Trade-off Accounts of Laws of Nature as Meta-Empirical Hypotheses

Niels Linnemann[1] and Robert Michels[2,3]

[1] *University of Geneva*

[2] *Goethe University Frankfurt*

[3] *LanCog, Centre of Philosophy, University of Lisbon*

Draft version (24 August 2026)

Abstract

Naturalistic trade-off accounts of laws stand and fall with their signature claim that laws of nature offer the best trade-off among a given set of theoretical virtues. So far, proponents of such accounts have not provided adequate evidence for this claim. Existing arguments support at most the conclusion that particular regularities, or laws drawn from the natural sciences, plausibly offer better trade-offs than typical examples of non-laws. We argue for a shift towards conceiving of trade-off accounts as testable meta-empirical hypotheses. We first expose the shortcoming just identified and then introduce a testing methodology that allows one to corroborate various hypotheses about the relation between laws and trade-offs among theoretical virtues, including, in particular, necessary and progressively stronger implications of the signature claim that lawhood is connected to the best trade-offs.

1 Introduction

A popular view holds that laws of nature compress facts about the world while striking an appropriate balance among theoretical virtues such as simplicity and strength. Put vividly, laws are the kinds of statements that can fit on a T-shirt while still capturing the world around us remarkably well. Within this tradition, systems of laws—or, more generally, systems from which laws can be derived—are selected because they achieve the best trade-off among virtues such as explanatory strength and simplicity relative to competing systems. Prominent examples include Humean accounts such as Lewis's Best System Account (BSA) and its many variants,[1] related compression-theoretic approaches that need not be committed to either side of the Humean/non-Humean divide,[2] and Loewer's influential non-Humean modification of Lewis's BSA, the Package Deal Account (PDA).[3]

Our concern is with any trade-off account that aspires to capture the extension of the laws recognised by the sciences, especially physics. Most, if not all, of the accounts listed above share this aspiration. We are concerned with two related questions: (i) whether current versions of these accounts have demonstrated a match between optimal trade-offs and actual paradigm examples of laws, and (ii) whether such accounts can take their naturalism one step further by making their central trade-off claim empirically

[1]These include Jaag and Loew's (2020) and Schrenk's pragmatic accounts (2008), Callender and Cohen's Better Best System Account (2010), Dorst's Best Predictive System Account (2018; 2019), Hicks's Epistemic Role Account (2018), Gómez Sánchez's modal variant (2020), Wilhelm's Best-Utility System Account (2022), and Esfeld and Deckert's Super-Humean views of laws (2017).

[2]Braddon-Mitchell's "lossy laws" proposal models best-system selection on lossy data compression, trading off compression, scope, and accuracy (Braddon-Mitchell, 2001). Wheeler's algorithmic theory identifies laws with algorithms that compress nature's data, while emphasising that actual laws involve a context-sensitive trade-off between compression and informativeness (Wheeler, 2019).

[3]We follow the usual division between Humean and non-Humean accounts of laws (Loewer, 2007; Loew and Jaag, 2020). Humeans often present themselves as occupying the naturalistic side of this divide because, unlike non-Humeans, they purportedly avoid scientifically external philosophical posits such as primitive modal connections, powers, or governing necessities. Recent work by Loewer (2020), however—see also (Lam and Wüthrich, 2023)—raises doubts about whether the assumption of a categorical mosaic of properties is itself genuinely naturalistic.

testable. We suggest a negative answer to the first question and a positive answer to the second. Existing trade-off accounts have so far failed to conclusively establish that their signature thesis applies to real laws, but this gap can potentially be closed in a strictly naturalistic manner.

To make our case, we draw on recent work in automated scientific discovery, especially symbolic regression, and argue that these emerging methods provide the means to test particular versions of trade-off accounts. By 'testable', we mean that there are methods that can (a) confirm or disconfirm whether a candidate law, or a system containing it, embodies the right kind of trade-off among theoretical virtues and (b) thereby support a trade-off account that links this particular trade-off to the relevant law or system of laws. We call this *meta-empirical* testing because it occurs one level above the empirical testing of the laws themselves: empirical data confirm or disconfirm laws, whose structural features in turn confirm or disconfirm the proposed trade-off account.

More specifically, our argument consists of three principal claims. First, standard philosophical arguments establish only that paradigm laws offer better trade-offs than cherry-picked contrast cases, but provide no conclusive reason to think that these trade-offs are optimal. Second, any trade-off account of laws can (and should) be understood as a meta-empirical hypothesis about actual laws or systems of laws, to be tested in the context of actual physical theories rather than merely defended through philosophical argument. Third, where the target laws are mathematical expressions, symbolic regression provides a primary testing ground for such hypotheses.

2 Trade-off accounts as meta-empirical accounts of laws of nature

Philosophical accounts of laws are meant to characterise laws systematically and, in particular, to distinguish them from non-laws. We should therefore require such accounts to classify independently identified paradigm cases from the sciences—for example, laws from physical theories—as laws. Indeed, a substantial strand of the metaphysics of science literature confronts accounts of laws with purported paradigm cases from physics and the special sciences.[4]

An influential challenge grounded in scientific practice comes from Maudlin (2007). Maudlin identifies a particular class of time-evolution equations, which he calls FLOTEs, as the central laws of physics. He uses them as test cases for philosophical accounts and concludes that only primitivism adequately captures their structure. The challenge arises because physical laws are typically expressed as equations, often differential equations, whereas philosophical accounts commonly represent laws as universal statements of the logical form 'For all x, if x is F, then x is G'.

One response is to revise the formulation of the scientific paradigm cases. Friend (2016), for example, argues that laws represented by equations can, or must, also be understood as expressing universal generalisations of the form just mentioned. Such a reformulation can accommodate the necessitarian account of Armstrong (1983) or the dispositionalist account of Bird (2007), which assigns laws a closely related conditional form (p. 46). A second response is to adopt an account that remains independent of the detailed form in which laws appear in scientific practice. Modal approaches such as Lange's counterfactual account (Lange, 2005), for instance, operate at a sufficiently

[4] The metaphysics of quantum gravity provides a prominent example: several contributions assess accounts of laws in light of theories in which spacetime may not be fundamental (Lam and Wüthrich, 2023; Cinti and Sanchioni, 2021).

abstract level to avoid objections based on the specific form of physical laws, albeit at the cost of leaving their relationship to scientific practice less determinate.

