Functionspaces, simplicity and curve fitting

Thomas Bonk[1]

Abstract

The number of adjustable parameters in a model or hypothesis is often taken as the formal expression of its simplicity. I take issue with this `definition´ and argue that comparative simplicity has a quasi-empirical measure, reflecting experts' judgements who track past use of a model-type in or across domains. Since models are represented by restricted sets of functions in a suitable space, formally speaking, a general `measure of simplicity´ may be defined implicitly for the elements of a function space. This paper sketches such a framework starting from intuitive constraints. It is shown how experts' judgements feed into this framework and how the usual definition can be recovered. A theorem by H. Akaike in the theory of model-choice has recently been used to shine new light on the relationship between the demand for simplicity and empirical success, or even `truth´. The approach favored here permits an alternative answer based on a reliabilist account of justification: if judgements of simplicity track past successful use of a model-type comparative simplicity is evidential and inductive.

Keywords Simplicity · Induction · Confirmation · Bayesianism · Model selection · AIC

What makes one hypothesis simpler than another? Does simplicity have *evidential* value, i.e. is it *rational* to believe (or accept) the simpler of two hypotheses, other things being equal? In the words of Forster and Sober: "the fundamental issue is to understand what simplicity has to do with truth." (1994, p.5) Many scientists, anecdotal evidence suggests, think the simpler hypothesis or model is evidentially privileged, and various *Bayesian* and *non-Bayesian* accounts aim to underwrite the

✉ Thomas Bonk
  bonk@lrz.uni-muenchen.de

1 Fakultät für Philosophie, Wissenschaftstheorie und Religionswissenschaft Ludwig-Maximilians-Universität, Geschwister-Scholl-Platz 1, D-80539 München, Germany

practitioners' judgment. Some philosophers grant simplicity at best a pragmatic, heuristic value in the pursuit of science (Quine, 1960) or remain profoundly skeptical about the very notion of simplicity (as in H. Putnam's statement: "I hate talk of 'simplicity'."; see also Howson & Urbach 1989).

Despite deep disagreements on how to account for the *evidential* nature of simplicity (if at all), there is now broad agreement on how to rank certain classes of quantitative hypotheses according to their simplicity. Objective *Bayesians* (Bandyopadhyay et al., 1996, 1999, 2014) and some *non-Bayesians* (Quine, 1960; Popper, 1980; Forster & Sober, 1994) agree in holding that the simplicity of quantitative functions between observables is – if not defined by – at least "measured" by the number of their adjustable parameters: the less – the simpler. With respect to two polynomials Bandyopadhyay et al. for instance express the "formal" notion of simplicity thus: "H1 is simpler than H2 because it has fewer adjustable parameters." (Bandyopadhyay et al. 2014; see also 1999 S398).

Addressing first the question regarding the definition of simplicity for quantitative hypotheses in Sect. 2 I begin by questioning the adequacy of the "measurement thesis" and point to a lack of independent reasons for it. Simplicity of a family of quantitative hypotheses, I suggest, varies with and "tracks" successful past use of the hypothesis. In Sects. 3 and 4 a formal (axiomatic) account of degree of simplicity in quantitative hypotheses based on elementary properties of function spaces to which the hypothesis' mathematical representations belong is proposed. The resulting simplicity ordering of families of hypotheses is compatible with, but independent of an ordering based on numbers of parameters. Any concrete simplicity ordering essentially depends in the present approach on the deliveries of experience and scientific practice which inform the scientist about the relative empirical success of a family in similar experimental circumstances. The discussion is restricted to parametric hypotheses in deterministic contexts: no effort is made to adapt it to richer treatments that include measurement error or to broader issues like simplicity as a feature of languages (cp. Goodman 1955), of theories or procedures, or to make connections with investigations of algorithmic or informational complexity (see f.e. Dasgupta 2011).

Turning to the second question regarding the evidential import of simplicity, in Sect. 5, it is shown how the novel account of simplicity sketched here offers the prospect of an independent *non-probabilistic* justification for the claim that the simpler hypothesis of a lot is the rational choice.

1 Degree of simplicity and empirical success

Why does it seem to many so plausible that the number of free parameters of a family of quantitative hypotheses "measures" its simplicity? As Forster observes: "So, the definition of simplicity is not a source of major disagreement." (Forster, 2001, p.90) Perhaps it is not a matter of major disagreement, but it should be one.

Popper's enthusiasm for this particular measure has a normative background: Popper held that the number of adjustable parameters is inversely related to the testability (and falsifiability) of a hypothesis (Popper, 1980, p. 140f). Hence, *if* simplicity of a function were measured by the number of free parameters, than *higher* simplicity

is proportional to greater falsifiability. This is a pleasing conclusion ("our theory explains *why simplicity is so highly desirable.*" loc.cit. p. 142), but logically it does not help to establish the premise.

