

Ratio via machina: Three standards of mechanistic explanation in sociology

Forthcoming in *Sociological Methods and Research*

Natalie B. Aviles, *University of California-San Diego* <naviles@ucsd.edu>

Isaac Ariail Reed, *University of Colorado at Boulder* <isaac.reed@colorado.edu>

ABSTRACT: Recently, sociologists have expended much effort in attempts to define social mechanisms. We intervene in these debates by proposing that sociologists in fact have a choice to make between three standards of what constitutes a good mechanistic explanation: substantial, formal, and metaphorical mechanistic explanation. All three standards are active in the field, and we suggest that a more complete theory of mechanistic explanation in sociology must parse these three approaches to draw out the implicit evaluative criteria appropriate to each. Doing so will reveal quite different preferences for explanatory scope and nuance hidden under the ubiquitous term ‘social mechanism.’ Finally, moving beyond extensive debates about realism and anti-realism, we argue prescriptively against ‘mechanistic fundamentalism’ for sociology, and advocate for a more pluralistic understanding of social causality.

INTRODUCTION

Mechanistic explanation has become ubiquitous in sociology, but its meaning remains obscure. On the one hand, a great deal of definitional work has emerged to formalize the concept of social mechanism. On the other, mechanism-talk has proliferated so quickly, and with such success, that references to mechanisms retain a certain ambiguity. In this paper we attempt to clarify the debate about mechanisms by proposing a three-fold typology of mechanistic explanation, backed by clear and consistent reasoning from the philosophy of science and social theory. We develop this typology alongside analysis of exemplary texts, bringing forward the evaluative standards implicit in each example. In articulating three distinct standards of what constitutes a mechanistic explanation we draw out a common intellectual impulse in the move toward mechanism-talk, while simultaneously articulating deep differences in what mechanisms are taken to reference and how sociologists are expected to use them in their explanations. This better reveals what is at stake in the many debates over mechanisms. Finally, in our own

engagement with these debates, we argue not for or against realism, nor for or against mechanisms, but rather against a ‘mechanistic fundamentalism’ that would flatten the social world into a limited number of efficient causal processes. In contrast to mechanistic fundamentalism, we articulate the beginnings of a pluralistic understanding of causation. This, we argue, both fits better much of what sociologists already find excellent about a variety of explanatory genres and, prescriptively, articulates a more productive way forward for the development of social theory.

PART I: THREE STANDARDS FOR MECHANISTIC EXPLANATION

It is widely understood among sociologists engaged in mechanism debates that a single definition of ‘mechanism’ does not exist in sociology (cf. Abbott 2007; Gross 2009; Gorski 2009; Hedström 2005; Hedström and Swedberg 1998). Nonetheless, with the exception of Abbott, these scholars attempt to address the lack of consensus by proposing their own single definition or theory of mechanisms. Breaking with these debates, we instead presume sociologists create accurate and actionable knowledge about the world through diverse practices that converge upon a concept that orients them to explanation in (minimally) similar ways. This approach leaves the door open to multiple productive interpretations of ‘mechanism’ that cohere in different approaches to analysis and explanation.

More precisely, we believe that that there are in fact three separate standards for mechanistic explanation that have emerged over the years—substantial, formal, and metaphorical—each of which directs a different kind of causal explanation based on different assumptions about the goals of explanatory practice. According to the substantial standard, a

mechanism is an explanation that anchors descriptions of causal processes in the operation of real entities whose essential properties and tendencies can be established via empirical research. For the formalist standard, a mechanism explains the relationship between two variables by stipulating an intervening causal pathway of sufficient depth relative to a common analytical focus in the discipline. Finally, in the metaphoric standard a mechanism identifies, in the interpretation of a case or cases, the stability and regularity of certain social phenomena in relation to the contingency, malleability, or irreducible complexity of other social phenomena that are part of the interpretation being made. In this first part of the paper, our purpose is to make clear what the implicit standards of mechanistic explanation are in each of the three types by analyzing explanations that adequately enunciate each standard, prioritizing work that has been recognized as exemplary within sociology. Juxtaposing these standard-bearers allows for discussion of what is at stake in accepting one standard over the others.

The Substantial Standard

In the substantial¹ standard, mechanisms are causal processes made up of real entities with stable tendencies, and mechanistic explanation accounts for empirical outcomes as the product of interactions between these entities and other attributes of the social world. The substantial standard of mechanistic explanation thus reserves the use of the term ‘mechanism’ for those instances in which an explanation (1) traces the process that leads from input to output via a set of interlocked, precisely determinable entities, and (2) uses concepts that refer directly to the basic causal capacities (or ‘natures’) of those entities. This standard is developed in careful philosophical detail by Nancy Cartwright.² In Cartwright’s view, mechanistic explanation captures components of an underlying ‘nomological machine,’ a causal structure comprising real

entities with stable causal capacities. The language of a mechanistic explanation thus must *match precisely* the underlying structure of the world, and this matching relies for its veracity on the way in which the concepts that move the explanation forward in language refer directly to the essential *natures* or *capacities* of the entities in the world. These capacities are, furthermore, not incidental to, but rather constitutive of, those entities.

In Cartwright's articulation, a 'mechanism' is simply a separable component of a nomological machine, which explains empirical phenomena by tracing the workings of structures comprising stable and determinable causal capacities. The goal of mechanistic explanation should thus be to capture how underlying causal structures create different parts of the world we experience. Instead of debating the invariance of laws, or their existence as linguistic constructs, Cartwright returns to the idea that what is captured by a well-designed experiment or an efficacious model is an inner tendency inherent in the nature of certain entities (capacities). The observed outcomes that result from the action of capacities are not always assured, as they can be redirected by other intervening events, but the stability of capacities means they can be relied upon to tend toward certain limited expressions.

