Patterns, Information, and Causation

Abstract: This paper articulates an account of causation as a collection of information-theoretic relationships between patterns instantiated in the causal nexus. I draw on Dennett's account of real patterns to characterize potential causal relata as patterns with specific identification criteria and noise tolerance levels, and actual causal relata as those patterns instantiated at some spatiotemporal location in the rich causal nexus as originally developed by Salmon. The rich causal nexus serves the role of 'pixels' in the Dennettian pattern ontology. I develop a representation framework using phase space to precisely characterize causal relata, including their degree(s) of counterfactual robustness, their causal profiles, causal connectivity, and to identify their privileged grain size or level. By doing so, I show how the philosophical notion of causation can be rendered in a format that is amenable for direct application of mathematical techniques from information theory such that the resulting informational measures are *causal* informational measures. This account provides a metaphysics of causation that supports interventionist semantics and causal modelling and discovery techniques.

Keywords: causation; information; patterns; counterfactuals; interventionism; causal modelling; Salmon; Woodward; Dennett

H.K. Andersen
Simon Fraser University
handerse@sfu.ca

Patterns, Information, and Causation

The asymmetry and directedness of causation and thermodynamics have been closely linked.[1] There are many ways to relate these arrows, but information theory offers a new avenue to explore the connection between the directions of thermodynamics and causation. Information theory, broadly speaking, is an expansion of many core ideas and techniques from thermodynamics. It is an incredibly powerful approach to many areas in physics, economics, and other sciences, and active research is broadening its application to new phenomena. Some of these new developments are extremely intriguing, in part because they are semi-philosophical in formulation, hinting at a new foundational ontology for physics in terms of information.[2]

There have been attempts to use various measures of information transfer as a way to sort out causal structure from data sets.[3] These have met with at best mixed results: informational relationships such as the Kullback-Leibler distance don't simply yield up causal relationships from data. Part of the problem with prior attempts to utilize information theory to find causal structure is that information theory is so broad that it can be applied to just about anything; the resulting informational relationships may not be between anything that could even possibly stand as causal relata. In order to get mileage out of information theory with respect to finding *causal* informational relationships, it must be applied to the right sorts of relata, namely, causal relata.

On the one hand, informational relationships will only be *causal* informational relationships when the relata are causal. On the other hand, it is not straightforward *to what* information theory should be applied such that the resulting relationships would be causal. Causation as discussed in philosophical debates is not yet in the right form for information theory to be directly applicable, nor is it immediately clear what aspects or elements of causation would be appropriate for such application. Central elements of information theory involve various informational and entropic relationships between probability distributions over various

[1] Hans Reichenbach, *The Direction of Time* (Berkeley, CA: University of California Press, 1956).
[2] See, for instance, Carlo Rovelli, "Relative Information at the Foundation of Physics," in Anthony Aguirre, Brendan Foster, and Zeeya Merali, eds., *It from Bit or Bit from It? On Physics and Information* (Cham: Springer, 2015), pp. 79-86.
[3] For example, Hlavá□ková-Schindler, K., Paluš, M., Vejmelka, M., & Bhattacharya, J. (2007). Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports*, 441(1), 1-46.

kinds of volumes. In order to use information theory in an account of causation, causation needs to be represented in the right format. That format involves volumes that can be partitioned over which the probability distributions can be put, such that the informational measures between volumes can be calculated. An apt representational framework, then, provides the key to apply these technical resources to a philosophical understanding of causation.

This paper begins with traditional philosophical approaches to causation and ends with the right materials to which information theory can be applied to yield an information-theoretic treatment of causation. This process involves incorporating processes and counterfactuals into a single unified account of causation, representing causal relata and relationships in ways that render them amenable to the application of mathematical tools from information theory. The task here is the philosophical preparation of the material, as it were, in order to apply the tools from information theory. This has two main directions: one is taking causation and articulating it such that we see how it can be treated with the tools of information theory while still being clearly a conceptual explication of the idea as it is found in philosophical discussion; another is taking information theory and showing how to deploy it for causation in particular. The focus of this paper is on the first direction, showing how mechanistic causal processes in conjunction with the interventionist counterfactual approach and a pattern ontology yield an appropriate target for application of information-theoretic mathematical tools. There are a variety of ways in which specific elements of information theory could be applied to determine degree or strength of causal connectedness. I take these to be methodological issues that ought to follow from, rather than precede, the clarification of the metaphysical character of causation.

The resulting view of causation, in its most basic form, is this. Causal relata *are* patterns instantiated in a rich causal nexus; causal relationships *are* informational relationships between those patterns. The notion of pattern is primitive, in the sense deployed by Sider.[4] It is not that individual patterns are primitive, but that the idea of a pattern, and patterns as what we seek when we look for causal relata, is primitive. With that laid out, the remaining considerations are largely methodological rather than metaphysical. Patterns are defined using identification criteria and noise tolerance levels. They are constituted by their description, and can, separately, be identified as occurring or failing to occur in particular spatiotemporal areas of the rich causal nexus. The physical details of what, exactly, the causal nexus is, is revealed by physics. It is a

[4] Theodore Sider, *Writing the Book of the World* (New York: Oxford University Press, 2011).