Neither response is especially attractive to a naturalist: the first revises the scientific paradigm cases, while the second retreats to an account that may be flexible enough to accommodate functional laws but whose central lawhood criterion is somewhat disconnected from scientific practice. A more radical naturalist option is to relinquish a substantive analysis of lawhood and adopt a form of primitivism. This is the path taken by Maudlin's FLOTE-based view and by broader constraint-based accounts such as those of Adlam (2022), Chen and Goldstein (2022), and Meacham (2025). The latter treat the functional role of laws as constraining the space of nomic possibilities rather than dynamically producing the world, as on Maudlin's version of primitivism. Given well-known issues, such as the 'form of the laws' problem mentioned above, primitivist accounts may seem worthy of serious consideration. However, the associated refusal to provide, or even entertain the possibility of, a substantive metaphysical analysis of lawhood is a radical move that not all philosophers are willing to make.

A further option remains available. Naturalists can formulate a substantive yet comparatively deflationary hypothesis about the structure shared by scientific laws and test that hypothesis directly against paradigm cases, or so we will argue. This approach differs both from primitivism[5] and from approaches that tie their lawhood criterion to a metaphysical feature, such as modality, which at best only indirectly tracks structural features of actual scientific laws.

Trade-off accounts are the most prominent, and perhaps the only developed, instances of this further option. They claim that known laws, or the systems from which those laws follow, achieve an optimal trade-off among theoretical virtues, whereas non-laws,

[5]Maudlin's primitivism nevertheless includes a syntactic and structural claim: laws take the form of time-evolution equations, or FLOTEs. It therefore has a meta-empirically testable component, since the claim that FLOTEs are the laws of nature must be assessed against our actual physical theories.

or the corresponding systems, fail to do so. This structural claim is meant to explain the distinction between laws and merely law-like accidents. Because trade-off accounts compare systems of laws with alternative ways of systematising the same phenomena, their central claim can be investigated. To vindicate a particular trade-off account, one must show that the relevant laws or systems compare favourably with their competitors and genuinely achieve the proposed optimum.

So far, attempts to show that actual physical laws or systems of laws satisfy a trade-off account have largely relied on plausibility arguments. Such arguments typically contrast a paradigmatic non-law, associated with one extreme of a trade-off, with a genuine candidate law that is then said to exemplify a better or optimal balance.[6] In the following, we argue that these existing arguments are problematic in two ways.

Consider a standard argument used to illustrate the workings of the classical trade-off account, Lewis's BSA (see, e.g., Beebee (2000) on God's Big Book of Facts). A familiar system of equations—often from Newtonian mechanics—is presented as much simpler, yet not problematically weaker, than a trivial system consisting of a long list of particular facts. From this contrast, rather than from an explicit comparison with the full range of relevant alternatives, it is concluded that the Newtonian system achieves the *optimal* trade-off between strength and simplicity. The problem is that this argument establishes at most that the candidate offers a *better* trade-off than an intentionally poor contrast case; it does not establish that the candidate belongs among the *best* systems. That is, this sort of argument is insufficient to show that the candidate law at issue has the characteristic property of laws according to the account's own standards.

[6]There is a long tradition of questioning whether the virtues involved in the trade-off can be made operationally usable. This problem arises, for example, when very different systems employ different vocabularies relative to which simplicity cannot be evaluated in the same way; see Cohen and Callender (2009) and the references therein. There are also questions about what could count as an 'optimal' balance (Linnemann and Michels, forthcoming). Even if these difficulties are set aside, however, someone (or something) would still have to carry out the intersystem comparison.

Roberts (2008, 7) has tacitly acknowledged this argumentative gap when presenting it in matters of plausibility. Reiterating the standard argument (here for the domain of planetary motion), he concludes: "So it is *plausible* that the best system in this simplified fictional universe will consist of Kepler's laws and all their consequences; at any rate, this system will be a serious contender in the race for 'best system."'

Admittedly, Roberts is a critic of the BSA, but the same kind of argument is common among proponents of the theory.

Most proponents of the BSA (or its variants) do not highlight this argumentative gap or the resulting leap of faith. A noteworthy exception is, however, Gómez Sánchez, who writes:

> I cannot offer decisive proof that generalizations like 'All ravens are black' will be necessitated by the axioms of the best system. Similarly, proponents of the best systems account of fundamental laws cannot prove that Schrodinger's equation (or whatever equation will come to replace it) will feature in the best system. All I will offer is a plausibility argument that statements like these are either in the best system or entailed by it. (Gómez Sánchez (2020), p. 453.)

Gómez Sánchez is indeed more transparent about the nature of her arguments. They nevertheless only make it plausible that the law candidates she discusses offer a *good* trade-off (and, at most, a better one than some bad alternatives). A demonstration that this trade-off is not only good but also 'wins' a concrete contest for the *best* system is once more absent.[7]

Variants of the BSA and the compression accounts mentioned above generally share

[7]Her arguments explicitly focus on high-level generalisations and non-fundamental or special-science laws, rather than physical laws that have some claim to being fundamental. Our focus is on non-probabilistic laws, but claims about the place of probabilistic laws in the BSA are likewise supported by plausibility arguments of this kind; see, e.g., Fenton-Glynn (2019).

this weakness. The second weakness is that the arguments are merely arguments from plausibility. They provide no concrete demonstration that actual laws or systems from physics result from an optimal trade-off. Instead, they support the intuition that those laws perform better than an extreme and deliberately unappealing alternative. One might regard this evidential modesty as part of the epistemically risky character of adopting a metaphysical 'stance' (Chakravartty, 2017), within a paradigm in which plausibility arguments may seem to be all that one can reasonably provide. We will see below, however, that one can strive to do better.

