International Studies in the Philosophy of Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cisp20

Explanations by Constraint: Not Just in Physics

Marc Lange

To cite this article: Marc Lange (2023) Explanations by Constraint: Not Just in Physics, International Studies in the Philosophy of Science, 36:4, 265-277, DOI: 10.1080/02698595.2023.2298085

To link to this article: https://doi.org/10.1080/02698595.2023.2298085

INTERNATIONAL STUDIES IN THE PHILOSOPHY OF SCIENCE
2023, VOL. 36, NO. 4, 265–277
https://doi.org/10.1080/02698595.2023.2298085

Explanations by Constraint: Not Just in Physics

Marc Lange

Department of Philosophy, University of North Carolina at Chapel Hill, NC, USA

ABSTRACT
Several philosophers have argued that 'constraints' constrain (and thereby explain) by virtue of being modally stronger than ordinary laws of nature. In this way, a constraint applies to all possible systems, for a variety of possibility that is broader (that is, more inclusive) than the variety we employ when we say that the ordinary laws of nature apply to all physically possible systems. Explanations by constraint are thus more broadly unifying than ordinary causal explanations. Philosophical examples of good candidates for constraints have heretofore been drawn almost exclusively from fundamental physics. This paper argues for the existence of such constraints (even multiple levels of them) in at least one human science (linguistics), not just in physics.

ARTICLE HISTORY
Received 4 July 2023
Accepted 19 December 2023

1. Introduction

A constraint (in the sense in which I shall use the term) is an explanatory principle in science that transcends the various principles that specify particular sorts of causal processes. Instead, a constraint imposes broad limits on the kinds of causes there could have been. For this reason, a constraint possesses a stronger variety of necessity than an ordinary law of nature possesses. A constraint would still have held even if the ordinary laws of nature that describe various particular kinds of causal influences had been different. (That was a counterfactual—indeed, a counterlegal.) A constraint explains not by virtue of describing the explanandum's causal history—and not even by describing the world's causal network as a whole. Instead, a constraint explains by describing the framework into which any possible causal interaction would have to fit.[1]

In this paper, I will briefly elaborate this conception of a constraint in terms of some examples of facts that have often been taken to be constraints in fundamental physics. I will then consider whether scientists working in other scientific fields, such as in the human sciences, have (at least implicitly) treated various other facts as constraints. I will argue that they have. I will examine some plausible examples in linguistics.

2. What are Constraints?

The law of energy conservation, for instance, is often taken by physicists to be a constraint.[2] Suppose we want to explain why a given isolated system has the same total

CONTACT Marc Lange mlange@email.unc.edu Chapel Hill, NC 27599, USA

energy as it did at some earlier moment. If energy conservation is a constraint, then to list the various kinds of forces that the system's components experience, and to show that each of these forces conserves energy, would misrepresent the reason why the system's energy is unchanged. If energy conservation is a constraint, then the fact that the isolated system's energy was conserved does not depend on the particular causes that happened to be operating within the system. It does not even depend on the types of causes that are possible according to the various fundamental force laws. If it is a constraint, then the principle of energy conservation transcends those petty causal details. The conservation law limits the kinds of forces that there could have been in that energy would still have been conserved even if there had been another species of fundamental force. (Here we have that counterlegal once again.) The source of energy conservation, on this picture, could be a spacetime symmetry principle (coupled with the constraint that all forces must operate within a Hamiltonian dynamical framework): namely, that (roughly speaking) causal laws cannot treat different moments differently.[3]

If energy conservation is a constraint, then the fact that electromagnetism, the weak nuclear force, and each of the other various fundamental kinds of forces conserves energy does not actually explain why energy is conserved. That is because such a putative explanation would incorrectly depict energy conservation as a kind of coincidental similarity among the various fundamental kinds of forces. Rather, if energy conservation is a constraint, then the order of explanatory priority runs in the opposite direction: the fact that energy must be conserved explains why each of the various actual fundamental kinds of forces conserves energy. The kinds of fundamental forces that are possible extend far beyond the kinds for which there are actual fundamental force laws. But the possible forces are limited by energy conservation. If energy conservation is a constraint, then energy conservation explains why the isolated system's energy remains unchanged. But this explanation does not work by describing relevant features of the system's causal history. For this reason, energy conservation does not supply *causal* explanations.

Many other physical principles are also often thought to be constraints on lower-level laws. Their status as constraints is reflected in their possessing greater necessity than the lower-level laws possess. They possess greater invariance under counterfactual antecedents than the force laws do. Even if the laws of dynamics had been different by including additional kinds of forces, the conservation laws (if they are constraints) would still have held. Likewise, even if photons had possessed some nonzero mass, the Lorentz transformations would still have held; the symbol 'c' in the transformations would simply not have stood for the speed of light (Lévy-Leblond 1976, 271). In the same way, the law of the parallelogram of forces would still have held even if there had been additional kinds of forces. The principle of relativity is often thought to be a meta-law restricting the forms that lower-level laws can take.[4] In classical physics, principles of kinematics and facts about the geometry of space seem to transcend the dynamics.[5] And mathematical facts have sometimes been thought to explain by constraining what there could have been. Mathematical truths have a stronger variety of necessity than even exalted laws of nature possess and so constrain what those natural laws could be. This idea was well expressed by Jean-Pierre Serre as quoted by Raoul Bott (Tu 2013, 414) when they were jointly awarded the Wolf Prize in Mathematics in 2000: 'While the other sciences search for the rules that God has chosen for this Universe, we mathematicians search for the rules that even God has to obey.'