metaphysical claim that genuine causation must be instantiated in the causal nexus somewhere, but the details of what constitutes the causal nexus in our actual world are ontic and subject to updating from physics. Likewise, which patterns we should use to most effectively track causation in that nexus is ontic and subject to ongoing revision based on considerations such as developments in the sciences. The overwhelming majority of patterns are counterfactually robust, in that they could have differed in their microphysical details in each token instantiation without thereby altering the relatum's causal profile. Illustrating the bounds of the counterfactual robustness as a volume in phase space illuminates how counterfactuals relate without reducing to microphysical causal processes. These volumes in phase space representing the counterfactual robustness zones of pattern-tokens instantiated in the rich causal nexus can be partitioned, and various probability distributions can be put over those partitions. This leaves us with the materials of causation in the right form for application of techniques from information theory.

I. Background

Pioneering work in causal modelling and search methodology[5] has been supplemented by the interventionist account of James Woodward,[6] explicitly situated as providing a semantics for the causal methodology. Woodward's work has been criticized for not giving an account of what causation really *is*, in some more fundamental sense, and instead 'merely' providing an account of how to recognize causation.[7] While it is fair to say that Woodward's account does not do this, this is not a criticism per se, since it was not a goal of his account to do so. What might with less sympathy be called circular may instead be construed as non-reductive. There is undoubtedly a circle of interdefinability in Woodward's account, where interventions are used to characterize causation while also involving thick causal concepts themselves. He explicitly notes the non-reductive character of his account: the goal is not to reduce the concept of causation to something else. Rather, it is to clarify causal explanation and the discovery and representation of

[5] Especially, Peter Spirtes, Clark N. Glymour, and Richard Scheines, *Causation, Prediction, and Search* (Cambridge, MA: The MIT Press, 2000), and Judea Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge: Cambridge University Press, 2000).
[6] James Woodward, *Making Things Happen: A Theory of Causal Explanation* (New York: Oxford University Press, 2005).
[7] See, for instance, Alexander Reutlinger, "Getting Rid of Interventions," *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, xliii, 4 (December 2012): 787-795.

causal structure by broadening the circle of terms that are interdefined. The technical definition of an intervention can shed light on other causal concepts such as making a difference without thereby reducing it to something non-causal. Woodward's goal was not a metaphysics of causation, it was a semantics for casual methodology.

This paper aims to provide what Woodward's account has been criticized for lacking. The account developed here provides the underlying metaphysics that supports the semantic account of Woodward, which in turn supports the causal search and modelling methodology. The resulting package of views, with now-sufficient conceptual and empirical resources, allows for the completion of the branching-off from philosophy that is already underway for causation to become an autonomous field of science. This paper follows in the footsteps of other work in philosophy of science, and draws on discussions in metaphysics, but is neither straightforwardly metaphysical nor part of the more standard contemporary philosophy of science discussion. It is perhaps more accurately construed as fitting into the rather old-fashioned tradition of natural philosophy.

There are, in contemporary discussions of causation, several distinct clusters of views. Two such clusters are difference-making accounts and mechanistic connection accounts.[8] The former includes the counterfactual theory of Lewis, the interventionist account of Woodward, as well as counterfactual accounts of explanation that include both causal and noncausal explanations.[9] There are major differences in how such accounts evaluate counterfactuals. Nevertheless, there are certain commonalities that distinguish this cluster: in particular, difference-making accounts do not require a physical chain of mechanisms or processes to connect cause and effect. The latter cluster of accounts includes broadly productive accounts of

[8] On this, see Ned Hall, "Two Concepts of Causation," in John Collins, Ned Hall, and L.A. Paul, eds., *Causation and Counterfactuals* (Cambridge, MA: The MIT Press, 2004), pp. 225-276. A third identifiable cluster involves the notion of powers or of capacities (see, for instance, Stephen Mumford and Rani Lill Anjum, *Getting Causes from Powers* (Oxford: Oxford University Press, 2011) or Nancy Cartwright, *Nature's Capacities and their Measurement* (Oxford: Clarendon Press, 1994), and another might include pluralism (see, for instance, Christopher Hitchcock, "Of Humean Bondage," *The British Journal for the Philosophy of Science*, liv, 1 (March 2003): 1-25, Peter Godfrey-Smith, "Causal Pluralism," in Helen Beebee, Peter Menzies, and Christopher Hitchcock, eds., *The Oxford Handbook of Causation* (Oxford: Oxford University Press, 2010: 326-337), or Nancy Cartwright, *Hunting Causes and Using Them: Approaches in Philosophy and Economics* (Cambridge: Cambridge University Press, 2007)). The way pluralism, powers, or other similar notions relate to this account will have to be addressed in a further paper.

[9] See, for instance, Alexander Reutlinger, "Explanation Beyond Causation? New Directions in the Philosophy of Scientific Explanation," *Philosophy Compass*, xii, 2 (February 2017): 1-11.

causation, such as the mechanistic causal processes and interactions of Salmon and Dowe, as well as the layered causal mechanisms of Glennan.[10] This approach is characterized by an emphasis on physical connections between causal relata; while the precise character of the connection(s) differs between accounts, the existence of such a connection is required for and usually constitutive of causation.