Forster and Sober hold that "The simplicity of a family of hypotheses (referred to by statisticians as a model) is measured by the number of adjustable parameters" (1994, p. 31). The "goodness of fit" of a parametric model in repeated measurements of a quantity (whose error is normally distributed) can be measured by its expected log-likelihood, or in Forster and Sober's terminology by its "predictive accuracy": the higher – the better. This quantity in turn decreases linearly with the number of the model's adjustable parameters. Consequently, models with the lesser number of free parameters are expected to do better in repeated experiments, other things being equal. Hence, *if* simplicity were measured by the number of free parameters, simpler models are rationally to be preferred in curve-fitting settings. Again, the conditional logically does not help to establish the premise that simplicity *is* generally and properly "measured" by the number of adjustable parameters of a model. Lacking a proper argument to this effect, claims that AIC finally vindicates the role of simplicity considerations in inductive reasoning (Forster & Sober, 1994, p.11) are premature.

More recently, Bandyopadhyay et al. too assert the measurement thesis of simplicity: "Recall our formal approach, where H_1 is simpler than H_2 because it has fewer adjustable parameters." (2014, Sect. 3.2) In their argument for an objective constraint on a priori probabilities in adjudicating between competing hypotheses, they go as far to suggest that "simplicity is a traditional part of the *a priori* basis for inference" (loc.cit.). I doubt that warranted inference requires an a priori basis, except perhaps in a vague psychological sense of "a priori". However that may be, the authors adopt the measurement thesis without offering a reason. On first sight, objective Bayesians should not care *which* formal notion of simplicity is at work, as long as it sufficiently motivates the assignment of priors to alternative models. But the account of simplicity matters, since as Sober (2000, p.436) pointed out, assigning a higher prior degree of subjective belief to the simpler model requires a justification – which is difficult to come by if simplicity is nothing more than the number of a model's adjustable parameters.

Forster & Sober (1994) carefully use "measure" of simplicity instead "definition" of simplicity. The measurement thesis cannot be a definition of (comparative) simplicity in the desired way, because the notion of "adjustable parameter" in itself is slippery (see the next paragraph) and because the measurement thesis may well turn out to be false, as I will argue.

For one, note that experts and laymen judge the comparative simplicity of a pair of functions *prior* to counting parameters. The function $sin(x)$ is simpler than the function $\sin(x) + \cos(x)$, the linear function is simpler that the quadratic etc. Simplicity must be an essentially parameter-*independent* quality. Many competing models studied in Burnham and Anderson (2002) have the same number of adjustable parameters, without by the way impeding the applicability of AIC. For instance, biologists studying the growth in length of female salmon consider these two models among others: $L(t) = L_\infty [1 - \exp(-k(t - t_0))]$ and $L(t) = L_\infty \exp(-\exp(-k(t - t_0)))$ with two model parameters k and L_∞ (each one is combined with four sub-models; Burnham and Anderson 2002, p. 142). Insofar as these models do not strike one as

equally simple, their relative simplicity must reflect (other) structural differences. Consequently, the notions "degree of simplicity" and "number of parameters" are conceptually distinct.

For many purposes the concept of the number of free, adjustable parameters of a family of functions is clear enough, yet counting and individuating free parameters of a quantitative hypothesis is not always a straight forward affair. The number of free parameters is a surprisingly elastic quantity: does $y_1 = \frac{a}{r^2}$ have *one* adjustable parameter while $y_1^* = a \cdot \frac{b}{r^2}$ has *two*? The difficulty vanishes by interpreting the product „$a \cdot b$" as *one* parameter (a move advocated by Sober 2000 p. 439). Reparametrization is formally possible, but of doubtful physical significance and methodological virtue. If a denotes, say, the mass of body 1 and b denotes the mass of body 2, should their product count as *one* genuine quantity? In Newton's law of universal gravity, for instance, this is not the case. Physical interpretation and certain formal characteristics *are* the preferred means for individuating and counting parameters. It is *ad-hoc* to treat distinct, physically differently interpreted parameters as *one* for no other purpose than fitting this class of cases to a preferred meta-theory of model-selection. Bandyopadhyay et al (2014) raise this issue, but instead of concluding that the measurement thesis is false, they severely restrict its application and resort to informal "pragmatic" criteria *of simplicity* for a very large class of cases.

Before turning to an alternative account of simplicity in Sect. 3, I will briefly look at one of the "pragmatic factors" that are supposed to guide judgements of simplicity, namely the degree to which one family of functions is easier "to handle" than a competing alternative (Bandyopadhyay et al. 2014 Sect. 3.2.): "Whether an equation is easy to handle plays a vital role in theory choice, and hence in the assignment of prior probabilities to theories. Working scientists rely on this reason frequently".

