On the Ostrogradski Instability; or, Why Physics Really Uses Second Derivatives

Noel Swanson*

April 24, 2019

Abstract

Candidates for fundamental physical laws rarely, if ever, employ higher than second time derivatives. Easwaran (2014) sketches an enticing story that purports to explain away this puzzling fact and thereby provides indirect evidence for a particular set of metaphysical theses used in the explanation. I object to both the scope and coherence of Easwaran's account, before going on to defend an alternative, more metaphysically deflationary explanation: in interacting Lagrangian field theories, it is either impossible or very hard to incorporate higher than second time derivatives without rendering the vacuum state unstable. The so-called *Ostrogradski instability* represents a powerful constraint on the construction of new field theories and supplies a novel, largely overlooked example of non-causal explanation in physics.

Contents

*Department of Philosophy, University of Delaware, 24 Kent Way, Newark, DE 19716, USA, nswanson@udel.edu

1 Introduction: Why does $F = ma$?

Nature, it seems, has an affinity for low-order differential equations. The various candidates for fundamental dynamical laws that fill physics textbooks rarely, if ever, employ anything other than first or second time derivatives. Newton's second law, Maxwell's equations, and the Einstein field equations are all second-order. The Schrödinger and Dirac equations are first-order, while the equations of motion derived from the standard model Lagrangian are second-order. Some emergent laws contain higher time derivatives, but these are ultimately thought to be explainable in terms of more fundamental, low-order laws.[1] The curious absence of high-order laws in the foundations of physics is made all the more curious by the litany of problems that such laws could potentially ameliorate.[2] The source of this absence is an intriguing, largely overlooked philosophical puzzle.

Easwaran (2014) sketches an enticing story that purports to resolve this puzzle. The story has two primary components. The first is a *reductionist thesis* according to which facts about velocity and acceleration are entirely grounded in facts about differences in position at different times. The second is a *causal thesis* according to which the laws of nature, plus present facts about position and velocity (and possibly other quantities like charge), causally determine facts about position at all times in the future. In conjunction with several auxiliary assumptions about the nature of causation, grounding, and the continuity of physical change, Easwaran argues that this

[1] In mechanical engineering, equations involving the third and fourth derivative of position, *jerk* and *snap*, are routinely employed to design everything from cams and motion controllers to elevators and rollercoasters (Eager et al., 2016). In some approaches to studying effective field theories it is common to add all higher-derivative terms consistent with the symmetries of the theory (Weinberg, 1995, ch. 12.3).

[2] Adding higher derivative terms to gravitational theories can help render them renormalizable. High-order alternatives to general relativity have also been proposed to avoid postulating dark energy. In particle physics, the Lee-Wick extension of the standard model uses higher derivative terms to help stabilize the Higgs mass in the face of divergent radiative corrections. Higher derivatives also arise naturally models of cosmic strings and stringy black holes. See Simon (1990) for a survey of various applications of higher derivative theories.

package of metaphysical views can explain why the fundamental laws take the low-order form that they do. In the absence of alternative explanations, he suggests that this story provides us with reason to adopt both the reductionist and causal theses.

Easwaran's paper is situated in the context of a broader debate about the metaphysics of velocity and acceleration that stretches back to Russell (1903). Several thinkers in this debate (Tooley, 1988; Arntzenius, 2000; Lange, 2005) have argued that the reductionist and causal theses are in fact incompatible. Easwaran's position is noteworthy for clearly articulating a way to reconcile them, as well as evincing positive support for their conjunction. In addition, it supplies a novel example of a mixed causal/non-causal explanation, incorporating both grounding and causal relations, as well as meta-nomological constraints in the sense of Lange (2016). If successful, it also supplies an example of how metaphysical issues can percolate up to the physical level, a situation where taking a stand on the metaphysical character of laws, causation, and quantities matters for physical explanation.

Alas, the story does not live up to all that it promises. On closer inspection, Easwaran's explanation only accounts for why there cannot be both low-order and high-order fundamental laws involving time derivatives of the same quantities. Apart from an unconvincing appeal to the naturalness or simplicity of low-order laws, no part of his story rules out a single set of fundamental high-order laws. There is an alternative physical explanation available that does not suffer from this significant limitation in scope and remains more metaphysically agnostic — generic higher-order Lagrangian field theories are energetically unstable. This instability, first discovered by Ostrogradski (1850) provides the key to a more powerful, unified story about the absence of high-order derivatives in the fundamental laws of nature. Moreover, this alternative story is neutral with respect to the reductionist and causal theses, leaving the scorekeeping in that metaphysical debate unchanged.

In §2, I review Easwaran's metaphysical explanation, drawing out some of its chief limitations. In §3, I introduce the Ostrogradski theorem, before using it to develop an alternative physical explanation in §4. In §5, I respond to objections and draw several related conclusions about the type of non-causal explanation offered by the Ostrogradski instability.

2 Easwaran's Metaphysical Explanation

Ockhamist considerations have long been cited in support of the reductionist thesis. Insofar as we can define velocity and acceleration in terms of suitable limits of ratios of spatiotemporal distances, we should reduce these derivative quantities to positional quantities. In Easwaran's preferred formulation, facts about velocities and accelerations are entirely grounded in facts about positions at different times. The standard mathematical procedure for defining derivative quantities takes an open, 2-sided limit.[3] Following this procedure, velocity and acceleration are revealed to be what Arntzenius (2000) calls *2-sided neighborhood properties*. Facts about the velocity and acceleration of an object at t are not grounded in fundamental facts about the object at t, but rather in facts about the position of the object in the interval $(t-\delta, t+\delta)$, for any $\delta > 0$.

If velocity and acceleration are 2-sided neighborhood properties, the reductionist thesis immediately comes into conflict with the causal thesis. Usually motivated by an anti-Humean conception of laws as entities that generate future states from present states, the causal thesis says that the laws of nature, plus present facts about position and velocity (and possibly charges), causally determine facts about position at all times in the future.[4] If velocity is a 2-sided neighborhood property, it is partially grounded in facts about positions at times in the future. Assuming that causes must precede their effects, velocity cannot causally determine facts about positions at all times in the future.[5]

[3]Formally, the velocity, v_t, is usually defined as the quantity (if any) that satisfies

$$\forall(\epsilon > 0)\ \exists(\delta > 0)\ \forall t'\ \left(|t' - t| < \delta \rightarrow \left|\frac{x'_t - x_t}{t' - t} - v_t\right| < \epsilon\right),$$

where t' ranges over all times and x_t and x'_t are the positions at t and t'.

[4]Easwaran leaves the notion of "causal determination" largely open. He does assume that it is irreflexive and that it interacts with grounding in a transitive and temporally oriented manner. Most notably, his assumption of *universal forward causation* requires that if A partly causally determines B, then there is a set S_A of sufficient grounds for A and a set S_B of sufficient grounds for B such that no member of S_A is temporally later than any member of S_B.

[5]This argument is too quick. Making it precise requires attention to subtleties surrounding the interaction between limits, causal determination, and grounding. There may well be other objections to Easwaran's story lurking in the shadows here. But for present purposes, the details do not matter.

To circumvent this difficulty, the causal reductionist can define velocity as the *past derivative* of position, using an open, 1-sided limit approaching t from the past.[6] This turns velocity into a *past neighborhood property*: facts about the velocity of an object at t are grounded in facts about the position of the object in the interval $(t-\delta, t)$, for any $\delta > 0$. Consequently, velocities at t are rendered suitable candidates to be causes of future positions. So far, so good, but Lange (2005) argues that trouble looms when we try to bring acceleration into the picture.