Though Cartwright developed her model from examples of physics experiments, she argues that something similar drives certain causal arguments in social science. Using economics as her inspiration, Cartwright finds parallels to the nomological machines of physics in the 'socio-economic machines' of social science models and experiments. Reasoning by analogy, Cartwright concludes that 'social capacities' of some kind *must* exist if regular social causes exist in this way. Thus, according to the substantial standard, a good mechanistic model that actually *explains* regularity in the social world must schematize real social capacities and their structural arrangements in a socio-economic machine (Cartwright 1999:53, 58-9). The key to

modeling mechanisms in the social sciences is to set up exactly the right circumstances so that the capacities we suspect of particular entities in the social world can express themselves clearly enough that we glean some sense of the underlying causal structure of social processes. Without a doubt, this is a daunting task.³

In sociology, critical realism has taken up this challenge most seriously. Though there is variation and inconsistency in how different critical realists understand the models of sociology to map onto those of the physical or natural sciences, critical realists share an assumption that sociological explanation ultimately must elaborate the intransitive ontological components (real structures, human natures) that furnish the social world. In the vernacular of critical realism, the ‘real’ refers to intransitive entities with causal capacities that exist irrespective of human understanding of them; the ‘actual’ refers to the operation (whether observable or not) of these entities in the world; and the ‘empirical’ refers to observations or experiences of the world (cf. Smith 2010:93). Causation inheres in the capacities of real entities that combine in causal mechanisms. Empirical observation is the imperfect vehicle of sociological inquiry that attempts to access real causes and mechanisms by hypothesizing actual processes based on observed outcomes. Critical realism thus stipulates that the mission of sociological inquiry should be to develop adequate causal explanations of social phenomena in terms of actual conjunctions of mechanisms comprising real entities with stable causal capacities and tendencies, a position consistent with the approach Cartwright describes.

Christian Smith exemplifies the critical realist approach to social mechanisms in *What is a Person?* (2010), wherein he grounds explanations of social causes in ‘a theoretical model of the ontology of the nature of human being’ (10). For Smith, the social world is populated by real entities, such as human beings, with certain causal capacities that interact to create higher-level

entities, such as social structures, that exhibit emergent causal capacities and tendencies of their own. Smith uses a critical realist formulation of personhood to argue for the ontology of human dignity. For Smith, ‘human dignity is an independent objective reality’ (2010:470); it is an emergent product of the causal capacities of personhood. Dignity exists as a real entity with specific causal capacities whether or not actors in a given historical context recognize it as such. Dignity is differently actualized in different contexts according to other triggering causal processes, but it is always present to some degree, and human actors everywhere experience empirical manifestations of dignity that enable sociologists to study its effects. Human dignity comprises one social capacity ingredient in mechanisms of solidarity that in turn actualize into specific (e.g. national) moral orders when triggered by certain cultural and structural forces. Importantly, Smith’s work also captures what is characteristically critical about critical realism, i.e. its attempt to preserve the possibility of normativity in scientific sociology. For Smith, critical realist sociology on topics like human dignity support moral action directed toward preserving the integrity of all persons.

It is worth pointing out that Smith’s explanation is consistent with a substantial understanding of causal mechanisms as composed of real capacities with determinable tendencies regardless of whether the reader finds his specific interpretation of human nature compelling. Nevertheless, while critical realist explanations such as Smith’s conform to Cartwright’s injunction that mechanistic explanation anchor causal processes in stable capacities with determinable tendencies, the assumption that adequate mechanistic explanations of empirically complex social phenomena must emanate from a small number of intransitive social capacities (that, for instance, inhere in human nature) may ultimately lack sufficient warrant. The tension is belied in frequent attempts among critical realists to reconcile ontological models of

human natures and structural capacities with the tenets of social constructionism and historicism indicated by the axiom that knowledge about the social world is ‘concept-dependent’ (cf. Gorski 2009).

These and other difficulties have troubled attempts to develop a deep, social realist ‘foundation’ to sociological science (cf. Alexander 1982-1983; Seidman 1991), and have led many sociologists to be somewhat skeptical of the utility of (various versions of) the substantial standard. In particular, the notion that what a strong causal explanation in sociology does is reveal the ‘real nature’ of an entity is rather distant from many causal claims in sociology, which tend to bound significantly the application of their hypotheses, and adopt a much more nominalist approach at a philosophical level. Given our reliance on her arguments above, it is worth noting that Cartwright expresses similar skepticism at certain points in her texts.⁴ What, then, is to be done?

The Formal Standard

The formal standard dispenses with the ontological goals of explanation that characterize the substantial standard. Instead, it develops a more pragmatic, analytical approach to modeling mechanistic causes. Rather than attempting to grasp the nature of entities and their interaction, it uses mechanistic language to discuss a recurrent problem in sociological research and theory: the way that some explanations may be judged to be insufficiently ‘deep’ because they fail to specify underlying causal pathways connecting inputs to outcomes. The second standard of mechanistic explanation thus engages and reconstitutes a longstanding sociological interest in explanation via

intervening variables. In so doing, it constructs models of causal processes or pathways in varying degrees of formality.

Critiquing approaches that attempt to explain outcomes by measuring the strength of association between variables, advocates of the formal standard like Morgan and Winship (2015) maintain that adequate explanation requires the additional depth provided by modeling a mechanism that hypothesizes how and why a given variable (D) affects another variable (Y). Thus the question ‘what is the mechanism?’ should be understood as an inquiry about intervening variables. If indeed investigators want to evaluate a theoretical account of a variable relationship, then ‘a more narrowly focused analysis of the putative causal pathways that relate D to Y must be undertaken’ (2015:330). Using Judea Pearl’s approach to causality, in which intervening variables are represented as constituting pathways from one variable to another in directed acyclic graphs, Morgan and Winship characterize the above example in the following way.

[FIGURE 1 ABOUT HERE]

If we think of mechanisms this way, they become particularly useful for addressing a frequent problem in social science, namely that the relationship between variable D and variable Y is affected by unobserved variables that cause both. If U is a cause of both D and Y , then the causal effect of D on Y can be, nonetheless, effectively estimated *via the intervening variable M* , if two things are true: M is *isolated*—that is, no observed or unobserved variables cause both D and M , and *exhaustive*— M intercepts all causal pathways from D to Y . It is, of course, possible to have a *set* of mechanisms (say, M , N , and another mechanism O) that satisfy these criteria, if the relationship between those mechanisms can itself be adequately rendered.

This standard of mechanistic explanation breaks sharply from the substantial standard. Because it has developed in contrast to some of the literature on ‘generative’ and ‘substantive’ mechanisms, and because it explicitly and assiduously avoids the issue of the ‘inner nature’ of social entities, we label this approach the formal standard. In this view, mechanisms are intervening causal pathways hypothesized relative to a well-defined research question about how to get from independent to dependent variables. In the formal approach, explanation via mechanisms frequently involves a process of ‘bottoming out’ which causal pathways matter relative to the community of inquiry’s interests. Machamer, Darden, and Craver explain:

Nested hierarchical description of mechanisms typically *bottom out* in lowest level mechanisms. These are the components that are accepted as relatively fundamental or taken to be unproblematic for the purposes of a given scientist, research group, or field. Bottoming out is relative: different types of entities and activities are where a given field stops when constructing mechanisms. The explanation comes to an end, and description of lower-level mechanisms would be irrelevant to their interests. (2000:13; cited in Morgan and Winship 2015:345)

Bottoming out allows sociologists to isolate and schematize analytically important causal processes in a way that does not require philosophical finality about the underlying natures or essential causal tendencies of entities.