Of course, the term 'constraint' is used in many senses in physics besides the sense in which I am using it. It has a particular sense in analytical mechanics, for instance, and it is also sometimes used to mean simply a well-established fact that any theory needs to fit in order to be a viable candidate for adoption. The sense of 'constraint' in which I am interested obviously differs from these two senses.

3. Constraints as Members of Stable Sets that Omit Some Laws

For the sake of definiteness, here is one picture (from Lange 2017) that aims to capture the foregoing distinction between constraints and ordinary laws. On this picture, it is part of scientific practice that the laws would all still have held under (roughly speaking) any counterfactual antecedent that is logically consistent with the laws. But their invariance under this particular range of counterfactual antecedents cannot be what makes the laws so important to science, since this range is itself picked out precisely by the laws and so invariance under this particular range is important only if the status of being a law is independently important. But there is a way for the laws' distinctive invariance under counterfactual antecedents to be what makes the laws so important; there is a way to specify the laws' distinctive persistence under counterfactual antecedents without invoking the distinction between laws and accidents in the first place. The solution is to recognise that (roughly speaking) the laws form a logically closed set of truths the members of which would all still have held under any counterfactual antecedent that is logically consistent with all of the set's members. That is, the set containing exactly the laws' logical closure is a 'stable' set. Forming a stable set is a property that can be specified (as I just did) without appealing to the laws. In this way, we can identify the sort of invariance under counterfactual antecedents that makes the laws special.

In forming a stable set, the laws would all still have held under every counterfactual antecedent that is logically consistent with them all. That is, the laws *would* all still have held under every counterfactual antecedent under which they *could* all still have held. In other words, the laws are collectively maximally invariant under counterfactual antecedents. Thus, they are inevitable, unavoidable—necessary. They *must* hold in that they would all still have held no matter what (i.e. under any condition under which they could logically possibly all still have held). In possessing a variety of genuine necessity ('natural' necessity), the laws differ from (for example) the truths that are logically entailed by all and only the facts about where in spacetime there are cats. We might say, in some context, that it is 'impossible' for me today to walk around my block without passing a cat. I 'must' pass a cat at some point as I walk around the block. But this 'necessity' is not a genuine variety of alethic modality because the facts giving the spacetime distribution of cats (at least on my block today) do not form a stable set. (Had there been a loud noise on the block, no cats would have been there.)

More generally, no set of truths containing an accident is stable, except for the set of all truths (which is trivially stable since no counterfactual antecedent is logically consistent with this set). It can be demonstrated that for any two stable sets, one must be a proper subset of the other (Lange 2009, 37). That is, the stable sets must form a natural hierarchy.

On this picture, a fact qualifies as a constraint by virtue of belonging to a set of truths that manages to achieve stability despite omitting some of the laws. Compared to the set containing all and only the laws, a set of constraints is a more exalted (that is, more exclusive) stable set in the natural hierarchy of stable sets. A constraint thereby possesses a stronger variety of necessity than an ordinary law of nature possesses. That is, the members of a set of constraints would all still have held even if the dynamical laws had been different in some respect that is logically consistent with that set of constraints. (Here we have that counterlegal once again.) Perhaps the conservation laws, the parallelogram of forces, the Lorentz transformations, and their colleagues form a set that is stable despite omitting the force laws. These constraints would all still have held, even if the force laws had been different.[6] That invariance is part of what makes them constraints; the dynamical laws have got to conform to them.

The facts of pure mathematics might belong to an even more exclusive set of constraints, constraining even the conservation laws and its mates. The mathematical facts thereby possess a variety of necessity that is even stronger than the conservation laws possess. Their variety of necessity entails that the facts of pure mathematics would have still held even if the conservation laws had not held.[7]