On a broadly causal interpretation of the laws of Newtonian physics, Lange proposes the following natural chain of dependence: forces cause acceleration, which cause changes in velocity, which cause changes in position. Forces in turn are grounded in the present masses, positions, charges, and possibly velocities of objects. If this is true, acceleration cannot be either a past-neighborhood or two-sided neighborhood property without the specter of retrocausation arising. It must be a *future-neighborhood property*, grounded in fundamental facts over the interval $(t, t+\delta)$. This can be achieved by defining acceleration as the *future derivative* of velocity, using an open, 1-sided limit approaching t from the future.[7] But if acceleration is a future neighborhood property, it cannot be a cause of changes in velocity, a past neighborhood property. Lange concludes that the causal reductionist position is untenable.

Easwaran proposes an alternative, more complicated chain of dependence that evades this conclusion: forces cause accelerations, which are future neighborhood properties grounded in facts about future positions and velocities. Positions and velocities are not caused by accelerations, but rather by whatever causes accelerations. Velocities are past neighborhood properties, and are causes of future positions. As in Lange's view, forces are grounded in the present masses, positions, charges and possibly velocities of objects.[8]

[6]The past velocity, v_t^p, is defined as the quantity (if any) that satisfies

$$\forall(\epsilon > 0)\ \exists(\delta > 0)\ \forall t', t''\ \left((t-\delta < t', t'' < t) \rightarrow \left|\frac{x_{t'} - x_{t''}}{t' - t''} - v_t^p\right| < \epsilon\right).$$

[7]The future velocity, v_t^f, is defined as the quantity (if any) that satisfies

$$\forall(\epsilon > 0)\ \exists(\delta > 0)\ \forall t', t''\ \left((t'' < t+\delta, t < t') \rightarrow \left|\frac{x_{t'} - x_{t''}}{t' - t''} - v_t^f\right| < \epsilon\right).$$

[8]Easwaran actually argues that fundamental forces will not be velocity dependent,

Identifying velocity with the past derivative of position and acceleration as the future derivative of velocity, causal reductionists can have their cake and eat it too.

Easwaran's proposal comes with an unexpected bonus. In order to fit the various pieces of this puzzle together consistently, the possible forms that the laws of physics can take are significantly constrained. He leverages these constraints into an explanation for why physics only makes use of first and second time derivatives.

Consider a law of the form,

$$\ddot{x} = F(x, \dot{x}, m, c) \ , \tag{1}$$

where F is a function that depends (possibly trivially) on the present positions, velocities, masses, and charges. (Newton's second law has this form.) Assuming the initial value problem is well-posed and has a unique solution, this law, along with the present state $(x_t, \dot{x}_t, m_t, c_t)$, determines the state at all other times t'. On Easwaran's picture, this mathematical fact is interpreted in causal terms: together, the present state and the law causally determine future states. In order for this to make sense, the present values of position, velocity, mass, and charge must be entirely grounded in facts about the present and past, while acceleration is entirely grounded in facts about the future. As long as velocity is a past derivative, and acceleration is a future derivative, all of this works out nicely.

Easwaran asks us to consider adding an additional law of the form,

$$x^{(3)} = Y(x, \dot{x}, \ddot{x}, m, c) \ , \tag{2}$$

that sets the third time derivative of position, *jerk*, equal to some fundamental force-like quantity, Y. The causal interpretation of (1) demands that acceleration must be a future derivative and velocity a past derivative, but the same interpretation of (2) requires jerk to be a future derivative and both acceleration and velocity to be past derivatives. So acceleration must be both a past and future neighborhood property, which cannot be.

In general, if a fundamental law causally determines the future by setting the nth temporal derivative of some quantity q, the present value of q along with its first $n-1$ derivatives must be grounded in the present and past. It

but nothing in his argument turns on this restriction, so I will ignore it for the sake of generality.

follows that the first $n-1$ derivatives have to be past derivatives while the nth derivative is a future derivative. Thus on Easwaran's causal reductionist view, there cannot be multiple fundamental laws that causally determine the future by setting the present value of different order derivatives of the same fundamental quantity on pain of contradiction.

Next, Easwaran appeals to the causal topology of time to explain why there are low-order dynamical laws in the first place. If there are, there cannot also be higher-order dynamical laws by the above argument. At this stage, though, the story becomes rather sketchy. Assuming that there is no causation at temporal distance and that fundamental quantities change continuously in time, it follows that the fundamental laws must involve both past and future neighborhood properties. Easwaran contends that derivatives are the simplest, most natural type of neighborhood property, "perhaps there is some alternative, but any other neighborhood property appears to be just as complicated" (p. 857). So the demands of causal topology render laws involving both past and future derivatives especially natural. A second-order law like (1) sets a second future derivative and uses a first past derivative as part of the initial conditions. A first-order law, like the Schrödinger equation, sets a first future derivative, and as long as some other aspect of the law involves a past neighborhood property, continuity is preserved. After canvasing these two cases, Easwaran abruptly concludes,

> these appear to be the two simplest ways to get the appropriate causal connections in both directions, and it is striking that the best candidate laws are of these forms. The causal reductionist view described above can give an explanation of this feature of the laws. (p. 857)

If so, it is an explanation of rather limited scope. While the view rules out multiple fundamental dynamical laws of different order, nothing, apart from the simplicity of first-order and second-order equations tells against a single higher-order dynamical law or a set of higher-order dynamical laws of the same order. When physicists contemplate the prospect of modifying the Einstein field equations or extending the standard model with by adding higher derivative terms, they are typically not considering adding new high-order laws to preexisting low-order ones, they are looking to completely replace them.

Even if it can be argued that general Ockhamist considerations favor low-order theories, there are other countervailing virtues that speak in favor of

high-order theories. For example, adding higher-order derivative terms to gravitational theories can render them renormalizable (Stelle, 1977). Maximality arguments favored by effective field theorists call for (in principle) adding as many terms to a Lagrangian as allowed by the theory's symmetries (Weinberg, 1995, ch. 12.3). There is active interest in searching for viable high-order theories in particle physics (Grinstein et al., 2008), string theory (Moura and Schiappa, 2006), and classical gravitational physics (Woodard, 2007). Such searches are not limited by the broad Ockhamist considerations that Easwaran alludes to, nor are they limited by commitment to the causal reductionist position he defends. If writing down consistent high-order theories were as simple as introducing slightly more complex equations or adopting a different metaphysical view of laws, we should expect textbooks to be filled with toy examples of such theories (especially given their theoretical utility). The fact that they are not, cannot be explained by Easwaran's story, and strongly suggests that something else is going on.

There is a further, deeper worry that threatens to undermine even the limited success of Easwaran's explanation. The causal interpretation of laws like (1) and (2) presupposes that the corresponding initial value problem is well-posed and has a unique solution.[9] Easwaran glosses over this point (I have tried to restore it to its proper place in the argument here), but once this seeming technicality is acknowledged, it becomes unclear exactly what situation we are being asked to imagine when we add equation (2) to equation (1). If equation (1) is sufficient to determine the future state by itself, and velocity, acceleration, and jerk are treated as reducible, then equation (2) appears to be epiphenomenal. Even if equation (2) is also sufficient to determine the future state by itself, the causal reductionist is hard-pressed to explain why there are two fundamental laws and a case of causal overdetermination, rather than a single law.