The situation has not changed with recent expansions of the literature on trade-off accounts. Consider, again, the BSA and its variants. Several authors argue that physical laws are better captured once further virtues are included. Wilhelm (2022), for instance, seeks to include computational usefulness as another virtue in the trade-off, in part to prevent boundary conditions from acquiring the status of laws. By including computational usefulness, or so he claims, a system that contains no initial conditions can still score higher overall than a system that also contains the initial conditions. This argument may establish that the inclusion of initial conditions affects whether a system is optimal, but it does not show that any particular system, with or without such conditions, achieves the optimum. [8] Similar points apply to the cognitive-effort-minimising variant of the BSA's trade-off proposed by Loew and Jaag (2018); Chen (2022)'s 'Reformed Humeanism', which adds a graded notion of naturalness; the modal reconceptualisation of strength or informativeness developed by Gómez Sánchez (2020), together with a corresponding notion of accuracy and degrees of naturalness and simplicity; and Dorst's (2018) pragmatic expansion of the set of theoretical virtues included in the BSA's

[8] Conversely, critics argue that physical laws do not exhibit the trade-off proposed by the BSA because strength takes priority over simplicity (Roberts, 2008; Woodward, 2014). These arguments display a similar lack of detailed engagement with the relevant physical laws. Roberts (2008), for instance, appeals to methodological remarks by Newton and Einstein and to toy models involving systems of varying simplicity and strength. This reverses the usual BSA intuition pump without replacing it with a systematic comparison of actual candidate systems.

lawhood-indicating trade-off. Plausible reasons for taking additional theoretical virtues into account are presented, but no conclusive evidence has been given that the resulting trade-off delineates laws from non-laws in a way that matches actual paradigm cases in both categories.

To sum up, the existing literature provides plausibility arguments suggesting that laws offer a better trade-off among theoretical virtues than certain paradigm non-laws. These arguments may well support the intuition that laws can be thought of as the result of an optimal trade-off among theoretical virtues, but a concrete demonstration—even in a controlled toy scenario—that this signature thesis of trade-off accounts holds is still missing.This neglect is surprising because the BSA and related compression views are popular among philosophers and practitioners of physics who describe themselves as naturalists. The most likely explanation is practical: technical means for systematically comparing actual candidate laws with respect to their associated trade-offs have only recently come to prominence in the sciences. As we will argue in the next section, these methods provide a way to test trade-off accounts meta-empirically. Suitably specified, such accounts can be subjected to attempts at meta-empirical falsification. If they survive sustained attempts at falsification, they gain an empirical advantage over rival accounts—an advantage that should matter greatly to anyone seeking a naturalistic account of laws. As Linnemann and Michels (forthcoming) point out, proponents of such accounts could press this advantage to mount an argument against competing theories of laws: if there is corroborating evidence for a systematic link between optimal trade-offs and lawhood, then proponents of philosophical theories of laws that ignore this link have some serious explaining to do.

3 Testing trade-off accounts through symbolic regression and variable discovery

In this section, we introduce recent advances in automated scientific discovery, in particular on symbolic regression and variable discovery, and argue how these techniques open up avenues for testing trade-off accounts meta-empirically.

A general point first: Empirical tests can only ever build on data that cover limited regions of space and time, and any empirical confirmation carries inductive risk already for this reason. The nature of data that can be used for testing trade-off accounts differs thus strongly from that of the idealised Humean mosaic encompassing all of space and time, which provides the basis from which laws-cum-regularities are derived in Lewis's original version of the BSA. Our proposal of testing consequently applies more naturally to pragmatic or epistemic versions of the BSA—of which one can easily think of as data-oriented—rather than to Lewis's original metaphysical formulation; the latter is not incompatible with our proposal though, either.

We propose to test whether functionally expressed laws are best trade-offs by letting the machine systematically survey sensible functionally expressed laws for relevant data and determine thus the system or systems which are credibly best for accounting for that data. The primary available tool for this purpose is symbolic regression in which one searches across possible symbolic functions for fitting given data. The richer the library of basis functions, the more closely the search can approximate the range of expressions admissible in physics.[9] As Linnemann and Michels (forthcoming, §3) explain for a philosophical audience, symbolic regression recovers explicit mathematical expressions from data by applying an operationalised trade-off to candidate functions (see, e.g., Schmidt and Lipson (2009); Udrescu and Tegmark (2020); Udrescu et al.

[9] Cf. to standard regression where one fits data with a narrowly predetermined family such as linear or exponential functions.

(2020); Cranmer (2023)). These candidates are typically generated by genetic algorithms, supplemented by neural-network preprocessing and other heuristics (such as the exploitation of symmetries); a brute-force search would not be feasible and typically practically impossible. Admittedly, the systematic survey might miss candidates because of the probabilistic nature of the search process—a limitation controllable through increasing the amount of independent runs—but also due to other processes.

Current implementations typically compare candidate expressions in terms of some well-chosen measure of accuracy of fit and some measure of descriptive simplicity; other virtues could also be used, provided that they can be operationalised. The expressions that are not dominated by any others across the relevant virtues form a Pareto front. To illustrate the nature of the Pareto front in the standard case of an accuracy–simplicity trade-off, no expression on the frontier can be made more accurate without becoming more complex, or simpler without becoming less accurate. Known physical expressions have hereby been found to correspond to particular convex corners on this Pareto front, where accuracy improves sharply at a comparatively small loss in simplicity (Udrescu and Tegmark, 2020; Udrescu et al., 2020). A sufficiently rich data set may yield several such points, potentially representing distinct laws or approximations at different descriptive levels (we will discuss this more through the example in the next section). These results provide an initial form of empirical support for a suitably data-oriented version of the BSA: laws correspond to distinguished trade-off points within the space of possible systematisations of physical data (Linnemann and Michels, forthcoming).

It is important to distinguish Pareto-optimality from bestness of a trade-off. A Pareto-optimal expression is non-dominated; whether it represents a best trade-off depends on a further principle for balancing the relevant virtues. We will propose convex-corner status as a possible stronger criterion for distinguishing points within the Pareto front.