4. Constraints in Linguistics

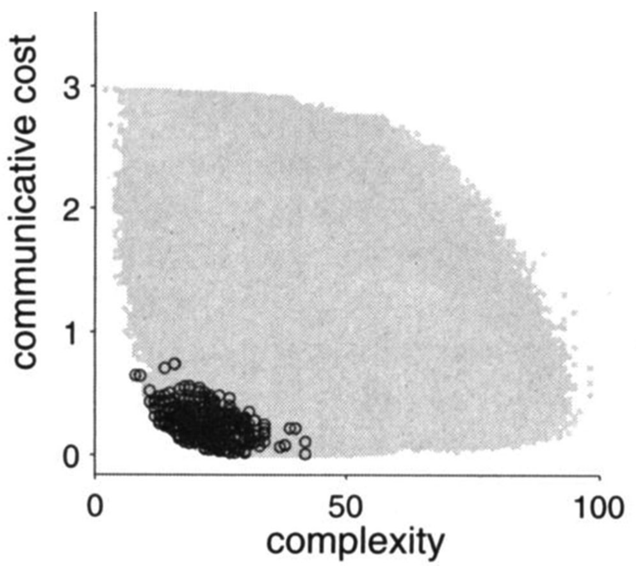
All of the various examples that I have given of facts that are often regarded as constraints (that is, as higher-level laws limiting the possible lower-level laws and as thereby providing non-causal explanations) have been putative laws of fundamental physics or facts of pure mathematics. Other scientific fields have their own laws (some of which may be accidents in fundamental physics, just as some laws of physics may be accidents of some special science). Each scientific field has its own range of interests. That range influences which facts and which counterfactual antecedents are relevant to the given field—and which falsehoods are close enough to the truth for that field's purposes. Accordingly, different sets qualify as stable for the purposes of different scientific fields (Lange 2002).

Are there different hierarchies of stable sets in different fields? In particular, are there serious candidates for constraints (and explanations by constraint) in sciences other than fundamental physics—especially in the human sciences? I will now argue that there are. Scientific fields besides fundamental physics have taken seriously the prospect of constraints—even multiple levels of constraints. I will now suggest that linguistics is one such field.

Different languages carve up human family relationships in different ways. For example, English uses two terms to refer to grandparents: the words 'grandfather' and 'grandmother'. I am told that Swedish and Mandarin Chinese use four distinct terms, allowing all four grandparents to be distinguished. This makes the terms more informative, which is beneficial to communication, but it also makes these languages more complicated to learn, to remember, and to use. The Australian language Kayardild reserves one term, *marrkathu,* for your father's sisters while grouping your mother's sisters with your mother as *ngamathu* (Passmore et al. 2023, 176). By contrast, English has the term *aunt* for all and only your parents' sisters and the female partners of your uncles. The number of possible systems for categorising even just siblings, parents, parent's siblings, and cousins has been calculated to include over 10 billion possible options. There are a tremendous number of systems that could be used for kin-categorisation but are used by no actual natural languages at all.

For over a century, anthropologists and linguists have contributed to the study of kinship-categorisation terminology by collecting data from various languages (past and present) and offering possible explanations of those data (Mollica and Piantadosi 2022, 767). In 2012, the linguists Charles Kemp and Terry Regier published an influential paper in *Science* that aimed to explain why certain possible kinship categorisations but not others are used by actual human languages. Kemp and Regier (2012) consider several hundred actual human languages and a huge number of merely possible kin-categorisation systems. For each one, Kemp and Regier measure the simplicity of its kin-categorisation system. A system 'is simple to the extent that it can be concisely mentally represented and therefore easily learned and remembered' (Kemp and Regier 2012, 1049). Its simplicity depends on the rules that are needed to define its various kinship relations in terms of a small set of primitives. Kemp and Regier also measure the informativeness of its kin categories. 'A system with a single category that includes all possible relatives would be simple but uninformative because this category does not help to pick out specific relatives. A system with a different name for each relative would be complex but highly informative because it picks out individual relatives perfectly' (Kemp and Regier 2012, 1049). Kemp and Regier's measure of informativeness reflects which kinds of kin information are generally more valuable for us to have. For instance, each person has parents and other ancestors (or else they would not have existed) but not everyone has siblings or grandchildren. So having additional information about ancestors baked into a language's kin-categories is generally more valuable than the same additional information about equally distant descendants. (Fans of David Lewis's best system account of laws (Lewis 1983; 1986; 1999) should love this appeal to 'the trade-off between simplicity and informativeness' (Kemp and Regier 2012, 1049)!)

Using these metrics of simplicity and informativeness, Kemp and Regier show that the overwhelming majority of actual languages have kin-categorisation systems with a nearly optimal combination of simplicity and informativeness. Here is a figure from their paper (Kemp and Regier 2012, 1052):



A possible kin-categorisation system is represented by a small grey region on the graph. Each actual system is represented by a black circle around the corresponding small grey region. These black circles are clustered towards the bottom left of the graph so densely that individual ones cannot be picked out. Kemp and Regier conclude that the fact that the kin-categorisation systems used by actual languages are clustered in this corner of the graph (that is, are all nearly optimal in their combination of simplicity and informativeness) is explained by the fact that languages tend to develop so as to increase the efficiency of communication by increasing informativeness and simplicity.

This is not intended to be a probabilistic explanation. The fact being explained is that certain possible kin-categorisation systems are not actualised in natural human languages, and the purported reason why is that they are not near-optimal (in terms of the simplicity/informativeness trade-off) and a system that is not near-optimal will not be actualised (because of the strong causal pressures favouring simplicity and informativeness).