This leaves two plausible situations we might be intended to consider: only (2) is sufficient to determine the future state, or (1) and (2) are jointly sufficient (but neither alone is sufficient). In either of these cases, though, since (1) does not causally determine the future state, there is no longer any obvious reason to give (1) Easwaran's preferred causal interpretation where the term on the right side causes the term on the left side. Either (2) is

[9]This condition might be relaxed for indeterministic theories where the laws and present state determine a probability distribution over future states. If the initial value problem is not well-posed, there is no such probability distribution. This is arguably a minimal requirement to talk about causal determination in any extended sense.

the only fundamental causal law, or we can combine (1) and (2) into a single fundamental causal law by substituting $F(x, \dot{x}, m, c)$ for $\ddot{x}$ in equation (2). In both cases, the future state depends on initial data consisting of $(x_t, \dot{x}_t, \ddot{x}_t, m_t, c_t)$, and so the natural causal reading is to interpret velocity and acceleration as past derivatives and jerk as a future derivative, caused by the force-like term Y. While we must give up on the idea that accelerations are caused by the force term F, in a theory where (1) no longer causally determines the future, this idea is unmotivated. If so, the first half of Easwaran's story evaporates, leaving only the unconvincing Ockhamist argument in favor of low-order theories.[10]

This objection is not necessarily decisive. It may be possible to provide motivation along the following lines: given a high-order law that causally determines the future, if it is possible to factor out a lower-order equation like (1), and if the force term, F, plays a certain kind of explanatory or predictive role in the theory (e.g., if there exist possible manipulations on F that alter $\ddot{x}$ in the sense of Woodward 2003), then we should interpret (1) as expressing a causal relation between F and $\ddot{x}$, even if (1) does not causally determine the future. It should be noted that the success of this maneuver will sensitively depend on the details of the functions, F and Y, and will not always be available. Moreover, it must still account for why a causal interpretation of (1) is preferable to an interpretation on which acceleration is grounded in the functional relation between position, velocity, mass, and charge expressed by F. Regardless, the objection puts additional pressure on a causal reductionist story already on the ropes. It would be preferable all-things-considered to have an alternative explanation not subject to these concerns. Fortunately, there is such a story in the offing.

3 The Ostrogradksi Theorem

The key to this alternative story lies in a deep no-go result for certain types of high-order Lagrangian theories.

Theorem (Ostrogradski)**.**[11] *If a non-degenerate Lagrangian, $\mathcal{L}(q, \ldots, q^{(n)})$, depends on the nth derivative of a single configuration variable q, with $n > 1$,*

[10]Moreover, nothing prevents opponents of causal reductionism from appealing to the same vague Ockhamist argument, nullifying the view's explanatory advantages.

[11]The theorem is usually attributed to Ostrogradski (1850), who did pioneering work on the Hamiltonian formulation of higher-order theories, however it is unclear if he recognized

then the energy function in the corresponding Hamiltonian picture is unbounded from below.

Our goal in this section will be to unpack this theorem, closely following the elegant derivation presented in Woodard (2007). Although the result directly extends to field theories, for ease of presentation in this section, we will focus on discrete Lagrangians.[12] Similarly, we will suppress tensor indices. The argument extends directly to configuration spaces of arbitrary dimension. The variable, q, may be interpreted as spatial position in some theories, but the same methods apply to any arbitrary configuration variable.

Suppose that $\mathcal{L}(q, \dot{q})$ only depends on q and $\dot{q}$, as typically assumed in textbook presentations of Lagrangian mechanics. Extremizing the action yields the familiar second-order Euler-Lagrange equations of motion:

$$\frac{\partial \mathcal{L}}{\partial q} - \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{q}} = 0 \ . \tag{3}$$

Non-degeneracy is a technical condition requiring the determinant of the Hessian matrix to be non-vanishing,

$$\det\left[\frac{\partial^2 \mathcal{L}}{\partial \dot{q}^2}\right] \neq 0 \ . \tag{4}$$

Non-degeneracy ensures that $\frac{\partial \mathcal{L}}{\partial \dot{q}}$ depends on $\dot{q}$ and that (3) has a well-posed initial value problem with a unique solution. It also entails that we can rewrite the equations of motion in Newtonian form,

$$\ddot{q} = F(q, \dot{q}) \ , \tag{5}$$

where the force function, F, depends on the inverse of the Hessian matrix. The state at any time is determined by initial data $(q_0, \dot{q}_0)$.

Non-degeneracy also entails that the Legendre transform is a local diffeomorphism between TQ and T^*Q. We can therefore use it to translate

it as a no-go result for higher-order field theories. The first work to do so appears to be Pais and Uhlenbeck (1950).

[12] For field theories, the Lagrangian is replaced by a Lagrangian density over a continuum of configuration variables indexed by spacetime region. The variational problem can be solved for each variable separately and the Ostrogradski theorem applies to each degree of freedom. As explained in §4, because there are so many coupled unstable degrees of freedom, this results in a serious physical problem for higher-order field theories.

between the Lagrangian and Hamiltonian descriptions of the system in a well-defined manner.[13] This translation identifies two canonical coordinates,

$$Q := q \qquad P := \frac{\partial \mathcal{L}}{\partial \dot{q}} \ , \tag{6}$$

and a Hamiltonian function,

$$\mathcal{H} := P\dot{q} - \mathcal{L} \ , \tag{7}$$

that satisfies Hamilton's equations and acts as the generator of time translations.

Suppose instead that we are given a higher-order Lagrangian $\mathcal{L}(q, \dot{q}, \ddot{q})$ that depends on q, $\dot{q}$, and $\ddot{q}$. Extremizing the action yields fourth-order equations of motion:

$$\frac{\partial \mathcal{L}}{\partial q} - \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{q}} + \frac{d^2}{dt^2}\frac{\partial \mathcal{L}}{\partial \ddot{q}} = 0 \ . \tag{8}$$

In this context, non-degeneracy requires that

$$\det\left[\frac{\partial^2 \mathcal{L}}{\partial \ddot{q}^2}\right] \neq 0 \ . \tag{9}$$

This entails that we can rewrite the equations of motion in a higher-order version of Newtonian form,

$$q^{(4)} = F(q, \dot{q}, \ddot{q}, q^{(3)}) \ , \tag{10}$$

[13] Since the Legendre transform is only guaranteed to be a local diffeomorphism, one might worry that the corresponding translation scheme will not establish complete physical equivalence between the Lagrangian and Hamiltonian pictures. This may well be, but one of the important physical properties that is preserved is the boundedness of the energy. Presentations of the Ostrogradski theorem typically begin with the Lagrangian picture and translate into the Hamiltonian picture since the latter is a more familiar setting for analyzing the energy spectrum, but if we want, we can stay on the Lagrangian side and derive the Ostrogradski instability directly for the Lagrangian energy function,

$$E_{\mathcal{L}} := (L_\Delta - 1)\mathcal{L} \ ,$$

where L_Δ is the Lie derivative with respect to the canonical Liouville vector field $\Delta := \dot{q}\ \partial/\partial\dot{q}$ which generates dilations long the fibers of TQ. $E_{\mathcal{L}}$ is a constant of motion and corresponds to the generator of time translations. For non-degenerate Lagrangians, the Legendre transform maps $E_{\mathcal{L}}$ onto the corresponding Hamiltonian function, so $E_{\mathcal{L}}$ is unbounded iff the Hamiltonian is.

and the state at any time is determined by initial data $(q_0, \dot{q}_0, \ddot{q}_0, q_0^{(3)})$. Translating via the Legendre transformation identifies four canonical coordinates,

$$Q_1 := q \qquad Q_2 := \dot{q} \qquad P_1 := \frac{\partial \mathcal{L}}{\partial \dot{q}} - \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \ddot{q}} \qquad P_2 := \frac{\partial \mathcal{L}}{\partial \ddot{q}} \ , \tag{11}$$

and a Hamiltonian,

$$\mathcal{H} := P_1 \dot{q} + P_2 \ddot{q} - \mathcal{L} \ , \tag{12}$$

that satisfies Hamilton's equations and acts as the generator of time translations.[14]