Symbolic regression presupposes that the relevant variables have already been identified. It therefore operates on data represented in a preselected 'coordinate system' (generally construed). Yet the representational framework within which laws are formulated arguably belongs to the systematisation itself. Recent technical developments make this requirement especially clear.Chen et al. (2022), for example, address this prior stage by learning fundamental state variables directly from raw video without prior knowledge of the system's physics. Their autoencoder allows for estimating the system's intrinsic dimension, roughly the number of latent coordinates required for reliable prediction, and identifying variables whose temporal stability can then be assessed. A natural further step would be to recover their dynamics as explicit symbolic equations. The combination of variable discovery and equation discovery through symbolic regression provides a fuller methodology for our programme. So, ideally, first, a system akin to that of Chen et al. (2022) estimates the intrinsic dimension of the recorded dynamics and learns compact latent variables sufficient for predicting well their evolution. This stage operationalises a trade-off between representational compactness and predictive adequacy. Second, symbolic regression is applied to the recovered variables to identify explicit equations governing them. Together, these stages could test from the ground up the hypothesis that the choice of physical variables and the laws relating them jointly optimises a recognisable simplicity–strength trade-off.[10]

For the moment, we focus on the second component which is sufficient for our immediate philosophical purpose: to improve on existing attempts to link optimal trade-offs with known laws. The first component extends the reach of the proposed

[10]In actual implementation, the two stages are more intertwined as not just the choice of dynamics depends on the choice of variables but also the appropriate choice of variables depends on the expected dynamics. The SINDy autoencoder (Champion et al., 2019) for instance looks simultaneously for variables and dynamics. (It takes in intrinsic dimension as a hyperparameter and works with a somewhat theory-laden predetermined function space whose members are only additively combinable.)

testing suite to practically 'theory-free' data sets involving unknown state variables (think of, e.g., the collection of video data). Without it—that is, by relying solely on symbolic regression—we are limited to the rediscovery of already known representations and correspondingly formatted suitable data sets. In the next section, we provide an example that likewise uses only the second component of the envisaged full testing suite.

The approach we propose departs significantly from how trade-off accounts are treated in the existing literature. There, they are usually presented and discussed only as generic claims to the effect that laws result from an optimal trade-off between systems based on a particular set of theoretical virtues. For the techniques we describe to be applicable, not only must the set of theoretical virtues be fixed, but the virtues must also be defined operationally. Once this has been done, the testing suite can be run on a given set of data. If the result aligns with the expected set of laws—that is, if the systems providing those laws are Pareto-optimal and, under the stronger hypothesis, occupy distinguished convex corners on the generated Pareto front—this provides meta-empirical evidence for the corresponding hypothesis. Repeatedly successful tests of increasingly specific trade-off accounts would also provide corroborating evidence for more generic trade-off accounts. For example, if several independently motivated operationalisations of simplicity and strength repeatedly produced the expected laws, that result would support the robustness of the chosen virtues. More generally, each successful run provides some corroboration of the relevant Pareto-optimality or convex-corner hypothesis and thereby of a necessary or stronger implication of the signature claim that laws belong to the best systems, although the strength of that corroboration depends on the breadth and independence of the candidate search.

Several qualifications are necessary. Current experiments require a partly specified observational setting; even the recovery of variables and laws presupposes some knowledge of the kind of system under observation. Moreover, the recovery of complex laws remains beyond present capabilities. The concern that a structured search merely reverse-engineers the desired result rather than discovering it should be addressed by evaluating the recovered expressions on data not used during training. Capabilities to recover systems of laws rather than single laws are still limited (see Section 5).

Like any empirical or meta-empirical hypothesis, a trade-off account may be refined iteratively. A negative result need not refute the entire approach: as when a physical theory confronts an anomaly, a mismatch between hypothesis and data should prompt an investigation into which component failed. The problem may lie in the chosen virtues, their operationalisation, or the way in which they are balanced. The crucial methodological constraint is that a revised hypothesis must be tested against genuinely new data rather than only the data that motivated the revision. Subject to this constraint, iterative refinement is legitimate. The situation resembles the development of scientific research programmes: a proposal may be adjusted in response to difficulties, but the programme is progressive only if each adjustment generates new empirical engagement rather than merely accommodating known problems (Lakatos, 1970).

It is worth stressing that using a method of automated discovery to test the structural properties of laws—whether they really exhibit the proposed trade-off structure—does not commit one to claiming that the physical theories of interest were actually discovered in this way, or even that their discovery from data would be practically feasible. Einstein, for example, discovered general relativity through a largely theoretical project, guided by the aim of reconciling special relativity with Newtonian gravity, by thought experiments such as the elevator experiment, and by desiderata associated with Mach's principle. Contemporary theories of quantum gravity are likewise sought and constructed—for better or worse—in the absence of new empirical data directly probing the relevant regime, through appeal to principles and the interpretation of existing theories. Our proposal is instead that symbolic discovery methods, even when supplied with prior knowledge of candidate laws, can be used to test the trade-off profiles of those candidates. In this way, they can help identify structural features that philosophical theories of lawhood seek to capture.

4 A first test case: kinetic energy relations as trade-off relations

We now provide a simple but paradigmatic test case. The test concerns two central questions about the expressions for kinetic energy conventionally used in classical and relativistic physics: first, whether they are Pareto-optimal, in the sense that they lie on a Pareto front and are not dominated by other expressions; and second, whether they are singled out among the Pareto-optimal expressions as distinguished convex corners. Readers less interested in the methodology than in the results may skip to the next section.

In special relativity, kinetic energy depends on mass and velocity:

$$K_r = mc^2 \left(\frac{1}{\sqrt{1-\beta^2}} - 1 \right), \qquad \beta := \frac{v}{c}.$$

In the low-velocity regime, this expression is well approximated by the Newtonian expression $K_N = mv^2/2 = mc^2\beta^2/2$. The zeroth-order approximation is the constant expression $K_0 = 0$. Since the symbolic-regression problem below uses dimensionless variables, we define

$$
\begin{aligned}
\kappa &:= \frac{K}{mc^2},\\
\kappa_r(\beta) &:= \frac{K_r}{mc^2} = \frac{1}{\sqrt{1-\beta^2}} - 1,\\
\kappa_N(\beta) &:= \frac{K_N}{mc^2} = \frac{\beta^2}{2},\\
\kappa_0(\beta) &:= \frac{K_0}{mc^2} = 0.
\end{aligned}
$$

Thus, K_r, K_N, and K_0 denote dimensional kinetic energies, whereas κ_r, κ_N, and κ_0 denote the dimensionless expressions evaluated in the experiment.