`Ease of manipulation' (cp. Quine 1960) is an aspect of the everyday use of "simple": it is easier to play checkers than chess, the former being the simpler game. A straight line segment is easier to draw unaided than a circle, the former being the simpler geometric figure. Similarly perhaps, the linear function is easier to calculate, to differentiate, to integrate, to keep "before" the mind and to remember than other functional relationships. Nevertheless, simplicity as `ease of manipulation' hardly "plays a vital role in theory choice" for three reasons. (a) Given the routine use of computer algebra systems nowadays (f.e. Mathematica) differences in formal manipulability between hypotheses of the kind used in modeling contexts are neglectable. (b) `Ease of manipulation´ is a matter of experience, skill and kind and degree of formalization. It depends on *who* is doing the manipulation and (historically speaking) when. On this account "anything goes" while most methodologists treat simplicity as an *objective formal feature* of a hypothesis or model or reference class of models. (c) `Manipulability´ cannot explain the *rationality* of choosing the simplest among a lot of alternative hypotheses for predictive purposes. Suppose we judge manipulability of a function by the ease the world's leading mathematician X displays in doing the relevant mathematics. Why should a biologist, considering the growth pattern of salmon, say, trust a methodological rule like "assign the model that X can most easily manipulate the highest probability" in order to point her reliably to the truth or to novel predictions? Nothing in the concept of formal manipulability has any connection to how the *world* is or might be.

I conclude that "being easier to handle" is unlikely to be a pragmatic operating factor in simplicity judgements as far as the modern sciences are concerned.

Judgments of (comparative) simplicity may be an outgrow or result of an *inborn disposition*. Particularly the way perception of shapes functions may manifest such a disposition (R. Arnheim). Subjects presented with, say, 10 points at equal distance from an unmarked center, spontaneously perceive the points as lying on the circumference of a *circle*, not as marking a 10-pointed irregular star. Yet, granted such a perceptual disposition exists (suggested by Quine 1960), it is hard to see how this feature operating on a basic *non-cognitive* level can account for, or even justify, inductive reasoning in advanced quantitative sciences of the sort considered here.

Having critically reviewed one prominent formal and two informal criteria, I turn to a more promising source of simplicity judgments. If one looks over well-worn examples of "simple" and "not so simple" phenomenological laws once again, then one conspicuous difference between them is the degree to which they are common. High-exponent, multi-summand and multi-bracketed families of functions are "uncommon" in the sense of being *rare* birds in the sciences: few, if any, known phenomenological laws have a really complex form. The family of linear functions, on the other hand, is ubiquitous and has been used frequently and successfully for inter- und extrapolating data. A large number of elementary quantities varies linearly with each other: on small intervals without exceptions (every functional dependence between two quantities can be *locally* linearized under mild formal requirements) and sometimes over the whole range. The dependence between distance traversed and time elapsed is the best known example; suffice it to say that there is no area of quantitative science that has not discovered a linear phenomenological law. True, the family of "parabolas" is as familiar to the practitioner and pupil alike as the linear family, still the latter has been vastly *more frequently applied successfully* than the former or the "cubic", and so on.

An electrical engineer who (today) observes oscillatory phenomena, naturally represents them in terms of *sin-* or *cos-* dependencies and chooses between certain forms of amplitudes. This type of transcendental functions has been applied hugely successfully in the past, particularly in cases where the underlying processes are of electromagnetic nature. It is no coincidence that trigonometric dependencies are now classified with the most simple families. Experts have gotten used to them and they are widely taught. Similar things can be said about the family of power laws, (now) universally used to model complex systems from astronomy to economy.

Being "common" by itself is of little relevance, except as a crude measure of the *frequency of application*, that is: of successful applications of a class of functions in point of successful prognosis, useful definitions and "absorption" into theories (this is somewhat analogue to the "entrenchment" of a predicate in Goodman (1983 p. 94f): certain families of functions are entrenched). Hence there is reason for the claim, that differences in the intuitive degree of simplicity of a quantitative hypothesis, or the family to which the hypothesis belongs, reflect the (perception of) frequency of successful past application. Very complex functional dependencies between observables are rare, if at all used successfully in the sciences; and *if* a complex form is used successfully, it is hardly *transferable* from one domain of application to another (as the linear, the exponentials, etc. are). Differences in the estimate of relative simplicity

between alternative hypotheses (by a scientist) reflect differences in the estimate of relative frequency of successful employment.

Talk of "frequency of successful use" is vague. It seems naïve to hope that one can quantify that notion by counting the number of successful applications and the number of not-so-successful applications of a family, and dividing by the total number of (known) applications. Is the counting restricted to a subject matter, an area of research, a discipline, or is it cross-disciplinary?