The mechanistic causal process account, developed by Salmon (*Causality and Explanation*) and Dowe (*Physical Causation*), offers a distinctive actualist account of causation. Yet this account faces a number of criticisms, two of which are especially germane here. It has difficulties picking out the right "grain size" with respect to size scale and organization; and it offers no way to understand why counterfactuals, especially interventionist counterfactuals, are both ubiquitous and so efficient in conveying information about causal relationships. Before this account can be integrated with intervenionism, these issues must be resolved.

With respect to the first problem, picking the right grain size, Williamson has shown how this mechanistic causal process approach lacks the resources to prevent causal drainage, such that causal efficacy reduces to some lowest level of microphysical processes.[11] Such strong causal reduction is descriptively unsatisfactory. Labelling vast swathes of scientific practice as entirely misguided requires exceptionally strong justification, given the ubiquity with which these higher-level causes are treated as genuine. It is also explanatorily problematic, in that a great deal of causally irrelevant microphysical information gets included in causal explanations.

With respect to the second problem, the mechanistic causal process account fails to replicate or even allow for a meaningful description of many key characteristics of causation. Part of Salmon's explicit motivation in developing his account was to eliminate the need for

---

[10] See, for instance, Wesley Salmon, *Scientific Explanation and the Causal Structure of the World* (Princeton, NJ: Princeton University Press, 1984) and *Causality and Explanation* (New York: Oxford University Press, 1998), Phil Dowe, *Physical Causation* (Cambridge: Cambridge University Press, 2000), Stuart Glennan, "Mechanisms, Causes, and the Layered Model of the World," *Philosophy and Phenomenological Research*, lxxxi, 2 (September 2010): 362-381. The 'new mechanisms' approach to explanation is not itself an account of causation (Holly Andersen, "A Field Guide to Mechanisms: Part I," *Philosophy Compass*, ix, 4 (April 2014): 274-283), and could be compatible with either strand here, for which reason I leave it aside in this discussion. The layered causal mechanisms view of Glennan is productive, like that of Salmon and Dowe, but it must be stressed that while the term 'mechanism' appears in both accounts, it picks out relevantly different features of the world in either (on this, see Holly Andersen, "A Field Guide to Mechanisms: Part II," *Philosophy Compass*, ix, 4 (April 2014):284-293).

[11] See Jon Williamson, "Mechanistic Theories of Causality, Part I," *Philosophy Compass*, vi, 6 (June 2011): 421-432.

counterfactuals; his target was Lewisian counterfactuals, however, which are different in important regards from counterfactuals in contemporary interventionist accounts. The process account is therefore stuck with an awkward dilemma: either interventionist counterfactuals reduce to microphysical mechanistic processes, even though we can't actually provide such reductions in the overwhelming majority of cases; or they fail to pick out genuine causal relationships, and we are on the hook to explain why they apparently work so well. This second horn has been chosen by Dowe in *Physical Causation*.

Both horns of this dilemma, furthermore, are unappealing for their strong metaphysically presumptive character, which is rather ironic given Salmon's motivation of the account as avoiding metaphysical commitments. Given the widespread applicability and usefulness of interventionism, it is a deep article of faith to claim that the 'real' causal story in such cases is entirely microphysical, even while acknowledging that the reduction in question is hopelessly beyond our epistemological grasp. It cuts the account off from the solid foundation of empirical evidence from the sciences in which it was originally intended to be grounded.

A number of authors have offered strategies for unification or consilience between process and difference-making accounts. Strevens offers a reconciliation between counterfactuals and processes, but does so in terms of causal explanation, rather than causation directly.[12] His kairetic account of explanation demonstrates how explanations, conceived of as propositional in structure and arranged into arguments, accommodates both physical process and counterfactual accounts of causation, by elimination of premises that are not required to deduce the conclusion. This winnowing process addresses the problem mentioned above of inclusion of too much explanatorily irrelevant detail in higher level causal explanations. But it unifies them into an account of causal *explanation*, not of causation, and the requirement that they be propositionally structured precludes the extension to causation. This is not a criticism, since his target is explanation rather than causation, but does distinguish his approach from the one here. Jackson and Pettit offer a distinction between causal efficacy and causal relevance.[13] Only efficacy is genuinely causal, however; causal relevance is important for causal *explanation*, but not causation per se. The account developed here recognizes both sides of that distinction as genuinely causal.

---

[12] Michael Strevens, *Depth: An Account of Scientific Explanation* (Cambridge, MA: Harvard University Press, 2008).

[13] Frank Jackson and Philip Pettit, "Program Explanation: A General Perspective," *Analysis*, l, 2 (March 1990): 107-117.

Ney comes most directly at the question of unification of difference-making and physical process theories.[14] She offers a reduction of difference-making to physical process based on the argument that the fundamental facts about causation are physical facts, not difference-making facts. The view offered here differs from hers in that a clean distinction between physical facts and difference-making facts turns out to be a false dichotomy, and the term 'fundamental' is equivocal. My view has the consequence that fundamental (in the sense of microphysical) physical facts are only degeneratively causal, and that fundamental (in the sense of the smallest set of metaphysically basic) facts about causation require both physical processes and counterfactuals.