A kinetic-energy expression is not the most immediate example of a law of nature; one might first think of differential equations or dynamical laws. Still, it is a robust physical relationship and can plausibly be treated as a derivative or secondary law-like expression. Our pragmatic reason for considering kinetic-energy expressions is that they are simple enough to be within reach of current symbolic-regression methods, as demonstrated, for example, by the AI Feynman 2.0 framework (Udrescu et al., 2020), which we also use here.

For clarity, let us distinguish between hypotheses, procedure, and observation and discussion:

> *Hypotheses:* (1) For some suitable accuracy–simplicity trade-off—that is, for suitably chosen notions of accuracy and simplicity—the dimensionless velocity–kinetic-energy expressions κ_r, κ_N, and κ_0 are Pareto-optimal: they are not dominated by any expression that is at least as accurate and at least as simple, and strictly better in one of these respects. (2) The proposed law candidates are not merely on the Pareto front but are singled out among the Pareto-optimal points as convex corners preceding accuracy plateaus, over which accuracy improves little as complexity increases.[11]

[11]The hypotheses will become clearer below.

Both hypotheses are already suggested to some extent in Linnemann and Michels (forthcoming), although not in a methodological context in which the proposed meta-empirical testing technique is set out clearly. Convex corners may indicate especially distinguished Pareto-optimal trade-offs: points that stand out because moving to the next Pareto-optimal point yields only a small improvement in one virtue at a comparatively large cost in the other. Hypothesis 1 treats Pareto-optimality as a necessary condition of lawhood; hypothesis 2 adds the stronger condition that law candidates occupy distinguished convex corners.[12]

> *Procedure:* First, generate data consisting of dimensionless velocities, $\beta = v/c$, and corresponding dimensionless kinetic energies, $\kappa = K/(mc^2)$, from the exact expression, with β sampled uniformly from $(0, 0.8]$. In one variant, sample β instead from a uniform distribution over $(0, 0.25]$; in two further variants, add 1% or 10% multiplicative measurement noise to the kinetic-energy values used for training. Second, use symbolic regression to find candidate expressions for the relation between β and κ.[13] Third, order the resulting candidate expressions by accuracy and simplicity (represented by their complementary notions inaccuracy and complexity in the plots in Figure 1). Following Udrescu et al. (2020), we measure simplicity by the description length of the expression and inaccuracy by the description length of its fitting error. Thus, instead of a conventional mean squared error, this measure reduces the influence of large errors and places greater weight on improving already well-fitted points. Concretely, the description length of the error for an expression f, relative to a test data set of N points, is

[12] One may understand the trade-off criterion as a biconditional without thereby reducing lawhood itself to trade-off status. We set this issue aside here because our focus is on the proposed trade-off structure rather than on the nature of lawhood more generally.

[13] We use the symbolic-regression machinery of AI Feynman 2.0 to generate candidate expressions; see Udrescu et al. (2020).

$A_d = 1/N \sum_{i=1}^{N} \log_2(1+|f(\beta_i)-\kappa_i|\cdot 2^{30})$.[14] Fourth, construct a Pareto Pareto front: the set of expressions for which one cannot improve accuracy without sacrificing simplicity, or improve simplicity without sacrificing accuracy. All other points are Pareto-dominated: for each, there is either a simpler expression with at least the same accuracy or a more accurate expression with at most the same complexity. Finally, compare the locations of κ_r, κ_N, and κ_0 with the Pareto front generated by the search, and construct a joint, reference-assisted Pareto front that includes these three expressions for visualisation.

Observation and discussion: The four variants are shown in the respective panels of Figure 1. Given the chosen axes, the joinnt Pareto front obtained by admitting both the reference expressions κ_r, κ_N, and κ_0 and the expressions found by the search takes a staircase-like form (dotted line). The frontier derived solely from the expressions found by the search (green line) does not dominate the reference expressions and, depending on the variant, partly coincides with the joint frontier. As the figure shows, κ_N was not recovered in our test runs.

Changing the sampling regime shifts the frontier, since approximations that perform well at low velocities may perform worse when higher velocities receive more weight. Adding noise changes which expressions are recovered during the search. Consequently, the exact dimensionless relativistic expression κ_r was not recovered from the noisiest data set.

The result supplies limited, search-relative corroboration of the first hypothesis: among the expressions recovered within the finite search budget,

[14] Udrescu et al. (2020) use a square rather than an absolute value here.

none dominates the three reference expressions, and the search-only frontier partly coincides with the reference-assisted frontier. This does not show that no dominating expression exists outside the explored candidate set. Moreover, the fact that no recovered expression lies below the *joint* frontier is true by construction; the evidentially relevant observation is that the recovered expressions do not dominate the independently specified reference expressions. The first hypothesis is of special interest because it expresses a necessary implication of the signature thesis shared by trade-off accounts.

Since the 'intermediate' Newtonian approximation was not found in *our* test runs, the evidence for the second hypothesis is weaker than that for the more general first hypothesis. The grey dotted line nevertheless illustrates that, when the Newtonian expression is inserted manually at the point determined by its accuracy and simplicity values, it forms a corner of the staircase. This establishes a structural fact about the expression's location relative to the explored candidates, not its independent recovery by the algorithm. The expression was, however, found in the runs reported by Udrescu et al. (2020). More testing is needed.

A further sensitivity, not shown in the immediate experiment here, concerns the notion of accuracy used during the search. If one uses a substantially different objective, such as a more global mean squared error criterion, the staircase structure need not be reproduced in the same way. Such an objective may favour expressions that reduce average error over the sampled range, even when they obscure simple regime-valid approximations. In exploratory PySR runs, for instance, the simplest constant expression was sometimes shifted away from 0: under a mean squared error objective, a small non-zero constant can dominate 0 by improving the average fit, although it is less natural as a physical limiting expression. This illustrates that the recovered frontier depends not only on the data and measurement error but also on the standard of accuracy used to guide the search.