The concept of "frequency of successful application" is vague but not ill-defined. Absolute frequencies (of successful fit) are difficult to determine and to determine with any certainty and confidence, yet in practice the scientist may rely on reasonable estimates of the frequency. It is an integral part of the *training of the scientist* to learn which family of function works for what type of problem and which doesn't. In any case, what matters are less estimates of the absolute (historic) frequencies, but the *relative proportions* of such frequencies for alternative functions that are being considered relevant in pursuit of a fitting model. Relevant information for this task is much easier to get by. Judgments of simplicity thus sum up in an informal way relevant information for the scientist: simplicity rankings track those frequencies, albeit more or less closely. (A fuller account should include - beside the "frequency of successful fit" - the degree to which a family has been transferred from one field of application to another.) In the majority of cases in which neither the scientist's personal experience nor her or any other "disciplinary matrix" provides for such frequencies their value is, strictly speaking, uniformly *zero*. This consequence is unnecessarily strict in view of the following consideration. Scientists cannot anticipate future developments: a complex exotic family of functions may one day turn out to be an adequate representation for the data in one area. Simplicity rankings of parametric families may track *actual* (historic) frequencies, as I have suggested so far, but may also be taken instead as forward *estimates* of *true* frequencies in the light of past use. As a matter of policy and open-mindedness the scientist's estimate for the true frequency should then never be set to zero. The choice of a sufficiently small *non-zero* value, subject to further consistency requirements (see below), will suffice for all deliberations. For the expository purpose of this section this point is left open.

These remarks do not answer objections to the notion of "frequency of success" fully, but I hope they help to make the thesis that simplicity of quantitative hypothesis varies with the frequency of successful applications ("tracking"), both more clear and plausible. Note two consequences. First, if the "tracking thesis" is true there need not be any *intrinsic* connection between the value of exponents, say, and the degree of simplicity of a function. *Any* function could have succeeded in the past (if nature favored it and actual scientific practice took the right turn), and hence become "simple" in the eyes of practitioners *regardless* of the number of its adjustable parameters, the absolute values of its exponents (if any), mean curvature or similar characteristics of functions. Second, judgments of simplicity by scientists on this account will change with the course of scientific history and practice. One may expect, for instance, that such judgments were generally less firm and had *less authority* (were less evidential) in the period of the rise of modern quantitative science, since there were few useful "frequencies" to track. One may expect, furthermore, that simplicity judgments differ both in firmness and with regard to rankings between disciplines,

and from research area to research area (though there may be convergence). A cursory look at the history of quantification in the sciences seems to bear out that simplicity was not an active criterion in formally representing sets of data, Occam's dictum not withstanding. The nearest case is the debate over the superiority of the heliocentric system. The increased "simplicity" (if measured by the number of parameters) of Copernicus's system, however, did nothing to persuade a scrupulous observer like Tycho Brahe of its truth, nor should it have done so *pace* Forster & Sober (1994, p.14). On the other hand, the continuous preference for circles and circular geometric representations (witness Galilei's belief that a falling body describes the arc of a circle, Dijksterhuis 1981, p. 349f) may have had less to do with an enduring attachment to Platonic conceptions of perfection, but with the perception that the geometric circle had an impressive empirical track-record as an instrument of data representation for planetary orbits.

Summing up: In the curve-fitting or model-selection context the formal concept of simplicity should not be reduced to "number of adjustable parameters", hence AIC does not vindicate simplicity as a guide to truth. Expert's judgements of comparative simplicity track past successful use, and although pragmatic factors may be at work in guiding scientists' judgements, "ease of manipulation" of formulas is not among them.

2 Degrees of simplicity for families of functions

In this section I outline a general method of assigning arbitrary functional dependencies (in a given space of functions) a degree or index of simplicity in a consistent and unique fashion. As it turns out, simplicity thus defined and certain features of families of functions, like the absolute value of exponents, are not necessarily linked. The previous section's suggestion that simplicity tracks frequency of successful past use conveniently complements the method (see Sect. 4).

Starting point is the observation that one tends to judge the sum of two functions to be less simple than each of them separately, provided the functions are in a sense independent of each other. The comparison of linear with quadratic functions (models) and higher-order polynomials in point of simplicity in curve-fitting contexts rests on an intuitive "rule of composition": a sum or product tends to produce more complex functions. Concepts like "composition" and "independence" have a natural and precise interpretation in the mathematics of function spaces, the "reservoir" of functional dependencies (hypotheses) between observables.

For the following only elementary properties of function spaces as vector spaces are needed. The space considered is complete, normed and has a (Schauder) basis, that is there is countable sequence of elements of the space such that all elements of the function space can be represented uniquely as a linear sum of independent elements of the basis. In particular, $S = L^2((a, b); \mathbb{R})$, the space of real square integrable function on a suitable interval (a,b) will do. S is equipped with an inner (weighted) product giving rise to the L^2 – norm $\|.\|_2$ and is convenient for representing quantitative relationships and models between observables. Widely used systems of (orthonormal) bases for the vector space S are *Legendre*-polynomials and the trigonometric

system *Fourier's*. The existence of a suitable set of specific functions in S that *spans* a set of candidate models, finite in number, is sufficient for the present purpose. The spanning set is usually finite and suggests itself by the models under investigation or perhaps by a look at the solution space of a differential equation.