Causal process tracing, an inferential technique developed for social science fields like political science, aims in a similar direction, with an epistemological rather than metaphysical orientation. Causal process tracing is an inferential solution for causal modelling in systems with variables on which we are unable to directly intervene, for logistical, ethical, or other reasons. For abstract variables for which we lack experimental means to evaluate counterfactuals, it supplements the variables with underlying causal processes that allow for further inferences about causal structure of unique systems.[15] The way in which counterfactuals can reveal clearly genuine but also incredibly abstract causal relationships--such as those posited between being a resource-rich country, having unstable governance, and civil war--can be enriched by finding ways in which those abstract relata are identified as an instance of the right sort.

Bringing these discussions together, it is apt to require of any proposed account aiming to unify process and interventionist counterfactual causation that such unification reveal something new about either cluster of approaches. The way in which counterfactuals emerge from productive and particular causal happenings in the physical world should help us understand such counterfactuals better, and such counterfactuals should give us a better handle on how to suss out the relevant processes from the noisy entropic world. As such, any

[14] Alyssa Ney, "Physical Causation and Difference-Making," *The British Journal for the Philosophy of Science*, lx, 4 (December 2009): 737-764.
[15] See, for instance, David Collier, "Understanding Process Tracing," *PS: Political Science and Politics*, xliv, 4 (October 2011): 823-830, Rosa W. Runhardt, "Evidence for Causal Mechanisms in Social Science: Recommendations from Woodward's Manipulability Theory of Causation," *Philosophy of Science*, lxxxii, 5 (December 2015): 1296-1307, Derek Beach and Rasmus Brun Pedersen, *Process-Tracing Methods: Foundations and Guidelines* (Ann Arbor, MI: University of Michigan Press, 2013).

underlying metaphysical view that unifies both approaches should have meaningful methodological consequences.

Salmon considered and rejected[16] one version of an information-theoretic approach to explanation, namely that of Greeno. Greeno offered a view of explanation similar to the statistical-relevance view, drawing on information theory to cash out the statistical relationships.[17] While Salmon thought this had some promise, he ultimately rejected it as an account of explanation: in agreement with Hanna,[18] Salmon says "... statistical relationships among observables have little, if any, explanatory force. ...the S-R basis needs to be supplemented with causal and theoretical considerations in order to be able to characterize genuine scientific explanations. It therefore seems to me, in effect, that Greeno's initial information-theoretic account (1970) had just the same strengths and weaknesses as the S-R model." (Salmon, *Scientific Explanation*, p. 100). Thus, at least one reason for Salmon's rejection of this early version was its inadequacy as an account of explanation, not causation. Furthermore, because it applied information-theoretic formulas to statistical relationships codifying knowledge, the account he considered construed information as a relationship between knower and a body of knowledge, not between parts of the world.

More recently, information and causation have been linked in terms of the measurement of causal specificity, the degree to which a cause is fitted specifically to a single effect.[19] In order to measure causal specificity, Griffiths et al. put probability distributions over the values that a variable can take, and use these probabilities to track how closely cause and effect are connected. Their approach also invokes information theory as a way of measuring ignorance about variable values rather than states of the world.[20] Their approach differs markedly from the one here in several ways: in tracking ignorance rather than causation directly, in only applying probability distributions to well-defined variable values rather than single cases, and in being methodological but not ontic in character. This fits with their goal of defining a measure for the purpose of tracking biological, in particular biochemical, specificity in genetics.

---

[16] *Scientific Explanation*, pp. 97-100

[17] James G. Greeno, "Evaluation of Statistical Hypotheses using Information Transmitted," *Philosophy of Science*, xxxvii, 2 (June 1970): 279-294.

[18] James F. Hanna, "On Transmitted Information as a Measure of Explanatory Power," *Philosophy of Science*, xlv, 4 (December 1978): 531-562.

[19] See, for instance, Paul E. Griffiths, Arnaud Pocheville, Brett Calcott, Karola Stotz, Hyunju Kim, and Rob Knight, "Measuring Causal Specificity," *Philosophy of Science*, lxxxii, 4 (October 2015): 529-555.

[20] See, for instance, Griffiths et al., "Measuring Causal Specificity," p. 533.

Finally, John Collier has offered an analysis of causation as the transfer of information, with a usage of 'information' that is broadly similar to the one developed here.[21] There are many interesting points of similarity and difference between our views, which must be explored in a further paper. For now, it is worth noting that he also draws on the conserved quantity transfer account, but does not rely on patterns as a way to pick out relata, instead conceiving of informational causal relationships as channels between a receiver and transmitter. He aims to eliminate counterfactuals and instead unify Salmon's processes with universals. And finally, part of his "minimal metaphysics" of causation involves the idea that causation is a computation whereby it is not particularly relevant what does the computation, but merely that a given computation is performed. One could, at the risk of oversimplification, contrast his use of information as broadly Neo-Platonic with my own as broadly Neo-Aristotelian in comparison.

Even while disagreeing with some of the details, the overall spirit of Salmon's approach to causation is one this paper follows.