Kemp and Regier also purport to explain in this way why near relatives tend more often than distant relatives to be split into multiple categories and why grandparents tend more often than grandchildren to be split. Again, there is more informational benefit in the actual splits because, for example, everyone has grandparents but not everyone has grandchildren. All of these proposed explanations are straightforwardly causal explanations. These proposed explanations are probabilistic since they invoke the probability that a speaker would benefit from being able to convey the information carried by a term for given kin-category. Facts like those I just mentioned (e.g. that near relatives are split more often into separate categories than are distant relatives) are to be 'explained as a consequence of the nonuniform distribution of need probabilities. … In particular, need probabilities are higher for near relatives than distant relatives and higher for ascending generations than descending generations' (Kemp and Regier 2012, 1052).

Although Kemp and Regier aim to explain why various possible kin-categorisation systems are absent from actual languages, they do not aim to explain why a given pair of actual languages have different near-optimal systems. They also do not aim to explain why a given language has one near-optimal kin-categorisation system rather than another. Kemp and Regier say that these explanations would require that more specific causes be considered, such as 'local social patterns of marriage and residence' (1053). Passmore and Jordan (2020, 1–2) likewise maintain that the joint optimisation of simplicity and informativeness in kinship terminology (as Kemp and Regier propose) 'explains the absence of many theoretical possibilities, but do[es] not explain why we observe any variation at all, nor the origin and maintenance of particular variants'. Rather, 'social structure, specifically, kinds of inheritance, descent, marriage and residence' and other 'local cultural practices' that embody 'cultural beliefs' will 'affect and are affected by how kin are categorized'. Researchers are now seeking 'not necessarily deterministic' (Passmore and Jordan 2020, 2) dynamical laws by which specific cultural features fix the ways that the two pressures invoked by Kemp and Regier (simplicity and informativeness) should be understood in a given case along with how 'the relative strength of these two pressures [should] be varied in different social contexts' (Kirby et al. 2015, 87). Many follow-up studies have recognised that Kemp and Regier's work 'has shown that … the trade-off between communicative efficiency and simplicity can explain at a coarse level the observed diversity in kinship systems' while emphasising that 'we need fine grained theories' to explain why there are different kin-categorisation systems in different languages and 'to explain evolutionary trajectories' of the kin-categorisation systems in individual languages (Mollica and Piantadosi 2022, 767).[8]

But now consider a different explanandum. Why are there no possible languages beyond the frontier of the grey zone (largely covered in black) in the bottom left corner of the diagram? That is, why is it impossible for a kin-categorisation system to do better in both simplicity and informativeness than the optimal systems in the figure? Why can't we get even closer to the origin in the diagram? What makes even better combinations of scores impossible?

This is not one of the facts that Kim and Regier aim to explain. They are concerned primarily to point out facts about the range of actual languages that are explained (in their view) by the tendency of languages to optimise simplicity and informativeness. That is, they aim to point out facts about human languages that are explained by how beneficial it is to increase the ease of language learning and the quantity of information communicated. But the explanandum that I have just mentioned (the optimisation frontier on the bottom left of the graph, i.e. the range of *possible* systems) is not set by the factors causing languages to tend to optimise simplicity and informativeness. Kemp and Regier never explicitly take the optimisation frontier as itself an explanandum; they take the range of actual systems, from within the range of possible systems, as an explanandum, but they do not take the range of possible systems as an explanandum. Nevertheless, their account generates the range of possible kin-categorisation systems (and uses it, in turn, to explain the range of *actual* systems). They arrive at that range by beginning with various constraints on human linguistic kin-categorisation systems: for instance, that each human being has exactly two biological parents, that human beings tend to live with their nearer biological kin, and so forth. By accounting for the range of possible systems, Kemp and Regier's approach explains the location of the optimality frontier and hence why no actual systems are found beyond that frontier. Suppose that human beings had reproduced entirely by binary fission, as bacteria do—or entirely by giving birth to genetically identical quadruplets, as the nine-banded armadillo does. Under this counterfactual antecedent, the range of possible kin-categorisation systems would have been different.

That was again a counterfactual conditional that just went by. The facts that determine which kin-categorisation systems qualify as possible are constraints in linguistics, by my lights. They have a stronger variety of necessity for linguistics than the facts about the causal factors that influence which possible kin-categorisation systems become actual. In other words, these constraints would still have held, had the dynamical laws of language change been different. These constraints are needed not merely to build a set that is stable for the purposes of linguistics and that includes the causal influences favouring simplicity and informativeness. In addition, these constraints taken all together, but without the causal influences favouring simplicity and informativeness, form a set that is stable for the purposes of linguistics. Linguists rightly hold these constraints fixed when linguists consider which kin-categorisation systems there would have been if there had been another principle of efficient communication besides simplicity and informativeness—and when linguists consider which kin-categorisation systems there would have been if relatives outside of the nuclear family had tended to live at considerable distances, reducing the benefit of distinguishing among aunts, uncles, and cousins. Counterfactual changes to the dynamical influences leave the constraints fixed; the constraints form a higher-level group of laws—a set that is stable for the purposes of linguistics. These constraints limit which human family trees qualify as possible, for a genuine variety of possibility.

