}$ and any nonnegative function $\nu : \mathcal{X} \to \mathbb{R}_{\geq 0}$, we define a *common form* model as

$$SC(x) := -\log_2 r \frac{p(x)}{\nu(x)} \tag{2}$$

with *specified complexity kardis* $\kappa(x) = r(p(x)/\nu(x))$.

Definition 3 (Canonical Specified Complexity Model)**.** Any common form model constrained such that $\nu(\mathcal{X}) \leq r$ is a *canonical specified complexity model*.

It should be stated explicitly that the constraint for canonical models is what allows us to bound the probability of observing extreme specified complexity values, and that ideally $\nu(\mathcal{X}) = r$, which allows us to obtain tight bounds. In addition, the scaling constant r should not depend on x. To define a canonical model, it suffices to define the kardis[1] $\kappa(x)$ and ensure that the constraints are satisfied for r, p, and ν. As a useful example, given any kardis $\kappa(x)$ from a canonical specified complexity

[1] From Greek *kardiá* (heart) and Ilocano *kardis* (pigeon pea, which is a small component contained in a larger pod).

model, we can define an augmented kardis (and thus augmented model) that incorporates a significance level term within the model. We do so as follows.

Definition 4 (r_α, $\kappa_\alpha(x)$, and $SC_\alpha(x)$)**.** Given $\kappa(x) = r(p(x)/\nu(x))$ from any canonical specified complexity model and any significance level $\alpha \in (0, 1]$, define

$$\begin{aligned} r_\alpha &:= \frac{r}{\alpha}, \\ \kappa_\alpha(x) &:= r_\alpha \frac{p(x)}{\nu(x)}, \text{ and} \\ SC_\alpha(x) &:= -\log_2 \kappa_\alpha(x) \\ &= SC(x) + \log_2(\alpha). \end{aligned}$$

It is easy to verify that any $SC_\alpha(x)$ thus defined is also a canonical specified complexity model, since $r_\alpha \geq r$ by the condition on α, which implies $\nu(\mathcal{X}) \leq r_\alpha$. Such significance level models will allow us to take arbitrary canonical form specified complexity models and use them directly for α-level hypothesis testing (Theorem 3).

Much like exponential family distributions have diverse surface forms yet can all be reduced to a single common form, existing specified complexity models, such as algorithmic specified complexity and semiotic specified complexity, can also be reduced to the common form of Definition 2. With little (and sometimes no) additional work, we can define canonical variants of these models. Having one canonical form is important for many reasons, not least of which because it allows us to prove general properties of many specified complexity models simultaneously, such as

Corollary 1. *(Level-α Property for Specified Complexity I) Given $X \sim p$ and significance level $\alpha \in (0, 1]$ and any canonical specified complexity model with kardis $\kappa(X)$,*

$$\Pr\left(\kappa(X) \leq \alpha\right) \leq \alpha.$$

We will discuss this result in Section 3, and give a proof in the Appendix.

1.2 Specified Complexity Model Recipe

We now give a recipe constructing new specified complexity models for any finite discrete $\mathcal{X}$. For $x \in \mathcal{X}$, define D as the set of all x such that x is the product of some process of interest. Let $\gamma(x)$ be any bounded function on $\mathcal{X}$ quantifying some aspect of objects in D (such as functional coherence, organization, irreducible complexity, meaning, relative compressibility, etc.), where larger values give stronger evidence of being produced by the process of interest. Thus defined, γ acts as a featurizer for x, outputting a scalar-valued feature correlated to its true label, the label being $\mathbb{1}\{x \in D\}$. We have assumed that larger values of γ indicate stronger correlation with membership, but if anti-correlated, we can define $\gamma(x) := -f(x)$ or $\gamma(x) := (f(x) + \epsilon)^{-1}$, where $f(x)$ is the original function and $\epsilon \geq 0$ is some small constant used to prevent undefined values. If needed, we can then rescale $\gamma(x)$ into the range $[0, 1]$ by using zero-one scaling:

$$\nu(x) := \frac{\gamma(x) - \gamma_{\min}}{\gamma_{\max} - \gamma_{\min}} \tag{3}$$

where we have defined $\gamma_{\min} := \min_{x' \in \mathcal{X}} \gamma(x')$ and $\gamma_{\max}$ analogously. If γ already has the desired range does not need rescaling, simply define $\nu(x) := \gamma(x)$.

Define $r := \nu(\mathcal{X})$ as a normalizing constant (or set r to any real-valued number exceeding this quantity). Given any proposed probability distribution p defined on $\mathcal{X}$ and following Equation 2, we define the generic specified complexity kardis as

$$\kappa(x) := r \frac{p(x)}{\nu(x)} \tag{4}$$

and the new specified complexity model as the negative log of the kardis, namely

$$SC(x) := -\log_2 \kappa(x). \tag{5}$$

Having verified that p is a probability distribution on $\mathcal{X}$, ν is a nonnegative function on $\mathcal{X}$ where $\nu(\mathcal{X}) \leq r$, we see that this recipe produces canonical specified complexity models.

2. STATISTICAL HYPOTHESIS TESTING

Statistical hypothesis testing [17–19] allows one to reject a proposed probability distribution (*null distribution*) as an explanation for a particular observation, whenever observations of that kind have low probability under the distribution. There are two broad classes of statistical tests [20, 21], those based on Neyman-Pearson hypothesis testing and those based on Fisherian p-values. These classes of tests have different forms, origins, and even statistical philosophies, but underlying both is the common idea that when collections of observations have low probability under a null distribution, rejecting that distribution as an explanation for the data leads to beneficial outcomes. Neyman-Pearson tests establish a procedure with good long-term statistical properties, controlling both false-positive and false-negative errors (Type-I and Type-II), whereas Fisher's p-values allow for the rejection of a null distribution without having to consider an alternative hypothesis. Hybrid tests [19, 20] can be formed by first choosing a significance level α, then computing a p-value and rejecting the null distribution if it is below α. Such tests, if performed consistently and repeatedly, will reject the null hypothesis with probability not exceeding α in the cases when it actually holds, so they are often used as a "best-of-both-worlds" approach in practice.[2] In general, a p-value is the smallest α at which

[2]However, others view this as a "worst-of-both-worlds" method. See, for example, the discussion in Lew [20].

we would reject the null hypothesis [22]. Reporting a test result using a p-value allows the reader to compare that value to any significance level they deem appropriate, leaving to them the decision to reject or not reject based on the observed data [19]. Because of their popularity and utility, we will focus our discussion of statistical hypothesis testing to p-values and the level-α tests derived from them, as described in Casella and Berger (Section 8.3.4) [19].

The reader should be aware, however, of the statistical controversy surrounding p-values (and statistical hypothesis testing more generally [23]), such as abuses like "p-hacking" and the potential for high false discovery rates even under ideal conditions [24]. The false discovery rate measures the proportion of times one is wrong when rejecting the null hypothesis, even if one rarely chooses to reject it. While the α-level controls how often you reject the null hypothesis when it actually holds (thus controlling Type-I error), it could be the case that you almost never reject when it **does not** hold (and you should), and thus the ratio comparing the number of times you *incorrectly* reject it to the total number of times you reject it overall could be quite large even for small α-levels. P-values are therefore imperfect and ubiquitous tools, with the potential for abuse or mishandling; they should be handled with the care and attention required for tool use in general.

2.1 P-Values

Given a set of observations $X^n = \{X_1 = x_1, \ldots, X_n = x_n\}$, we need some way of measuring how likely observations of this type are under the proposed null distribution. If the outcome is probable under the distribution, we have no reason to reject it as an explanation for the data; if the observed data (and outcomes like it) are sufficiently unlikely under the null distribution, then we reject it as an explanation for the data. To do so, one needs to first define a *test statistic* T, which is a real-valued function of X^n, and a *significance level* α, which defines the level of improbability for which we will reject the null hypothesis. The test statistic is typically some statistic of the observed data (hence the name), such as the average of the observations, the number of heads in a series of coin flips, or some other property of the data. Although typically $n > 1$ (i.e., the test statistic is a function of more than just one observation), for simplicity we will consider the special case of $n = 1$, which will also allow for easier comparison with specified complexity models later.

Different hypothesis tests lead to different definitions of "extreme" test statistic values. For some tests, large $T(X)$ are considered extreme and will lead to rejection, whereas in other cases small $T(X)$ will do so. Other tests, such as two-tailed tests, reject large magnitude $|T(X)|$ values. A p-value, denoted $p_{\text{val}}(x)$, is defined as the probability (under the null distribution) of the set of all observations that have test statistics at least as extreme as the value $T(x)$ observed. Assuming that large test statistic values are extreme, they can be defined formally as follows:

Definition 5 (P-Value for Large Extreme Test Statistics)**.** For $X \sim P_\theta$,

$$p_{\text{val}}(x) = P_\theta \left(T(X) \geq T(x) \right).$$

Similar definitions can be given for small extreme statistic values, or large magnitude extreme values.

Under p-value statistical testing, the null distribution is rejected as an explanation for the data whenever the p-value is less than the specified significance level. The following property holds for p-value statistical tests, whenever testing for large or small extreme test statistic values, since they are distributed according to a Uniform-$(0, 1)$ distribution whenever the test statistic has a continuous distribution [22]:

Definition 6 (Level-α Property)**.** Given $X \sim P_\theta$, we say the *level-α property* holds for random variable $Y := f(X)$ if

$$\Pr(Y < \alpha) \leq \alpha.$$

We note that p-values are themselves random variables, being functions of the random variable X. By the level-α property, the probability under the null distribution of seeing a p-value of level α is therefore no greater than α. P-values for which the level-α property holds are called *valid* [19]. Because small p-values are by definition unlikely under the proposed distribution, observation of them may justify rejecting the distribution in question as an explanation for the data. Surprisingly, we can show this level-α property also holds for canonical specified complexity models, and so the same set of implications follow for such models.

2.2 P-Value Specified Complexity Models

A p-value statistical test can be cast into specified complexity model form by simple algebraic manipulation. Begin with any such test that rejects a null distribution when the p-value is smaller than α, namely,

$$p_{\text{val}}(X) < \alpha. \tag{6}$$

If we rearrange the sides of the inequality and take negative logs, we obtain

$$-\log_2 \left(\frac{p_{\text{val}}(X)}{\alpha} \right) > 0, \tag{7}$$

which is a common form specified complexity model test (letting $r = 1$), though not a canonical one (since α is not normalized over $\mathcal{X}$). However, for every test of

this form there exists an equivalent canonical specified complexity model test on a new two-element space, $\tilde{\mathcal{X}}$, defined as follows.[3] Let $\tilde{\mathcal{X}} := \{1, 0\}$, where the first element is identified with the extreme region of $\mathcal{X}$, and the second, with the complement of that region. The null distribution, test statistic, and original observation induce the following probability distribution on $\tilde{\mathcal{X}}$:

$$\tilde{p}(\tilde{x}) := \begin{cases} p_{\text{val}}(X) & \text{if } \tilde{x} = 1, \\ 1 - p_{\text{val}}(X) & \text{if } \tilde{x} = 0. \end{cases} \tag{8}$$

Since the p-value is the probability of observing such extreme values under the null distribution, the event of observing $\tilde{X} = 1$ has this same probability under $\tilde{p}$. Thus, the original test is transformed into an equivalent test of a weighted coin flip (for a Bernoulli random variable $\tilde{X}$) to see if the outcome of heads is likely given its weight. For this new test the null distribution being directly tested is no longer the original distribution, but is now the induced Bernoulli distribution, $\tilde{p}$, with parameter $p_{\text{val}}(X)$. This new distribution serves as the complexity component of the specified complexity model. It is a proxy for the original null distribution, and rejecting it results in rejecting the original distribution as well (and vice versa).

The α term can take one of two interpretations: it can be interpreted as either a specification function, which assigns probability mass α to $\tilde{X} = 1$, or as a number of replicational pseudo-trials $r = 1/\alpha$, which are additional chances of seeing the event in question [4]. We will discuss both interpretations shortly.

Returning to the original non-canonical p-value specified complexity model on $\mathcal{X}$, it follows directly from the level-α property that

$$\Pr\left(-\log_2\left(\frac{p_{\text{val}}(X)}{\alpha}\right) > 0\right) \leq \alpha. \tag{9}$$

Furthermore, we can prove a conservation result for these models, for any desired α-level:

Theorem 1. *(Conservation of P-Value Specified Complexity) Under the same conditions as Definition 6, for $X \sim P_\theta$ and $\alpha \in (0, 1]$,*

$$\Pr\left(-\log_2\left(\frac{p_{val}(X)}{\alpha}\right) > b\right) \leq \alpha 2^{-b}.$$

Two-mass Interpretation Consider the ratio $p_{\text{val}}(X)/\alpha$. As we have seen, the $p_{\text{val}}(X)$ term is the probability mass of the extreme region determined by the test statistic and null distribution. The α term can be interpreted as a probability mass, being the mass under a different distribution (a *specification* distribution) for the same observation of "heads," using the weighted coin analogy. We can associate α with a default level of "fit" under the specification distribution. When the heads outcome has a higher probability under this alternative distribution than under the null distribution, we reject the null distribution in favor of the alternative. The test condition $p_{\text{val}}(X) < \alpha$ is equivalent to

$$\frac{p_{\text{val}}(X)}{\alpha} < 1, \tag{10}$$

so we are essentially performing something like a modified likelihood ratio test under this interpretation, with the caveats that we are not considering maximization over a constrained and an unconstrained class of parameters, nor are we allowing for general c cutoff values.