> *Statistical and causal relations constitute the patterns that structure our* world – the patterns into which we fit events and facts we wish to explain. *Causal processes* play an especially important role in this account, for they *are the mechanisms that propagate structure and transmit causal influence* in this dynamic and changing world. In a straightforward sense, we may say that these processes provide the ties among the various spatiotemporal parts of our universe… *They are the channels of communication by which the physical world transmits information about its own structure.* (Salmon, *Causality and Explanation*, p. 66; emphasis added.)

The idea of patterns as structuring the world, and causal processes as transmitting information about structure in ways that tie the world together spatiotemporally, will be deployed in a new way in this paper.

II. Pattern ontology

The notion of a pattern is itself very intuitive and familiar, but in a way that can be misleading when it comes to a pattern *ontology*, committing to patterns as what actually exists, rather than patterns as a identifier or classification for other, more familiar, ontological items.

---

[21] John D. Collier, "Causation is the Transfer of Information," in Howard Sankey, ed., *Causation and Laws of Nature* (Dordrecht: Kluwer Academic Publishers, 1999), pp. 215-245, "Information, Causation and Computation," in Gordana Dodig-Crnkovic and Mark Burgin, eds., *Information and Computation: Essays and Scientific and Philosophical Understanding of Foundations of Information and Computation* (Hackensack, JN: World Scientific, 2011), pp. 89-106.

When Dennett originally introduced the idea of a pattern ontology,[22] conceiving of information as a relationship between parts of the world, rather than as a kind of synonym of knowledge, was very unfamiliar to most readers. Accordingly, the radical implications of such an ontology were underappreciated. The field of information theory has developed rapidly since his paper: an explosion of new work and techniques has occurred in the last fifteen years or so, along with an explosion in the capabilities of modelling software. As such, the particular technical definition of pattern that Dennett introduced, involving algorithmic compressibility, requires refinement. It is also worth noting that Dennett introduced the idea with a specific application to beliefs. In this way, the paper undersold his own view: his proffered pattern ontology for belief only works if one also takes a pattern ontology for *everything*. This paper thus substantially extends Dennett's own use of a pattern ontology.

Denett introducespatterns using Conway's Game of Life, set in a pixellated computer world. The basic constraints on the world is simply that pixels are either on or off, and the algorithm that determines the state of each pixel through time steps according to a very basic set of rules involving the states of its neighboring pixels. If, for instance, a certain number of a given pixel's neighboring pixels are on at time 1, then that given pixel will be on at time 2. Thus, there is a bottom level of the game-world where all pixels are governed deterministically by a set of equations giving the time evolution of their states from a set of initial conditions. The initial conditions are just the distribution of on and off states for all pixels at time 0. At this pixel level description of the world, there are discernible but quite basic patterns, necessarily confined to individual pixels' behavior through time (in other words, they turn on and off). The bit map is a description of the entire state of the world that gives the exact state of each pixel. It is exhaustive, in that each pixel is completely described, and unwieldy, in that it takes a great of informational space to give such a detailed description.

However, moving up from the pixel level, there are stable and identifiable types of patterns that 'live' in Conway's Game of Life. These stable, trackable patterns can be described using a new vocabulary that is not tied to individual pixels. Gliders are such a species of pattern: they are a stable configuration of pixels that repeat a cycle of state changes so that after one cycle, the glider has 'moved' across the screen and enters another cycle to keep 'moving'. Eaters are another species in the Game of Life; when they encounter other pixel patterns like gliders, they destroy their stability, thus 'eating' the glider. This is what Dennett calls the design level:

[22] Daniel C. Dennett, "Real Patterns," *The Journal of Philosophy*, lxxxviii, 1 (January 1991): 27-51.

there are patterns like eaters and gliders that maintain stability over time and across the screen, despite the fact that they are nothing over and above sets of pixels flashing on and off according to deterministic algorithms. At the pixel level, nothing moves; pixels can turn on and off but cannot relocate. At the design level, movement is possible – the same glider can move across pixels and be consistently tracked or re-identified over time. Design levels can get arbitrarily complicated: it is possible to build a Turing machine in Conway's Game of Life, made out of eaters and gliders.

Even though Conway's Game of Life is a fully deterministic system with simple iteration rules, it can be very computationally costly to work out future states of the system based on the rules plus the bit map. Making predictions about future states is very costly in terms of efficiency, but very accurate because the pixel level is fully deterministic. Dennett's insight is that one can use the design levels instead as a way of tracking what is going on in the world, with a certain kind of trade-off. It is, computationally, vastly more efficient to keep track of design level patterns like eaters and gliders, in order to do things like predict what will be printed on the Turing machine tape in the next several seconds. These vast improvements in efficiency come at the cost of a fairly small reduction in accuracy. The predictions are no longer 100% accurate, since there is discarded information which might turn out to be relevant. However, for many purposes, this small cost in accuracy is worth the improvement in efficiency. We can keep track of these patterns that are composed of nothing mysterious – it is still just pixels – but which move and behave in slightly indeterministic but computationally simple ways.[23]

The pattern *ontology* involves thinking of the Game of Life world as one comprised of patterns, rather than merely of pixels. Gliders and eaters really are there, in a somewhat deflationary but crucially non-reductive construal of 'really there'. Gliders genuinely exist as they cruise across the screen. Any pattern that can be reliably picked out and tracked through time

[23] While Dennett introduced patterns using Conway's Game of Life, there are newer games that perhaps better illustrate the extraordinary capacity of a pattern ontology in a simulated world as an analogue for reality. Minecraft would in many ways be a better example to introduce a pattern ontology for our world, since Minecraft is laid out explicitly as a translation of our world. The Minecraft world is three dimensional, plus time, as opposed to the flat two dimensionality plus time of Conway's Game. Even though both games are played through screens, Minecraft's world is not confined to that screen as Conway's is; players must 'look through' the screen and keep track of a three dimensional map to play successfully. Minecraft also illustrates how rich the pattern ontology can be. I rely on Conway's Game rather than Minecraft in this paper to follow Dennett's presentation more clearly.