At first glance, equations (7) and (12) are remarkably similar. In particular, both depend linearly on the canonical momenta. Since the momenta can take arbitrary negative values, it seems that both (7) and (12) will be unbounded from below. But careful consideration of the constraints imposed by non-degeneracy reveals that they have strikingly different properties. In both cases, non-degeneracy entails that the definitions of the canonical coordinates, (6) and (11), can be inverted. In the first case, this means that $\dot{q}$ can be rewritten as a function of P and Q, and the Hamiltonian (7) takes the form:

$$\mathcal{H} = P f(P, Q) - \mathcal{L} \ . \tag{13}$$

So if $f(P, Q)$ has a suitable form, the linear dependence on P can be removed and the Hamiltonian is bounded from below. In the second case, it is $\ddot{q}$ that can be rewritten as a function of Q_1, Q_2, and P_2, and the Hamiltonian (12) takes the form:

$$\mathcal{H} = P_1 Q_2 + P_2 f(Q_1, Q_2, P_2) - \mathcal{L} \ . \tag{14}$$

As before, the linear dependence on P_2 can be removed if $f(Q_1, Q_2, P_2)$ has a suitable form, but the linear dependence on P_1 cannot be removed. Thus the higher-order Hamiltonian (14) is unbounded from below. This is the source of the Ostrogradski instability.

[14] See Miron et al. (2002) for a systematic treatment of the relevant technical machinery for Lagrangians depending on time derivatives of arbitrary order, $\mathcal{L}(q, \ldots, q^{(n)})$. The generalized Lagrangian state space, $T^n Q$, is the *n-osculator bundle* over configuration space. The dual Hamiltonian phase space is defined as $T^{*n} Q := T^{n-1} Q \times T^* Q$. If the Lagrangian is non-degenerate, the Legendre transform is a local diffeomorphism between these two spaces and maps the generalized Lagrangian energy function onto the corresponding Hamiltonian.

In the general case, given a non-degenerate Lagrangian $\mathcal{L}(q, \ldots, q^{(n)})$, the Euler-Lagrange equations are given by,

$$\sum_{i=0}^{n} \Big(- \frac{d}{dt} \Big)^i \frac{\partial \mathcal{L}}{\partial q^{(i)}} = 0 \ , \tag{15}$$

the Legendre transform identifies $2n$ canonical coordinates,

$$Q_i := q^{(i-1)} \qquad P_i := \sum_{j=i}^{n} \Big(- \frac{d}{dt} \Big)^{j-i} \frac{\partial \mathcal{L}}{\partial q^{(j)}} \ , \tag{16}$$

and the Hamiltonian has the form,

$$\mathcal{H} := \sum_{i=0}^{n} P_i q^{(i)} - \mathcal{L} \ . \tag{17}$$

Non-degeneracy entails that $q^{(n)}$ can be written as a function of $Q_1, \ldots, Q_n$ and P_n, and so the Hamiltonian is linear in the first $n-1$ momentum coordinates,

$$\mathcal{H} := P_1 Q_2 + \ldots + P_{n-1} Q_n + P_n f(Q_1, \ldots Q_n, P_n) - \mathcal{L} \ . \tag{18}$$

Thus adding higher-derivative terms increases the number of phase space dimensions in which the Hamiltonian is unbounded.

The argument sketched here relies directly on the assumption of non-degeneracy. If the Lagrangian is degenerate, the determinant of the Hessian matrix vanishes, and greater care must be taken to define the proper Legendre transformation. In this context Pons (1989) extends the Ostrogradski theorem, showing that the relevant Hamiltonians will naively contain linear momentum terms just like (18). The catch is that degenerate theories necessarily include additional constraints which can reduce the number of physical phase space dimensions and cancel out these linear terms. This opens up a possible avenue for evading the no-go result which will be explored in the next section.

4 A Physical Explanation

The Ostrogradski theorem reveals a linear momentum dependence present in any non-degenerate higher-order Lagrangian theory. It is tempting to jump

from the fact that the Hamiltonian is unbounded to the conclusion that such theories must be physically pathological, but that would be too quick. If the Ostrogradski Hamiltonian (18) describes a closed system, it is a constant of the motion, and therefore the total energy, even if negative, is conserved. So the problem cannot be simply that the energy decays in such theories. If the system interacts with another system, then problems can certainly arise — while the total energy remains constant, the subsystems could be excited to arbitrarily negative (or positive) energy. Such runaway solutions have been explored in the context of classical particle mechanics, where there are various strategies for eliminating or taming them (Simon, 1990). While the jury is still out, it is far from clear that runaways render a theory unphysical.

There is, however, a broad class of theories where the Ostrogradski instability is a serious defect: interacting field theories of both the classical and quantum variety. Woodard (2007, 2015) sketches a general argument that any field theory with a Hamiltonian like (18) has an unstable vacuum state. It is customary to view a free scalar field theory as a system of coupled harmonic oscillators, $\phi(x)$, at each spacetime point. Each oscillator obeys the second-order equations of motion,

$$\ddot{\phi} + \omega^2 \phi = 0 \ , \tag{19}$$

where ω is the frequency. Pais and Uhlenbeck (1950) consider the simplest higher-order generalization of this case, a system of coupled oscillators obeying fourth-order equations of motion:

$$\frac{1}{\omega^2}\phi^{(4)} + \ddot{\phi} + \omega^2 \phi = 0 \ . \tag{20}$$

For both equations of motion we can decompose solutions into positive and negative frequency modes. The Hamiltonian for (19) is quadratic in the momentum variable, and thus bounded from below. As a result, both frequency modes carry positive energy. The Hamiltonian for (20), in contrast, suffers from the Ostrogradski instability and is linear in one momentum variable. Consequently, positive frequency modes carry positive energy, and negative frequency modes carry negative energy. Woodard argues that this is a generic feature: solutions to the equations of motion for field theories with Hamiltonians like (18) will be sums of positive and negative energy modes.

In a free theory, this is unproblematic. The positive and negative energy modes of each oscillator do not couple, and the field configuration is stable.

In an interacting theory, however, coupling between the positive and negative energy modes entails that any apparently stable field configuration can decay further by producing positive and negative energy excitations. Moreover, because there are so many ways for such decays to occur, they are entropically favored. There is one way for the system to remain stable and an infinite number of ways for it to decay into pairs of excitations with arbitrarily positive and negative energy. And multiple decays are favored over single decays. Woodard concludes:

> [. . .] such a system instantly evaporates into a maelstrom of positive and negative energy particles. Some of my mathematically minded colleagues would say it isn't even defined. I prefer to simply observe that no theory of this kind can describe the universe we experience in which all particles have positive energy and empty space remains empty. (p. 413)

Woodard's argument gives us a physical reason to reject a large class of higher-order theories. In axiomatic approaches to quantum field theory, vacuum stability is guaranteed by the *spectrum condition*, which rules out Ostrogradski Hamiltonians. In general relativity, various energy conditions, *weak*, *strong*, *dominant*, play a similar role. Although rarely made explicit in classical field theory, vacuum stability is a plausible constraint on physical possibility, prima facie. Even if vacuum stability is not treated as a constraint, we can still appeal to Woodard's more anthropic line of reasoning. No field theory with an Ostrogradksi Hamiltonian could possibly describe a stable world like ours.[15]

[15]Is this really explanatory? That depends on what we take the explanandum to be. If we are puzzled about the absence of higher-order theories in physics textbooks, then a deflationary epistemic explanation seems perfectly adequate. It may turn out that no deeper meta-law or principle requires the vacuum to be stable. Nonetheless, the world could not be remotely like what we observe if it were described by a higher-order Lagrangian field theory. Compare this to the question, why do the laws of physics involve interactions? No deeper meta-law or principle rules out free field theories, but the world could not be remotely like what we observe if there were no interactions. (In fact, the Ostrogradksi instability reveals that free field theories are more like our actual laws than higher-order field theories in some sense.) Both explanations have a similar (weak) anthropic character, though the Ostrogradski explanation is far less trivial. We expect that the broadly observable features of the world will be sensitive to the presence or absence of interactions. It is genuinely surprising to discover that they are even more sensitive to the presence or absence of higher derivatives.