Pseudo-trials Interpretation A second interpretation of the α term involves moving it from underneath the p-value, to give a quantity $r = 1/\alpha \geq 1$. Returning again to the weighted coin analogy, we can view r as the total number of coin flips we have, which gives us more chances of seeing the event in question (namely, a flip that lands on heads). It is a number of pseudo-trials, which can be thought of as hallucinated coin flips. We have observed heads, and we want to know if that outcome was likely given the number of flips r and the weight on heads (probability $p_{\text{val}}(X)$). Whenever r is an integer, we have

$$\frac{p_{\text{val}}(X)}{\alpha} = r \cdot p_{\text{val}}(X) = \mathbb{E}[Y] \tag{11}$$

where $Y \sim \text{Binom}(r, p_{\text{val}}(X))$ (i.e., Y is a binomially-distributed random variable, with parameters $N = r$ and $p = p_{\text{val}}(X)$, and with expected number of heads equal to $Np = r \cdot p_{\text{val}}(X)$). Since we reject the null hypothesis whenever $p_{\text{val}}(X)/\alpha < 1$, we see that the test is equivalent to testing if the expected number of heads is fewer than one under the proposed distribution. If it is, we reject the null distribution as an explanation for the data, since we observed a heads outcome.

Under either interpretation, we can see that p-value tests are equivalent to specified complexity model hypothesis tests. Furthermore, for any specific p-value test there exists an equivalent two-element *canonical* specified complexity model test. Taking the two-mass interpretation, define $\tilde{\nu}(\tilde{X} = 1) := \alpha$, and we see that testing

$$p_{\text{val}}(X) < \alpha \tag{12}$$

is equivalent to testing

$$-\log_2\left(\frac{\tilde{p}(\tilde{X} = 1)}{\tilde{\nu}(\tilde{X} = 1)}\right) > 0, \tag{13}$$

[3] I have been told this formulation formally corresponds to a *coupling* [25].

on the two-element space $\tilde{\mathcal{X}}$, with $\tilde{p}(\tilde{X}=1)$ implicitly a random variable of X. Under the pseudo-trials interpretation, define $\tilde{\nu}(\tilde{X}=1) := 1$ and $r := 1/\alpha$, so that we obtain the equivalent test

$$-\log_2\Big(r\tilde{p}(\tilde{X}=1)\Big) > 0, \tag{14}$$

again defined in terms of a random variable $\tilde{X} \sim \tilde{p}$.

For either definition of $\tilde{\nu}$, letting $\tilde{\nu}(\tilde{X}=0) = 1 - \tilde{\nu}(\tilde{X}=1)$, we have $\tilde{\nu}(\tilde{\mathcal{X}}) = 1 \leq r$. Because of this, we see that both specified complexity models in the new tests are canonical specified complexity models. Thus, p-value testing is reducible to canonical specified complexity model testing on a new space $\tilde{\mathcal{X}}$.

3. MAIN RESULTS

Having defined canonical specified complexity models and shown their relationship to p-value statistical tests, we next state a series of results for this family of models, with full proofs given in the Appendix. We demonstrate that one can upper bound the probability of observing extreme specified complexity values for arbitrary probability distributions, without needing the distribution in question to have a tractable analytical or parametric form, in contrast to traditional p-values. This feature is a consequence of simultaneously observing a low probability value under the distribution and a high specification value under the specification function: since the specification values are normalized, few elements can have large values. Although many elements can have low probability values (thus making the occurrence of observing any such low-probability event probable, given enough of them), few can have low probability while being highly specified. Because specified complexity is an information quantity (namely, it is logarithmic in form), the resulting bound takes the form of a one-sided conservation result.

3.1 Canonical Conservation Bound

Theorem 2 (Conservation of Canonical Specified Complexity)**.** *Let $p(x)$ be any discrete or continuous probability measure on space $\mathcal{X}$, let $r \in \mathbb{R}_{>0}$ be a scaling constant, and let $\nu : \mathcal{X} \to \mathbb{R}_{\geq 0}$ be any nonnegative integrable function where $\nu(\mathcal{X}) \leq r$. Then*

$$\Pr\left(-\log_2 r\frac{p(X)}{\nu(X)} \geq b\right) \leq 2^{-b}, \tag{15}$$

where $X \sim p$.

Remark. Any canonical specified complexity model will satisfy the conditions of the above theorem by definition, so it holds for all such models.

Let us pause to understand the import of Theorem 2. Given an observed specified complexity value $SC(x)$, the key question becomes:

> *What is the probability of observing a specified complexity value at least as extreme ('special') as $SC(x)$?*

Theorem 2 allows us to upper bound this probability by $2^{-SC(x)}$. This result is somewhat surprising, since it holds for arbitrary discrete or continuous distributions with arbitrary tail behavior, without making any strong parametric assumptions on the form of p. To gain intuition as to why specification functions bound tail behavior, consider what it means to have a specified complexity value at least as extreme as $SC(x)$. By Definition 3, we have

$$SC(x) = -\log_2 r\frac{p(x)}{\nu(x)}. \tag{16}$$

Thus, for $x' \in \mathcal{X}$, $SC(x') \geq SC(x)$ implies $p(x') \leq p(x)$ or $\nu(x') \geq \nu(x)$ (or both). While many x' elements in $\mathcal{X}$ can have $p(x') \leq p(x)$, which could even make the occurrence of *some* small probability event all but certain (as is true for uniform distributions on large spaces), it cannot be the case that an arbitrarily large proportion may have $\nu(x') \geq \nu(x)$, since $\nu(\mathcal{X}) \leq r$. Thus, increasing the proportion of large ν values increases the scaling constraint r, which implies a larger kardis $rp(x)/\nu(x)$ and smaller $SC(x)$ value, which controls the tightness of the bound.

If we try to adversarially devise p and ν functions so as to violate the bound, we can see that the integration constraint still protects us, through the following informal argument. For any observed $SC(x)$ value from a canonical specified complexity model, we can attempt to choose p and ν such that the probability

$$\Pr(SC(X) \geq SC(x)) \tag{17}$$

is larger than the bound given by Theorem 2. To accomplish this, we must make this probability large (i.e., a large proportion of p's mass must reside on regions where the SC values exceed $SC(x)$) and the bound tight (i.e., the observed $SC(x)$ must be large). This gives us the best chance of violating the bound. To address the first goal, we can make the probability of the set of extreme values as large as possible by assuming that $x = \text{argmax}_{x' \in \mathcal{X}}\, p(x')$ and that such an argmax exists. Then, $p(x') \leq p(x)$ for all $x' \in \mathcal{X}$. Because of this, ν can remain constant or decay with p, as long as $p(x')/\nu(x') \leq p(x)/\nu(x)$, which will guarantee $SC(x') \geq SC(x)$ for all $x' \in \mathcal{X}$, thus maximizing the probability $\Pr(SC(X) \geq SC(x))$ by making it equal to 1.

To address the second goal, we assume the observed $SC(x)$ is as large as possible under the conditions already established. In such a case, ν will have contours that decay at exactly the same rates as the contours of p, by the following argument. If ν decayed any slower (i.e.,

there were some $\nu(x')$ larger than absolutely necessary) r would be larger than absolutely necessary, making $SC(x)$ smaller than need be, and thus violating our assumption that $SC(x)$ is as large as possible. If ν decayed any faster than p, there would exist some $x' \in \mathcal{X}$ with positive measure such that $p(x')/\nu(x') > p(x)/\nu(x)$, which would contradict our assumption that the probability $\Pr(SC(X) \geq SC(x))$ is maximized. Because the contours of ν follows those of p exactly at all points, this makes ν a scaled copy of p, namely, $\nu(x) = cp(x)$. Maximizing $SC(x)$ requires making $rp(x)/\nu(x)$ as small as possible. Either we can minimize r or $p(x)/\nu(x)$ or both. To minimize r, set $r = \nu(\mathcal{X})$, its smallest possible value. To minimize the ratio $p(x)/\nu(x)$ requires maximizing the separation between values of p and ν, which means maximizing the scalar multiplier c. The multiplier must be nonnegative, since ν is a nonnegative function (as is p). The multiplier must also not equal 0, since $SC(x)$ would be undefined in that case. For any positive multiplier c less than r,

$$r\frac{p(x)}{\nu(x)} = r\frac{p(x)}{cp(x)} = \frac{r}{c} > 1. \tag{18}$$

This would imply a negative $SC(x)$ for all $x \in \mathcal{X}$, making the bound from Theorem 2 exceed 1, thus trivially satisfying it. Therefore, we can assume $c \geq r$. Furthermore, our integration constraint on ν implies the scalar multiplier must not exceed r, since p integrates to one, which implies

$$r \geq \nu(\mathcal{X}) = cp(\mathcal{X}) = c. \tag{19}$$

Thus, the scalar multiplier must equal r. This implies $SC(x) = -\log_2(r/r) = 0$. By Theorem 2, the bound corresponding to $SC(x) = 0$ is $\Pr(SC(X) \geq 0) \leq 2^0 = 1$, so our bound is again satisfied even under this adversarial case.

The addition of specification is what allows us to control these probabilities. Considering probabilities in isolation is not enough. While unlikely events can happen often (given enough elements with low probability), specified unlikely events rarely occur. This explains why even though every sequence of one thousand coin flips is equally likely given a fair coin, the sequence of all zeroes is a surprising and unexpected outcome whereas an equally long random sequence of heads and tails is not. Specification provides the key to unlocking this riddle.

3.2 Level-α Property for Specified Complexity Tests

Several results follow from Theorem 2. One of these we have already encountered, namely that canonical kardii, like p-values, possess the level-α property that allows them to bound the mass of extreme values under distribution p. The following two corollaries show how this applies to kardii and specified complexity values, respectively, with proofs given in the Appendix. The second result gives us a simple way to construct hypothesis tests from canonical specified complexity models.

Corollary 1 (Level-α Property for Specified Complexity I)**.** *Given $X \sim p$ and significance level $\alpha \in (0, 1]$, let $\kappa(x)$ be the kardis from any canonical specified complexity model. Then*

$$\Pr(\kappa(X) \leq \alpha) \leq \alpha.$$

Corollary 2 (Level-α Property for Specified Complexity II)**.** *Let $SC(x)$ be any canonical specified complexity model. Then for $X \sim p$ and significance level $\alpha \in (0, 1]$,*

$$\Pr(SC(X) \geq -\log_2 \alpha) \leq \alpha.$$

Corollary 2 gives us a practical hypothesis test for all canonical specified complexity models, where we reject the null hypothesis whenever test statistic $T(x)$ (defined as $SC(x)$) exceeds $-\log_2 \alpha$. Table 1 gives cutoff values for common α levels. As can be seen from the table, observing specified complexity values exceeding ten bits corresponds to an α-level of less than 0.001.

Table 1: Test statistic cutoff by α-level.

α	$T(x)$
.1	3.33 bits
.05	4.33 bits
.01	6.65 bits
.001	9.97 bits
.0001	13.29 bits
α	$-\log_2 \alpha$ bits

Augmented canonical specified complexity models also possess the level-α property, and thus can also be used to define hypothesis tests. We first state a more general result (with proof given in the Appendix) that allows us to demonstrate the property of such models as a direct corollary.

Theorem 3 (α-Exponential Bound)**.** *Given $X \sim p$, significance level $\alpha \in (0, 1]$, and any augmented canonical specified complexity model $SC_\alpha(x)$ as defined in Definition 4,*

$$\Pr(SC_\alpha(X) \geq b) \leq \alpha 2^{-b}.$$

Corollary 3 (Level-α Property for Specified Complexity III)**.** *Let $SC_\alpha(x)$ be any augmented canonical specified complexity model as defined in Definition 4. Then for $X \sim p$ and significance level $\alpha \in (0, 1]$,*

$$\Pr(SC_\alpha(X) \geq 0) \leq \alpha.$$

Proof. Let $b = 0$ and invoke Theorem 3. □

In other words, the probability (under p) of observing positive values for an augmented canonical specified complexity model with significance level α is no greater than α.

3.3 Common Form Model Bound

Although by Definition 3 all canonical specified complexity models are normalized such that $\nu(\mathcal{X}) \leq r$, we can easily obtain a tail-bound for unnormalized common form models using Theorem 2, with a full proof given in the Appendix.

Corollary 4 (Common Form Model Bound)**.** *For any common form model SC, where r and $\nu(\mathcal{X})$ are as in Definitions 1 and 2, we have*

$$\Pr\left(SC(X) \geq b\right) \leq 2^{-b} r^{-1} \nu(\mathcal{X}).$$

In the special case of $r = 1$, this simplifies to

$$\Pr\left(SC(X) \geq b\right) \leq 2^{-b} \nu(\mathcal{X}).$$

With the above corollary one can make use of non-canonical specified complexity models, since the size of the specification function can be controlled directly in the bound. This gives a result that may be applied to previously published specified complexity models without needing to first define canonical form variants.

3.4 Combining Models

Given that multiple specified complexity models exist, and others may be defined using the recipe in Section 1.2, one naturally wonders if two or more canonical models can be combined to produce yet another canonical model. Namely, given specified complexity models $SC_1(x), \ldots, SC_m(x)$, can we form a combined model $SC'(x)$ subject to the bounds given previously? Assume all models share a common $p(x)$ function, so that there is a single random variable $X \sim p$ for all SC_i. There are two simple ways to create a combined model: 1) combine specification functions $\nu_i(x)$ into a hybrid specification function $\nu'(x)$ (with which a canonical kardis can be defined), or 2) average the specified complexity models together. We give bounds for both types of combined models, where the first results in a new canonical model and the second is no longer a canonical model since the product probability function on the single random variable X no longer integrates to 1 in general. For the second type of combined model, we give a looser bound on the probability of observing extreme combined specified complexity values under distribution p.

Theorem 4 (Combined Models I)**.** *Let $SC_1(x), \ldots, SC_m(x)$ be canonical specified complexity models with corresponding specification functions $\nu_1(x), \ldots, \nu_m(x)$, scaling constants $r_1, \ldots, r_m$, and common probability function $p(x)$. Define a set of mixture variables*

$$\Lambda = \{\lambda_i : i = 1, \ldots, m, 0 \leq \lambda_i \leq 1\},$$

such that $\sum_{i=1}^{m} \lambda_i = 1$. Then for any such Λ and defining $r' := \max_i r_i$,

$$SC'(x) := -\log_2 r' \frac{p(x)}{\sum_{i=1}^{m} \lambda_i \nu_i(x)}$$

is a canonical specified complexity model.