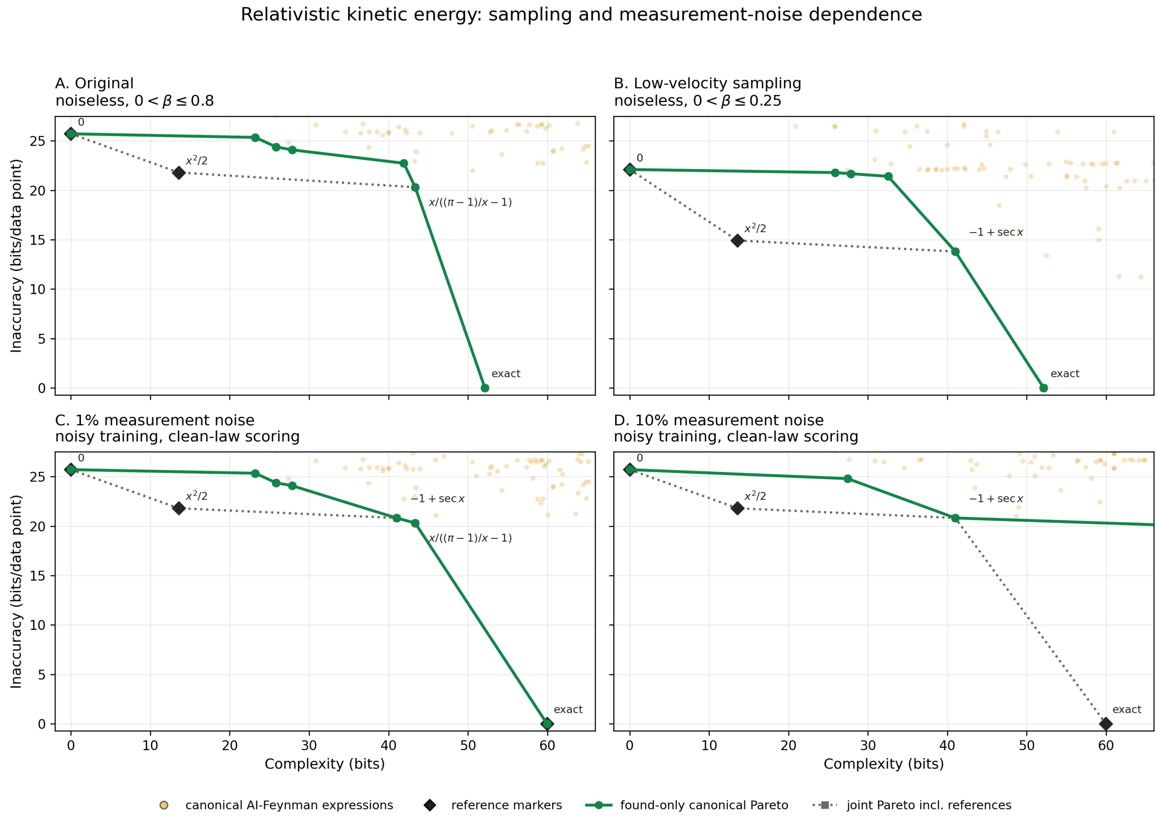

Figure 1: Four plots of inaccuracy versus complexity, in the sense described in the text, for expressions found as proxies for the dimensionless relativistic kinetic energy κ from synthetic data. Yellow points show expressions recovered during the search, after equivalence cleaning: where several expressions are algebraically equivalent, only the simplest, least complex representative is kept. Black diamonds mark externally supplied reference expressions, including the zero model, the Newtonian approximation, and the exact relativistic expression. The green curve is the Pareto front over found expressions. The grey dotted curve is a reference-assisted joint frontier obtained only if the black reference markers are also admitted; it is therefore not an algorithmic discovery by itself. The panels compare a noiseless broad-domain search, $0 < \beta \leq 0.8$, a noiseless low-velocity search, $0 < \beta \leq 0.25$, and searches trained with 1% and 10% multiplicative measurement noise in κ. The noisy searches are evaluated here on clean held-out data, so the panels compare how noise in the training data changes the expressions recovered relative to the same underlying law.

In sum, the example illustrates our proposed testing methodology and its practical application, albeit in a limited context. We focused on two relatively general hypotheses that Linnemann and Michels (forthcoming) had already advanced as conjectures on the basis of the findings in Udrescu and Tegmark (2020).

5 Some first lessons for trade-off accounts despite limitations

Current automated discovery systems remain limited to comparatively simple physical models; most systems of laws used in scientific practice cannot yet be tested to determine whether they achieve an optimal trade-off among virtues relative to their competitors. The concrete example considered so far may suggest that only simple collections of candidates, each effectively consisting of a single axiom, can be tested. Some methods, however, already discover and select among *systems* of equations (and variables) rather than merely among individual equations. SINDy (Brunton et al., 2016) and the SINDy autoencoder (Champion et al., 2019), for example, can recover coupled systems such as the Lorenz dynamics and predator–prey models while discriminating among alternatives with different trade-off profiles. Pareto front methods could then be used to examine the relevant trade-offs among virtues.[15] These limitations should nevertheless be kept in perspective. Any empirical investigation concerns a restricted domain; what matters is whether that domain can yield initial hypotheses and some preliminary corroboration of them.

Simple cases can already teach us a great deal about identifying systems of laws

[15] One might also ask whether a single equation can itself constitute a system of laws and thus suffice to show that the trade-off is relative to systems. The number of axioms comprising a system depends heavily on the language of representation. In standard mathematical notation, the Lorenz dynamics may be written as a single vector-valued law and hence, perhaps, as one axiom. Friend's reformulation, by contrast, can turn a single law into a system by restating it as several first- or second-order conditionals.

through trade-offs. In particular, they allow us to formulate and test hypotheses about which virtues matter and how. The example suggests that the appropriate measure of accuracy may need to balance global and local performance. Here, global accuracy is accuracy across the entire data set, whereas local accuracy is accuracy within a selected subdomain. A criterion concerned only with global accuracy may favour expressions that perform well on average but are not highly accurate in any physically significant regime. Many familiar physical laws and approximations, by contrast, are exact or highly accurate only within delimited regimes. A trade-off intended to capture such laws must therefore place sufficient weight on local accuracy.

Second, the example reveals a possible shape of the trade-off curves and thereby clarifies how some Pareto-optimal points may be distinguished from others. Linnemann and Michels (forthcoming) have already argued that convex corners on these curves should be regarded as candidates for laws: they occur immediately before further gains in one virtue become disproportionately costly in terms of another.