(subject to the conditions to be discussed shortly), and which allows one to make predictions that are better than chance, is as real as any other pattern. The bit map pixel changes also exist, and are simply a very boring, mathematically degenerate, pattern. I'll call this Laplace's Pattern, since it would be the pattern Laplace's Demon would surely use if Conway's Game were a demon-haunted world. It is not the only real pattern, however. There is a kind of profligacy to the realism about patterns here. For a complicated world there could be a vast number of different ways of picking out such patterns that give us predictive grasp on the system. But it is not a troubling profligacy, because the degree of realism is very, very, minimal: there is not much commitment involved in saying that some pattern is 'really' there. And it turns out to be rather hard to find patterns that genuinely meet the criteria, such that concerns about rampant proliferation of patterns are misplaced.

Further, it is objective whether a given pattern occurs, not merely epistemic or even perspectival. "A pattern exists in some data – is real –if *there is* a description of the data that is more efficient than the bit map, whether or not anyone can concoct it" (Dennett, "Real Patterns," p. 34, emphasis in original). Patterns themselves are defined in terms of identification criteria. We can define or change those criteria, and change the noise tolerance for picking out a pattern. Once those parameters are set, however, it is fully objective whether a given portion of the screen contains a glider with at most 5% noise. Pattern ontolog*ies* (not, Ontology) can be perspectival when different interests or goals lead to genuinely different ways of carving up the world into patterns, different interconnected patterns that together 'cover' the whole.

The more efficiently we describe a pattern, the faster we can identify whether it occurs. If there are tasks for which speed is relevant, we might prioritize efficiency of description and accept reduced accuracy as a worthwhile compromise. Conversely, accuracy might be highly valued for a different task, and so efficiency might be lowered in order to gain in terms of accuracy. Yet both an efficient but noisy pattern, and a different inefficient but accurate pattern, may be 'really' there in the same area of the screen, even if the two patterns are not identical. Compare someone using pattern A, with a highly efficient description and 20% noise, and someone else using pattern B, with a low efficiency description and 5% noise, to describe the same set of pixels. Even though A and B are different, "… if both patterns are real, they will both get rich. That is to say, so long as they use their expectation of deviations from the 'ideal' to temper their odds policy, they will do better than chance – perhaps very much better" (Dennett, "Real Patterns," p. 35).

This has a consequence worth drawing out. Returning to the previous example, if we stipulate the pattern to be a "glider, no more than 5% noise," there is an objective fact about whether or not there is such a pattern in a given patch of pixels. There may be more than one pattern in that patch, however; perhaps there is also "eater, with 5% noise". These patterns may be picking out overlapping pixels as part of distinct and genuinely real patterns: several pixels might be part of the eater and part of the glider at a given instant. If we pick one of those pixels, we might ask: which of the two patterns does it really belong to? Shouldn't the one, or the other, but not both patterns present in that very patch of pixels? How can both be genuinely and equally real if they are double-counting the same pixels? If one is committed to a pattern ontology, then both patterns can be present, and be genuine, and overlap, potentially to a substantial degree. Yet they are not identical, even with substantial overlap; each pattern picks out a chunk of the pixels *as* an instance of a type of pattern that differs, and at least some of chunk of pixels thus picked out differ. These patterns are identified in this particular area, but have different ways of 'going on' as rules by which to find more tokens. The patterns themselves might overlap in single instances, and be equally real and equally present in that patch, because the pattern itself is also construable as a collection of such instances, and the defining criteria for each pattern can differ markedly and the collection of instances differ markedly.

There is no strict trade-off between efficiency and accuracy. There are patterns that turn out to have incredibly high efficiency as well as a very low noise level. There may be different patterns that have a very high noise level but are also really inefficient. We could use them if we wanted to make our jobs hard--nothing prevents us from attempting to find, and perhaps even succeeding in finding, perversely constructed patterns. But our hypothetical willingness to engage in perversely unnecessary contortions in constructing an ontology doesn't tell us much if anything about the world. It is analogous to the case of representing a sphere using rectangular coordinates instead of spherical coordinates. It is no doubt possible, but would certainly make the job harder, and has no interesting philosophical, mathematical, or physical significance. There are clear advantages to finding patterns that manage to both be highly accurate and also highly efficient.