Theorem 5 (Combined Models II)**.** *Let $SC_1(x), \ldots, SC_m(x)$ be canonical specified complexity models sharing a common probability function $p(x)$. Define a set of mixture variables*

$$\Lambda = \{\lambda_i : i = 1, \ldots, m, 0 \leq \lambda_i \leq 1\},$$

such that $\sum_{i=1}^{m} \lambda_i = 1$. Then for any such Λ,

$$\Pr\left(\sum_{i=1}^{m} \lambda_i SC_i(X) \geq b\right) \leq m 2^{-b}.$$

4. ADDITIONAL RESULTS

In this section we show how several existing specified complexity models, such as algorithmic specified complexity [5], irreducible complexity [26], and functional information [27], can be recast into common form, with close variants defined as canonical specified complexity models. Doing so allows us to prove conservation results bounding the probability of observing extreme values under those models. As canonical models, they can also be used to form statistical hypothesis tests as in Section 3.2. Demonstrating the commonality of such a diverse group of models highlights the utility of the canonical model abstraction, and suggests that any model seeking to solve similar problems will converge to a form similar to canonical specified complexity.

4.1 Semiotic Specified Complexity

We begin with a form of specified complexity outlined in Dembski (2005) [4], which was proposed as an improvement and mathematical refinement to the model suggested in Dembski (1998) [1]. Although recent work in algorithmic specified complexity seems aimed at superseding this model [5, 6], we include it here for historical purposes. As shown in the Appendix, semiotic specified complexity is a common form model (though not canonical), so we can invoke Corollary 4 and obtain the following bound, along with a conservation bound on a closely related canonical variant.

Corollary 5 (Bounds on Semiotic Specified Complexity (SSC))**.** *Let $p(x)$ be any probability distribution on a finite discrete space $\mathcal{X}$ and define*

$$SSC(x, s) := -\log_2 \left[10^{120} \varphi_s(x) p(x)\right]$$

where $\varphi_s(x)$ is the number of patterns available to semiotic agent s which have descriptive complexity no greater than that of x. Then for $X \sim p$, we have

$$\Pr\left(SSC(X,s) \geq b\right) \leq 2^{-b} 10^{-120} \sum_{x \in \mathcal{X}} \varphi_s(x)^{-1}.$$

Additionally, whenever $|\mathcal{X}| \leq 10^{120}$, the following conservation bound holds:

$$\Pr\left(SSC(X,s) \geq b\right) \leq 2^{-b}.$$

4.2 Algorithmic Specified Complexity

As mentioned previously, the conservation result from Ewert et al. [6], which initially inspired the work here, follows as a corollary of the general result given in Theorem 2. Algorithmic specified complexity is a canonical specified complexity model, allowing us to apply Theorem 2 directly.

Corollary 6 (Conservation of Algorithmic Specified Complexity (ASC))**.** *For discrete space $\mathcal{X}$, define*

$$ASC(x,c,p) := -\log_2 p(x) - K(x|c),$$

where $K(x|c)$ is the conditional Kolmogorov complexity of x given context c. Then for $X \sim p$,

$$\Pr\left(ASC(X,c,p) \geq b\right) \leq 2^{-b}.$$

Proof. Let $r = 1$ and $\nu(x) = 2^{-K(x|c)}$, so that

$$ASC(x,c,p) = -\log_2 r\frac{p(x)}{\nu(x)}. \tag{20}$$

Because ν is a probability measure on $\mathcal{X}$ (being a universal algorithmic probability distribution), this implies $\nu(\mathcal{X}) = 1$, making it less than or equal to r. The result then follows from Theorem 2. □

Remark. Under the conditions of Corollary 6, given any test statistic $T(x) = ASC(x,c,p)$, we obtain

$$\Pr\left(ASC(X,c,p) \geq T(x)\right) \leq 2^{-T(x)}. \tag{21}$$

Thus, one can reject the null hypothesis at all significance levels $\alpha \geq 2^{-T(x)}$.

4.3 Defining Quantitative Irreducible Complexity

Irreducible complexity as defined by Behe [26] describes

> ...a single system which is composed of several well-matched, interacting parts that contribute to the basic function, and where the removal of any one of the parts causes the system to effectively cease functioning.

In his book [26], Behe proposes irreducible complexity as an indicator of intentional design by intelligent agents. Behe further clarified the definition, adding [28]:

> An irreducibly complex evolutionary pathway is one that contains one or more unselected steps (that is, one or more necessary-but-unselected mutations). The degree of irreducible complexity is the number of unselected steps in the pathway.

and

> Demonstration that a system is irreducibly complex is not a proof that there is absolutely no gradual route to its production. Although an irreducibly complex system can't be produced directly, one can't definitively rule out the possibility of an indirect, circuitous route. However, as the complexity of an interacting system increases, the likelihood of such an indirect route drops precipitously.

The somewhat qualitative nature of the original definition has typically led to an 'all-or-nothing' view of irreducible complexity [29, 30]. It can be argued, however, that the original formulation more naturally suggests a continuum of 'irreducibility.' We attempt to provide such a quantitative, non-binary formulation of irreducible complexity here. We make no claim that this is an optimal or otherwise definitive model, but is offered simply as a relevant application of the canonical view of specified complexity.

Within the definitions and clarifications above, a few zero-one indicators of irreducible complexity are present, such as the need for 'several' (more than two) parts, the parts being 'well-matched' and 'interacting', and all parts contributing necessarily to the basic (core) function of the system such that the removal of any of them also removes that functionality. In addition to these indicator features, there are scalar-valued features, such as the complexity of the system (e.g., the number of 'unselected steps', length of 'circuitous route[s]'), the count of interacting components, and the functional specificity of such 'well-matched' interacting parts. Thus, irreducible complexity contains a complexity component (complex system with several interacting parts, all routes must be circuitous with unselected steps) and a specification component (parts contribute to an identifiable function, the core is irreducible, all parts are well-matched). The first aspect ensures irreducibly complex systems are unlikely under proposed distributions governing the emergence of biological systems, and the second captures aspects of human engineered systems, correlating irreducible complexity with intentional design.

Following Behe, let us say a system is *irreducible* if it contains at least three well-matched components that contribute necessarily to a core function, such that removal or mutilation of any of the components causes the system to cease functioning with respect to the core function. More formally, given a finite discrete space

$\mathcal{X}$, a modular system $x \in \mathcal{X}$ containing components $y_1, \ldots, y_\ell$ (with each y_i being drawn from a shared space $\mathcal{Y}$ of possible options), and a core function $g : \mathcal{X} \to [0, 1]$, let

$$C_g(x) = \{y_1, \ldots, y_k\} \tag{22}$$

be defined as the *irreducible core* of the system relative to function g, such that $g(C_g(x)) > 0$ and $g(C_g(x) \setminus \{y_i\}) = 0$ for all $y_i \in C_g(x)$ (i.e., the core can perform the function and removing any component from the irreducible core causes the function to cease, which is identified here by a functional value of zero).[4] If neither x nor any subset of x can perform the function, the irreducible core is simply the empty set. Our definition considers the components of x as they are; we do not allow for modifications to the components (such as fusing two components together or altering one so that it performs the role of two components while eliminating the second). We implicitly assume interactions among core components remain as they were in x, but will consider changes in interaction patterns that might preserve core function when later measuring interaction specificity. We assume function g is scaled so that maximum function occurs when $g(x) = 1$ and $g(x) = 0$ corresponds to no detectable core function activity. To make estimation easier, the space $\mathcal{X}$ might be restricted to some manifold of the true space, such as the subspace of systems containing k components or fewer (where k corresponds to the number of core components in x).

For a configuration x with irreducible core $C_g(x)$, we can derive the following features from $C_g(x)$ to form specification function ν. Let k be the number of components in the core, which will act as a scalar multiplier for our function, so that when there are many necessary components, the specification function will take larger values. Let m be the number of *functionally relevant* interactions between the components of $C_g(x)$, where an interaction is functionally relevant if removing it causes a change in the value of $g(x)$. Interactions that do not contribute to or otherwise cause detectable differences in the core function are functionally irrelevant, and are thus excluded.

In addition to the number of interacting components, our specification function will also include terms for part specificity and interaction specificity, corresponding to the 'well-matched' qualifier of Behe's original definition. Well-matched components of a system are often an indication of intentional purpose, allowing humans to perceive a coherent whole from many separate pieces, where the wholeness suggests a higher-level intentional cohesion or concept. Figure 1 shows how strong interface matching

[4]Note, this implicitly assumes closure of the irreducible core operation within the space $\mathcal{X}$, such that $\mathcal{X}$ also contains any system that could be constructed from x's components or their alternatives.

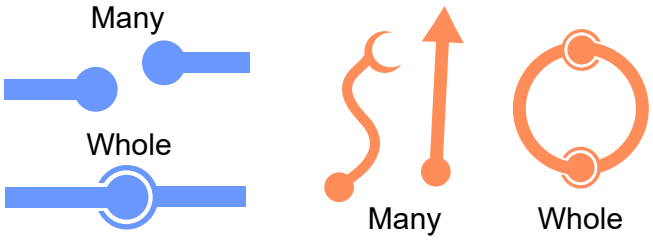

Figure 1: Whole vs. Many. Tight interface coupling suggests connection and coherent wholeness. When interfaces are not tightly matched, we perceive parts as individual items rather than a unified, coherent whole. **doi:**10.5048/BIO-C.2018.4.f1

is suggestive of cohesion. In the figure, the left hand side contains two blue-tinted sub-figures. At the top, we perceive two separate pieces, whereas the image on the bottom is perceived as a unified whole, even though it is also comprised of two separate pieces, albeit with well-matched interfaces connecting them. On the right, there is another contrasting pair of peach-tinted images, where the pieces without matching interfaces are quickly perceived to be separate entities (Many), whereas the pieces sharing matching interfaces are perceived as a single ring (Whole), rather than as the two separate pieces that comprise it. In both cases, the well-matched interfaces strongly suggest cohesion, which is one of the ways humans identify functional wholes rather than collections of individual pieces. At minimum, they signal a correlation that might be the result of a common cause or top-down, intentional coordination (especially when resulting in functional systems) [31]. Furthermore, well-matched components contribute to the improbability of a configuration by limiting the number of viable alternative options. Thus, they contribute simultaneously to the complexity and specificity of a system.

For each component y_i in core $C_g(x)$, let $w(y_i)$ be the proportion of components from the space $\mathcal{Y}$ of possible options that maintain core function when replacing y_i in the system, namely,

$$w(y_i) := \frac{|\{a : a \in \mathcal{Y}, g(C_g(x_{y_i,a})) > 0\}|}{|\mathcal{Y}|}, \tag{23}$$

where $x_{y_i,a}$ denotes the configuration that replaces component y_i with component a, and $\mathcal{Y}$ denotes the subspace of possible alternate components (e.g., a component library, lexicon, or alphabet). Define

$$\gamma_1(x) := \prod_i \frac{1}{w(y_i)} \tag{24}$$

as the product of the reciprocal of all $w(y_i)$ proportions, so that when parts are very specific within the system (not many alternatives maintain function), the $w(y_i)$ proportions are small and thus the γ_1 value is large. Under the conditions that $y_i \in \mathcal{Y}$ for all y_i and assuming a finite

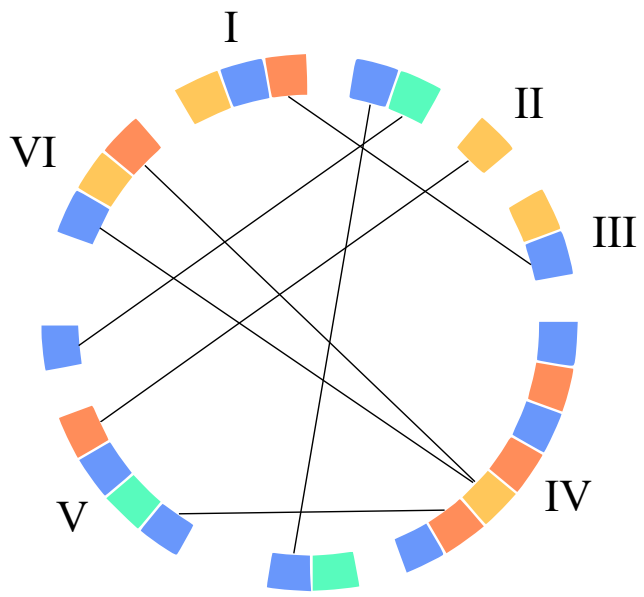

Figure 2: Components and their relevant interaction surfaces. Objects I-VI are components of a system, and each subdivision (colored box) corresponds to a functionally relevant interaction surface, such that component I has three such surfaces, component II has one, and component V has an interaction surface cardinality of four. Each surface can possibly interact with one or more other surfaces, represented as links (black lines) between surface nodes.
doi:10.5048/BIO-C.2018.4.f2

and discrete $\mathcal{Y}$ space, the proportions $w(y_i)$ will always be nonzero, and thus γ_1 will always be well-defined.

To measure interaction specificity, we consider for each y_i the functionally relevant interaction surfaces and the proportion of possible interactions that maintain core function. Let the *interaction surface cardinality* be the number of distinct ways a component can interact with other components in the space such that the interactions have different effects on the core function. For example, if a component can be attached to another at n possible points, but these only result in two possible functional level outcomes (say, high or low function), then the interaction surface cardinality is two (and not n). Thus, the interaction surface cardinality measures the number of interaction behaviors of a component, rather than merely the number of raw possibilities for interaction. Figure 2 represents this graphically, with components I-VI each having some number of functionally relevant interaction surfaces (behaviors), where a 'surface' can be thought of as some (possibly noncontiguous) region which results in similar functional performance when attached to any of the points within that region. In the figure, component I has three distinguishable interaction surfaces (and thus an interaction surface cardinality of 3), and interacts with component III, itself having an interaction surface cardinality of 2. Interactions are represented as black links within the figure, joining interaction surfaces of different components, which are represented as colored boxes.