The structure of S invites one to think of its elements as "composed" in the sense of being expressed (represented) as a sum of elements of the chosen, fixed basis or span. The elements of the basis in turn are not so composed, are not decomposable into more "elementary" entities of the space: they are intuitively the *simplest elements* of S. Let the letter "c" denote the *index* or *degree* of simplicity of an arbitrary function f in S relative to a fixed basis B ($f = \sum_k a_k \cdot f_k$, $\{f_k\}_{k \in \mathbb{N}}$ is a basis for f). The index satisfies the following conditions:

1) c(-) is a positive number in the interval [0,1] ("1=maximally simple").
2) Elements f of S, that only differ in the value of those coefficients of the development $f = \sum_k a_k \cdot f_k$, which are non-zero, have the same degree of simplicity.

The first condition sets upper and lower bounds for c(-). The second condition says, that c(-) has the same value on certain classes of functions. For instance, the polynomials $y_1 = x^2 + 3x^3$ and $y_2 = 25x^2 - 103\ x^3$ in $L^2((a,b);\mathbb{R})$ (or in C(a,b)) are intuitively speaking equally simple, and this is what the second condition intends to capture. Classes of functions, whose members have the same zero coefficients, like y_1 and y_2, are equivalence classes and together with the class consisting of the "zero-element" of S form a partition of S.

The second condition is perhaps controversial: Jeffreys proposed to rank differential *equations* (in one real variable) according to the absolute *values* of their coefficients, apart from degree and order (Jeffreys, 1961, p. 47; cp. Howson, 1988). It is the class, however, not an individual member, that is assigned an index c. Condition 2 *is* problematic but for a reason that becomes evident in connection with functions that cannot be represented by a *finite* sum in the chosen basis; more on this below.

The third condition expresses the "composition property" introduced a few paragraphs above. It says that a function which is a sum of more than one element of the basis cannot be simpler than any combination of those elements of the basis B, that enter with non-zero coefficients into the development of the function. The elements of B are to be the "simplest" elements of the space S.

3) The index c of an arbitrary element f of S is always smaller or equal to the product of the indices of the elements of the basis of f in S (with fixed B).

In Sect. 4 I consider a final requirement.

The following examples of value distributions for c(-) are meant to illustrate the consequences (and consistency) of the "axioms" stated. For this purpose suppose the model f in question can be represented in the vector space of polynomials of degree N, P_N, spanned by the set of monomials $\{f_k\}_{k \leq N}$.

Let c(-) be defined on P_N as follows:

$$c(f) = \begin{cases} (2^{\frac{1}{N}} - 1)^n & \text{where } n \text{ is the number of elements of the span} \\ & \text{that enter the development of } f \text{ with non-zero coefficients} \\ 0 & \text{if } f \text{ is the "zero-element"} \end{cases}$$

The index satisfies all three conditions. c is in [0,1]. c(f) is independent of the absolute values of the (non-zero) coefficients of f. The product of the indices of the elements of the span that enter the development of f with non-zero coefficients is larger or equal to c(f). (Note, if the subspace is partitioned as suggested, and a representative g is selected from each of the $2^N - 1$ partitions, than the indices of the representatives sum up to 1.) Besides the finiteness of P_N, the example is limited in an important respect: it represents all elements of B as equally simple. These restrictions on c(-) are lifted easily:

Let $f_k := x^k (k \in \mathbb{N}_0)$, then the set $\{f_k\}$ spans the space of all polynomials P on the interval [a,b].

Let the indices of the elements of the set $\{f_k\}$, $c(f_k) = c_k$, be in [0, 1] such that. $\prod_{k=0}^{\infty} (c_k + 1) = 2$; and define for f in P:

$$c(f) = \begin{cases} \Pi_k c(f_k) & \text{where } k \text{ ranges over the indices of those and only those elements of the span} \\ & \text{that enter the development of } f \text{ with non-zero coefficients} \\ 0 & \text{if } f \text{ is the "zero-element"} \end{cases}$$

The first two conditions are evidently satisfied. So is the third condition: the product of the indices of those elements of the span which appear in f with non-zero coefficients, $c(f_k)$, is equal or larger than the index of f: $\prod_{k=0}^{N} c(f_k) \geq c(f)$, from the definition of c(-). If f in P is an infinite power series, then $\lim_{N \to \infty} \prod_{k=0}^{N} c(f_k)$ converges to 0. In this case c(f) =0, and the third requirement is again satisfied.