The idea of a pattern can be used in a variety of ways. It has recently enjoyed something of a renaissance, especially in philosophy of science, although the term has also been employed for a variety of philosophical purposes prior to Dennett's use of it (see the Salmon quote in the

previous section). The idea of a pattern, however, is so flexible and useful that, like information theory, it can be dangerously vague. There can be many uses of the notion of patterns that do not involve commitment to the real or objective existence of patterns. For instance, Potochnik uses the idea to identify different types of explanatory patterns.[24] The same explanatory pattern might be used even though the precise explananda differ substantially. In her usage, though, patterns are not objective phenomena, but instead are characteristic features of explanations that can be identified across different explanations.

This brings us to the key question in order to provide a pattern ontology *for causation*. In Conway's Game of Life, the pattern ontology works as a mild form of realism because there is a very well-defined answer to the question *of what* they are patterns: they are patterns of pixels. In order to treat causal relata as patterns, we need to clarify *of what* they are patterns: what is the analogue to pixels in the actual world such that patterns in *that analogue* could be what we recognize as causal relata?

III. Counterfactually robust patterns in the rich causal nexus

The answer to this question can be drawn from Salmon's mechanistic causal process theory. The nexus of conserved quantities that propagate and can be transferred via interaction in the Salmon-Dowe account provides excellent material for the 'pixels' of a causal nexus. The lowest causal level has edges that are continuously propagating conserved quantities and nodes where these lines intersect in exchanges of conserved quantities. These 'pixels' can be kept track of individually, just as actual pixels in the Game of Life can be tracked.[25] Causal relata are patterns in this rich causal nexus.

This section lays out some details of how patterns can be instantiated in the causal nexus in part by laying out a representational device or framework for modelling causal relata and

[24] Angela Potochnik, *Idealization and the Aims of Science* (Chicago, IL: University of Chicago Press, 2017).

[25] There may be physical quantities such that their exchange is conserved across some but not all interactions. Rather than pointing at such cases as counterexamples to the entire causal process framework, I think we should treat these examples as interesting ways to develop and refine the account. For instance, one could compare what happens if such a quantity were removed entirely, such that it contributed no edges or nodes to the nexus, versus retaining the interactions where the quantity is conserved as nodes in the nexus but not the interactions where conservation fails. Highlighting the empirical difference this would make to models of systems where this is relevant can provide grounds to select one or the other treatment of the quantity with respect to the nexus.

relations in the nexus. I will rely on phase space as a way to represent the exact state of the causal nexus (it might, but need not, be the exact microphysical specification; coarser grained representations might not be microphysical). This is then used to define the notion of counterfactual robustness and the idea of a counterfactual robustness zone that is the linkage between the causal process nexus and counterfactuals. The goal is to motivate this approach for those less familiar with phase space as a representational device. This example will be simplified, perhaps tediously so, although not in ways that affect the main point, and will leave aside for further discussion elsewhere interesting questions about continuous versus discrete representations, ergodicity, and the choice of phase space rather than e.g. state space or configuration space.

Phase space is a common representational device in thermodynamics and statistical mechanics. Many readers may have encountered phase space defined for the particles in a box of gas: each particle has three degrees of freedom for position and three for momentum, so the total dimensionality for the phase space is 6 times the number of particles. Information representing the exact location plus momentum for each individual particle is contained in the point in phase space, and the movement of all particles is tracked individually by the changing 'location' in phase space of the point. The total volume of phase space for such a box of gas is given by the totality of accessible points for the particles, their range of possible locations and momenta. The temperature of such a box of gas is then given by various volumes in phase space: all points within that volume correspond to the same temperature. A volume here just is a region in the phase space, in which points can be identified. The temperature must be represented by volumes since there are so many microstates of the box compatible with each macrostate of temperature; each volume just is the collection of all the points that give rise to the same macrostate. Put another way, any given macrostate of temperature has *some* precise microstate, but would have the same macrostate/temperature with a different microstate: if two particles were momentum-switched or location-switched, or if one particle were given the same magnitude of momentum but with the opposite direction. As long as the point representing the microstate of the box is anywhere within that volume of phase space, the box has the same temperature.

Applying this to the rich causal nexus, the total momentary state of the causal nexus can be represented exhaustively by specifying the values for the relevant degrees of freedom for each edge (a causal process) and node (a causal interaction) in the nexus. This total state

specifies a phase space of N dimensions, where N is equal to the numbers of causal processes at that moment times the number of degrees of freedom for each causal process (the degrees of freedom is not constant for each causal process, because some causal processes may bear multiple conserved quantities varying over time). One point in phase space provides the exact physical specifications for the entire nexus at that moment. The time evolution of the nexus then traces out a path through phase space, where each point on the continuous trajectory is equivalent to a full specification of the state of the nexus at successive moments in time. Each point is thus unique as a state of the system; for any two points, no matter how close, something must differ about some part of the microphysical causal nexus. The more edges and nodes differ in their values, the further away those points are in phase space (leaving aside the details of the measure of such distance for now). Tracking the evolution of the nexus through time traces out a trajectory through the accessible volume of phase space.