Several remarks are in order. First, Woodard's stability argument relies on the assumption that the theories in question are field theories, at least to good approximation. It is because there is such a vast number of coupled oscillators that decays into positive and negative energy modes are so likely. The same entropic reasoning does not apply to a lone pair of coupled oscillators, but the argument still goes through for effective field theories approximated by a sufficiently large number of oscillators.[16] Second, the argument relies on the Lagrangian character of the laws. It is because the equations of motion are derived by extremizing a Lagrangian that the Hamiltonian must be linear in certain momentum variables. In principle, nothing rules out the possibility of a higher-order field theory whose equations of motion do not have this character. (Insofar as we can view such a field theory as a system of coupled oscillators, however, they cannot be Pais-Uhlenbeck oscillators (20), the simplest, most natural higher-order generalization of the simple harmonic oscillator.) Third, although quantization can help stabilize certain systems like the hydrogen atom, it will not help eliminate the Ostrogradski instability. The instability of the classical model of the Hydrogen atom comes from the ability to position the electron (with fixed momentum) arbitrarily close to the nucleus. Such configurations represent a tiny corner of phase space and quantization effectively excises this corner. In contrast, the Ostrogradski instability arises from a linear momentum dependence in nearly half of the phase space dimensions. Quantization cannot excise such a large quadrant of phase space.

Assuming that nature is described by a Lagrangian field theory, the only way to avoid the Ostrogradksi instability is to declare a large sector of phase space off limits. This requires introducing constraints that reduce the effective dimensionality of phase space, thereby rejecting the original assumption that the Lagrangian is non-degenerate. Degenerate theories are not uncommon. Any Lagrangian theory with gauge symmetries or odd-order equations of motion will be degenerate. But not just any degenerate Lagrangian will work. The degeneracy must give rise to enough constraints of just the right sort, tuned in just the right way, to excise the entire unstable momentum

[16]There is an important caveat here. In some cases it is possible for an ultraviolet cutoff to stabilize the momentum sector, in which case higher derivative terms can be consistently added to effective field theories. See Eliezer and Woodard (1989) for a development of this idea. Since the cutoff does not apply to the underlying exact theory, however, the Ostrogradksi instability remains a constraint on the fundamental physical laws, as long as they describe sufficiently many coupled degrees of freedom.

sector.

Constraints are given by vanishing functions of phase space coordinates. In the constrained Hamiltonian formalism (Henneaux and Teitelboim, 1992), constraints are separated into a hierarchy, *primary*, *secondary*, *tertiary*, etc., depending on how they are generated. If a Lagrangian is degenerate, the conjugate momenta identified by the Legendre transform are not independent of each other. Primary constraints arise from the definition of these momenta. Secondary constraints then arise from requiring the primary constraints to be preserved under the dynamics. Tertiary constraints arise from requiring the secondary constraints to be preserved under the dynamics, and so on. The *constraint surface* is the submanifold of phase space where all constraints in the hierarchy are satisfied. Non-trivial secondary or higher constraints are needed in order for the dimension of the constraint surface to be smaller than the original phase space. The hierarchy can be further divided into two classes. The *first class* includes any constraint whose Poisson bracket with all constraints vanishes on the constraint surface. The *second class* includes all other constraints. These have a non-vanishing Poisson bracket with at least one other constraint on the constraint surface.

Constraints can either arise naturally as the generators of gauge symmetries or be inserted into the theory by hand. Either way, viewed as limits on physical possibility, they can reduce the effective dimensionality of phase space. Starting with a degenerate higher-order theory, one must first calculate the full hierarchy of constraints. Plugging the constraints into the original Lagrangian and extremizing yields the physical equations of motion on the constraint surface. At this stage two important questions arise: is the reduced theory stable, and is it still a high-order theory?

Since the Ostrogradski Hamiltonian (18) is unstable in $n-1$ momentum dimensions, at least $n-1$ secondary or higher constraints are needed to excise the entire unstable sector. Furthermore, Pons (1996) proves that the Lagrangian for the reduced dynamics is non-degenerate iff the original theory has only first-class constraints. So if a degenerate theory has only first-class constraints and the reduced dynamics are high-order, the original version of the Ostrogradski theorem applies and the theory remains unstable. So a necessary condition for the removal of the Ostrogradksi instability is the existence of at least some second-class constraints and $n-1$ secondary or higher constraints.

Motohashi et al. (2016) and Klein and Roest (2016) explore a number of instructive cases where it is possible to fully remove the Ostrogradski in-

stability. The simplest examples are degenerate Lagrangians of the form $\mathcal{L}(\ddot{q}, \dot{q}, q, \dot{x}, x)$, describing the coupling of a fourth-order system with configuration variable q, to a non-degenerate second-order system with configuration variable x. They show that a necessary and sufficient condition for avoiding the Ostrogradski instability in this case is the following:

$$\frac{\partial^2 \mathcal{L}}{\partial \ddot{q}^2}\frac{\partial^2 \mathcal{L}}{\partial \dot{x}^2} - \Big(\frac{\partial^2 \mathcal{L}}{\partial \dot{x} \partial \ddot{q}}\Big)^2 = 0 \ , \tag{21}$$

which expresses the vanishing of the determinant of the generalized Hessian matrix. (21) is equivalent to the existence of a second-class primary constraint,

$$\Xi := P_2 - f(Q_2, Q_1, P_x, Q_x) \approx 0 \ , \tag{22}$$

where P_1, P_2, Q_1, Q_2 and P_x, Q_x are the canonical coordinates associated with the fourth-order and second-order system respectively, and ≈ 0 means that the function vanishes on the constraint surface. Requiring $\{\Xi, \mathcal{H}\} \approx 0$ generates a secondary constraint that reduces the phase space dimension by one, eliminating the unstable momentum sector. When all the dust settles, though, the reduced equations of motion are only second-order.

These examples can be generalized by considering Lagrangians of the form $\mathcal{L}(\ddot{q}_i, \dot{q}_i, q_i, \dot{x}, x)$, describing the coupling of multiple fourth-order systems with configuration variables q_i, $i = 1, \ldots, k$, to a non-degenerate second-order system. For instance, the q_i might represent k coupled Pais-Uhlenbeck oscillators in a model of an effective fourth-order field theory. In this case, the analogue of (21) is no longer sufficient for avoiding the Ostrogradski instability. It does give rise to k primary constraints, Ξ_i, analogous to (22), but these are no longer guaranteed to generate a sufficient number of secondary (or higher) constraints needed to eliminate the k unstable dimensions from phase space. Motohashi et al. prove that a sufficient (if rather strong) condition is the vanishing of all Poisson brackets of the primary constraints with each other, $\{\Xi_i, \Xi_j\} \approx 0$. If so, the theory is rendered stable, but once again the reduced equations of motion are second-order.