Of course, I am not claiming that the explanations proposed by Kemp and Regier are correct or that the facts that I have just mentioned provide further explanations alongside those proposed by Kemp and Regier. It may be that there are no universal drivers of language development (such as simplicity and informativeness, according to Kemp and Regier) that are responsible for patterns in actual kin-categorisation systems. Perhaps the only way to explain why certain kin-categorisation systems are present in some actual human languages—and other systems are absent from all actual human languages—is to explain causally why English has this particular system, Sudanese has that system, Iroquois has this other system, and so forth, proceeding through every actual human language individually.

But this makes no difference to my point, which is that Kemp and Regier's proposed explanations have been taken seriously and that their proposals implicitly take certain facts as constraints on kin-categorisation systems (in that the necessity associated with those facts is stronger than the necessity possessed by the causal principles governing language development). These constraints carve out a range of *possible* human languages and thereby explain why all *actual* human languages are possible in that sense—for instance, why no actual human language goes beyond the optimality frontier in the diagram. Because of those constraints, it is impossible to go beyond that frontier. These are non-causal explanations in linguistics, since linguistics is not interested in why human beings reproduce as they do rather than like bacteria or armadillos. The facts about human kin relations are constraints on any possible human kin-categorisation system.

5. A Hierarchy of Constraints—Not Just in Physics

I have just given an example of one level of constraints within linguistics. But there may be even higher levels—that is, explanatory principles in linguistics that possess a variety of necessity even stronger than the variety possessed by the constraints that I have just mentioned. These stronger principles constrain not merely all possible *human* languages, but all possible *natural* languages.

Consider a merger. That is, consider a case where within a given language, two phonemes (each one appearing in various different words) become one, so that subsequent language-learners learn only the single merged sound and the different words now share the same phoneme. Consider the principle, which has long been widely accepted in historical linguistics, that such mergers are irreversible 'by linguistic means' (Hickey 2004; Labov 1994, 144, 311). That is, when apparent de-mergers take place, the apparent merger was only partial (a 'near merger'), since there remained in the population some individuals for whom the two sounds did not merge completely, or the de-merger occurred because new non-merged speakers joined the population from outside of it (Thomas 2006; Trudgill et al. 2003). This principle of merger-irreversibility does not depend on the petty details of the ways that human beings acquire, process, or use language. The standard, widely accepted explanation of this principle (Labov 1994) is that to reverse a merger, a population would have to learn which of the words in the merged class had which of the original sounds. But these facts are not evident in speech occurring after the merger. There is no possible cause for a population to treat the words with the merged sound that were formerly pronounced in one way differently from the words with the merged sound that were formerly pronounced in the other way.[9]

Here we have a limitation on the causes of language change. This limitation is imposed not by human anatomy or human psychology or even the fundamental laws of physics. Rather, the limitation is imposed by the fact that a natural language is learned from current speech and so a vanished feature of the language's past makes no difference to the way in which the language is now learned. Mergers would still have been irreversible even if the tendencies favouring strength and informativeness had been replaced by other tendencies—or, for that matter, even if human beings had reproduced like bacteria or armadillos. The irreversibility of mergers transcends those comparatively contingent causal details.

Suppose, then, that we want to explain why a given merger was not in fact reversed (or why no merger has ever been reversed) 'by linguistic means'. This is explained by the impossibility of merger reversal. This explanation does not work by giving an inventory of the various particular causes influencing the language after the merger and then by showing, one by one, that each of these causes lacked the power to unmerge the sounds. The merger's persistence did not depend on the particular causes that happened to be operating. Merger irreversibility is part of the framework in which all causes of natural-language change have to operate. Accordingly, the explanation here is not a causal explanation; it does not work by supplying information about the language's causal history.

Explaining why a given merger persisted is like explaining why a given isolated system conserved its energy. The fact being explained does not depend on the particular causes that happened to be operating within the system. The principle of energy conservation transcends those petty causal details. Likewise, the principle of merger irreversibility imposes limits on the kinds of causes there could have been.

It is no coincidence of the various actual kinds of physical forces that none of them changes the total energy. It is likewise no coincidence of the various actual kinds of causes of natural-language change that none of them has the power to produce unmergers. Rather, unmergers are impossible because a language has no memory. That is a restriction on all possible causal mechanisms for language change. They all proceed from the system as it is now, irrespective of how it was in the past. Two hypothetical but realistic linguistic communities that are alike now, but arrived at their current state from different linguistic pasts, must be treated alike by the causal mechanisms of language change. This principle restricts the kinds of causes of language change that there could be, and so it is more strongly necessary than that language change tends to favour increased simplicity and informativeness. It is more strongly necessary than even the constraints imposed by the sorts of biological family relationships that human beings have. Compared with the irreversibility of mergers, the details of human anatomy are accidental.

That each of the various causes of language change fails to reverse mergers does not explain why mergers go unreversed. Rather, that mergers are irreversible explains why each of the various causes of language change fails to reverse mergers.

This is akin to what we saw earlier about the order of explanatory priority if energy conservation is a constraint. The fact that each of the various actual fundamental kinds of forces conserves energy does not explain why energy conservation holds, since that alleged explanation would incorrectly depict energy conservation as a coincidence. Rather, the order of explanatory priority runs in the opposite direction: if energy conservation is a constraint, then that constraint explains why each of the various actual kinds of forces conserves energy.