The relativity of bottoming out also reveals the way in which the formal standard is more pragmatist, rather than realist, in its philosophical orientation. The point is to connect explanation to the concerns of a given community of inquiry and the well-defined research questions it has produced. In those areas of sociology that have tended to think of mechanisms as intervening variables, one could argue that this kind of bottoming out relative to the concerns of the research

community has become part of the formal standard for mechanistic explanation.⁵ However, because the search for intervening variables and the search for a lower level are not identical pursuits, departures from this version of the formal standard are possible. These can go in one of two directions. First, one might abandon the search for a lower level, and simply pursue intervening variables at any level of analysis. If there are good practical reasons to do this (in terms of policy interventions in the world, for example), this may indeed accord well with the spirit of the formal standard, and in particular with its pragmatist orientation. However, second, one might *demand* that bottoming out reach to a given level, based on a specific rendering of the social world writ large. That is, one might propose that explanations *always* bottom out at the level of individual human beings understood as having natures that dispose them to act in a certain way. This second departure is, in our view, a departure that returns to the substantive standard, and we discuss the implications of this approach at several points in Parts II and III of this paper.

For Morgan and Winship, whose approach is pragmatic, the formal understanding of mechanisms is tied into a broader program wherein social causality is accessed via the potential outcomes framework. However, the formal standard appears in many other approaches, even if those approaches are less strict in their avoidance of ‘ontological’ debates about social reality, and less clear about the pragmatist basis for their philosophy of science. For example, it is also exemplified in Peter Hedström’s early programmatic work advocating for an analytical approach to social mechanisms.⁶ According to Hedström, mechanisms are abstract models, analytical constructs to be tested against empirical phenomena or simulations, and emphatically not to be evaluated based on realist interpretations of the necessarily simplified and incomplete assumptions they make about the social world (Hedström and Swedberg 1998: 14-15). For

Hedström, the mechanisms analytical sociologists identify should resemble the kind of simplified model represented by ‘Coleman’s boat,’ wherein macro-level social phenomena are explained in terms of intervening individual actions oriented toward situational conditions in ways that generate transformations in aggregate patterns, as in Figure 2 (Hedström and Swedberg 1998; Hedström and Ylikoski 2010).

[FIGURE 2 ABOUT HERE]

Hedström (1998) illustrates this approach by constructing a model of rational imitation, a theory that actors imitate the actions of other actors based on their belief that others’ actions indicate the most effective strategies for achieving a desired outcome. Hedström proposes rational imitation as an individual-level phenomenon that produces quasi-general aggregate mechanisms, such as cognitive dissonance reduction or self-fulfilling prophecies, by intervening between macro-level variables. Hedström, drawing upon Merton’s example of the self-fulfilling prophecy as a mechanistic explanation of bank runs, creates a simulation to model the aggregate dynamics of individual acts of rational imitation. Far from assuming rational imitation derives from an innate capacity of real human actors, Hedström’s model tests different heuristics for action that range from completely imitative to completely atomistic, and function to test the parameters of the hypothesized mechanism rather than to mirror real phenomena. Hedström’s model of the self-fulfilling prophecy explains how certain macro-level associations (bank runs) occur when certain situational conditions (rumors of insolvency and occurrences where people have withdrawn their funds from the bank) trigger intervening action-formation mechanisms (actors’ beliefs that the actions of others indicate the truth of insolvency rumors and that they should therefore act similarly) which then manifest transformational mechanisms (cascading numbers of imitators withdrawing their funds, rendering the bank insolvent after all).

Though proposed as a candidate explanation for empirical phenomena, Hedström's simulation does not purport to directly represent 'real-life situations' (1998:325). It is a simplified abstraction of a more generic process whose empirical manifestation is subject to all manner of complex contingencies unspecifiable in the model. According to the formal standard, a 'theoretical model is in principle constructed in such a way that it includes only those elements believed to be essential for the problem at hand. The target of the theoretical analysis, then, is this model and not the reality that the model is intended to explain' (Hedström and Swedberg 1998:14). An adequate formal explanation of mechanistic causation ultimately explains social phenomena as the outcome of unobservable intervening mechanisms abstracted from empirical examples through careful causal modeling.

The Metaphorical Standard

It is worth remarking that, despite prolific attempts to define what mechanisms are in social science inquiry, many sociologists who couch their explanations in mechanisms advance neither strong ontological stances on the nature of social reality nor construct formal analytical causal models to support their claims (cf. Vaidyanathan et al. 2015). Nevertheless, the idea of mechanism appears to do some systematic conceptual work for the sociologists who deploy mechanistic explanation outside of the substantial or formal standards. In some areas of sociology, the rhetoric of 'mechanism' matters, not because it *denotes* an ontological entity or an analytical model, but because it *connotes* a particular imagery that is useful for organizing observations. For many sociological explanations, we propose, the core imagery evoked by the term 'mechanism' is a key element of an explanation—namely, that something in the social world works regularly and efficiently, *as if it were a machine*. What this entails, then, is usually a

claim about some parts of the world working more like a machine than other parts. This is the metaphorical standard.

In the metaphorical view, ‘mechanism’ is simultaneously a sensitizing *and* explanatory concept, though it is explanatory in a different way than discussed in the sections on substantial and formal approaches to mechanisms (indeed it is possible that strong proponents of the substantial and formal standards would not recognize what follows here as ‘explanation’). The concept is sensitizing in that it primes the researcher to pick out attributes of a broader, intertwined, and messy social world that are *more machine-like* than other processes that surround them. To explain how the mechanism metaphor enforces a coherent standard for explanation in sociology, we must take a brief detour into the theory of metaphor.

The linguistic turn in philosophy, pivoting on Wittgenstein, opened up a purportedly denotative scientific discourse to literary scrutiny, including work on metaphor.⁷ As we understand it, metaphorical analysis inspired by the linguistic turn is not just about the format of scientific argument, but, in the standard for sociological explanation we propose here, speaks to a heuristic process of generating sociological explanations (cf. Gerhart and Russell 1984). In other words, we imagine metaphor as basic to the process whereby the sociologist sorts, colligates, and construes evidence in a case. Literary theorist and philosopher Paul Ricoeur gives an account of this kind of sorting.