Any f in P that is represented by an infinite power series has the same index c=0 − f is *maximally not simple*. For instance, the degree of simplicity of the local Taylor expansion of $sin(x)$ in (a,b) has index 0. This is an apparently *counterintuitive* result: *sin* and *cos* are among the most simple functions known. The reply is that *sin* is indeed *not* simple *if* its representation is an infinite power series. Conversely, since the map sin(x) appears "simple" for the representation of certain phenomena, a polynomial basis is just *not adequate* for describing those phenomena. (Equivalent terms need not appear equally simple.) Hermite polynomials turn up not only in quantum mechanics of the harmonic oscillator but also rather unexpectedly in models of the population density of kangaroos (Burnham Anderson 2002, p.257). The next section explores the question of how a scientist may choose a space of "plausible" models (expressed as a "basis" or a spanning set) in a given experimental context.

In concluding the section it is worthwhile pointing out an instructive transformation of the index c(f):

$$\zeta(f) =_{\text{DF}} - log\, c(f)$$

hence ζ in [0, ∞ [. Condition 3) gives the inequality:

$$\zeta(f) \geq \sum_k \zeta(f_k),$$

where the sum is over all those elements of the basis *B (of L^2)* that enter the finite development of *f* with non-zero coefficients. A natural interpretation of ζ is as a degree of *complexity* of *f* relative to the chosen basis *B*. The inequality says intuitively correct: the degree of complexity of a sum of independent functions is greater or equal to the sum of the degree of complexity of each of the independent functions. The definition allows a comparison with Jeffreys' "complexity of an equation".

3 On choosing a "model space"

The space of models reflects in a precise way background information and assumptions imported into modeling the phenomenon in question. The choice of a function space, a basis or a span, represents this information. What can be said generally about choosing a basis? Which rules govern the assignment of degrees of simplicity to the basis?

Three factors guide the choice of a mathematical basis for the quantitative representation of a phenomenon: (a) the data themselves; (b) considerations of what has worked well for similar looking data-sets; (c) theoretical expectations regarding the causes (the explanation) of the phenomena (set of measurements) in question.

"Raw data" usually suggest or exclude certain *general* requirements for mathematical representations: periodicity, boundedness from above, monotonicity, symmetries, etc. The curve-fitter will tend to choose a basis (and "basic" models) accordingly. She builds up model spaces from families of functions that have "proved" themselves in the past (although perhaps in different contexts), guided by the considerations mentioned. As noted above, this repertoire is changing and perhaps growing with time and practice. The hope is that any of these by itself will suffice for the task, or else a finite *combination* of them will do the job: a linear sum, or a more general operation (like summing up their inverses). The *fewer* independent functions are needed for interpolation and prognosis the "simpler" (and convincing) the result will appear. Finally, *theoretical explanations* of the phenomenon usually suggest suitable candidates for the mathematical basis of description: if the curve-fitter believes she is observing a decay process she may avoid trigonometric functions and opt for linear combinations of exponentials. Pragmatic considerations like these only partially determine the choice of a basis for *S* since an *infinite* sequence of independent functions is required. New and novel data may prompt a switch to a new or differently structured model space at any time. While the orthodox "paucity-of-parameters"-criterion suggest an invariant, universal measure of simplicity, the present account takes care of the fact that simplicity orderings may change over time even without prompt by an expanding data set.

Secondly, how are degrees of simplicity assigned to the elements of the basis once it is chosen? In the light of what was argued for in the previous sections I suggest that the degrees of simplicity of the elements of the basis track the frequency of past successful application. This suggestion gives an "experiential" meaning to the

$c(f_k)$. It comes at a price: most elements of any basis will *never* have been "applied successfully" at all (that is directly, as an integral law). The frequency of successful use is *zero*. In consequence, a finite segment only of any (model) space S has *non-zero* degrees of simplicity. (The issue is more complicated; see Ellis 1966 about dependencies between the form of phenomenological laws and scaling properties of the observables. To avoid a threatening conventionalism, I assume here that the set of observables along with their "scales" have been fixed prior to any inductive generalization.)

This feature of the present account of simplicity actually fits scientific practice: if the data force the curve-fitter "to go beyond" what is expressible as, say, a linear sum of a few well-tested functions, so that c(*f*)=0 , then she will reject the initial model space – and hence theoretical assumptions governing its choice – as unsuitable. For instance, if she finds that the growing set of data can be better expressed directly by the trigonometric *sin* than by any *finite*, fine-tuned sum of polynomials, she will switch to a basis that includes *sin*; hence c(*f*) ≠ 0.