These examples show that constructing a stable, high-order Lagrangian theory is a delicate balancing act. The constraints are entered by hand and tuned to have a particular functional dependence with each other.[17] To date,

[17] A necessary condition for stability is $\det\{\Xi_i, \Xi_j\} \approx 0$ (Motohashi et al., 2016), so at least this much cancelation is required.

no examples have been constructed where the constraints arise from the action of a natural-looking gauge group. Nor have the techniques been extended to field theories; in the continuum limit, $i \to \infty$, and the task of tuning all of the primary constraints becomes even more daunting.[18] Moreover, in all such examples, the reduced equations of motion are ultimately second-order. They are not really higher-derivative theories at all.

Perhaps there are ways to construct genuine high-order interacting field theories from degenerate Lagrangians that we have not yet discovered. Or perhaps there is a generalization of the Ostrogradski theorem ruling out all such theories as pathological.[19] Although our understanding remains partial at this stage, all current evidence points towards the idea that stable high-order theories are extremely difficult, if not impossible, to construct. To sum up: if nature is described by an interacting Lagrangian field theory with a stable vacuum, then higher than second-order equations of motions are either impossible or very special, requiring just the right interplay between constraints to eliminate the Ostrogradski instability without reducing the dynamics to second-order laws.

5 Laws, Meta-Laws, and Non-Causal Explanation

We have good reason to believe that our world is described (to close approximation) by Lagrangian field theories. The Ostrogradski theorem reveals that higher than second time derivatives cannot easily be incorporated into this theoretical framework, if at all. This tells against any set of fundamental high-order laws, not just mixtures of high-order and low-order laws like Easwaran's explanation. Even if it turns out to be possible to exploit the degeneracy loophole to construct stable high-order field theories, it appears inevitable that such theories will have artificially tuned constraints, render-

[18] Valencia Villegas (2017) makes some progress towards generalizing the approach of Motohashi et al. to high-order scalar field theories. His analysis reveals that there are a number of additional unexpected complications that arise for stabilizing such theories. For example, if a high-order field system is coupled to a low-order field system, Ostrogradski stability requires a lower bound on the coupling parameter, $\alpha > 1/m$, where m is the mass of the low-order field.

[19] Motohashi et al. (2016) sketch one such argument, but it is based entirely on their strong sufficient condition for stability, $\{\Xi_i, \Xi_j\} \approx 0$, and therefore not entirely satisfactory.

ing them less natural than low-order laws. This represents a significant improvement over the vague appeal to simplicity in Easwaran's explanation. In addition, the Ostrogradski explanation better tracks scientific practice. Physicists are actively interested in the viability of high-order field theories and they view the Ostrogradski instability as a significant no-go result to contend with in this arena. Along with renormalizability, it represents one of the most powerful, general theoretical constraints on the construction of new field theories (Simon, 1990; Woodard, 2015). In contrast, Easwaran's story relies on contentious metaphysical principles that do not play a major role as theoretical constraints in scientific practice. The Ostrogradksi story is therefore a more appealing explanation from a broadly naturalistic standpoint.

Prima facie, the Ostrogradski explanation is compatible with both the causal and reductionist theses, as well as their negations. Insofar as Lagrangian mechanics employs standard tools from differential geometry, it is plausible that there are various points where the argument in §3-4 presupposes that velocity, acceleration, and higher time derivative quantities can be defined by appropriate limits. But nothing in the explanation turns on viewing such derivative quantities as either neighborhood properties or instantaneous properties. Similarly, while the contours of the explanation may shift depending on the background view of laws, nothing about it commits us to one view over another. On Humean views, the explanandum is a fact about the best system summarizing the categorical facts. Like any fact about the laws, the absence of high-order time derivatives is ultimately explained by the underlying mosaic of categorical facts. What the Ostrogradski explanation then shows is how certain structural patterns in this mosaic depend on other structural patterns. The best system cannot include high-order equations of motion if it is also stable, Lagrangian, and field-theoretic.[20] On non-Humean views, the explanandum is a fact about a certain species of modal facts. Whether these modal facts are primitive (e.g., Maudlin 2007), or reducible to other non-categorical facts (e.g., Cartwright 1999, Lange 2009), the Ostrogradski explanation reveals how they cannot involve higher than second time derivatives while also retaining their stable Lagrangian field-theoretic character.

One might object that Lagrangian laws themselves are incompatible with

[20]Whether the field-theoretic character of the laws is a categorical or non-categorical fact will depend on the particular version of Humeanism under consideration.

the causal thesis. The worry is that although we can derive causal equations of motion from Lagrangian principles, fundamentally speaking, Lagrangian laws determine the evolution of the system by the principle of extremal action, requiring boundary conditions in both the future and the past. There are two things to say in response. First, it is not clear that an interpretation of Lagrangian mechanics that treats the action principle as fundamental is necessarily incompatible with the causal thesis. As Easwaran presents it, the thesis is silent about how exactly the laws and present state "causally determine" future states. One feasible option is that the action principle grounds the space of possible dynamical histories, then the actual present state causes future states subject to this modal constraint. Another option is that the action principle grounds the equations of motion, and these along with the actual present state jointly cause future states. While the fundamental laws may not be considered causes of future states on either of these readings, it is unclear if the causal thesis is actually committed to this claim (or should be).

Second, one can opt for an interpretation of Lagrangian mechanics where the equations of motion, rather than action principles, are viewed as more fundamental. Different Lagrangian functionals can give rise to the same equations of motion, and physicists typically interpret such Lagrangians as physically equivalent.[21] From this angle, action principles simply look like a convenient starting point to derive dynamical laws expressed by differential equations with a particular form. This idea can be developed into an elegant, coordinate-free formulation of Lagrangian mechanics utilizing the intrinsic geometric structure of tangent bundles (de León and Rodrigues, 1989). In this framework action principles play a secondary role, and there is no trouble reconciling Lagrangian laws with the causal thesis.

It is perhaps unsurprising that the Ostrogradski explanation remains agnostic about the metaphysical principles that drive Easwaran's explanation. Even if we doubt that there is a sharp dividing line between physical and metaphysical hypotheses, the inputs into the Ostrogradski theorem have historically been treated as the former, part of the raw data that philosophical theories about laws and explanation endeavor to capture. The theorem ex-

[21]It is well-known that Lagrangians related by point transformations give rise to the same equations of motion, but there are more general symmetries hiding in the Lagrangian formalism. For example, the transformation sending $\mathcal{L} \to \mathcal{L} + \hat{\theta}$, where $\hat{\theta}$ is the natural function on TQ defined by a closed 1-form on Q, is always a symmetry of the Euler-Lagrange equations.

poses a very general problem with high-order Lagrangian theories that is independent from the deeper metaphysical hypotheses that Easwaran employs. This is a further Ockhamist advantage, the Ostrogradski explanation commits us to less metaphysical baggage.[22]

There are a number of lingering questions. How metaphysically deflationary is the story sketched in §3-4? Is it really explanatory? If so, exactly what kind of explanation does it provide? In order to help answer these questions, we turn to two important ideas in the recent philosophical literature on non-causal explanation.