Linguists investigate whether other principles have the same transcendent status as the principle of merger irreversibility. One plausible candidate, considering how natural languages are transmitted from generation to generation, is that communicability must be preserved between generations. A language cannot change too rapidly compared to the rate at which it can be learned and used. There is some very strong variety of possibility for which it is impossible for English to have changed a century's worth per day—so that the contract you signed a week ago was drafted in the English of Chaucer (c.1342–1400), you lectured ten days ago in the language of the Anglo-Saxon monk Ælfric (c.955–c.1010), and the news was broadcast two weeks ago in the language of Beowulf (c.700–750). Such rapid change would make a language inoperative between generations and so it could not be transmitted from one generation to another.

I conclude, then, that various hypotheses positing constraints (and various proposed scientific explanations appealing to constraints) are sometimes deservedly given serious consideration not only in fundamental physics, but also in at least one of the human sciences as well.

Notes

1. In referring to 'causal interactions', 'causal history', and so forth, I do not intend to presuppose any particular account of what causal relations consist in.
2. For example, Wigner (1972, 13) says in his Nobel Prize presentation speech: '[F]or those [conservation laws] which derive from the geometrical principles of invariance it is clear that their validity transcends that of any special theory—gravitational, electromagnetic, etc.—which are only loosely connected … .' Likewise, Feynman (1967, 59) says, 'When learning about the laws of physics you find that there are a large number of complicated and detailed laws, laws of gravitation, of electricity and magnetism, nuclear interactions, and so on, but across the variety of these detailed laws there sweep great general principles which all the laws seem to follow. Examples of these are the principles of conservation … .' Similarly, Bergmann (1962, 144) remarks that the conservation laws are 'general laws applying uniformly to every assembly of mass points regardless of the particulars of the force laws.' I take all of these remarks to be characterising various conservation laws as constraints on the dynamical laws. Lange (2009; 2017) gives further examples.
3. The association between symmetries and conservation laws (within a Hamiltonian framework), codified in Noether's first theorem, does not suffice to establish the explanatory priority of symmetries over conservation laws because (as Noether showed) the association runs in both directions; Noether's theorem is 'symmetric' in this way. Accordingly, some philosophers have held that the symmetries do not help to explain the conservation laws. For example, Albert (2015, 14) says that 'what actually explains [the conservation of energy] are the fundamental physical laws of the actual world'. Likewise, Brown and Holland (2004, 10) maintain that 'the real physics is in the Euler-Lagrange equations of motion for the fields, from which the existence of dynamical symmetries and conservation principles, if any, jointly spring'. (Later Brown (2022) offered a pragmatic account of

explanation, using it to reject the explanatory priority of symmetry principles.) However, these views run contrary to a broad agreement among physicists that symmetries explain conservation laws. (See, among many others, Wigner 1964, 959; Weinberg 2004; Zee 1986.) Rather than run contrary to this broad agreement, Lange (2007; 2009) has aimed to derive the symmetry principles' explanatory priority over the conservation laws from the symmetry principles' status as meta-laws (whereas the conservation laws are first-order laws). Meta-lawhood is associated with a broader range of counterfactual invariance and hence a stronger variety of necessity than the associated first-order laws possess. Roughly speaking, the symmetries' explanatory priority depends on its being the case that the symmetries would still have held, even if the conservation laws (or the Hamiltonian framework) had failed to obtain.

4. 'STR is not a theory in the usual sense but is better regarded as a second-level theory, or a theory of theories that constrains first-level theories' (Earman 1989, 155).
5. For further discussion and references, see Lange (2017, 144–145).
6. The key role played by such counterlegals in Lange's account of laws and of 'explanations by constraint' has provoked many potential objections. (See, for example, Andersen 2018; Morrison et al. 2019; Saatsi 2018; Skow 2018; Woodward et al. 2011.) Questions that might be asked include how these counterfactuals can be ascertained empirically, whether these counterfactuals can account for explanatory asymmetries, whether natural laws are preserved in science under as wide of a range of counterfactual antecedents as Lange maintains they are, whether Lange should instead have construed 'explanations by constraint' as associated with entire systematic patterns of counterfactual dependence, how necessity is related to stability under nested counterfactuals, whether some 'explanations by constraint' are too superficial to constitute genuine explanations, and what is responsible for these counterfactuals' truth. Although Lange (in Lange 2018; Lange 2019; Woodward et al. 2011, 45–52) offers replies to these questions, I do not believe that their details are crucial to the prospects (under discussion here) of extending something like Lange's notion of 'explanation by constraint' outside of fundamental physics.
7. Ben-Menahem (2018) also looks at various principles in physics that I would characterise as having often been taken (by physicists and philosophers alike) to be constraints (such as symmetry principles, conservation laws, and determinism). As I am doing, she takes them (if they are indeed constraints) to be explanatory; she highlights how these constraints would explain why certain things never occur and also how these constraints would help to delimit 'physical possibility' (110). Unlike Lange, she does not characterise the principles that she calls 'constraints' in terms of counterfactuals or as modally stronger than ordinary natural laws (16); rather, she leaves open how they should be understood metaphysically. (She calls them 'causal constraints' because they constrain the possible causal processes (15); Lange takes them as explaining non-causally, but this seems like a merely terminological difference.) Whereas Lange's 'constraints' directly constrain first-order laws, Ben-Menahem's directly constrain events. Adlam (2022) also characterises certain putative laws as 'constraints' and 'as primarily concerned with describing what is possible or impossible' (5). But unlike Lange, she does not use counterfactuals to cash out their status. Furthermore, her concerns are very different from mine. She is concerned with fashioning a non-Humean account of lawhood that would apply if (as some current physical theories say) there are no fundamental laws (such as the various force laws) that 'act moment-by-moment on states' (22) to govern the time-evolution of individual physical systems interacting with one another (7, 12). Instead, the 'constraints' with which Adlam is concerned are 'genuinely global, atemporal laws' that 'govern holistically and all-at-once' (11), such as variational principles and consistency conditions. Since Adlam's 'constraints' are not constraints on the various specific dynamical laws that take 'the standard Newtonian time-evolution form' (2) in applying to particular time-asymmetric causal (12) interactions, Adlam's constraints are not higher-order laws; they are all of the laws. This is very different, then, from the sorts of constraints with which I am concerned (and that I am aiming to show may arise in explanations outside of fundamental physics). Physicists like Wigner and