Metaphor, Ricoeur (2003) argues, generates truth statements that exist somewhere between poetic innovation and empirical observation. This is because metaphors are by their nature a ‘split reference’ to invention and reality, with one foot in the realm of creative fiction and the other in the terra firma of the world. To fully grasp their meaning, readers must interpret metaphors as having a split copula; that is, they must read the verb ‘to be’ that holds together a

metaphoric statement as ‘it is *and* it is not.’ Such a reading highlights the ‘semantic impertinence’ at the heart of all metaphorical statements, i.e. the fact that they are absurd if taken literally (178). Semantic impertinence justifies Ricoeur’s use of the tension theory of metaphor, the idea that metaphors cohere not only because they are similar to their referents, but because they are at the same time implicitly different from them. The tension is reconciled when a metaphor transcends the initial shock of its literal absurdity to indicate something else that is similar, a process that enables users of the metaphor to elaborate reality in new ways.

In sociology, we propose, metaphorical tension is useful in explaining the ways in which social life, when thickly described, is both like and not like a machine. This insight guides much historical sociological research into how and when social life becomes more regular and regulated, the way in which such machine-like regularity works, and the effect of this regularity on the human subjects who engage, support, participate in, and resist it. Thus, thinking metaphorically about mechanisms enables the development of more innovative and intellectually powerful interpretations and case-based explanations.

The metaphorical use of mechanism is central to the causal explanation proposed by historical sociologist Richard Biernacki in *The Fabrication of Labor* (1995). Biernacki’s explanation offers an extensive and detailed account of how labor was conceptualized and enacted in two different settings for early industrialization: Britain and Germany. He argues that different cultural categories germane to each country gave labor and its remuneration different manifestations in each case—in Britain, as the amount of labor crystallized in the end product, and in Germany, as the disposition over the laborer’s labor power over a given period of time. Note, however, that this classically ‘humanist’ or ‘interpretive’ explanation gained torque precisely because these interpretations of the effect of cultural categories were made with

explicit reference to a mechanistic part of the social world that contrasted with them: the *piece-rate mechanism*. Biernacki posits that ‘the economic environment limited the mode of payment’ (45) so that piece-rate remuneration for weavers was the norm in *both* Britain and Germany. The piece-rate mechanism was a stable, regular, and general phenomenon undergirding wool textile manufacture in both Britain and Germany that emerged from the practical and utilitarian economic actions of mill owners using technologies of comparable structure and efficiency. In other words, metaphorically speaking, it operated like a machine to produce consistent actions across disparate contexts.

Note that Biernacki’s explanation does not posit the piece-rate mechanism as an entity with capacities, nor does it rest upon a formal causal model that stipulates piece-rate payment as an intervening causal pathway. Rather, Biernacki uses more metaphorical language to defend his argument: the piece-rate mechanism was a ‘Rorschach test for industrial culture’ (44) in that it enabled different cultural categories to constitute unique pay scales around it. This to us indicates that the role of mechanism in Biernacki’s explanation is neither substantial nor formal, but rather functions within the conceptual apparatus of the explanation as a point of focus, such that the piece-rate mechanism is *counterposed* to its ‘precipitation’ into how labor is measured on the shop floor in each case, via the cultural categories used to *interpret* and thus fabricate labor.

In this instance of metaphorical usage, the mechanism is what is ‘held constant’ across the two cases and then ‘interpreted’ on the ground; this would indeed appear to be one way in which this sort of explanation tends to work—it can also be found, for example, in Salzinger’s (2003) interpretation of factory work, in which the more mechanistic process of feminine labor as a vector of globalization manifests according to different interpretations on each factory floor. But the use of the mechanism metaphor does not have to point to a ‘constant’ in this way. For

example, a paired comparison might detail how a certain case congealed a mechanistic process while another, contrasting case did not. Wilson (2011) makes this argument when he compares reflection and refraction models of tax administration in relation to domestic and colonial systems of governance. Alternatively, the advent of mechanistic social life might be a matter of degree. Gorski (2003) makes this sort of argument in comparing the relative strength or intensity of social discipline in the Netherlands and Prussia.

What distinguishes metaphorical mechanisms from the other two types is that the criteria used to evaluate the presence and/or strength of a mechanism emerge *in relation to the attributes of specific contexts or cases* and not in relation to conjunctions of intransitive capacities or models of intervening causal pathways. The metaphorical standard, then, trades parsimony and transposable theory for nuance and explanation-via-case-description. This difference prompts the next part of this paper, which looks more generally at differences between the standards and their implications for debates about sociological explanation.

PART II: THE MECHANISMS DEBATES, RECONSIDERED

We now propose to use these three standards to clarify what is at stake in (some of) the mechanisms debates in sociology. Consider two propositions.

Proposition 1: Debates about ‘realism’ are really debates between the substantial and formal standards

One extended strand of debate in sociology concerns the degree to which a ‘realist’ program that adheres to the substantial standard of mechanistic explanation is applicable to sociology, *even as a regulatory ideal that is never fully reached in practice*. Can we say that, in

sociology, what we must access are the social capacities whose causal tendencies structure the world? There are a variety of arguments for such a position. For example:

--John Goldthorpe (2007) argues that Rational Action Theory may not describe individual actors in their full, elaborated reality, but does capture a core tendency of decision-making action that is causally significant, and thus can be the basis of explanation.

-- Hedström (2005), departing from his earlier formalism, argues that sociologists should construct analytical models to capture the ‘desires, beliefs, and opportunities’ (DBOs) that motivate social actors as real entities prone to certain activities, and that the tendencies created by DBOs, in interaction, explain aggregate social outcomes.

--Daniel Little (2012) argues that social realities at the meso- and macro-level emerge from interactions between purposeful entities (persons) at the micro-level, and these emergent entities have ‘natures’ or ‘capacities’ of their own.

All of these arguments share a commitment to a substantial standard, and criticism tends, for all the discussion of realism and anti-realism, to focus in on the issue of whether ‘ontology’ or ‘entities with properties’ should really drive the theorization of sociological mechanisms.