The first idea is a non-causal generalization of central themes from Woodward's interventionist account of causal explanation. Woodward (2003) argues that causal explanations aim to answer counterfactual "what-if-things-had-been-different" questions (*w-questions*, for short) by citing how one variable changes under possible interventions on other variables in a causal model. On Woodward's account, interventions are causal processes that surgically change the value of a variable, shielding that variable from other influences so that the change is only due to the intervention itself (an idealized experimental manipulation). Successful explanations identify features of the causal model that make a difference to whether or not the explanandum occurs. Such causal difference makers will be invariant under a range of possible interventions on the system in question. As many proponents of non-causal explanation have noted, the causal modeling framework and the interpretation of interventions as causal processes play a rather minor role in Woodward's account of explanation. In any sort of model or theory, if it is possible to isolate modular variables and surgically change their values by purely mathematical or conceptual interventions, we can coherently trace chains of counterfactual dependence between them. On this generalized interventionist view, non-causal explanations aim to answer w-questions just like causal explanations. The difference is that the relevant variables need not be part of a causal model, and the explanans involves citing how one

[22]If the reductionist and causal theses turn out to be true, but not metaphysically necessary (a possibility Easwaran leaves open), then this also translates into a certain scope advantage. The Ostrogradski explanation covers possible worlds where the metaphysical character of the laws is different, unifying a diverse set of cases not covered by Easwaran's explanation. Of course, if his explanation were successfully patched up, it would apply to causal reductionist worlds with both Lagrangian and non-Lagrangian laws, unifying a different set of cases outside of the scope of the Ostrogradski explanation. So this advantage alone would not be decisive.

variable changes under possible non-causal interventions on other variables, identifying invariant non-causal difference makers.[23]

The second idea is Lange's influential analysis of non-causal explanations involving constraints and meta-laws. Lange (2009, 2016) argues that just as the laws of nature can be viewed as modal constraints governing the non-nomic categorical facts, there are more abstract modal constraints that govern the laws themselves. These may include mathematical, conceptual, and metaphysical constraints, as well as *meta-laws*, metaphysically contingent laws about laws. General symmetry principles as among the best candidates for the latter. The conservation of energy can be explained by the fact that the laws of nature are time-translation invariant. On Lange's telling, this is a meta-law; even if the laws described different forces or types of matter, they would still be time-translation invariant. The explanatory credentials of meta-laws come from their greater-than-physical grade of necessity. In general, we appeal to more necessary modal constraints to explain less necessary constraints, but not vice versa. Lange goes on to defend a unified non-Humean view of laws and meta-laws, but we need not follow him down this path. The ability to accommodate the explanatory role of constraints and meta-laws is arguably a desideratum for any successful account of laws and explanation.[24]

The story sketched in §3-4 can be viewed most directly as a non-causal explanation in the generalized interventionist sense discussed above. It iso-

[23]For examples of views in this direction, see Bokulich (2011), Saatsi and Pexton (2012), Rice (2015), and Reutlinger (2016). It is also a major theme in many of the essays in Reutlinger and Saatsi (2018).

[24]This is somewhat contentious. Humeans might worry that their concept of laws cannot naturally accommodate meta-laws, leading to skepticism about the latter. (Humean meta-laws would be part of the best system summarizing the nomic facts, but it is unclear exactly what such a summary should look like.) Similarly, non-Humeans like Maudlin who view laws as primitive modal facts about temporal evolution might be skeptical of meta-laws that are not obviously dynamical. If the Ostrogradski explanation relies on meta-laws, then the story may require taking a stand on the metaphysics of laws and meta-laws after all. Perhaps, but this skepticism can be resisted. Meta-laws can be interpreted as very general kinematic constraints consistent with non-Humean views like Maudlin's, while Yudell (2013) sketches a possible strategy for extending the Humean account. Furthermore, as we will go on to see, there is a deflationary reading of the Ostrogradski explanation available on which the constraints are mathematical and conceptual necessities rather than meta-laws. Interpreting certain inputs into the Ostrogradski theorem as meta-laws may modally strengthen the explanation (and this may come with certain metaphysical costs), but the deflationary reading can be adopted by every party in the debate.

lates independent features of the laws of nature and explores changing these features via conceptual and mathematical interventions to answer a range of different w-questions. How would things be different if the laws contain higher than second time derivatives? How would they be different if the Lagrangian is degenerate, or if the laws are not Lagrangian at all? In the process, we have identified certain non-causal factors — the laws are Lagrangian, field-theoretic, describe non-trivial interactions, and have a stable vacuum solution — that make a difference to whether or not the equations of motion can include higher-derivative terms. Moreover, these difference makers are invariant under a broad range of interventions changing the particular form of the Lagrangian, which forces and interactions are present, the matter content of the fields, and the background spacetime structure.

The story can also be cast as a constraint explanation in Lange's sense, although exactly which constraints are operative is subject to debate. On a minimalist reading, the Ostrogradski theorem acts as a mathematical constraint. In conjunction with certain conceptual constraints pertaining to the physical interpretation of the mathematics, the argument in §4 yields a necessary conceptual truth: if an interacting non-degenerate Lagrangian field-theory has a stable vacuum state, then it cannot include higher than second time derivatives. (Making this rigorous will require showing that Woodward's informal argument in §4 can be turned into a deductively valid, mathematically precise argument.) If it turns out that there is a generalization of the Ostrogradski theorem covering degenerate Lagrangians, then there is a broader conceptual truth of this type that we should appeal to instead. If not, and it turns out that stable degenerate theories are possible but fine-tuned, then the explanation takes on a different character altogether. The mathematical and conceptual constraints do not rule out high-order theories tout court, but render them less likely or less natural than low-order theories in some sense requiring further elaboration. (Making this precise will require a choice of topology on the space of Lagrangian field theories allowing for the definition of a suitable probability measure or a more general measure characterizing *generic* theories.)

This reading reveals a pattern of mathematical and conceptual constraints which supports a deflationary resolution to our puzzle. Physicists are interested in stable, interacting Lagrangian field theories, and it is either very hard or impossible for such theories to have higher-order equations of motion. Insofar as we have reason to believe that our world is described by such theories, we have reason to believe that the laws cannot include higher

than second time derivatives. But is there a deeper sense in which the laws must have this form? There are more robust readings available on which one or more of the assumptions in the Ostrogradski theorem are interpreted as meta-laws. Just as the laws are held fixed under a suitably broad range of counterfactual suppositions about the categorical facts, the meta-laws are held fixed under a suitably broad range of counterfactual suppositions about the laws. Even if the laws were different, they would still be constrained by the meta-laws.

A strong case can be made that the Lagrangian character of the laws is itself a meta-law. In practice, physicists consider all sorts of counterfactual variations of the laws while holding their Lagrangian character fixed. (It is rare to find mention of non-Lagrangian laws at all.) Like Hamiltonian mechanics, Lagrangian mechanics represents an extremely fruitful, unified framework for constructing a range of wildly different theories that nonetheless share important structural commonalities. These structural properties have survived multiple scientific revolutions that have otherwise radically reshaped our view of what the laws of nature might look like. Moreover, they are the sort of frameworks that allow us to ask well-posed counternomic questions in the first place. How would things be different if the laws included higher than second time derivatives? That is simply too broad of a question to have a determinate answer. But if the laws are Lagrangian, then we can say something more definite about the form that these higher-derivative laws might take, and in certain circumstances rule them out.