Feynman (note 2) evidently construe symmetry principles and conservation laws as constraints in the sense I am describing, i.e., as constraining lower-level, dynamical laws. (Chen and Goldstein (2022) develop a primitivist view of laws much like Adlam's according to which laws constrain the physical possibilities atemporally.)

8. Despite its limitations, many studies have taken Kemp and Regier's work as a model for explaining other linguistic phenomena, including color terms and indefinite pronouns. See, for instance, Denić, Steinert-Threlkeld, and Szymanik (2022), Kirby et al. (2015) and the references cited therein.
9. For further discussion of apparent merger reversals and how 'by linguistic means' in the irreversibility principle ('Garde's Principle') needs to be interpreted, see Clark, Watson, and Maguire (2013) and the other papers that it introduces.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

Marc Lange http://orcid.org/0000-0002-0704-7376

References

Adlam, Emily. 2022. "Laws of Nature as Constraints." *Foundations of Physics* 52 (1): 28 (1–41).
Albert, David. 2015. *After Physics*. Cambridge: Harvard University Press.
Andersen, Holly. 2018. "A Pragmatist Challenge to Constraint Laws." *Metascience* 27 (1): 19–25. https://doi.org/10.1007/s11016-017-0251-5.
Ben-Menahem, Yemima. 2018. *Causation in Science*. Princeton, NJ: Princeton University Press.
Bergmann, Peter G. 1962. "The Special Theory of Relativity." In *Encyclopedia of Physics*, Vol. 4, edited by S. Flügge, 109–202. Berlin: Springer-Verlag.
Brown, Harvey. 2022. "Do Symmetries 'Explain' Conservation Laws? The Modern Converse Noether Theorem vs. Pragmatism." In *The Philosophy and Physics of Noether's Theorems: A Centenary Volume*, edited by J. Read and N. J. Teh, 144–168. Cambridge: Cambridge University Press.
Brown, Harvey, and Peter Holland. 2004. "Dynamical vs. Variational Symmetries: Understanding Noether's First Theorem." *Molecular Physics* 102 (11–12): 1133–1139. https://doi.org/10.1080/00268970410001728807.
Chen, Eddy Keming, and Sheldon Goldstein. 2022. "Governing Without a Fundamental Direction of Time: Minimal Primitivism About Laws of Nature." In *Rethinking the Concept of Law of Nature*, edited by Yemima Ben-Menahem, 21–64. Cham: Springer.
Clark, Lynn, Kevin Watson, and Warren Maguire. 2013. "Special Issue on Phonological Mergers in English." *English Language and Linguistics* 17 (2): 229–234. https://doi.org/10.1017/S1360674313000014.
Denić, Milica, Shane Steinert-Threlkeld, and Jakub Szymanik. 2022. "Indefinite Pronouns Optimize the Simplicity/Informativeness Trade-Off." *Cognitive Science* 46 (5): e13142. https://doi.org/10.1111/cogs.13142.
Earman, John. 1989. *World Enough and Space-Time*. Cambridge: MIT Press.
Feynman, Richard. 1967. *The Character of Physical Law*. Cambridge: MIT Press.
Hickey, Raymond. 2004. "Mergers, Near-Mergers and Phonological Interpretation." In *New Perspectives on English Historical Linguistics*, edited by Christian Kay, Carole Hough, and Irené Wotherspoon, 125–137. Amsterdam: John Benjamins.

Kemp, Charles, and Terry Regier. 2012. "Kinship Categories Across Languages Reflect General Communicative Principles." *Science* 336 (6084): 1049–1054. https://doi.org/10.1126/science.1218811.