For each interaction j, let u_j be the proportion of all alternative interactions (including itself) which do not eliminate core function, namely,

$$u_j := \frac{|\{l : l \in B, g(C_g(x_{j,l})) > 0\}|}{|B|}, \tag{25}$$

where B is the set of all possible interactions and $x_{j,l}$ represents system x where interaction j has been replaced by a new interaction l, connecting two other surfaces.[5] Thus, u_j measures the proportion of possible links that can be swapped for interaction j without eliminating core function. One can change the way two components interact with each other (by altering which interaction surfaces are connected), swap one of the components for a different one, or change the pair of interacting components for a different pair. All these possibilities are captured by u_j. Define

$$\gamma_2(x) := \prod_j \frac{1}{u_j} \tag{26}$$

as the product of the reciprocal of all u_j proportions, so that when interactions are very specific this value will be large.

Given k, m, $\gamma_1(x)$, and $\gamma_2(x)$ we define the *irreducible specificity* function ν as follows.

Definition 7 (Irreducible Specificity Function)**.**

$$\nu(x) := \mathbb{1}\{k \geq 3\} \cdot k^{\lambda_1} \cdot m^{\lambda_2} \cdot \gamma_1(x)^{\lambda_3} \cdot \gamma_2(x)^{\lambda_4}$$

where k is the number of irreducible core components, m is the number of functionally relevant interactions among core components, $\gamma_1(x)$ is the part specificity of core components, $\gamma_2(x)$ is the interaction specificity for core component interactions, and $\mathbb{1}\{k \geq 3\}$ denotes the indicator function that equals 1 when $k \geq 3$ and 0 otherwise. The variables $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \in [0, \infty)$ are hyperparameters controlling the relative contributions of each element to $\nu(x)$. Default values are $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 1$.

Definition 8 (Irreducible Specificity Scaling Constant r)**.**

$$r := \sum_{x' \in \mathcal{X}} \nu(x')$$

where ν is the irreducible specificity function of Definition 7, 9, or 10.

[5] It should be stated explicitly that the connections are abstract interactions, not necessarily physical attachments. We use the language of physical attachments and surfaces as visualization tools, not as literal descriptions, though the abstract connections can in fact be physical attachments at times.

It should be noted that any decision to expand the space $\mathcal{X}$ which results in $\nu(x)$ values that are some scalar multiple larger (or smaller) than they should be will be taken care of by the scaling action of r. For example, let $\nu'(x)$ be the function for an unnecessarily large $\mathcal{X}'$ such that $\nu'(x) = k\nu(x)$, where $\nu(x)$ is the true value for the correctly sized space $\mathcal{X}$, $r' = \sum_{x' \in \mathcal{X}} \nu'(x')$, and $r = \sum_{x' \in \mathcal{X}} \nu(x')$. Then,

$$\frac{r'}{\nu'(x)} = \frac{\sum_{x' \in \mathcal{X}} \nu'(x')}{\nu'(x)} \tag{27}$$

$$= \frac{\sum_{x' \in \mathcal{X}} k\nu(x')}{k\nu(x)} \tag{28}$$

$$= \frac{\sum_{x' \in \mathcal{X}} \nu(x')}{\nu(x)} \tag{29}$$

$$= \frac{r}{\nu(x)} \tag{30}$$

which will hold for any nonzero scalar multiple k.

For the complexity component of our specified complexity model, we take the negative base-2 logarithm of the probability of x under some proposed distribution induced by the generative processes under consideration, namely, $-\log_2 p(x)$, were p is the proposed distribution. Historically, irreducible complexity has been considered primarily for biological phenomena, though the concept may be applicable to technological systems as well [30, 32]. For biological systems, one will have some probabilistic model that attempts to explain the emergence of the system as a result of process acting in nature. Although the true distribution of nature is unknown (and possibly unknowable), one might estimate $\hat{p}(x)$ using frequentist methods or upper bound it by considering only the appearance of some necessary subsystem, such as a single gene sequence, for which the probabilistic resources and selection landscape are much better understood. Quantitative irreducible complexity is always relative to a proposed distribution, whether selectionist or neutral. Proposing probability increasing mechanisms including natural selection and self-organization requires additional work, since irreducibly complex systems rule out directly selectable pathways, leaving only indirect, circuitous routes, and self-organization proposals require empirical observations that the structures in question self-organize under biologically realistic conditions. Because irreducibly complex systems cannot be optimized by gradually improving the core function in an incremental manner, this feature may prove useful in ruling out the probability increasing benefits of natural selection in such scenarios.

Given an irreducible specificity function $\nu(x)$, we formally define **quantitative irreducible complexity** as

$$QIC(x) := -\log_2 r\frac{p(x)}{\nu(x)} \tag{31}$$

$$= -\log_2 \left[\left(\sum_{x' \in \mathcal{X}} \nu(x') \right) \frac{p(x)}{\nu(x)} \right] \tag{32}$$

which equals $-\infty$ when x is not an irreducible system (i.e., does not have at least three components in its irreducible core or does not have at least one interaction among components).

This view of irreducible complexity as a special case of specified complexity does not originate with us, but has a much earlier history, being stated explicitly as early as 2002 [3]:

> Irreducibly complex biological systems exhibit specified complexity. Irreducible complexity is therefore a special case of specified complexity.

While the general view presented here is not original,[6] this present research is the first to formalize irreducible complexity as a canonical specified complexity model, and prove a corresponding conservation of information result (see Corollary 7).

4.3.1 Simplified Variants

It may be desirable in some cases to use simplified versions of the irreducible specificity function that are easier to estimate. To do so, one can change the λ hyperparameters to zero to exclude the effect of any element from the computation. Two examples are given.

Definition 9 (Simplified Version I)**.** Setting $\Lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (1, 1, 1, 0)$, we obtain

$$\nu(x) = \mathbb{1}\{k \geq 3\} \cdot k \cdot m \cdot \gamma_1(x)$$

where k is the number of irreducible core components, m is the number of functionally relevant interactions among core components, and $\gamma_1(x)$ is the part specificity of core components.

The specification function of Definition 9 is the same as that given in Definition 7, except that it omits the interaction specificity term by setting its hyperparameter value to zero. This term may be difficult to estimate at times, yet it can serve as a signal of design activity (since designers routinely create tightly-matching interfaces for interacting parts), so should be included whenever possible. We can further simplify the specification function by removing the part specificity term, as follows.

[6] For example, see Chapter 5 of Dembski's No Free Lunch for a quantitative approach for estimating the irreducible complexity of a specific biological system [3].

Definition 10 (Simplified Version II)**.** Setting $\Lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (1, 1, 0, 0)$, we obtain

$$\nu(x) = \mathbb{1}\{k \geq 3\} \cdot k \cdot m$$

where k is the number of irreducible core components and m is the number of functionally relevant interactions among core components.

Although this variant is easy to estimate, it does not account for how well-matched components are, ignoring both their part specificity and their interaction specificity. In all cases, we define the scaling constant r as in Definition 8.

4.4 Conservation of Quantitative IC

Given a canonical form quantitative irreducible complexity model as defined in Section 4.3 (under any settings of the λ hyperparameters), it follows that one is unlikely to observe large values for such models under any proposed distribution p whenever p is the true distribution generating the observations. We state this as a corollary with proof given in the Appendix.

Corollary 7 (Conservation of Quantitative Irreducible Complexity (QIC))**.** *Let $p(x)$ be any probability distribution on a discrete finite space $\mathcal{X}$ and define*

$$QIC(x) := -\log_2 r\frac{p(x)}{\nu(x)}$$

where $\nu(x)$ and r are defined as in Section 4.3. Then for $X \sim p$,

$$\Pr\left(QIC(X) \geq b\right) \leq 2^{-b}.$$

Thus, the tail probability under proposed distribution p of observing a quantitative irreducible complexity value at least as large as that of an observation $X = x$ is no greater than $2^{-QIC(x)}$.

4.5 Functional Information

Hazen et al. [27] introduced *functional information* $I(E_{\mathsf{X}})$ to measure, in bits, the surprisal of observing a given level of function for living and digital structures. The authors define functional information as

$$I(E_{\mathsf{X}}) = -\log_2 F(E_{\mathsf{X}}), \tag{33}$$

where, for a fixed function X, $F(E_{\mathsf{X}})$ is the proportion of all possible configurations of the system exhibiting a degree of function at least E_{X}. Since we have used X to denote a configuration random variable instead of denoting a function, we will make the following change of symbols to remain consistent with the notation in this paper. Let $g : \mathcal{X} \to \mathbb{R}$ represent an arbitrary function, let $\mathcal{X}$ be the space of all possible configurations of the system and let $x \in \mathcal{X}$ be an element (a specific configuration) within that space. We then translate from their notation as follows:

$$g := \mathsf{X}, \tag{34}$$
$$F_g(x) := F(E_{\mathsf{X}}), \text{and} \tag{35}$$
$$I[F_g(x)] := I(E_{\mathsf{X}}). \tag{36}$$

Note that E_{X} is some level of function which implicitly has been observed for a configuration x. Having made this translation, we will allow X to refer to the random variable taking values of $x \in \mathcal{X}$, and use our notation in what remains.

Following Hazen et al. [27] and assuming that g increases with increasing degrees of function, for discrete finite spaces we define $M_g(x) := |\{x' \in \mathcal{X} : g(x') \geq g(x)\}|$, which gives

$$F_g(x) = M_g(x)/|\mathcal{X}|. \tag{37}$$

Functional information is a common form model for discrete finite spaces $\mathcal{X}$, which we state in the following proposition and prove in the Appendix:

Proposition 1. Functional information $I[F_g(x)]$ is a common form specified complexity model with

$$p(x) = |\mathcal{X}|^{-1},$$
$$\nu(x) = F_g(x)^{-1}, \text{and}$$
$$r = |\mathcal{X}|.$$

In general it is not canonical, since $\nu(\mathcal{X})$ may exceed r.

The fraction of functional configurations $F_g(x)$ acts as the specification component in the model. We will refer to $F_g(x)$ as the *functional specificity* of configuration x with respect to g. Given that functional specificity is a well-defined quantity that can be estimated from observations, we may consider using it as the specification component for a canonical specified complexity model. We do so next.

4.5.1 Functional Specified Complexity

One of the drawbacks of the functional information model is that while it tells you the relative size of the set of functional level[7] $g(x)$, it cannot give any indication of how likely one is to encounter such an extreme level of function under a nonuniform distribution on configurations. For example, assume you have a space of ten thousand configurations where all but one have identical levels of function for g, and one configuration achieves a much higher level of function. Further, assume that the configuration of high functionality is the one that is almost always observed in the wild. Then $I[F_g(x)]$ would be large even though high levels of function are actually

[7] Here, as in other parts of the paper, we abuse notation slightly so that $g(x)$ denotes the level of function for configuration x with respect to g.

quite common for sampled configurations. This behavior is by design, since functional information is measured with respect to levels of function, not with respect to individual configurations or their likelihoods. However, it would be useful to allow for arbitrary nonuniform probability functions, so as to better capture how "special" a given configuration is with regards to both its likelihood and its particular level of function.[8]

Let $F_g(x)$ be the functional specificity as defined in Section 4.5 and let $p(x)$ be any probability function on discrete finite space $\mathcal{X}$. We then define *functional specified complexity* as follows.

Definition 11 (Functional Specified Complexity)**.** For function g, functional specificity $F_g(x)$, and probability function $p : \mathcal{X} \to [0, 1]$, the *functional specified complexity kardis* is

$$\kappa(x) := |\mathcal{X}|(1 + \ln|\mathcal{X}|)\frac{p(x)}{F_g(x)^{-1}}. \tag{38}$$

Given the functional specified complexity kardis, the *functional specified complexity (FSC)* is thus

$$FSC(x) := -\log_2\left[|\mathcal{X}|(1 + \ln|\mathcal{X}|)\frac{p(x)}{F_g(x)^{-1}}\right] \tag{39}$$

$$= -\log_2 r\frac{p(x)}{\nu(x)} \tag{40}$$

where we have defined $r = |\mathcal{X}|(1 + \ln|\mathcal{X}|)$ and $\nu(x) = F_g(x)^{-1}$.

The relation between functional specified complexity and functional information is as follows (proof given in the Appendix).

Theorem 6 (Relation of $FSC(x)$ to Functional Information)**.** *For functional information $I[F_g(x)]$ and $FSC(x)$ with probability distribution p, the following relation holds:*

$$FSC(x) = I[F_g(x)] + I(x) + c$$

where $I(x) = -\log_2 p(x)$ is the surprisal of x under distribution p and $c = -\log_2|\mathcal{X}|(1 + \ln|\mathcal{X}|)$.

Theorem 6 establishes that functional specified complexity is the sum of the functional specificity surprisal of configuration x and its Shannon surprisal under distribution p (plus a constant that does not depend on x), and therefore captures how "special" or "surprising" a configuration is with regards to both its likelihood and its level of function, as we had desired. We will next see that this form is also a canonical specified complexity model, which means the results from Section 3 hold for them as well (with proof given in the Appendix).

[8]Another simple option is to measure functional specificity as the proportion of *observed* configurations with functional level greater than or equal to $g(x)$, making it an estimated p-value under the true distribution of nature.