Bringing regression into focus also raises the question of what regression is for. It is typically directed towards prediction, and the virtues used to evaluate it reflect that aim: simplicity and an appropriate balance between local and global accuracy can be understood as safeguards against overfitting and underfitting. Yet prediction itself requires further specification. Does the relevant success consist in in-distribution or out-of-distribution prediction—interpolation or extrapolation—or both? How should distant extrapolation be valued relative to local extrapolation? The answers should affect the appropriate notion of accuracy. We may also ask whether regression, and hence laws, theories, and ultimately science, has primary goals other than prediction. Explanation and understanding are usually included among the goals of science. Our simple example suggests that predictive considerations alone can recover or single out familiar physical expressions. If this result could be generalised, it could support a provocative thesis; at present, however, the example is too limited to establish it.

Moreover, the results suggest that some elements commonly associated with the claim that laws result from an optimal trade-off—in particular, the Humean ontology of the BSA—may be, in a certain sense, dispensable. The feasibility of working with data, which may or may not derive from a Humean mosaic, suggests that this part of the BSA is inessential, at least if the focus is on the fit between actual scientific laws and philosophical lawhood criteria. If trade-off accounts operating on data are (meta-)empirically testable, then, from a naturalistic perspective, the usual focus in discussions of the BSA and its pragmatic variants may be misplaced. Historically, the trade-off among theoretical virtues was introduced into the BSA to refine the regularity view and protect it against familiar counterexamples to the naive regularity account. On this narrative, the BSA is essentially and primarily a regularity theory. Our proposal gives us reason to re-evaluate that perspective. The BSA remains a regularity theory, but the link between optimal trade-offs and lawhood may come closer to the core of the account than its historical role as a device for protecting regularism from counterexamples suggests.[16][17]

In sum, the extensional claim of a trade-off account can already be tested in simple cases, although applying the approach to more complex physical theories remains a distant goal. Trade-off accounts can nevertheless now be formulated concretely enough to generate hypotheses for more advanced tests, which lie, however, beyond the scope of this programmatic paper. We expect this line of research to develop alongside symbolic regression and equation discovery more generally.

[16]This conclusion is independent of objections to the particular proposals advanced by Loewer (2007) and Lam and Wüthrich (2023).

[17]We leave open whether similar things can be said in the case of compression-based accounts of laws.

6 Conclusion

We have identified an apparent argumentative gap in discussions supporting trade-off accounts of laws. Proponents of such accounts only anecdotally argue that actual laws show a *better* trade-off than plausibly worse systems of candidate laws. For example, they appeal to Newtonian laws as compared with a system listing all obtaining facts, or to a successful scientific theory as compared with its reformulation as a single universally quantified axiom involving an artificial predicate that contains the theory's entire content. Such arguments fail, however, to show that concrete paradigm examples of laws are identifiable through a *best* trade-off.

Drawing on recent advances in automated scientific discovery, we have suggested that this argumentative gap can begin to be filled by computational methods. On our proposal, trade-off accounts can be tested for their fit with actual science; the combination of automated state-variable discovery and symbolic regression promises to provide evidence about whether scientific laws are Pareto-optimal and whether they occupy distinguished convex corners. The adequacy of trade-off accounts thereby becomes a meta-empirical question whose investigation using symbolic regression and other methods of automated scientific discovery can now begin. We have only sketched this programme and illustrated some of its potential; we do not claim to have executed it beyond a suggestive proof of concept.

On our programmatic proposal, philosophical arguments retain a legitimate and important, albeit epistemically subordinate, role: they can suggest which trade-offs—which combinations of theoretical virtues and quantifiable definitions of them—are worth exploring, rule out manifestly inadequate proposals, and guide their refinement. Our proposal is not intended as an attack on the metaphysics of laws. Rather, it helps to clarify the boundary between comparatively less and more naturalistically committed forms of metaphysics. There is an irony in the fact that an approach to laws such as the BSA, whose proponents reject allegedly untestable metaphysical commitments such as primitive modality, has itself rarely been subjected to systematic testing and, in some versions, cannot be tested without further specification. We believe that our positive proposal, though still at the proof-of-concept stage, will open up new possibilities for constructive work in this direction.

Acknowledgements

We thank Matej Krátký, Annica Vieser, Lisa Vogt, and Christian Wüthrich for discussions of the ideas developed in this draft. Claude Code and ChatGPT (Codex, Prism) were used for language editing. Codex was used to create the plot in Figure 1, including implementing a variant of the AI Feynman 2.0 symbolic regression techniques by Udrescu et al. (2020) and post-processing the results. The authors bear sole responsibility for the content. N.L. thanks the Swiss National Science Foundation for support (grant number: 105212_207951). R.M.'s work on this paper was supported by the FCT (CEEC IND4ed, DOI: `https://doi.org/10.54499/2021.03171.CEECIND/CP1702/CT0015`). N.L. and R.M. furthermore acknowledge the FCT's support of the project 'Indeterminacy in Science' (DOI: `https://doi.org/10.54499/2023.15136.PEX`).

References

Emily Adlam. Laws of nature as constraints. *Foundations of Physics*, 52(1):1–41, 2022. doi: 10.1007/s10701-022-00546-0.

D. M. Armstrong. *What Is a Law of Nature?* Cambridge University Press, Cambridge, 1983. doi: 10.1017/CBO9781139171700.

Helen Beebee. The non-governing conception of laws of nature. *Philosophy and Phenomenological Research*, 61(3):571–594, 2000. doi: 10.2307/2653613.

Alexander Bird. *Nature's Metaphysics: Laws and Properties.* Oxford University Press, Oxford, 2007. doi: 10.1093/acprof:oso/9780199227013.001.0001.

David Braddon-Mitchell. Lossy laws. *Noûs*, 35(2):260–277, 2001. doi: 10.1111/0029-4624.00296.

Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016. doi: 10.1073/pnas.1517384113.

Craig Callender and Jonathan Cohen. Special sciences, conspiracy and the Better Best System Account of lawhood. *Erkenntnis*, 73(3):427–447, 2010. doi: 10.1007/s10670-010-9241-3.

Anjan Chakravartty. *Scientific Ontology: Integrating Naturalized Metaphysics and Voluntarist Epistemology.* Oxford University Press, Oxford, 2017. doi: 10.1093/oso/9780190651459.001.0001.