However, given that determining the frequency of successful use of a family of functions is not an exact science, it is perhaps sagacious to adopt the maxim *never* to set $c(f_k)$ equal to zero ($f_k \in B$), and always to work with an arbitrary small but non-zero "place holder" value. In consequence, *all f* in the present space of models (with the exception of the zero-element and those with infinite developments in B) have non-zero degrees of simplicity.

For the purpose of illustration, consider $y_1 = a\sin(bx)$ and $y_2 = a\left(1 - \exp\left(-x^2\right)\right)$ · sin(b·x) as competing two-parameter models. Both are in $S = L^2((-1,1); \mathbb{R})$. If the basis of S is provided by the set of Fourier functions, then y_1 has a finite development, but not y_2. Consequently, y_2 appears as less simple than y_1 in accordance with intuition. If instead the set, say, of Legendre Polynomials is chosen as a orthonormal basis of S, than y_1 and y_2 appear as equally simple, though perhaps with degree of simplicity zero. The current proposal shows why the number of adjustable parameters is not always and not *per se* a useful measure of simplicity.

Finally, given the interpretation of c(-) as frequencies it is a requirement that the $c(f_k)$ *sum up to 1*. Of course, frequency counting requires subtle classifications and groupings of phenomena in "relevant" and "irrelevant" and "similar" and "non-similar" classes (the vagaries of which enhance the point about not letting $c(f_k)$ take the value zero). The "summation condition", announced earlier as a final condition on c(-), says:

4) Let S be partitioned (π) in the way indicated under 2); then the indices of the representatives g_k selected from each partition in π sum up to 1: $\sum_{k \in \pi} c(g_k) = 1$.

Both examples toward the end of Sect. 3 satisfy the summation condition. The condition is motivated solely by the particular "empirical" interpretation adopted here for the index c . It is formally independent of 1) to 3) as the core of an "axiomatic" characterization of a simplicity index for functions.

(Note that the summation condition and the 3. condition (above) work against each other in infinite function spaces: $\sum_{k \in \pi} c(g_k) = 1 \Rightarrow c(g_k) \to 0$ for k→∞ ($g_k \in B$) as a

necessary condition for convergence of the sum. Hence $\lim_{k\to\infty}\prod_k c\,(g_k)=0$, for any f in S that has no *finite* development.)

4 Simplicity and induction

The present account of simplicity may serve as a self-contained, *non-probabilistic* basis for justifying the reliance on the simplest hypothesis as a guide to the future. Besides, it helps overcoming difficulties in a Bayesian account of the "curve-fitting" problem. I begin with the second claim.

Simplicity orderings are an important factor in objectivist Bayesian probability kinematics. I have commented in Sect. 2 on two methods proposed (Jeffreys, 1961, Bandyopadhyay et al. 2014, 1999) in order fix the priors by ranking competing hypotheses in the light of counting "adjustable parameters" or by "ease of use". Neither solves the problem of justifying belief in the simplest hypothesis among a lot of alternatives in the sense of giving an independent reason (or explanation) for why the simpler one should be the more probable one. By assigning the simplest hypothesis the highest probability Bayesians *presuppose* a (hypothetical) "solution" of the problem in question. The present account fills in the gap: suppose there are (only) two "competing" models f and g in S, with basis B, that both deductively fit the data up to now. The curve-fitter's background information (W) includes information about *relative* frequencies of successful applications for f and g, and she evaluates the a priori probabilities of both in the light of *W*. Let h denote the relative frequencies of successful applications in her "field" and making use of David Lewis' "Principal Principle" one finds.:

p(f | W)=p(f | W'& h(f))=h(f)=c(f).
p(g | W)=p(g | W' & h(g))=h(g)=c(g).

Hence p(f | W)>p(g |W) if and only if c(f)>c(g), that is if f is simpler than g. The same argument goes through in the case of countable many alternatives. No further conditions on *a priori* probabilities are called for.

However, the present treatment of simplicity furnishes an independent, non-probabilistic justification for believing or "accepting" the simplest of a lot. Suppose then, that relative to S and B and a set of data, a simplicity function c(-) is given. In order to justify the choice of the most simple family as rationally superior to any other choice, one would need a claim (T) like:

"The simplest family in S, compatible with the data, is the mostlikeliest to be true."

assuming that (controversially perhaps) scientists pursue truth. Alternatively: "The simplest family of functions in S, compatible with the data, is the most rational to believe (to `choose´)." (The difference does not matter for the question at stake here.)

The key to justifying (T) is that simplicity judgments track the frequency of past successful applications (Sect. 2), hence (T) is schematically equivalent to (T*):

> "The most frequently successfully applied family of functions in *S*, compatible with the data, is the most likeliest to be true."