Theorem 7 (Canonical Functional Specified Complexity)**.** *For any probability distribution $p(x)$ on discrete finite space $\mathcal{X}$, $FSC(x)$ is a canonical specified complexity model with*

$$\nu(x) = F_g(x)^{-1}, \text{ and}$$
$$r = |\mathcal{X}|(1 + \ln|\mathcal{X}|).$$

4.6 Conservation of FSC

Corollary 8 (Conservation of Functional Specified Complexity (FSC))**.** *Define $FSC(x)$ as in Definition 11. Then for $X \sim p$*

$$\Pr(FSC(X) \geq b) \leq 2^{-b}.$$

Proof. The proof follows immediately from Theorems 2 and 7. □

5. PRACTICAL CONSIDERATIONS

5.1 Estimating $p(x)$

As stated in Section 4.3, every canonical specified complexity model is defined relative to a proposed distribution. This is the distribution rejected whenever anomalously large specified complexity values are witnessed for observations purportedly generated under that distribution. To compute the specified complexity kardis for an observation x one does not require knowledge of the full distribution p, but only knowledge of $p(x)$, the distribution value for that observation. Since one does not typically have complete knowledge of the true generating distribution for non-artificial systems (or even for sufficiently complex artificial systems), this value will have to be estimated or upper bounded in most cases. To upper bound $p(x)$, simplifying assumptions are made which would increase the likelihood of observation x relative to $p(x)$. To estimate $p(x)$ (which we denote as $\hat{p}(x)$) for discrete combinatorial systems, one can sample from the generative process and use frequentist methods to estimate $\hat{p}(x)$ using the ratio of observations of x to the number of samples, namely,

$$\hat{p}(x) = \frac{\text{number of observations of } x}{\text{total number of samples}}$$

which will approach the true value asymptotically under standard mild statistical assumptions. For continuous systems, nonparametric methods of density estimation (such as kernel density estimation [33, 34]) may be used. The sampling distribution used to estimate $\hat{p}(x)$ is the distribution that will be rejected or not given the specified complexity test, so one must ensure the actual generative process of interest is used for generating the samples. Similarly, if using analytical approximation methods to bound the value $p(x)$, one should ensure that the approximation still reflects the true process reasonably well. As a practical example, if one is attempting to reject an

algorithmically produced system, one should compute the likelihood under that algorithmic process, not under some other distribution such as a uniform distribution (which may be easier to compute but might lower the odds of observing x), unless using that other distribution as a baseline for computing a minimum plausibility requirement (see Section 6.1). However, if one cannot compute or estimate the odds of a configuration x being produced by a particular probabilistic process, there is no need to reject the process at all, since it never actually rises to the level of an explanation. Probabilistic explanations require probabilities; without attached probabilities they remain conjectures. Claiming that a probabilistic process boosts the likelihood of observing some particular result remains without force unless one can provide some actual numbers for this claim, namely, what $p(x)$ is under the proposed distribution induced by the process (or some reasonable lower bound). Thus, the burden of providing such estimates resides with those proposing the probabilistic process as an explanation in the first place.

As general advice, statistics and machine learning are fields that specialize in estimating likelihoods and probabilities for complex systems, and should be looked to for guidance in this area. In other words, when attempting to estimate specified complexity values for real-world systems, first find a statistician.

5.2 Estimating r

Another difficulty that arises for canonical models is the computation of r, the normalizing scaling constant, which must be computed over the entire space $\mathcal{X}$. In some cases computation can be avoided entirely, such as when using pre-normalized specification functions such as probability distributions, which integrate or sum to 1 (resulting in $r = 1$). In the case of general specification functions on discrete finite spaces, one can estimate r using independent uniform random sampling on the space. The law of large numbers guarantees that

$$\frac{1}{n}\sum_{i=1}^{n}\nu(x_i) \overset{n\to\infty}{=} \mathbb{E}[\nu(X)] \tag{41}$$

where the expectation is taken with respect to a uniform random distribution on $\mathcal{X}$, namely, $X \sim u(x)$. Thus,

$$\frac{1}{n}\sum_{i=1}^{n}\nu(x_i) \overset{n\to\infty}{=} \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}}\nu(x). \tag{42}$$

Define

$$\hat{r} := |\mathcal{X}|\left(\frac{1}{n}\sum_{i=1}^{n}\nu(x_i)\right) \tag{43}$$

where n is the number of independently identically distributed samples taken drawn from $u(x)$, x_i denotes the ith sample, and $|\mathcal{X}|$ is the size of $\mathcal{X}$. Then,

$$\hat{r} \overset{n\to\infty}{=} |\mathcal{X}|\left(\frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}}\nu(x)\right) = \sum_{x\in\mathcal{X}}\nu(x) = r. \tag{44}$$

Thus, we can form an asymptotically consistent estimator $\hat{r}$ by independently sampling configurations on any discrete finite space $\mathcal{X}$ uniformly at random and averaging their rescaled observed ν function values. As the number of samples increases, so does the accuracy of the estimate of r.

5.3 Multiple Specification Testing

Although one can bound the probability of observing extreme specified complexity values using any independent specification function, if one repeatedly tries different specification functions until a large specified complexity value is observed, this amounts to "SC-hacking" (similar to "p-hacking") and violates the assumptions undergirding our bounds. The bounds are valid for any *single* independent (but arbitrary) specification function; introducing multiple possible specification functions raises the probability of observing a large SC value. For example, assume for a given $p(x)$ value and finite discrete space $\mathcal{X}$, we observe an SC value of b, using some specification function ν_0 that is also a probability distribution on $\mathcal{X}$. Seeing that b is too small, we then choose another specification function uniformly at random from all possible probability distributions on $\mathcal{X}$ (i.e., we implicitly assume $\nu(\mathcal{X}) = 1 = r$). We continue to repeat the process until a sufficiently large SC value b' is observed. Then the probability of observing the event $SC(X) \geq b'$ was not $2^{-b'}$ or less, since we artificially increased the number of chances we had of observing such an event. Given enough trials (and a threshold $b \leq -\log_2 p(x)$), the asymptotic probability of observing b' that exceeds b approaches 1 as we repeatedly sample more and more specification functions. Under the stated assumptions, the probability of uniformly at random choosing a specification function such that $b' > b$ in a single trial is

$$\Pr(\{\nu : b' \geq b\}) = 1 - \nu_0(x),$$

and so the probability of finding such a function given N trials is

$$1 - \nu_0(x)^N \to 1$$

as $N \to \infty$. Thus, one must control for multiple specification testing just as one must control for multiple hypothesis testing. A simple solution is to always use the same specification function for the domain in question, which fixes the number of "specification trials" to one, while also ensuring the specification function is independent of future observations from $\mathcal{X}$, being held constant regardless of those outcomes.

5.4 Mechanized Specification Functions

In the same year (1936) that Alan Turing introduced his model of computation (now known as the *Turing machine*) [35], Emil Post described a model essentially equivalent to Turing's, which he called "Formulation 1" [36]. Although equivalent in power, Post's system made use of a "worker" who would manually move left or right and manipulate symbols in boxes. Turing had the insight to realize the worker could be replaced with a finite state machine, giving rise to a completely mechanized form of computer. Similarly, previous work in specified complexity such as Dembski's semiotic specified complexity [4] has relied on the presence of a semiotic agent (presumably human) whose own cognitive context would be used to compute specified complexity values. Given the recent advances in machine learning and AI, it seems plausible that such computations can be mechanized to provide automated systems for computing such values. To estimate or compute specified complexity values requires having a reliable estimate of $p(x)$ (presenting perhaps the most difficult step), an estimate of r (which can be estimated using uniform random sampling on $\mathcal{X}$, for finite spaces), and an automated specification function ν. Having covered details related to estimating $p(x)$ in Section 5.1, and defined an asymptotically consistent estimator for r in Section 5.2, we will restrict our discussion in this section to various strategies for mechanizing specification functions.

The functional specificity of Hazen et al. [27] used to create functional specified complexity (see Section 4.5.1) provides a promising route towards automated specification functions. By creating systems to measure and rank the functional response of configurations for a given function g, one could compute direct estimates for the specification values without further human intervention. These systems would only require accurate measurements of the functional response for each observed configuration, of which we already have many viable technologies. Thus, functional specified complexity models will perhaps be the first to be fully mechanized.

Axe et al.'s *Stylus* [37, 38] system presents another possible (but non-general) route to mechanization. Given the character recognition system present in Stylus, one could define a specification function based on the conformation of vector structures to legible Chinese. Given that most of the possible vector structures do not form legible Chinese characters, the summation of ν should remain small for reasonably sized spaces. Since Stylus is an already existing system, mechanizing Stylus-based specification function should be a straightforward research task, which will result in the first fully computable and mechanized specified complexity model.

Machine learning systems present another possible route towards automated specified complexity systems. Given any classifier or regression system mapping to non-negative values, one can define a specification function using that learner. For binary classifiers, if the negative class represents objects produced under the null distribution and the positive class represents objects produced by some alternative process, observing large positive specified complexity values can be used to rule out a negative class label with an accompanying probability bound. Since such classification systems are already automated, reusing them as specification functions presents a frictionless path towards mechanized specification functions. The same should hold for deep-learning systems and other AI systems, which can be trained to have large responses to classes of interest (such as the class of artifacts created by adversarial humans, in security scenarios).

Lastly, language models used in natural language processing provide yet another possible path, since n-gram and other models can be trained to provide probability distributions over the space of character strings, thus computing nonnegative values for text strings that could act as specification values. For a natural language processing system trained on the output of human English speakers, such a system could be used to distinguish real human utterances from text generated by a well-defined random probabilistic process. Importantly, such a system could perform such functions without the need of a human agent to compute the specified complexity values.

Whether any of the above ideas is developed into actual working specification systems, it should be clear that there is no great logical barrier to creating automated specification functions. The current lack of such systems presents a research opportunity rather than an insurmountable roadblock.

5.5 Towards Detecting Design

While the machinery of specified complexity hypothesis tests allow us to reject proposed distributions as explanations for observed phenomena, inferring design from such a result requires additional constraints on specification functions. When the specification function $\nu(x)$ is a probability distribution on the space $\mathcal{X}$ that is correlated with design in some way, giving large values only to features of designed objects, then the specified complexity test becomes a form of likelihood-ratio test, where the alternative hypothesis is that the object is more likely to have been produced as a product of intentional design. Only then does rejecting the null hypothesis in favor of the alternative serve as evidence for design hypotheses. Without this correlation, rejection of the null hypothesis does not necessarily provide evidence of design.

While such correlation-based specification functions could occasionally lead to wrong inferences, this is not necessarily a fatal objection. Such correlation-based systems would provide evidence of design, not proof; the more evidence we gather, the more confident we

become in our conclusions, as is the universal nature of science. All scientific systems (and practitioners) have the potential to err.

6. EXAMPLES

6.1 Explanations: Plausibility Requirements

Canonical specified complexity models can be used to evaluate the plausibility of probabilistic explanations, by establishing how much a proposed process must boost probabilities over a baseline to avoid rejection and be considered a valid explanation of the observed data. Uniform models prove useful in this regard, since they allow us to establish an easily computable baseline against which to compare the plausibility of any non-uniform model. We do so as follows.

For a discrete, finite space, let $SC_u(x)$ denote a canonical specified complexity value for an observation x under a uniform distribution $u(x)$, namely

$$SC_u(x) := -\log_2 r\frac{u(x)}{\nu(x)} \tag{45}$$

and assume

$$\Pr(SC_u(X) > SC_u(x)) < \alpha,$$

where α is a hypothesis test significance level, so that the uniform distribution can safely be rejected as an explanation for the data. You then propose a mechanism which boosts the probability of x above that of the uniform model, by some factor of $s >> 1$. Thus,

$$p(x) = su(x). \tag{46}$$

Let $SC_p(x)$ be the specified complexity value of the canonical model using the proposed distribution $p(x)$ for observation x, namely,

$$SC_p(x) := -\log_2 r\frac{p(x)}{\nu(x)}. \tag{47}$$

For your proposed mechanism to be viable it must avoid rejection under the specified complexity hypothesis test. Thus, for the new distribution to even be considered as a possible explanation, it is required that

$$\Pr(SC_p(X) > SC_p(x)) > \alpha \tag{48}$$

where $X \sim p$ and α acts as a minimum plausibility baseline, since any model with a tail probability less than α will be rejected as an explanation for the observed data. As a hypothesis test significance level, smaller α values will make us more certain of our decision to reject the distribution in question, avoiding the false rejection of true explanations, but will also make it easier for a false explanation to escape rejection. Typically this α will be set below 0.05, often taking values of 0.01 or 0.001, since we usually want to avoid false rejection.

Given the requirement that $\Pr(SC_p(X) > SC_p(x)) > \alpha$ and the fact that our model is a canonical specified complexity model, we can invoke the bound from Theorem 2 to infer

$$\alpha < \Pr(SC_p(X) \geq SC_p(x)) \tag{49}$$

$$\leq 2^{-SC_p(x)}. \tag{50}$$

Thus,

$$\alpha < 2^{-SC_p(x)}. \tag{51}$$

Combining Equations (51), (47) and (46), we can rearrange to obtain

$$s > \frac{\alpha\nu(x)}{ru(x)}. \tag{52}$$

By Equation (52), we see that in order to avoid rejection one needs to demonstrate that an explanation boosts the probability of observation x by at least a factor of $\alpha\nu(x)/(ru(x))$ over the uniform probability. Whenever we observe large specification values $\nu(x)$ this increases the burden of proof on the person proposing the mechanism. Whenever $u(x)$ is extremely small (as it often is) the burden will be large.

Equation (52) is a quantifiable constraint that can be used to formally differentiate explanations that are causally adequate from those which are not: unless one can demonstrate quantitatively that a mechanism boosts probability by at least a factor of s, that mechanism cannot be considered a plausible explanation. Simply put, s is the entry fee for a probabilistic mechanism to even enter the tournament of competing explanations. Without demonstrating that this burden has been met (or at very least demonstrating that $p(x)$ boosts the probability on the scale of s), $p(x)$ cannot yet be considered a plausible explanation of observation x.

While α may be adjusted slightly to reduce this burden, it cannot be adjusted substantially without giving the appearance of special pleading to protect an unlikely explanation. Setting it too small guarantees that one will only reject explanations at ridiculously small α-levels, which means one will almost always fail to reject, rendering the hypothesis test relatively powerless. Setting α much smaller than 0.00001 will draw suspicion; setting it below 10^{-10} will likely draw heckles. While protecting against false rejection, the α-level also directly reflects the plausibility of the explanation itself; although setting it low may help to avoid outright rejection of a hypothesis, such a move concedes that the proposed explanation is so implausible that it cannot even meet an anemic probability requirement like 0.00001. The one proposing the explanation will want to set α high, to show that we will fail to reject even at liberal α values, which in turn means that the explanation is a strong one, rendering the observed data likely under the model. On the other

hand, if a model fails to demonstrate it meets the minimum plausibility requirement even for small α values, we can be confident in our refusal to consider it. Since the burden of demonstration falls on the one proposing the mechanism, α values should remain reasonably large, no smaller than 0.001 as a rule-of-thumb, to demonstrate the bare minimum of viability for the explanation. If the observed data remains extremely unlikely under a mechanism whose sole purpose is making it *more* likely, then the explanation isn't very compelling. What good is an explanation that fails to make reasonably likely the one thing it is supposed to explain? So while a smaller α might help ease the burden, it can never fully eliminate it.