Morgan and Winship (2015) argue that the attempt by Goldthorpe and other proponents of rational choice theory to ground statistical research in an ontology of the actor and, as a result, a certain basic set of generative mechanisms, can lead to perverse consequences for the field. In particular, they point out that if scholars become advocates for a certain key mechanism based on a favored but arbitrary ontology of social life, and if they support the existence of such a mechanism through various highly indirect entailments, then the result may be ‘mechanism

warlordism’ wherein ‘the mechanisms of the most industrious scholars—those who can dream up the largest number of hypotheses to affirm, who can recruit the largest number of students to do the same, and who can attract the largest amount of funding to collect the data on their hypotheses...could receive the most affirmation’ (Morgan and Winship 2015: 343). The threat of mechanism warlordism supports their articulation of the formal standard.

It is revealing and, we think, positive evidence for the utility of the typology proposed here, that Isaac Reed (2011), arguing from a very different tradition of thought from Morgan and Winship, arrives at a similar criticism of what Reed calls ‘realism’ and we call the substantial standard. His argument against realism amounts to an argument that, by starting with a certain ontology, investigators ‘short circuit’ the project of understanding social life as it is subjectively perceived and meaningfully interpreted at hand, instead beginning with an ontology arbitrarily set by the theoretical school within which the investigator works. This argument is mobilized for a very different agenda—Reed is interested in historically bounded, thickly described case studies, not intervening variables—but the issue with the substantial standard is the same.

Both arguments, in other words, dispute the pursuit of entities with inherent ‘natures’ as appropriate for sociological explanation. Another way of saying the same thing is this: there may be ‘ontological’ consistencies to the entities that make up social life as we study it in sociology (e.g. the nature of persons, certain properties of social facts like money, and so on), but it is extremely unlikely that, by attaining precision about this ontology, we will be able to sufficiently explain the observed variations that primarily occupy sociologists.

To summarize, from the perspective of the formal standard, attempts to elaborate sociological epistemology around generative mechanisms, the micro-foundations of sociology, or critical realism are insufficiently pragmatist in their understanding of social science. In

aspiring to the substantial standard, such approaches attempt to replace the contingent process whereby a community decides whether an explanation is ‘insufficiently deep’ and thus requires more work on mechanisms, with an ontological guarantor that requires all explanations to function in a single language. In response, those in pursuit of the substantial standard may argue that the pragmatist approach is undertheorized, which is to say that it has not made clear the presuppositions about actors and action upon which it rests.

This sort of back-and-forth between substantial and formal standards is, we think, what is actually going on in much of the mechanism-talk that has become part and parcel of working sociology (and is near ubiquitous in conference talk where research is evaluated and criticized). Much of the confusion about the term mechanism, we hazard, results from the fact that these two standards of explanation can *sometimes* gel together nicely, but oftentimes do not. If, for example, by positing an isolated and exhaustive intervening variable we are also articulating in a clear way a testable picture of certain entities that exist and drive process, then these two meanings come together nicely. But in much sociological research, this keen match is unavailable.

Proposal 2: An essential divide in the use of the term ‘mechanism’ in sociological explanation is the value placed upon parsimony

The metaphorical standard reveals, however, how much is held in common in the debates between the substantial and formal standard. Broadly, the substantial and the formal standards value a relatively straightforward understanding of parsimony as the key to explanatory strength, and as such they reject the more local and thick-descriptive approach prominent in many explanations that follow the metaphorical standard. Put bluntly, those following the metaphorical standard elevate nuance as a guiding epistemic norm in place of parsimony. Thus, although the

metaphorical standard and the formal standard share skepticism about ontological claims about entities and natures, they break over their understanding of the goals of explanation and the purpose of theory in social science.

In particular, as an explanatory practice, the metaphorical standard suggests a focus on the phenomenology of the interpretation of evidence. Returning to Ricoeur, metaphors are a ‘seeing-as’ in the Wittgensteinian sense, the intuition of a coherent experience from ambiguous context (2003:252). This leads to a productive view with regards to explanation that echoes the ‘thick description’ advocated by Geertz: it ‘unleashes the power that certain fictions have to redescribe reality’ in creative ways (Ricoeur 2003:5). The implication is that neither an underlying ontological structure *nor* an isolated and exhaustive intervening process needs to be fully stipulated to obtain a good mechanistic explanation. Rather, the analyst evokes the imagery of mechanism to draw a key contrast within the case. The metaphoric standard thus encourages a hermeneutic attitude, where explanations are evaluated based on how their application of the term ‘mechanism’ and its connotations illuminates regularities without reifying them or reducing all phenomena of interest to the mechanism.

In the metaphorical standard, the interpretation of empirical observations through the metaphor of mechanism ‘works’ to produce good knowledge claims because that which is identified as a mechanism is grasped in its regularity and consistency *in comparison and relationship to* aspects of social life that are not like machines with regular interlocking parts—local cultures, historical contingencies, human idiosyncrasies. It is here that the semantic impertinence of metaphor becomes indispensable to good metaphorical explanation: ‘interpretation of metaphor is not possible unless one first perceives the incompatibility of the non-figurative meaning of the lexeme with the rest of the context’ (Ricoeur 2003:215). The

tension between the regularity and stability implied by the metaphor of mechanism and the experience of a messy and irregular social world must be evident in metaphorical explanations, both to enable sound evaluation and to defend against hypostatization.

In a paper delivered at New York University on ‘Analytical Sociology and Pragmatism,’ Neil Gross identified this issue of parsimony as central to the debates swirling around the Analytical Sociology (AS) movement. In particular, he argued for certain details and historically-bounded accounts as indeed identifying ‘substantive’ mechanisms, in opposition to the extreme levels of abstraction he find in AS. Gross saw clearly a debate about parsimony that in fact extends well beyond Analytical Sociology. With reference to certain historical details central to Robert Scott’s account of pilgrimage cures in the Middle Ages—e.g. that in a Christian world where illness was taken to be the result of sin, going on a pilgrimage took a person out of an emotionally suffocating environment—Gross argued that Analytical Sociologists had mistaken historical details essential to a causal story for “minute detail” that had to be discarded; in contrast he argued for a more historically grounded and nuanced comprehension of mechanisms. ‘This is the essence of the epistemic problem for mechanisms in sociology—how abstract is too abstract? How much should be sacrificed at the altar of parsimony?’⁸