Kathleen Champion, Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, 116(45):22445–22451, 2019. doi: 10.1073/pnas.1906995116.

Boyuan Chen, Kuang Huang, Sunand Raghupathi, Ishaan Chandratreya, Qiang Du, and Hod Lipson. Automated discovery of fundamental variables hidden in experimental data. *Nature Computational Science*, 2:433–442, 2022. doi: 10.1038/s43588-022-00281-6.

Eddy Keming Chen. Fundamental nomic vagueness. *Philosophical Review*, 131(1):1–49, 2022. doi: 10.1215/00318108-9415127.

Eddy Keming Chen and Sheldon Goldstein. Governing without a fundamental direction of time: Minimal primitivism about laws of nature. In Yemima Ben-Menahem, editor, *Rethinking the Concept of Law of Nature: Natural Order in the Light of Contemporary Science*, pages 21–64. Springer, Cham, 2022. doi: 10.1007/978-3-030-96775-8_2.

Enrico Cinti and Marco Sanchioni. Humeanism in light of quantum gravity. *Synthese*, 199(3–4):10839–10863, 2021. doi: 10.1007/s11229-021-03270-w.

Jonathan Cohen and Craig Callender. A Better Best System Account of lawhood. *Philosophical Studies*, 145(1):1–34, 2009. doi: 10.1007/s11098-009-9389-3.

Miles Cranmer. Interpretable machine learning for science with PySR and SymbolicRegression.jl. *arXiv preprint arXiv:2305.01582*, 2023. doi: 10.48550/arXiv.2305.01582.

Chris Dorst. *The Best Predictive System Account of Laws of Nature*. PhD thesis, University of North Carolina at Chapel Hill, 2018.

Chris Dorst. Towards a best predictive system account of laws of nature. *The British Journal for the Philosophy of Science*, 70(3):877–900, 2019. doi: 10.1093/bjps/axy016.

Michael Esfeld and Dirk-André Deckert. *A Minimalist Ontology of the Natural World*. Routledge, New York, 2017. doi: 10.4324/9781315305239.

Luke Fenton-Glynn. Imprecise chance and the best system analysis. *Philosophers' Imprint*, 19, 2019.

Toby Friend. Laws are conditionals. *European Journal for Philosophy of Science*, 6: 123–144, 2016. doi: 10.1007/s13194-015-0131-z.

Verónica Gómez Sánchez. Crystallized regularities. *The Journal of Philosophy*, 117(8): 434–466, 2020. doi: 10.5840/jphil2020117827.

Michael Townsen Hicks. Dynamic Humeanism. *The British Journal for the Philosophy of Science*, 69(4):983–1007, 2018. doi: 10.1093/bjps/axx006.

Imre Lakatos. Falsification and the methodology of scientific research programmes. In Imre Lakatos and Alan Musgrave, editors, *Criticism and the Growth of Knowledge*, pages 91–196. Cambridge University Press, Cambridge, 1970.

Vincent Lam and Christian Wüthrich. Laws beyond spacetime. *Synthese*, 202(3):77, 2023. doi: 10.1007/s11229-023-04305-0.

Marc Lange. Laws and their stability. *Synthese*, 144(3):415–432, 2005. doi: 10.1007/s11229-005-5874-1.

Niels Linnemann and Robert Michels. Laws of nature as results of a trade-off: Rethinking the Humean trade-off conception. *The Philosophical Quarterly*, forthcoming. doi: 10.1093/pq/pqaf027.

Christian Loew and Siegfried Jaag. Making best systems best for us. *Synthese*, 197(6): 2525–2550, 2018. doi: 10.1007/s11229-018-1829-1.

Christian Loew and Siegfried Jaag. What Humeans should say about tied best systems. *Analysis*, 80(2):273–282, 2020. doi: 10.1093/analys/anz047.

Barry Loewer. Laws and natural properties. *Philosophical Topics*, 35(1/2):313–328, 2007. doi: 10.5840/philtopics2007351/214.

Barry Loewer. The package deal account of laws and properties (PDA). Unpublished manuscript, 2020. URL `https://sites.rutgers.edu/barry-loewer/wp-content/uploads/sites/195/2020/05/Loewer-2020-The-Package-Deal-Account.pdf`.

Tim Maudlin. *The Metaphysics Within Physics*. Oxford University Press, Oxford, 2007. doi: 10.1093/acprof:oso/9780199218219.001.0001.

Christopher J. G. Meacham. Constraint accounts of laws. *Ergo*, 12, 2025. doi: 10.3998/ergo.7427.

John T. Roberts. *The Law-Governed Universe*. Oxford University Press, Oxford, 2008. doi: 10.1093/acprof:oso/9780199557707.001.0001.

Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *Science*, 324(5923):81–85, 2009. doi: 10.1126/science.1165893.

Markus Schrenk. A theory for special science laws. In Helen Bohse and Sven Walter, editors, *Selected Papers Contributed to the Sections of GAP.6: 6th International Congress of the Society for Analytical Philosophy, Berlin, September 2006*, pages 121–131. Mentis, Paderborn, 2008. ISBN 978-3-89785-589-2.

Silviu-Marian Udrescu and Max Tegmark. AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16):eaay2631, 2020. doi: 10.1126/sciadv.aay2631.

Silviu-Marian Udrescu, Andrew Tan, Jiahai Feng, Orisvaldo Neto, Tailin Wu, and Max Tegmark. AI Feynman 2.0: Pareto-optimal symbolic regression exploiting graph modularity. In *Advances in Neural Information Processing Systems*, volume 33, pages 4860–4871, 2020. doi: 10.48550/arXiv.2006.10782.

Billy Wheeler. Compressibility and the algorithmic theory of laws. *Principia*, 23(3): 461–485, 2019. doi: 10.5007/1808-1711.2019v23n3p461.

Isaac Wilhelm. Tractability and laws. *Synthese*, 200(4):318, 2022. doi: 10.1007/s11229-022-03638-6.

James Woodward. Simplicity in the best systems account of laws of nature. *The British Journal for the Philosophy of Science*, 65(1):91–123, 2014. doi: 10.1093/bjps/axt026.