Lastly, although we only considered a uniform distribution and discrete space, this same method easily extends to continuous and other spaces by simply replacing the uniform distribution with another low-probability, easily computable distribution. There is nothing special about using the uniform distribution as a baseline, other than it confers small probabilities on configurations in large spaces and is typically easy to compute. Any other distribution with these properties can also be used.

6.2 Electronic Coin Flipper

Imagine we have built an electronic coin flipper, which should result in fair coin flips, namely, the probability of heads should equal the probability of tails, $p(H) = p(T) = 0.5$. We would like to test the hypothesis that our coin flipper is fair using a canonical specified complexity model. To do so, we must choose which specification function we will use (the hypothesized distribution $p(x)$ being given). We have some options, but for simplicity we will use the functional specification function of Section 4.5.1, since we do not need to estimate r and because it is relatively easy to use.

Let $M_g(x)$ be the number of sequences in our space $\mathcal{X}$ of possible observations which are at least as surprising as our observation, having at least as many heads in their sequences (or as many tails, since the complement sequence can be just as surprising). The function $g : \mathcal{X} \to \mathbb{R}_{\geq 0}$ therefore measures the degree to which the sequence diverges from the expected number of heads, which can be formally computed by

$$g(x) = |k - \ell/2| \tag{53}$$

where k is the number of heads observed in a sequence of ℓ flips.

Since $\nu(x) = F_g(x)^{-1}$ where $F_g(x)$ is the proportion of sequences of the same length with at least as many heads (or tails), we have

$$\begin{aligned}\nu(x) &= F_g(x)^{-1} \\ &= \frac{|\mathcal{X}|}{M_g(x)} \\ &= \frac{2^\ell}{M_g(x)}.\end{aligned}$$

Our trials are independent and identically distributed, so under the fair coin hypothesis we have

$$p(x) = 2^{-\ell} \tag{54}$$

and following Section 4.5.1 our kardis becomes

$$\begin{aligned}\kappa(x) &= |\mathcal{X}|(1 + \ln|\mathcal{X}|)p(x)\frac{M_g(x)}{|\mathcal{X}|} \\ &= \frac{(1 + \ell \ln 2)M_g(x)}{2^\ell}.\end{aligned} \tag{55}$$

Our specified complexity model is therefore given by

$$FSC(x) = \ell - \log_2(1 + \ell \ln 2) - \log_2 M_g(x). \tag{56}$$

For our example, assume that we observe strange behavior, and all of our flips but one are heads. Thus,

$$M_g(x) = 2\binom{\ell}{1} + 2\binom{\ell}{0} = 2\ell + 2,$$

since there are exactly ℓ sequences of one tails possible and a sequence with all tails but one would result in the same g value, giving us a multiplier of 2, and since there are two sequences with larger g values, the all heads and the all tails sequences. Thus,

$$FSC(x) = \ell - \log_2(1 + \ell \ln 2) - (1 + \log_2(\ell + 1)). \tag{57}$$

For $\ell = 6$, Equation (57) yields roughly -0.17, which is not large enough to reject the fair coin flipper hypothesis. However, if $\ell = 13$ and we observe 12 heads, we obtain a value larger than 4.86, which is enough to reject the fair coin hypothesis at an $\alpha = 0.05$ significance level. At $\ell = 23$, we can reject at an $\alpha = 0.0001$ level.

When comparing our specified complexity model bounds to exact probabilities, we find that our canonical specified complexity bound adds a roughly 0.69ℓ factor to what we would obtain by using the parametric binomial model to directly compute the probability of seeing a sequences with at least as many heads (or at least as many tails, taking the complement) as the sequence observed. We state this in the following corollary, proven in the Appendix.

Corollary 9 (Relation to Optimal Parametric Bound)**.** *Let α be the tail probability bound derived from using Equation (56) in conjunction with Theorem 2 and let*

ϵ be the probability under the fair-coin binomial model of observing a sequence with at least as many heads (or taking the complement of the sequence, at least as many tails). Then,

$$\alpha = (1 + \ell \ln 2)\epsilon.$$

By the above corollary, we lose some efficiency compared to using the precise probabilities when the true underlying distributions are known; yet we gain flexibility in general, since we can compute canonical specified complexity values without knowing the full distribution or assuming it has some parametric form.

6.2.1 Insufficiently Sharp Specification Functions

Keeping with our coin flipper example, we will now look at what happens when we choose our specification function unwisely, such that $\nu(x)/r$ is never large. We will call such functions *blunt* or insufficiently sharp specification functions. Although the bounds for canonical specified complexity models will hold using *any* specification function that is nonnegative and respects the integration constraint for r, not all specification functions are equally powerful for rejecting spurious distributions. As an extreme example, consider a uniform specification function on finite $\mathcal{X}$, where $\nu(x) = c$ for all $x \in \mathcal{X}$ and $c > 0$. This implies $r = c|\mathcal{X}|$ and $\nu(x)/r = 1/|\mathcal{X}|$. Then for any uniform distribution $p(x) = 1/|\mathcal{X}|$, the specified complexity kardis for all x is

$$\kappa(x) = r\frac{p(x)}{\nu(x)} \tag{58}$$

$$= \frac{1/|\mathcal{X}|}{1/|\mathcal{X}|} \tag{59}$$

$$= 1. \tag{60}$$

Taking the negative log base 2 of $\kappa(x)$ will always result in a specified complexity value of 0, giving the trivial tail probability bound of 1. Thus, using such a blunt specification function, one will never be able to reject the uniform distribution for any observation, even if the observation is a sequence of one million heads in a row.

While a completely flat specification function is worthless, some reasonable-looking specification functions may also be problematic. As an example, let us use as our specification function the raw function g from Equation (53), which measures the absolute-value difference between the number of heads observed and the number expected, instead of a functional specificity $F_g(x)$ based on it. Thus,

$$\nu(x) = |k - \ell/2| \tag{61}$$

where x is a sequence of coin flips of length ℓ, k is the number of heads observed in the sequence, and $\ell/2$ is the expected number of heads for a fair coin. Since the trials are still independent and identically distributed, we again have

$$p(x) = 2^{-\ell} \tag{62}$$

for any sequence x of length ℓ. To finish defining our kardis we need to compute (or estimate) $r = \sum_{x\in\mathcal{X}} \nu(x)$, where $\mathcal{X}$ is the space of all possible sequences of coin flips of length ℓ. For simplicity, we will assume ℓ is an even integer, so that combinatoric rearrangement gives us

$$\sum_{x\in\mathcal{X}} \nu(x) = \sum_{k=0}^{\ell} |k - \ell/2| \binom{\ell}{k} \tag{63}$$

$$= \sum_{k=0}^{\frac{\ell}{2}} (\ell/2 - k)\binom{\ell}{k} + \sum_{k=\frac{\ell}{2}+1}^{\ell} (k - \ell/2)\binom{\ell}{k} \tag{64}$$

$$= 2\sum_{k=0}^{\frac{\ell}{2}} k \binom{\ell}{k + \frac{\ell}{2}} \tag{65}$$

$$= \frac{1}{2}(\ell + 2)\binom{\ell}{\frac{\ell}{2} + 1} \tag{66}$$

$$= r. \tag{67}$$

Comparing r to $\nu(x)$, an issue arises. Since

$$\nu(x) \leq |\ell - \ell/2| = \ell/2$$

this implies

$$\frac{\nu(x)}{r} \leq \frac{\ell}{2}\left[\sum_{k=0}^{\ell} |k - \ell/2| \binom{\ell}{k}\right]^{-1} \tag{68}$$

$$= \frac{\ell}{2}\left[\frac{\ell + 2}{2}\binom{\ell}{\frac{\ell}{2} + 1}\right]^{-1} \tag{69}$$

$$\leq \binom{\ell}{\frac{\ell}{2} + 1}^{-1}. \tag{70}$$

Assume we observe 50 heads in a row; under this specification function our specified complexity value would be

$$-\log_2 r\frac{p(x)}{\nu(x)} = -\log_2\left(2^{-50}\binom{50}{26}\right) < 3.22 \tag{71}$$

which means we cannot reject the fair coin flipper hypothesis at even an $\alpha = 0.1$ level. Observing 100 heads in a row only results in a specified complexity value of about 3.68, which is not enough to reject the null hypothesis at an $\alpha = 0.05$ level, despite witnessing such an extreme observation. Thus, using overly broad specification functions that give significant "mass" to too many elements in $\mathcal{X}$ results in near powerless tests. Specification functions must concentrate their mass on few elements to be useful, so that at least some observations might lead to rejection of the null hypothesis.

7. CONCLUSION

Specified complexity models have a long history, with the seminal notion of ordered complexity dating back to at least Erwin Schrödinger and his concept of *aperiodic crystals* [39, 40]. While several specified complexity models have been proposed in the literature, these models all share common elements which make them amenable to formal reduction. We have proposed such a reduction here, showing that several models (including p-value test models, semiotic specified complexity, algorithmic specified complexity, and functional information) are common form models, and reformulating slight variants of them into a mathematical form called canonical specified complexity. Furthermore, we have proven several properties of canonical models, such as proving the conservation of complex specified information for these models and giving other bounds on their tail probabilities for observing extreme values, which can be used to form statistical hypothesis tests. Canonical models are therefore tools for rejecting proposed distributions as explanations for observed phenomena, similar to Fisherian p-values. Furthermore, they do not require knowledge of the full analytical distribution to be used, but simply computation based on the observed configuration in question and a normalizing constant. Although it is a nontrivial task to form reliable estimates of specified complexity values for real-world systems, the task is by no means impossible nor should it even prove any more difficult than standard estimation tasks in machine learning and statistical inference. Canonical models can be used in technological settings, such as in anomaly and cheating detection, and can plausibly be automated using standard machine learning techniques, such as using neural networks or probabilistic language models as specification functions. The purpose of this manuscript is to formally demonstrate the mathematical unity of various proposed specified complexity models and suggest ways in which they might be used in technological and other empirical research.

APPENDIX: PROOFS

Theorem 1. *(Conservation of P-Value Specified Complexity) Under the same conditions as Definition 6, for $X \sim P_\theta$ and $\alpha \in (0, 1]$,*

$$\Pr\left(-\log_2\left(\frac{p_{val}(X)}{\alpha}\right) > b\right) \leq \alpha 2^{-b}.$$

Proof.

$$\Pr\left(-\log_2 \frac{p_{\text{val}}(X)}{\alpha} > b\right) = \Pr\left(\log_2 p_{\text{val}}(X) < \log_2 \alpha - b\right) \quad (72)$$

$$= \Pr\left(p_{\text{val}}(X) < \alpha 2^{-b}\right) \quad (73)$$

$$\leq \alpha 2^{-b}, \quad (74)$$

where the final inequality follows from the level-α property. □

We next prove a well-known result from probability theory which will be useful in deriving later results.

Lemma 1 (Markov's Inequality)**.** *Let $(\mathcal{X}, \Sigma, p)$ be a probability space, $f : \mathcal{X} \to \mathbb{R}_{>0}$ be a nonnegative measurable function such that $\mathbb{E}[f(X)]$ exists, and $\epsilon \in (0, \infty)$. Then*

$$\Pr\left(f(X) \geq \epsilon\right) \leq \frac{\mathbb{E}[f(X)]}{\epsilon}$$

where $X \sim p$.

Proof. Define $W = \{x \in \mathcal{X} : f(x) \geq \epsilon\}$. Then,

$$\mathbb{E}[f(X)] = \int_{\mathcal{X}} f(x)dp(x) \quad (75)$$

$$\geq \int_{W} f(x)dp(x) \quad (76)$$

$$\geq \int_{W} \epsilon dp(x) \quad (77)$$

$$= \epsilon \int_{W} dp(x) \quad (78)$$

$$= \epsilon \Pr\left(f(X) \geq \epsilon\right). \quad (79)$$

Dividing both sides by ϵ yields the result. □

Armed with Markov's inequality, we proceed to prove our main theorem, a generalization of an older theorem by Milosavljević [41]. It may be easier to understand in the special case of discrete random variables from Theorem 2, since integration is replaced with simple weighted summation. We offer two proofs, with the second proof similar to that given in [6] for algorithmic specified complexity, especially its measure theoretic reformulation given by [42]. Note that the second proof holds for general probability measures p, extending the result to cases that are neither fully discrete nor continuous.

Theorem 2 (Conservation of Canonical Specified Complexity)**.** *Let $p(x)$ be any continuous or discrete probability measure on space $\mathcal{X}$, let $r \in \mathbb{R}_{>0}$ be a scaling constant, and let $\nu : \mathcal{X} \to \mathbb{R}_{\geq 0}$ be any nonnegative integrable function where $\nu(\mathcal{X}) \leq r$. Then*

$$\Pr\left(-\log_2 r\frac{p(X)}{\nu(X)} \geq b\right) \leq 2^{-b}, \quad (80)$$

where $X \sim p$.