PART III: AGAINST MECHANISTIC FUNDAMENTALISM, OR, THE DAPPLED WORLD OF SOCIOLOGY

Let us suppose that the reader has accepted a minimal version of our argument thus far: that there are three ways to think of mechanistic explanation that describe the world of sociological practice that includes the term ‘mechanism.’ (We say this is a minimal version of our argument

because, for example, one reader could believe it and also believe that *no* substantial ontologies of social mechanisms live up to the substantial standard, while another could believe it and hold that Hedström's and Goldthorpe's sociologies *do* meet this standard.) Stated as such, it remains an argument for more *clarity* in what standards sociologists are applying to themselves when they conduct explanations. It remains, as it were, an attempt at clear talk about good practice. What follows is, by contrast, more normative talk about sociological practice, which is to say, an account of how sociologists *should* think about mechanisms. In this regard, we think that the problem with mechanisms in sociology is not realism or anti-realism, but rather a tendency toward *mechanistic fundamentalism*. In this section, we clarify what we mean by this term, and propose an alternative—the dappled world approach to social causality—we think better captures the variation in the social world sociologists observe and explain. In doing so, we adapt some of the arguments of Cartwright, but apply them specifically to sociology and sociology's engagement with the implications, for explanation, of history and culture.⁹

Mechanistic fundamentalism is the argument that (1) the social world is made up of a finite set of mechanisms comprising elements of a stable 'social ontology' presumed to underwrite social phenomena, and (2) the pursuit of causal explanation in sociology amounts to the delineation of a very small set of such mechanisms that can account for an extremely wide variety of behaviors, outcomes, etc. (in other words, given agreement over this ontology, the epistemic value of parsimony then becomes paramount). Mechanistic fundamentalism is founded on the following two moves that are analogous to classic arguments in the philosophy of science:

Downward reduction is the expectation that the workings of the social world will 'bottom out' at the level of the individual person, who is taken as the fundamental building block of social explanation. As Abbott points out, extant discussions of mechanisms lead

quickly to attempts to delineate the ‘primitives’ of mechanisms—whether they are Goldthorpe’s (2007) rational actors, Hedström’s (2005) DBOs, or Neil Gross’s (2009) pragmatist A-P-H-Rs. It is worth mentioning that this is a direct analogy to the stacked world of biology-chemistry-physics, with the foundational particle physics at the bottom of the stack.

Crosswise reduction assumes that models constructed under highly controlled or abstracted conditions hold true across all empirical phenomena, including those occurring outside of the conditions stipulated in an explanation. In sociology, crosswise reduction is betrayed in the assumption that a small number of general processes must undergird all empirical phenomena, and that it is the task of explanation to reconstruct empirical phenomena to fit the abstract models upon which these observed phenomena are presumptively based.

The arguments for and against these two sorts of reductionism dot the literature, and we do not go into them here. Rather, we wish to point out that, understood in this way, the opposite of mechanistic fundamentalism is not some sort of ‘constructivism,’ ‘relativism,’ or ‘anti-realism,’ but, following Cartwright, a *pluralism* about the sorts of causal models needed to explain social phenomena. That is to say, the problem with mechanistic fundamentalism is not that it wishes to build a science of social reality that has as its goal explanation, nor is it that it expects scientific discoveries to be responsible to evidence. Rather, the problem is that it tends towards *irresponsibility* towards evidence in so far as it combines the substantial standard with a pursuit of parsimony as the primary epistemic value. To see how this is the case, imagine the following thought experiment.

Imagine that the social world exists, and there is in it a certain number of causal configurations, inner workings, etc. Social scientist F goes out into this world, and, insisting on parsimony above all else, stops developing models (verbal or mathematical) of what he sees after he has, based on his observations, built 20 such models. After that, each empirical phenomenon he observes, with the phenomena separated by relatively intuitive categories (barfights, marriages, arrests, etc.), is fitted into one of his 20 models. In contrast, social scientist E goes out into the world, and uses the same intuitive, informal criteria to decide when he has stopped observing one phenomenon and started to observe another, but for each and every new phenomenon builds a new theoretical model.

We expect that most sociologists—whether ‘realist’ or ‘constructivist,’ ‘naturalist’ or ‘humanist’—would agree that somewhere *between* these two extremes is a more desirable research program for knowledge production in social science. Social mechanisms are neither the fundamental particles of the social universe *nor* are they perfect empirical descriptions. And this means not only that we should avoid reducing the world to *certain specific mechanisms* (that is, avoid the victory of a given mechanism warlord), but that alongside the development of ‘mechanism’ as a causal imagery, we should develop *other explanatory genres based on different causal imagery*. (Prominent examples inside sociology include ‘field’ and ‘network.’) Doing so will require an engagement with the denotations and connotations of these terms as sensitizing as well as explanatory concepts. In other words, it will require engagement with and elaboration of the metaphorical standard, not only for ‘mechanism,’ but for other key causal terms as well. (For work in this spirit applied to the concept of ‘field,’ cf. Martin 2003.)

To do this requires us to see the social world as ‘dappled’—as consisting of sundry causes that are best captured by different terms, one of which is ‘mechanism.’ In this, we again

follow Cartwright (2004), who advocates ‘thick’ causal imagery that communicates information about specific kinds of causes in terms that are of greater explanatory use than the fugitive general causal laws we often attempt to abstract from them. We argue that including a metaphoric standard allows us to establish clearer criteria for thick causal accounts that draw from the ‘content-rich’ causal imagery sociologists use to convey these accounts in the first place. These criteria, derived as they are from specific uses of concepts, remain appropriate to the specific causes they describe, and, as Cartwright insists, should not be reduced to more fundamental standards.¹⁰

There is some argument for labeling this sort of pluralism a brand of ‘realism,’ in the sense that the metaphorical approach to mechanisms shares with a variety of different realisms the idea that, through its theoretical development, it can enable a better grasp on, and explanations of, social reality. However, the argument for elaborating a metaphorical approach to sociological explanation as a fruitful way forward could also be taken as ‘anti-realist,’ in the sense that it tends towards an interest in the complex connotations and tensions produced by metaphorical concepts, and thus away from a sheer denotative interpretation of how sociological understanding works. But perhaps the point is that realism/anti-realism is not the right dichotomy. Instead, at issue is the tension between parsimony and nuance, between fundamentalism and pluralism.