Kirby, Simon, Monica Tamariz, Hannah Cornish, and Kenny Smith. 2015. "Compression and Communication in the Cultural Evolution of Linguistic Structure." *Cognition* 141:87–102. https://doi.org/10.1016/j.cognition.2015.03.016.

Labov, William. 1994. *Principles of Linguistic Change, Vol. 1: Internal Factors*. Oxford: Blackwell.

Lange, Marc. 2002. "Who's Afraid of *Ceteris-Paribus* Laws? Or: How I Learned to Stop Worrying and Love Them." *Erkenntnis* 57 (3): 407–423. https://doi.org/10.1023/A:1021546731582.

Lange, Marc. 2007. "Laws and Meta-Laws of Nature: Conservation Laws and Symmetries." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 38 (3): 457–481. https://doi.org/10.1016/j.shpsb.2006.08.003.

Lange, Marc. 2009. *Laws & Lawmakers*. New York: Oxford University Press.

Lange, Marc. 2017. *Because Without Cause*. New York: Oxford University Press.

Lange, Marc. 2018. "Reply to My Critics: On Explanations by Constraint." *Metascience* 27 (1): 27–36. https://doi.org/10.1007/s11016-017-0252-4.

Lange, Marc. 2019. "Replies to Heron and Knox, Morrison, and Strevens." *Philosophy and Phenomenological Research* 99 (3): 739–748. https://doi.org/10.1111/phpr.12648.

Lévy-Leblond, Jean-Marc. 1976. "One More Derivation of the Lorentz Transformations." *American Journal of Physics* 44 (3): 271–277. https://doi.org/10.1119/1.10490.

Lewis, David. 1983. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61 (4): 343–377. https://doi.org/10.1080/00048408312341131.

Lewis, David. 1986. *Philosophical Papers: Volume II*. New York: Oxford University Press.

Lewis, David. 1999. "Humean Supervenience Debugged." In *Papers in Metaphysics and Epistemology*, edited by David Lewis, 224–247. Cambridge: Cambridge University Press.

Mollica, Francis, and Steven Piantadosi. 2022. "Logical Word Learning: The Case of Kinship." *Psychonomic Bulletin & Review* 29 (3): 766–799. https://doi.org/10.3758/s13423-021-02017-5.

Morrison, Margaret, Michael Strevens, John Heron, and Eleanor Knox. 2019. "Book Symposium on *Because Without Cause*." *Philosophy and Phenomenological Research* 99 (3): 720–738. https://doi.org/10.1111/phpr.12649.

Passmore, Sam, Wolfgang Barth, Kyla Quinn, Simon Greenhill, Nicholas Evans, and Fiona Jordan. 2023. "Kin Against Kin: Internal Co-Selection and the Coherence of Kinship Typologies." *Biological Theory* 16:176–193. https://doi.org/10.1007/s13752-021-00379-6.

Passmore, Sam, and Fiona Jordan. 2020. "No Universals in the Cultural Evolution of Kinship Terminology." *Evolutionary Human Sciences* 2:e42 (1–14).

Saatsi, Juha. 2018. "A Pluralist Account of Non-Causal Explanations in Science and Mathematics." *Metascience* 27 (1): 3–9. https://doi.org/10.1007/s11016-017-0249-z.

Skow, Bradford. 2018. "Of Strawberries and Energy Conservation." *Metascience* 27 (1): 11–18. https://doi.org/10.1007/s11016-017-0250-6.

Thomas, Erik. 2006. "Vowel Shifts and Mergers." In *Encyclopedia of Language and Linguistics*, 2nd ed., edited by Keith Brown, 484–494. Oxford: Elsevier.

Trudgill, Peter, Daniel Schrier, Daniel Long, and Jeffrey Williams. 2003. "On the Reversibility of Mergers: /W/, /V/ and Evidence from Lesser-Known Englishes." *Folia Linguistica Historica* 24:23–46.

Tu, Loring W. 2013. "Remembering Raoul Bott (1923–2005)." *Notices of the American Mathematical Society* 60 (4): 398–416. https://doi.org/10.1090/noti977.

Weinberg, Steven. 2004. "The Making of the Standard Model." *European Physical Journal C: Particles and Fields* 34 (1): 5–13. https://doi.org/10.1140/epjc/s2004-01761-1.

Wigner, Eugene. 1964. "Symmetry and Conservation Laws." *Proceedings of the National Academy of Sciences* 51 (5): 956–965. https://doi.org/10.1073/pnas.51.5.956.

Wigner, Eugene. 1972. "Events, Laws of Nature, and Invariance Principles." In *Nobel Lectures: Physics, 1963–1970*, 6–19. Amsterdam: Elsevier.

Woodward, Jim, Barry Loewer, John Carroll, and Marc Lange. 2011. "Counterfactuals All the Way Down?" *Metascience* 20 (1): 27–52. https://doi.org/10.1007/s11016-010-9437-9.

Zee, A. 1986. *Fearful Symmetry*. Princeton, NJ: Princeton University Press.