(T*) says that *past successful applications* are a *reliable* guide to the truth in the choice of functional forms: family *f* (in S) has been "frequently successfully applied" in the past, so it will likely make a "successful fit" in the present, similar case too. This is a piece of garden-variety inductive reasoning, enumerative induction. It is the most simplistic of inductive modes, perhaps more often false than right (see Wagenitz 2003 for interesting examples). Nevertheless, many ingenious justifications for enumerative induction have been offered in the past. One sticks out as initially attractive and intuitive: enumerative induction itself has (frequently) led to true conclusions in the past, hence enumerative induction will likely lead to a true conclusion (or true predictions) in the present case as well. This sort of justification for enumerative induction is itself "enumerative" and appears badly circular. On a closer look the circularity is of a special kind ("rule circularity"), and set in the context of an *externalist theory* of justification, may not be vicious at all (Papineau, 1992; H. Mellor 1991). This is not the place to fill in the details and argue this case, but one point should have become clearer: *if* enumerative induction is justifiable, then comparative simplicity of a quantitative hypothesis has evidential value. Much the same can be said if one takes induction in the sciences to be broadly "eliminative" instead of (naively) enumerative. Papineau (1993, p.166) argues that "physical simplicity" is the key to eliminative induction:

> "if the constituents of the world are indeed characterized by the relevant kind of physical simplicity, then a methodology which uses observations to decide between alternatives with this kind of simplicity will for that reason be a reliable route to the truth."

Papineau leaves the content of the notion of "physical" simplicity, and hence the question of how this particular feature can take on such a crucial role, open. If simplicity is understood in the sense indicated in this paper this lacuna is filled and the account strengthened.

The task of justifying (T) is a version of the familiar problem of induction. If the considerations above are on the right track, one cannot well consistently embrace the rationality of inductive reasoning *and* be a sceptic about simplicity's evidential role (I suspect Putnam 1972, from which the quote at the beginning was taken, is a case in point). There is no *special* problem regarding the "evidential" value of simplicity of hypotheses.

References

Bandyopadhyay, P., Boik, R., & Basu, P. (1996). The curve fitting problem: A Bayesian approach. *Philosophy of Science*, *63*(supplement), S264–S272

Bandyopadhyay, P., & Boik R. (1999). The curve fitting problem: a Bayesian rejoinder. *Philosophy of Science*, *66*, S390–S402

Bandyopadhyay, P., Bennett, J. & Higgs, M. (2014). How to Undermine Underdetermination. *Foundations of Science*, *20*(2), 107–127

Burnham, K. P. & Anderson D. (2002). *A practical information-theoretic approach. Model selection and multimodel inference*. Springer

Dasgupta, A. (2011). Mathematical foundations of randomness. In: *Philosophy of Statistics*. North-Holland. pp. 641–710

Dijksterhuis, E. J. (1981). *The mechanization of the world picture. Pythagoras to Newton*. Princeton University Press

Ellis, B. (1966). *Basic concepts of measurement*. Cambridge University Press

Forster, M., & Sober, E. (1994). How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *The British Journal for the Philosophy of Science*, *45*, 1–35

Forster, M. (2001). The new science of simplicity. In A. Zellner H. Kreuzenkamp & M. McAleer (Eds.), *Simplicity, inference and modelling* (pp. 83–119). Cambridge University Press

Goodman, N. (1955). Axiomatic measurement of simplicity. *The Journal of Philosophy*, *LII*, 702–722

Goodman, N. (1983). *Fact, fiction and forecast*. Harvard University Press

Howson, C. (1988). On the consistency of Jeffreys's simplicity postulate, and its role in Bayesian inference. *The Philosophical Quarterly*, *38*(150), 68–83

Howson, C., & Urbach, P. (1989). *Scientific reasoning. The Bayesian approach*. Ill.: Open Court

Jeffreys, H. (1961). *Theory of probability* (3.ed.). Clarendon Press

Mellor, D. H. (1991). The warrant of induction. *Matters of metaphysics*. Cambridge University Press

Papineau, D. (1992). Reliabilism, induction and scepticism. *The Philosophical Quarterly*, *42*, 1–20

Papineau, D. (1993). *Philosophical naturalism*. Blackwell

Popper, K. R. (1980). *Logic of discovery* (10th ed.). Hutchinson

Putnam, H. (1972). Other minds. In R. Richard, & I. Scheffler (Eds.), *Logic and art. Essays in honor of Nelson Goodman.* 1972, Bobbs-Merrill

Quine, W. V. (1960). On simple theories in a complex world. In: *Ways of paradox* (rev.ed., 1976). Harvard University Press. pp. 228–245

Sober, E. (2000). Simplicity. In Newton-Smith (Ed.), *A companion to philosophy of science* (pp. 433–441). Blackwell: W.H

Wagenitz, G. (2003). Simplex sigillum veri, auch in der Biologie? *Nachrichten der Akademie der Wissenschaften zu Göttingen*, II. Math.-Physikalische Klasse. Göttingen: Vandenhoeck und Ruprecht. pp. 99–114