CONCLUSION

Surely, to the extent that mechanisms have become ubiquitous resources in causal explanation, we should be more precise about what we mean when we talk about mechanisms. Efforts to

construct a single definition of social ‘mechanism’ have faltered because there are three different ways of using the concept of mechanism to construct explanations, and these uses converge far too infrequently to justify any of the universal, monosemic definitions sociologists have so far proposed. The expositions offered in Part I elaborate three separate standards to capture the nuance across different kinds of mechanism-talk and save their contributions to the discipline where they inevitably fail to overlap with the demands of other standards. More to the point, it suggests that *each of the three standards presupposes a different approach to the construction and evaluation of explanation, and conflating these standards in an effort to enforce a single universal definition of ‘mechanism’ will lead to misunderstandings of explanatory practice in sociology*. This paper attempts to reduce those misunderstandings, and to offer a way forward in ongoing discussions of the scope and utility of mechanistic explanation in sociology.

REFERENCES

- Abbott, Andrew. 2007. 'Mechanisms and relations.' *Sociologica* 1(2): 1-22.
- Abbott, Andrew. 1988. 'Transcending General Linear Reality.' *Sociological Theory* 6(2): 169-186.
- Alexander, Jeffrey. 1982-3. *Theoretical Logic in Sociology*, Vol. 1-4. London: Routledge.
- Bechtel, William and Robert C. Richardson. 1993. *Discovering Complexity*. Princeton: Princeton University Press.
- Biernacki, Richard. 1995. *The Fabrication of Labor: Germany and Britain, 1640-1914*. Berkeley: University of California Press.
- Cartwright, Nancy. 2009. 'If no capacities then no credible worlds. But can models reveal capacities?' *Erkenntnis* 70(1), 45-58.
- Cartwright, Nancy. 2004. 'Causation: One word, many things.' *Philosophy of Science* 71(5): 805-820.
- Cartwright, Nancy. 1999. *The Dappled World: A study of the boundaries of science*. Cambridge: Cambridge University Press.
- Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge: Harvard University Press.
- Crespo, Ricardo F. 2013. *Theoretical and Practical Reason in Economics: Capacities and Capabilities*. Dordrecht: Springer Netherlands.
- Gerhart, Mary and Allan M. Russell. 1984. *Metaphoric Process: The creation of scientific and religious understanding*. Fort Worth, TX: TCU Press.
- Goldthorpe, John H. 2007. *On Sociology*. Second edition. Stanford: Stanford University Press.
- Gorski, Philip. 2009. 'Social 'mechanisms' and comparative-historical sociology: A critical realist proposal.' Pp. 147-194 in *Frontiers of Sociology*, edited by P. Hedström and B. Wittrock. Leiden: Brill.
- Gorski, Philip. 2003. *The Disciplinary Revolution: Calvinism and the Rise of the State in Early Modern Europe*. Chicago: University of Chicago Press.

- Gross, Neil. 2009. 'A pragmatist theory of social mechanisms.' *American Sociological Review* 74(3): 358-379.
- Hedström, Peter. 2005. *Dissecting the Social: on the principles of analytical sociology*. Cambridge: Cambridge University Press.
- Hedström, Peter. 1998. 'Rational imitation.' In: *Social Mechanisms: An analytical approach to social theory*. Hedström, Peter and Richard Swedberg (eds.) Cambridge: Cambridge University Press.
- Hedström, Peter and Richard Swedberg, eds. 1998. *Social Mechanisms: An analytical approach to social theory*. Cambridge: Cambridge University Press.
- Hedström, Peter and Petri Ylikoski. 2010. 'Causal mechanisms in the social sciences.' *Annual Review of Sociology* 36(1): 49-67.
- Hirschman, Daniel and Isaac Ariail Reed. 2014. 'Formation Stories and Causality in Sociology.' *Sociological Theory* 32(4): 259-282.
- Lichterman, Paul and Isaac Ariail Reed. "Theory and Contrastive Explanation in Ethnography." *Sociological Methods and Research*. Forthcoming. Published Online First 2014.
- Little, Daniel. 2012. 'Explanatory Autonomy and Coleman's Boat.' *Theoria* 27(2): 137-151.
- MacCormac, Earl. R. 1976. *Metaphor and myth in science and religion*. Durham: Duke University Press.
- Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. 'Thinking about mechanisms.' *Philosophy of Science* 67(1): 1-25.
- Martin, John Levi. 2003. 'What Is Field Theory?' *American Journal of Sociology* 109(1): 1-49.
- Menzies, Peter. 2012. 'The causal structure of mechanisms.' *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43(4): 796-805.
- Morgan, Stephen L. and Christopher Winship. 2015. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Second edition. Cambridge: Cambridge University Press.
- Pepper, Stephen C. 1961. *World Hypotheses*. Berkeley: University of California Press.
- Peirce, Charles S. 1992[1877]. "The Fixation of Belief." Pp. 109-123 in Nathan Houser and Christian Kloesel, eds., *The Essential Peirce: Selected Philosophical Writings, Vol. 1*. Bloomington: Indiana University Press.
- Reed, Isaac Ariail. 2011. *Interpretation and Social Knowledge: On the Use of Theory in the Human Sciences*. Chicago: University of Chicago Press.

Ricoeur, Paul. 2003 [1975]. *The Rule of Metaphor: The Creation of Meaning in Language*. London: Routledge.

Salzinger, Leslie. 2003. *Genders in Production: Making workers in Mexico's global factories*. Berkeley: University of California Press.

Seidman, Steven. 1991. 'The end of sociological theory: The postmodern hope.' *Sociological Theory* 9(2): 131-146.

Smith, Christian. 2010. *What is a Person? Rethinking Humanity, Social Life, and the Moral Good from the Person Up*. Chicago: University of Chicago Press.

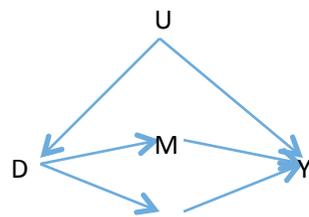
Steinmetz, George. 1998. 'Critical realism and historical sociology. A review article.' *Comparative Studies in Society and History* 40(1): 170-186.

Talisse, Robert. 2004. "Towards a Peircean Politics of Inquiry." *Transactions of the Charles S. Peirce Society* XL(1): 21-38 (Winter).

Vaidyanathan, Brandon, Michael Strand, Austin Choi-Fitzpatrick, Thomas Buschman, Meghan Davis, and Amanda Varela. 2015. "Causality in contemporary American sociology: An empirical assessment and critique." *Journal for the Theory of Social Behaviour* Published Online February 11, 2015: DOI: 10.1111/jtsb.12081, 1-24.