Lange (2009) contrasts meta-laws with *byproducts* of the laws, properties of the laws that hold in virtue of whatever the laws happen to be. Byproducts do not constrain the laws, they are explained by the laws. For instance, it seems that we should appeal to the laws to explain why nature has non-trivial interactions. This feature is most plausibly interpreted as a byproduct. Similarly, the non-degeneracy assumption is likely a byproduct, although one which appears to be an eliminable part of the explanation. The situation for the stability and field-theoretic assumptions are less clear. Although stability is often viewed as an axiom for field theories, it need not be interpreted as a constraint on physical possibility. Instead, it can be interpreted as a constraint on epistemic possibility in line with Woodard's preferred reading of the no-go argument in §4. Unstable field theories cannot describe universes like ours where "all particles have positive energy and empty space remains empty." In this case, it is more natural to view stability as a byproduct

explained by the laws rather than a constraint.[25] But a strong argument that unstable field theories are physically pathological could tip the scales in the other direction. It is similarly unclear that the laws must be field-theoretic in any modal sense other than epistemic. The plethora of non-field theories in both classical and quantum physics suggests not. On the other hand, physicists often view Lagrangian field theory as a sub-framework within the broader Lagrangian framework. Interpreting field-theoretic assumptions as meta-laws might be coherently motivated by metaphysical arguments against action at a distance or by more localized concerns such as the need to unify quantum mechanics and relativity.[26]

As before in the debate over laws, the contours of the explanation offered by the Ostrogradski instability might shift depending on how these questions are answered. Even if none of the main assumptions turn out to be meta-laws, though, the minimalist reading provides a compelling deflationary explanation. A wholesale skeptic about meta-laws can still acknowledge the explanatory force of the mathematical and conceptual constraints entailed by the Ostrogradski theorem. In addition, it appears that any version of the story sketched in §3-4 will have a residual anthropic component. Nothing forces the laws to have non-trivial interactions, but in worlds like

[25]There is a worry here. If stability is explained by the form of the laws, then we cannot appeal to it to explain the lack of higher derivatives in the equations of motion without getting the order of explanation backwards. From this perspective, it is the form of the laws, including the absence of higher derivatives, that explains stability, not vice versa. Although the outlook here is unclear, I think the Ostrogradski story remains explanatory, regardless. If both vacuum stability and the absence of high-order derivatives are byproducts, there may simply be no determinate fact about grounding relations between them. In this case stability along with the relevant mathematical/conceptual constraints can explain the lack of higher derivatives, or vice versa. Note that the derivations are not symmetric: stability entails that there are no higher derivatives, but the fact that there are no higher derivatives does not entail stability unless the Lagrangian has the right form to eliminate the linear dependence on P in equation (13). Even if we interpret stability as partly grounded in the absence of higher derivatives, and therefore only the second derivation turns out to be metaphysically explanatory, then the minimalist reading of the Ostrogradski story still plausibly offers a type of epistemic explanation: given our evidence, including the fact that our world is stable, the fundamental laws cannot include higher derivatives. Indeed, if stability is a byproduct partly grounded in the absence of higher derivatives, I suspect that this is the only sort of explanation for the absence of high-order laws that can be given.

[26]A number of informal arguments (see, for example, Weinberg 1995, ch. 1) suggest that any interacting relativistic quantum theory must be a field theory, but so far these arguments have not been made rigorous.

ours where there are interactions, additional constraints apply. The debate is over the character of these constraints and the extent of this residual anthropic component. Are the central assumptions in the Ostrogradski theorem meta-laws or byproducts? Future investigation into this question, as well as into extensions of the theorem for degenerate Lagrangians, stand to enrich the philosophical literature on non-causal explanation and further illuminate why physics really uses second derivatives.

Acknowledgements

I would like to thank Kenny Easwaran, David Baker, John Burgess, Ned Hall, Hans Halvorson, and Tristan Smith for valuable discussion, comments, and criticism.

References

Arntzenius, F. (2000). Are there really instantaneous velocities? *Monist 83*, 187–208.

Bokulich, A. (2011). How scientific models can explain. *Synthese 180*, 33–45.

Cartwright, N. (1999). *The Dappled World: A Study of the Boundaries of Science*. Cambridge University Press.

de León, M. and P. Rodrigues (1989). *Methods of Differential Geometry in Analytical Mechanics*, Volume 158 of *North-Holland Mathematics Studies*. North-Holland.

Eager, D., A.-M. Pendrill, and N. Reistad (2016). Beyond velocity and acceleration: Jerk, snap, and higher derivatives. *European Journal of Physics 37*, 065008.

Easwaran, K. (2014). Why physics uses second derivatives. *British Journal for the Philosophy of Science 65*, 845–62.

Eliezer, D. and R. Woodard (1989). The problem of nonlocality in string theory. *Nuclear Physics B 325*, 389.

Grinstein, B., D. O'Connell, and M. B. Wise (2008). The Lee-Wick standard model. *Physical Review D 77*, 025012.

Henneaux, M. and C. Teitelboim (1992). *Quantization of Gauge Systems*. Princeton University Press.

Klein, R. and D. Roest (2016). Exorcising the Ostrogradsky ghost in coupled systems. *Journal of High Energy Physics 2016*(130).

Lange, M. (2005). How can instantaneous velocity fulfill its causal role? *Philosophical Review 114*, 433–68.

Lange, M. (2009). *Laws and Lawmakers:*. Oxford University Press.

Lange, M. (2016). *Because Without Cause: Non-Causal Explanation in Science and Mathematics*. Oxford University Press.

Maudlin, T. (2007). *The Metaphysics Within Physics*. Oxford: Oxford University Press.

Miron, R., G. Hrimiuc, H. Shimada, and S. Sabau (2002). *The Geometry of Hamilton and Lagrange Spaces*. Kluwer Academic Publishers.

Motohashi, H., K. Noui, T. Suyama, M. Yanaguchi, and D. Langlois (2016). Healthy degenerate theories with higher derivatives. *Journal of Cosmology and Astroparticle Physics 2016*(033).

Moura, F. and R. Schiappa (2006). Higher-derivative-corrected black holes: Perturbative stability and absorption cross section in heterotic string theory. *Classical and Quantum Gravity 24*, 361.

Ostrogradski, M. (1850). Mémoire sur les équations différentielles relatives au problème des isopérimètres. *Memoirs of the Academy of St. Petersbourg VI*(4), 385.

Pais, A. and G. Uhlenbeck (1950). On field theories with nonlocalized action. *Physical Review D 79*(145-65).

Pons, J. (1989). Ostrogradski's theorem for higher order singular Lagrangians. *Letters in Mathematical Physics 17*(181-9).

Pons, J. (1996). Plugging the gauge fixing into the Lagrangian. *Interational Journal of Modern Physics A 11*, 975–88.

Reutlinger, A. (2016). Is there a monist theory of causal and non-causal explanations? The counterfactual theory of scientific explanation. *Philosophy of Science 83*(733-45).

Reutlinger, A. and J. Saatsi (Eds.) (2018). *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanation*. Oxford University Press.

Rice, C. (2015). Moving beyond causes: Optimality models and scientific explanation. *Noûs 49*(589-615).

Russell, B. (1903). *Principles of Mathematics*. Cambridge University Press.

Saatsi, J. and M. Pexton (2012). Reassessing Woodward's account of explanation: Regularities, counterfactuals, and non-causal explanation. *Philosophy of Science 80*, 613–24.

Simon, J. (1990). Higher-derivative Lagrangians, nonlocality, problems, and solutions. *Physical Review D 41*(12), 3720–33.

Stelle, K. (1977). Renormalization of higher-derivative quantum gravity. *Physical Review D 16*, 953.

Tooley, M. (1988). In defense of the existence of states of motion. *Philosophical Topics 16*, 225–254.

Valencia Villegas, J. (2017). *Higher-Order Time Derivative Theories. Interpretation, Instability, and Possible Stabilization*. Ph. D. thesis, National University of Columbia.

Weinberg, S. (1995). *The Quantum Theory of Fields*, Volume I. Cambridge University Press.

Woodard, R. P. (2007). Avoiding dark energy with 1/R modifications of gravity. In L. Papantonopoulos (Ed.), *The Invisible Universe: Dark Matter and Dark Energy*, Volume 720 of *Lecture Notes in Physics*. Springer, Berlin, Heidelberg.

Woodard, R. P. (2015). The theorem of Ostrogradski. *Scholarpedia 10*, 32243.

Woodward, J. (2003). *Making Things Happen*. Oxford University Press.

Yudell, Z. (2013). Lange's challenge: Accounting for meta-laws. *British Journal for the Philosophy of Science 64*, 347–69.