Proof. When $b \leq 0$ the result holds trivially. Assume $b > 0$ and define $W = \{x \in \mathcal{X} : p(x) > 0\}$. Then

$$\Pr\left(-\log_2 \frac{rp(X)}{\nu(X)} \geq b\right) = \Pr\left(\log_2 \frac{\nu(X)}{rp(X)} \geq b\right) \quad (81)$$

$$= \Pr\left(\frac{\nu(X)}{p(X)} \geq 2^b r\right) \quad (82)$$

$$\leq (2^b r)^{-1}\mathbb{E}\left[\frac{\nu(X)}{p(X)}\right]. \quad (83)$$

where the final line follows from Markov's inequality. In the continuous case, we have

$$\mathbb{E}\left[\frac{\nu(X)}{p(X)}\right] = \int_{\mathcal{X}} \frac{\nu(x)}{p(x)} p(x)dx \tag{84}$$
$$= \int_{W} \nu(x)dx \tag{85}$$
$$\leq \nu(\mathcal{X}), \tag{86}$$

where the inequality follows from the nonnegativity of ν. In the discrete case, we also obtain

$$\mathbb{E}\left[\frac{\nu(X)}{p(X)}\right] = \sum_{x\in\mathcal{X}} \frac{\nu(x)}{p(x)} p(x) \tag{87}$$
$$= \sum_{x\in W} \nu(x) \tag{88}$$
$$\leq \nu(\mathcal{X}). \tag{89}$$

Thus,

$$\mathbb{E}\left[\frac{\nu(X)}{p(X)}\right] \leq \nu(\mathcal{X}), \tag{90}$$

and we obtain

$$\Pr\left(-\log_2 \frac{rp(X)}{\nu(X)} \geq b\right) \leq (2^b r)^{-1} \mathbb{E}\left[\frac{\nu(X)}{p(X)}\right] \tag{91}$$
$$\leq 2^{-b} r^{-1} \nu(\mathcal{X}) \tag{92}$$
$$\leq 2^{-b}, \tag{93}$$

where the final inequality follows from the condition placed on function ν. □

Alternative Proof. Define

$$R = \{x \in \mathcal{X} : -\log_2 (rp(x)/\nu(x)) \geq b\} \tag{94}$$
$$= \{x \in \mathcal{X} : rp(x)/\nu(x) \leq 2^{-b}\} \tag{95}$$
$$= \{x \in \mathcal{X} : \nu(x) \geq 2^b rp(x)\}. \tag{96}$$

By the nonnegativity of ν, Eqs. (94) and (96), and the fact that $X \sim p$, we have

$$\nu(\mathcal{X}) \geq \nu(R) \tag{97}$$
$$\geq 2^b rp(R) \tag{98}$$
$$= 2^b r \Pr\left(-\log_2 r\frac{p(X)}{\nu(X)} \geq b\right). \tag{99}$$

Dividing through by $2^b r$ we obtain

$$\Pr\left(-\log_2 r\frac{p(X)}{\nu(X)} \geq b\right) \leq 2^{-b} r^{-1} \nu(\mathcal{X}) \tag{100}$$
$$\leq 2^{-b} \tag{101}$$

where the final inequality follows from the condition placed on function ν. □

Corollary 1 (Level-α Property for Specified Complexity I)**.** *Given $X \sim p$ and significance level $\alpha \in (0, 1]$, let $\kappa(x)$ be the kardis from any canonical specified complexity model. Then*

$$\Pr\left(\kappa(X) \leq \alpha\right) \leq \alpha.$$

Proof.

$$\Pr\left(\kappa(X) \leq \alpha\right) = \Pr\left(r\frac{p(X)}{\nu(X)} \leq \alpha\right) \tag{102}$$
$$= \Pr\left(\log_2 r\frac{p(X)}{\nu(X)} \leq \log_2 \alpha\right) \tag{103}$$
$$= \Pr\left(-\log_2 r\frac{p(X)}{\nu(X)} \geq -\log_2 \alpha\right) \tag{104}$$
$$\leq \alpha, \tag{105}$$

where the final inequality follows from application of Theorem 2 with $b = -\log_2 \alpha$. □

Corollary 2 (Level-α Property for Specified Complexity II)**.** *Let $SC(x)$ be any canonical specified complexity model. Then for $X \sim p$ and significance level $\alpha \in (0, 1]$,*

$$\Pr\left(SC(X) \geq -\log_2 \alpha\right) \leq \alpha.$$

Proof. The proof follows immediately from Theorem 2 and fact that $2^{-(-\log_2 \alpha)} = \alpha$. □

Theorem 3 (α-Exponential Bound)**.** *Given $X \sim p$, significance level $\alpha \in (0, 1]$, and any augmented canonical specified complexity model $SC_\alpha(x)$ as defined in Definition 4,*

$$\Pr\left(SC_\alpha(X) \geq b\right) \leq \alpha 2^{-b}.$$

Proof.

$$\Pr\left(SC_\alpha(X) \geq b\right) = \Pr\left(-\log_2\left(\frac{\kappa(X)}{\alpha}\right) \geq b\right) \tag{106}$$
$$= \Pr\left(-\log_2 \kappa(X) \geq -\log_2 \alpha + b\right) \tag{107}$$
$$= \Pr\left(\kappa(X) \leq \alpha 2^{-b}\right) \tag{108}$$
$$\leq \alpha 2^{-b}, \tag{109}$$

where the final inequality follows from application of Corollary 1. □

Corollary 4 (Common Form Model Bound)**.** *For any common form model SC, where r and $\nu(\mathcal{X})$ are as in Definitions 1 and 2, we have*

$$\Pr\left(SC(X) \geq b\right) \leq 2^{-b} r^{-1} \nu(\mathcal{X}).$$

In the special case of $r = 1$, this simplifies to

$$\Pr\left(SC(X) \geq b\right) \leq 2^{-b} \nu(\mathcal{X}).$$

Proof. Define $r_1 := r^{-1}\nu(\mathcal{X})$ and $r_2 := \nu(\mathcal{X})$. Subtracting $\log_2 r_1$ from both sides, we obtain

$$\Pr\left(SC(X) \geq b\right) = \Pr\left(-\log_2 r\frac{p(X)}{\nu(X)} \geq b\right) \tag{110}$$

$$= \Pr\left(-\log_2 rr_1\frac{p(X)}{\nu(X)} \geq b - \log_2 r_1\right) \tag{111}$$

$$= \Pr\left(-\log_2 r_2\frac{p(X)}{\nu(X)} \geq b - \log_2 r_1\right). \tag{112}$$

Since $\nu(\mathcal{X}) \leq r_2$ by construction, we invoke Theorem 2 to obtain,

$$\Pr\left(SC(X) \geq b\right) = \Pr\left(-\log_2 r_2\frac{p(X)}{\nu(X)} \geq b - \log_2 r_1\right) \tag{113}$$

$$\leq 2^{-b+\log_2 r_1} \tag{114}$$

$$= 2^{-b}r^{-1}\nu(\mathcal{X}). \tag{115}$$

□

Theorem 4 (Combined Models I)**.** *Let $SC_1(x), \ldots, SC_m(x)$ be canonical specified complexity models with corresponding specification functions $\nu_1(x), \ldots, \nu_m(x)$, scaling constants $r_1, \ldots, r_m$, and common probability function $p(x)$. Define a set of mixture variables*

$$\Lambda = \{\lambda_i : i = 1, \ldots, m, 0 \leq \lambda_i \leq 1\},$$

such that $\sum_{i=1}^m \lambda_i = 1$. Then for any such Λ and defining $r' := \max_i r_i$,

$$SC'(x) := -\log_2 r'\frac{p(x)}{\sum_{i=1}^m \lambda_i\nu_i(x)}$$

is a canonical specified complexity model.

Proof. Since $SC_1(x), \ldots, SC_m(x)$ are all canonical models that share a common $p(x)$, this implies $p(x)$ is a probability distribution on $\mathcal{X}$, $\nu(x) := \sum_{i=1}^m \lambda_i\nu_i(x)$ is a nonnegative function (being the positively weighted sum of nonnegative $\nu_i(x)$ functions), and $r' > 0$ (as are all r_i, including the maximum). Therefore, it suffices to show that $\nu(\mathcal{X}) \leq r'$. We have

$$\nu(\mathcal{X}) = \sum_{i=1}^m \lambda_i\nu_i(\mathcal{X}) \tag{116}$$

$$\leq \sum_{i=1}^m \lambda_i r_i \tag{117}$$

$$\leq \sum_{i=1}^m \lambda_i r' \tag{118}$$

$$= r', \tag{119}$$

since $\sum_{i=1}^m \lambda_i = 1$. The first inequality follows since $\nu_i(\mathcal{X}) \leq r_i$ for all $i = 1, \ldots, m$. By Definition 3, $SC'(x)$ is therefore a canonical model. □

Theorem 5 (Combined Models II)**.** *Let $SC_1(x), \ldots, SC_m(x)$ be canonical specified complexity models sharing a common probability function $p(x)$. Define a set of mixture variables*

$$\Lambda = \{\lambda_i : i = 1, \ldots, m, 0 \leq \lambda_i \leq 1\},$$

such that $\sum_{i=1}^m \lambda_i = 1$. Then for any such Λ,

$$\Pr\left(\sum_{i=1}^m \lambda_i SC_i(X) \geq b\right) \leq m2^{-b}.$$

Proof.

$$\Pr\left(\sum_{i=1}^m \lambda_i SC_i(X) \geq b\right) \leq \Pr\left(\max_{i=1,\ldots,m} SC_i(X) \geq b\right) \tag{120}$$

$$= \Pr\left(\bigvee_{i=1,\ldots,m} (SC_i(X) \geq b)\right) \tag{121}$$

$$\leq \sum_{i=1}^m \Pr\left(SC_i(X) \geq b\right) \tag{122}$$

$$\leq \sum_{i=1}^m 2^{-b} \tag{123}$$

$$= m2^{-b}, \tag{124}$$

where the first inequality follows from taking a convex mixture, the second inequality from a union bound and the final inequality follows from Theorem 2 and the fact that all models are canonical models. □

Corollary 5 (Bounds on Semiotic Specified Complexity (SSC))**.** *Let $p(x)$ be any probability distribution on a finite discrete space $\mathcal{X}$ and define*

$$SSC(x, s) := -\log_2\left[10^{120}\varphi_s(x)p(x)\right]$$

where $\varphi_s(x)$ is the number of patterns available to semiotic agent s which have descriptive complexity no greater than that of x. Then for $X \sim p$, we have

$$\Pr\left(SSC(X, s) \geq b\right) \leq 2^{-b}10^{-120}\sum_{x\in\mathcal{X}} \varphi_s(x)^{-1}.$$

Additionally, whenever $|\mathcal{X}| \leq 10^{120}$, the following conservation bound holds:

$$\Pr\left(SSC(X, s) \geq b\right) \leq 2^{-b}.$$

Proof. Let $r = 10^{120}$ and $\nu(x) = 1/\varphi_s(x) = \varphi_s(x)^{-1}$, so that

$$SSC(x, s) = -\log_2 r\frac{p(x)}{\nu(x)}.$$

Thus, SSC is a model with common form. Invoking Corollary 4, we obtain

$$\Pr\left(SSC(X, s) \geq b\right) \leq 2^{-b} r^{-1} \nu(\mathcal{X}) \tag{125}$$

$$= 2^{-b} 10^{-120} \sum_{x\in\mathcal{X}} \varphi_s(x)^{-1}. \tag{126}$$

We note that because $\varphi_s(x) \geq 1$ (since there is always at least one pattern with equal or lower descriptive complexity, namely, the pattern itself), we have

$$\sum_{x\in\mathcal{X}} \varphi_s(x)^{-1} \leq \sum_{x\in\mathcal{X}} 1 \tag{127}$$

$$= |\mathcal{X}|. \tag{128}$$

Thus, $|\mathcal{X}| \leq 10^{120}$ implies $\sum_{x\in\mathcal{X}} \varphi_s(x)^{-1} \leq 10^{120}$, which implies

$$10^{-120} \sum_{x\in\mathcal{X}} \varphi_s(x)^{-1} \leq 1 \tag{129}$$

and the conservation result follows. Thus, both $|\mathcal{X}| \leq 10^{120}$ and $\sum_{x\in\mathcal{X}} \varphi_s(x)^{-1} \leq 10^{120}$ are sufficient conditions for conservation of semiotic specified complexity. □

Remark. It should be noted that although Landman claims [8] that the number of possible states should be used as the replicational resources (scaling constant) r rather than the number of possible events, it should be clear that such a claim is mistaken, since the replicational resources represent the number of attempts a system is given to produce a given result (if every operation on every bit were a sampling attempt), which corresponds to the total number of sampling events possible since the Big Bang (and not to the number of possible states). In either case, since we are concerned with proving bounds for Dembski's model, and he uses the scaling constant of 10^{120}, we do so here as well.

Corollary 7 (Conservation of Quantitative Irreducible Complexity (QIC))**.** *Let $p(x)$ be any probability distribution on a discrete finite space $\mathcal{X}$ and define*

$$QIC(x) := -\log_2 r\frac{p(x)}{\nu(x)}$$

where $\nu(x)$ and r are defined as in Section 4.3. Then for $X \sim p$,

$$\Pr\left(QIC(X) \geq b\right) \leq 2^{-b}.$$

Proof. By the definition of $p(x)$, the nonnegativity of ν, and the fact that $r = \nu(\mathcal{X})$, we obtain the result by direct application of Theorem 2. □

Proposition 1. Functional information $I[F_g(x)]$ is a common form specified complexity model with

$$p(x) = |\mathcal{X}|^{-1}, \tag{130}$$

$$\nu(x) = F_g(x)^{-1}, \text{and} \tag{131}$$

$$r = |\mathcal{X}|. \tag{132}$$

In general it is not canonical, since $\nu(\mathcal{X})$ may exceed r.

Proof.

$$I[F_g(x)] = -\log_2 F_g(x) \tag{133}$$

$$= -\log_2 1/F_g(x)^{-1} \tag{134}$$

$$= -\log_2\left(|\mathcal{X}|\frac{|\mathcal{X}|^{-1}}{F_g(x)^{-1}}\right) \tag{135}$$

$$= -\log_2 r\frac{p(x)}{\nu(x)}. \tag{136}$$

$I[F_g(x)]$ is therefore a common form specified complexity model where p is a uniform distribution and the specification function ν equals to $1/F_g(x)$.