Wilson, Nicholas Hoover. 2011. 'From reflection to refraction: State administration in British India, circa 1770–1855.' *American Journal of Sociology* 116(5): 1437-77.

FIGURES



N

Figure 1: Directed acyclic graph representing a mechanism that explains the causal pathways from D to Y. Directly adapted from Morgan and Winship 2015: 332

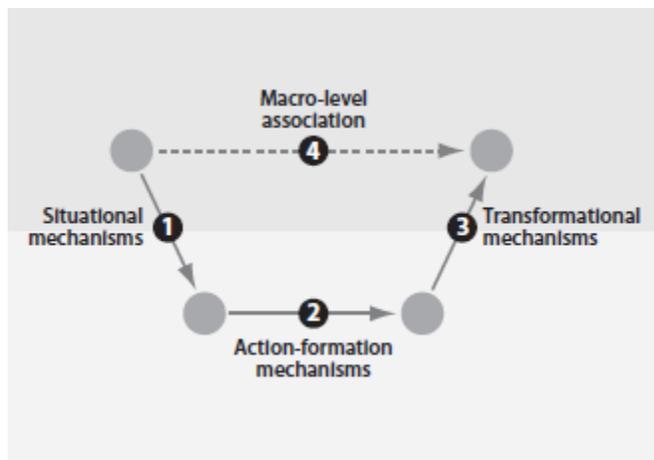


Figure 2: Hedström and Ylikoski’s typology of social mechanisms based on Coleman’s boat.

Adapted directly from Hedström and Ylikoski 2010:59.

NOTES

¹ We use the term “substantial” because this standard includes mechanisms that posit causal capacities inherent to particular substances or entities.

² We have selected Cartwright’s theory of mechanisms because it is the most comprehensive, synthesizing prominent structuralist, process, and interventionist positions on mechanistic causation, represented separately by philosophers like Bechtel and Richardson (1993), Machamer, Darden, and Craver (2000), and Menzies (2012), into a coherent ontological position with clear epistemological implications for causal explanation.

³ Cartwright herself tests the uncertain construct of social capacities by analyzing economic models, such as game-theoretic simulations. She finds that, to the extent these models meet the substantial standard, they do so by sacrificing external validity (1999:149). Ultimately, Cartwright finds that social capacities are extremely difficult to pin down, in part because they are constituted and reconstituted by human agents reacting to historical circumstances, individual impulses, and any number of other irregular or impermanent factors (cf. Crespo 2013).

⁴ While Cartwright by no means dismisses sociologists’ ability to construct causal explanations, she is generically skeptical about social scientists’ ability to *convincingly* construct mechanisms that reveal real social capacities. This is due in part to a lack of plausible general principles from which social scientists can work to isolate the tendencies of social capacities by eliminating extraneous forces and bracketing structural assumptions built into their models (Cartwright 2009:50).

⁵ The philosophical reference point here is C.S. Peirce’s concept of the community of inquiry. For discussions of the concept see Peirce 1992 [1877]; Talisse 2004; Lichterman and Reed, forthcoming.

⁶ Though initially a strong proponent of a purely analytical approach to social mechanisms, Hedström’s subsequent research around the desires-beliefs-opportunities (DBO) model can be interpreted as presuming a substantialist basis for the DBOs it postulates. Hedström’s ontological drift highlights how the potential for slippage between realist and analytical approaches bedevils many efforts to define a single standard for mechanistic explanation, a problem that further justifies the need to self-consciously parse the different explanatory standards of each approach.

⁷ An early forerunner was Stephen Pepper (1961), who considered mechanism to be a prime example of a ‘root metaphor,’ the basic framework that generates comprehensive ‘world hypotheses’—grand and all-encompassing schemes that guide cognition in science and philosophy. For Pepper, metaphor pre-sorts the kinds of processes that can be imagined as ontological furniture of the world in the first place.

⁸ This talk was delivered at New York University, December 9, 2013.

⁹ In the same text wherein she developed her understanding of nomological machines, Cartwright developed her criticism of standard philosophy of science in a new direction. ‘It is not *realism*,’ she wrote, ‘but *fundamentalism* that we need to combat’ (Cartwright 1999:23, emphasis original). Fundamentalism holds that the things described in laws or models are not only true, but true everywhere. In application, the fundamentalist position posits that every observable instance of any given phenomenon can ultimately be reduced to one of a few basic laws. For Cartwright, the problematic fundamentalism which drives to reduce the world to the laws of physics has as its preferable counterpoint her notion of the ‘dappled world.’ The dappled world is one where those purportedly universal, fundamental laws of physics describe only a limited range of phenomena and only under certain conditions; the rest of the world as we know it is populated by various ‘natures’ which we know to tend toward certain activities but

cannot hope to predict in the same way we can those few nomological machines that are described in natural law. Those situations not governed by universal laws can be subsumed under these laws ‘only by distortion, whereas they can often be described fairly correctly by concepts from more phenomenological laws’ (1999:31). Where natures can be described more accurately in phenomenological terms, they should ‘be admitted into the body of knowledge on their own merit’ (1999:24). It is important for understanding our argument that Cartwright’s notion of ‘natures’ or ‘capacities’ that make up nomological machines is *part of her* dappled world hypothesis—in other words, the world of causality has to be understood as dappled *because* different things in the world have different natures. This is why Cartwright calls her position one of ‘metaphysical nomological pluralism’—it does not abandon general causes, and assumes that our descriptions of causes can get very close to describing the real world. However, it assumes that there are many species of cause, and that the task of the (physical, natural, and social) sciences is to catalog the great variety of causes in ways that help us explain our world.

Our argument is that a metaphorical understanding of mechanisms aids in *doubling down* on this ‘dappled world’ hypothesis. In the human sciences, the construal of the world is such that even within the set of causes that concern ‘sociology,’ the world is extraordinarily dappled, because objects, kinds, and processes are formed and then unformed in a world of cultural and historical variation of ontology (cf. Hirschman and Reed 2014). In this context, there is a parallel to the fundamentalism in the philosophy of science that attempts to reduce all causal natures to the laws of physics.

¹⁰ This argument is based in our own affinity for the metaphorical standard. However, a more pluralistic approach to mechanisms is possible for sociologists who are not inclined towards the metaphorical standard, and do not wish to extend the group of causal metaphors beyond mechanism. Particularly from within the formal standard, one could posit “sundry mechanisms of different kinds.” We thank anonymous reviewer 1 of the paper for pointing this out.