We then have

$$\sum_x \nu(x) = \sum_x 1/F_g(x) \tag{137}$$

$$= \sum_x 1/(M_g(x)/|\mathcal{X}|) \tag{138}$$

$$= |\mathcal{X}| \sum_x 1/M_g(x), \tag{139}$$

where $M_g(x)$ is the number of configurations in $\mathcal{X}$ with functional level at least to that of x. Because $M_g(x) \leq |\mathcal{X}|$, this sum will exceed $|\mathcal{X}|$ in all but one case, the special case where all configurations have identical function levels so that

$$|\mathcal{X}| \sum_x 1/M_g(x) = |\mathcal{X}| \sum_x 1/|\mathcal{X}| \tag{140}$$

$$= |\mathcal{X}|. \tag{141}$$

Thus, the model is not canonical in general. □

Theorem 6 (Relation of $FSC(x)$ to Functional Information)**.** *For functional information $I[F_g(x)]$ and $FSC(x)$ with probability distribution p, the following relation holds:*

$$FSC(x) = I[F_g(x)] + I(x) + c \tag{142}$$

where $I(x) = -\log_2 p(x)$ is the surprisal of x under distribution p and $c = -\log_2 |\mathcal{X}|(1 + \ln |\mathcal{X}|)$.

Proof.

$$FSC(x) = -\log_2\left(|\mathcal{X}|(1+\ln|\mathcal{X}|)\frac{p(x)}{F_g(x)^{-1}}\right) \tag{143}$$

$$= -\log_2 F_g(x) - \log_2\left(|\mathcal{X}|(1+\ln|\mathcal{X}|)p(x)\right) \tag{144}$$

$$= I[F_g(x)] - \log_2 p(x) - \log_2|\mathcal{X}|(1+\ln|\mathcal{X}|) \tag{145}$$

$$= I[F_g(x)] + I(x) + c. \tag{146}$$

□

Theorem 7 (Canonical Functional Specified Complexity)**.** *For any probability distribution $p(x)$ on discrete finite space $\mathcal{X}$, $FSC(x)$ is a canonical specified complexity model with*

$$\nu(x) = F_g(x)^{-1}, \text{ and} \tag{147}$$

$$r = |\mathcal{X}|(1+\ln|\mathcal{X}|). \tag{148}$$

Proof. Since $FSC(x) = -\log_2\kappa(x)$ it suffices to show that $\kappa(x)$ is a canonical kardis. We have from the definition of $FSC(x)$ that $\nu(x) = F_g(x)^{-1}$, which implies

$$\sum_x \nu(x) = \sum_x 1/F_g(x) \tag{149}$$

$$= \sum_x 1/(M_g(x)/|\mathcal{X}|) \tag{150}$$

$$= |\mathcal{X}|\sum_x 1/M_g(x) \tag{151}$$

where $M_g(x)$ is once again the number of configurations in $\mathcal{X}$ with functional level at least to that of x.

Let $x_{(1)}$ denote the configuration with the smallest $M_g(\cdot)$ value, $x_{(2)}$ be the configuration with the next smallest value, etc., so that

$$M_g(x_{(1)}), M_g(x_{(2)}), \ldots, M_g(x_{|\mathcal{X}|})$$

is a well-defined list of ascending order.

Considering the sum $\sum_x 1/M_g(x)$, we then have

$$\sum_x \frac{1}{M_g(x)} = \frac{1}{M_g(x_{(1)})} + \frac{1}{M_g(x_{(2)})} + \ldots + \frac{1}{M_g(x_{(|\mathcal{X}|)})} \tag{152}$$

$$\leq \frac{1}{1} + \frac{1}{M_g(x_{(2)})} + \ldots + \frac{1}{M_g(x_{(|\mathcal{X}|)})} \tag{153}$$

$$\leq \frac{1}{1} + \frac{1}{2} + \ldots + \frac{1}{M_g(x_{(|\mathcal{X}|)})} \tag{154}$$

$$\leq \sum_{i=1}^{|\mathcal{X}|} 1/i \tag{155}$$

$$\leq (1+\ln|\mathcal{X}|), \tag{156}$$

where the final inequality comes from the fact that $\sum_{i=1}^{|\mathcal{X}|} 1/i$ is the partial sum of harmonic series, and thus has a well-known upper bound of $(1+\ln N)$ where N is the upper limit of the partial sum.

Thus,

$$\sum_x \nu(x) = \sum_x 1/F_g(x) \tag{157}$$

$$= \sum_x 1/(M_g(x)/|\mathcal{X}|) \tag{158}$$

$$= |\mathcal{X}|\sum_x 1/M_g(x) \tag{159}$$

$$\leq |\mathcal{X}|(1+\ln|\mathcal{X}|) \tag{160}$$

$$= r. \tag{161}$$

Since $p(x)$ is a probability distribution on $\mathcal{X}$, ν is a nonnegative function on $\mathcal{X}$ such that $\nu(\mathcal{X}) \leq r$, and $r > 0$ whenever $|\mathcal{X}| \geq 1$ (which holds for all nonempty discrete spaces), the kardis $\kappa(x)$ is canonical, and $FSC(x)$ is therefore a canonical form specified complexity model, proving the result. □

Corollary 9 (Relation to Optimal Parametric Bound)**.** *Let α be the tail probability bound derived from using Equation (56) in conjunction with Theorem 2 and let ϵ be the probability under the fair-coin binomial model of observing a sequence with at least as many heads (or taking the complement of the sequence, at least as many tails). Then,*

$$\alpha = (1+\ell\ln 2)\epsilon.$$

Proof. Define

$$k' := \max\{k, \ell - k\}$$

and let

$$\epsilon = 2\sum_{i=k'}^{\ell} \text{Binom}(\ell, 0.5, i)$$

be the probability under the fair-coin binomial model of observing a sequence with at least as many heads (or taking the complement of the sequence, at least as many tails).

Combining Equation (56) with Theorem 2, we obtain

$$\alpha = 2^{-FSC(x)} \tag{162}$$

$$= 2^{\log_2\kappa(x)} \tag{163}$$

$$= \kappa(x) \tag{164}$$

where $\kappa(x)$ is as defined in Equation (55).

By the definition of $M_g(x)$ in Section 6.2, we have

$$M_g(x) = 2\sum_{i=k'}^{\ell}\binom{\ell}{i} \tag{165}$$

which is the number of configurations in $\mathcal{X}$ which have g values at least as extreme. By Equations (55) and (165),

we have □

$$\alpha = \kappa(x) \tag{166}$$
$$= (1+\ell\ln 2)2^{-\ell}\left[2\sum_{i=k'}^{\ell}\binom{\ell}{i}\right] \tag{167}$$
$$= (1+\ell\ln 2)\left[2\sum_{i=k'}^{\ell}\binom{\ell}{i}2^{-\ell}\right] \tag{168}$$
$$= (1+\ell\ln 2)\left[2\sum_{i=k'}^{\ell}\binom{\ell}{i}2^{-i}2^{-(\ell-i)}\right] \tag{169}$$
$$= (1+\ell\ln 2)\left[2\sum_{i=k'}^{\ell}\binom{\ell}{i}0.5^{i}(1-0.5)^{\ell-i}\right] \tag{170}$$
$$= (1+\ell\ln 2)\left[2\sum_{i=k'}^{\ell}\mathrm{Binom}(\ell,0.5,i)\right] \tag{171}$$
$$= (1+\ell\ln 2)\epsilon. \tag{172}$$

1. Dembski WA (1998) The Design Inference: Eliminating Chance through Small Probabilities. Cambridge University Press (Cambridge). doi:10.1017/CBO9780511570643
2. Dembski WA (2001) Detecting Design by Eliminating Chance: A Response to Robin Collins. Christian Scholar's Review 30:343–358.
3. Dembski WA (2002) No Free Lunch: Why Specified Complexity Cannot Be Purchased Without Intelligence. Rowman & Littlefield (Lanham).
4. Dembski WA (2005) Specification: The Pattern that Signifies Intelligence. Philosophia Christi 7:299–343. doi:10.5840/pc20057230
5. Ewert W, Dembski WA, Marks II RJ (2012) Algorithmic Specified Complexity. Engineering and Metaphysics doi:10.33014/isbn.0975283863.7
6. Ewert W, Marks RJ, Dembski WA (2013) On the Improbability of Algorithmic Specified Complexity. In: System Theory (SSST), 2013 45th Southeastern Symposium on, 68–70. IEEE. doi:10.1109/SSST.2013.6524962
7. Montañez GD (2016) Detecting Intelligence: The Turing Test and Other Design Detection Methodologies. In: Proceedings of the 8th International Conference on Agents and Artificial Intelligence, 517–523. doi:10.5220/0005823705170523
8. Landman HA (2008). Dembski's Specified Complexity: A Simple Error in Arithmetic. https://bit.ly/2UFi18k. Accessed: 2018-12-13.
9. English T, Greenwood GW (2008) Intelligent Design and Evolutionary Computation. In: Design by Evolution. Springer pp 7–30. doi:10.1007/978-3-540-74111-4_2
10. Olofsson P (2008) Intelligent Design and Mathematical Statistics: A Troubled Alliance. Biology & Philosophy 23:545–553. doi:10.1007/s10539-007-9078-6
11. Elsberry W, Shallit J (2011) Information theory, evolutionary computation, and Dembski's 'complex specified information'. Synthese 178:237–270. doi:10.1007/s11229-009-9542-8
12. Milosavljević A (1993) Discovering sequence similarity by the algorithmic significance method 284–291.
13. Dembski WA (2004) The Design Revolution: Answering the Toughest Questions about Intelligent Design. InterVarsity Press (Downers Grove).
14. Dembski W (2013). Before They've Even Seen Stephen Meyer's New Book, Darwinists Waste No Time in Criticizing *Darwin's Doubt*. https://bit.ly/2Gq4MFi. Accessed: 2018/02/20.
15. Ewert W (2013). Information, Past and Present. https://bit.ly/2Gqt7qq. Accessed: 2018/02/20.
16. Luskin C (2010). Intelligent Design Proponents Toil More than the Critics: A Response to Wesley Elsberry and Jeffrey Shallit. https://bit.ly/24UnDho.
17. Neyman J, Pearson ES (1933) On the Problem of the Most Efficient Tests of Statistical Hypotheses. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 231:289–337. doi:10.1098/rsta.1933.0009
18. Fisher RA (1932) Statistical Methods for Research Workers (4th ed.). Oliver & Boyd (Edinburgh).
19. Casella G, Berger RL (2002) Statistical Inference (2nd ed.). Duxbury Press (Pacific Grove).
20. Lew MJ (2012) Bad Statistical Practice in Pharmacology (and Other Basic Biomedical Disciplines): You Probably Don't Know P. British Journal of Pharmacology 166:1559–1567. doi:10.1111/j.1476-5381.2012.01931.x
21. Loveland JL (2011) Mathematical Justifcation of Introductory Hypothesis Tests and Development of Reference Materials .
22. Wasserman L (2013) All of Statistics: A Concise Course in Statistical Inference. Springer Science & Business Media (New York). p 156–157
23. McShane BB, Gal D, Gelman A, Robert C, Tackett JL (2017) Abandon Statistical Significance. arXiv preprint arXiv:1709.07588 .
24. Colquhoun D (2014) An investigation of the false discovery rate and the misinterpretation of p-values. Royal Society open science 1:140216. doi:10.1098/rsos.140216
25. Thorisson H (2000) Coupling, Stationarity, and Regeneration. Springer (New York).
26. Behe MJ (1996) Darwin's Black Box: The Biochemical Challenge to Evolution. Simon and Schuster (New York).
27. Hazen RM, Griffin PL, Carothers JM, Szostak JW (2007) Functional Information and the Emergence of Biocomplexity. Proceedings of the National Academy of Sciences 104:8574–8581. doi:10.1073/pnas.0701744104

28. Behe M (2005). Evidence for Intelligent Design from Biochemistry. https://bit.ly/2Jb47oH. From a speech delivered at Discovery Institute's "God & Culture Conference" (August 10, 1996). Accessed: 2017-11-23.
29. Miller KR (2004) The flagellum unspun–the collapse of Irreducible Complexity. Philosophy of Biology: An Anthology 439–449.
30. Ewert W (2014) Digital Irreducible Complexity: A Survey of Irreducible Complexity in Computer Simulations. BIO-Complexity 2014. doi:10.5048/BIO-C.2014.1
31. Axe D (2016) Undeniable: How Biology Confirms Our Intuition That Life Is Designed. HarperCollins (New York). p 146–162.
32. Bartlett J (2010) Irreducible Complexity and Relative Irreducible Complexity: Foundations and Applications. Occasional Papers of the BSG 15:1–10.
33. Parzen E (1962) On estimation of a probability density function and mode. The Annals of Mathematical Statistics 33:1065–1076. doi:10.1214/aoms/1177704472
34. Rosenblatt M (1956) Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics 832–837. doi:10.1214/aoms/1177728190
35. Turing A (1936) On Computable Numbers, with an Application to the Entscheidungsproblem. J. of Math 58:345–363.
36. Post EL (1936) Finite Combinatory Processes–Formulation 1. The Journal of Symbolic Logic 1:103–105. doi:10.2307/2269031
37. Axe DD, Dixon BW, Lu P (2008) Stylus: A system for evolutionary experimentation based on a protein/proteome model with non-arbitrary functional constraints. PloS one 3:e2246. doi:10.1371/journal.pone.0002246
38. Axe DD, Ewert W (2016) Stylus Experiments Made Easy–A Free App for Personal Computers. BIO-Complexity 2016. doi:10.5048/BIO-C.2016.1
39. Schrödinger E (1944) What Is Life? The Physical Aspect of the Living Cell and Mind. Cambridge University Press (Cambridge).
40. Gleick J (2011) The Information: A History, a Theory, a Flood. Pantheon Books (New York). Ch. 9.
41. Milosavljević A (1995) Discovering dependencies via algorithmic mutual information: A case study in dna sequence comparisons. Machine Learning 21:35–50. doi:10.1007/BF00993378
42. English T (2015). Deobfuscating a theorem of Ewert, Marks, and Dembski. https://bit.ly/2JNcuHr. Accessed: 2018/04/15.