Patterns are defined by giving criteria by which they can be recognized and identified. The identification criteria that constitute a given pattern provide the handles by which they can be represented in this format. An adequate characterization of a pattern provides criteria which allow us to definitively 'check' whether that pattern occurs in a given region of the nexus.[26] Patterns have a kind of robustness in tokening, which Dennett illustrates with the example of the pattern Bar Code. Bar Code can be defined ato have greater or lesser tolerance for noise; some black pixels could be swapped with some white ones, while remaining a token of the very same pattern. Each consecutive tokening of the pattern may have a different noise level, but still token *the same pattern*, according to the pattern identification criteria. A single tokening of the Bar Code pattern could have been different in a number of pixels while remaining a token, in the very same place, of the very same pattern.

Analogously, when we pick out patterns in the causal nexus, except under extraordinarily unusual circumstances,[27] there are multiple ways in which the 'pixels' of the causal nexus could have been different while the very same pattern was still instantiated. For instance, one edge in the nexus could have had any value within a given range for momentum while instantiating the

[26] Dennett uses algorithmic compressibility as a criterion for pattern recognition, but there may be issues with this precise definition. Given how much the idea of compressibility, and specific ways to compress, have developed in the last twenty years, it would be surprising if his original suggestion still happened to be the most fruitful pattern definition to have come out of computer science.

[27] Such cases will primarily occur in physics, involving mathematically-defined, noise-intolerant patterns. There is much more to be said on this in further work.

same pattern within the defined noise tolerance range. One interaction could be swapped with another, with no overall effect on the pattern thus tokened. Call this *counterfactual robustness*: it is the possibility of counterfactual variation in specific values for the underlying nexus for a given tokening of a causal relatum. Any token causal relatum where some microphysical details of the actual state of the nexus could have been slightly different, while still tokening the same pattern, is counterfactually robust.

Counterfactually fragile relationships exist between *points* in phase space connected by the time evolution of a system. The causal relationships connecting points are fragile because any variation in the exact microstate of the nexus changes the point and thus destroys the original connection. Causal relationships between *volumes* of phase space are counterfactually robust. A great deal more variation in the microstate of the causal nexus is required, enough to leave the boundaries of one volume, to destroy a causal relationship that may exist between two volumes.[28] The extent of the robustness is defined by the pattern identification criteria themselves, including the noise tolerance levels, in that it is these factors that determine how much could be changed within the bounds of *that* specified noise level for *that* particular pattern rather than another.

Counterfactually robust causal relata are the analogue in the causal nexus of design levels in the Game of Life. Such relata involve some noise tolerance and efficiency improvements over the bit-level description of the pixels. The modal boundaries of counterfactual robustness for a given relatum will be set by the identification criteria for the pattern, its noise tolerance, and the state of the causal nexus in that spatiotemporal area. Whether a given pattern occurs in a given spatiotemporal section of the causal nexus will be an entirely objective matter, once the description and noise tolerance are set.

The *counterfactual robustness zone* is given by the microstate that happened to occur plus the nearby portions in phase space, where the microstate differs from the actual one but still instantiates that same pattern with the same noise tolerance and efficiency. Information about counterfactual variation in individual tokenings in the causal nexus is encoded into the boundaries of a volume of points that is required to represent relata with any nonzero counterfactual robustness. Each time a pattern is instantiated in the causal nexus, there is, trivially, some exact microphysical state of the relevant portion of the causal nexus that is the

[28] This distinguishes counterfactual robustness from multiple realizability, which is a related but separate notion involving types rather than tokens. Similarly, counterfactual robustness is not merely the supervenience of a volume on the points that comprise it.

microstate for that pattern. However, that exact microstate is not itself adequate to fully represent the causal relatum in question, if that pattern has any degree of counterfactual robustness. There are multiple ways in which the exact microstate of the nexus could have been different without changing the relatum this defined. Each such other possible configuration of the causal nexus that is consistent with the original causal relatum needs to be included in order to adequately represent the pattern tokening in question. This involves a volume in phase space, one which includes but is not limited to the point representing the actual exact microstate.

The *causal profile* of a relatum is defined by the totality of causal interactions into which a tokened pattern enters, as cause or effect. The causal profile is to a large extent delineated by the pattern identification criteria and noise tolerance. All causal relations that an identified volume can enter into, considering both the trajectories that enter the volume and those that exit the volume, collectively constitutes this causal profile. The causal profile can be thought of like a 'fingerprint' that identifies a particular volume and distinguishes that volume from other volume individuations that may be very similar but not identical. Two volumes that mostly overlap will be distinguishable based on the total causal profile; even though they will share many of the same causal relations, they will not share all of them.

Changing the counterfactual robustness zone thus alters the causal profile; essentially, it picks out a different relatum. If we change the noise tolerance of a given pattern, so that greater error is allowed in tokenings, then the set of tokens of that given pattern thereby gets larger, but additionally, the causal profile of each token member of the original smaller set also changes. Each token would have a wider range of values for a given edge or node, yet still count as the same token. This expands the boundaries of the counterfactual robustness zone for each token, while also expanding the set of tokens that count as tokens of the pattern. Similarly, if we 'redescribe' the pattern using different criteria, it will be enormously hard, if not impossible, to pick out *exactly* the same set of tokens, because the redescription alters the modal boundaries of each token, thereby increasing the counterfactual robustness volume required to represent it. Thus, changing the way in which a given token is individuated will change its causal profile, even holding fixed the actual state of the nexus.

This has the significant consequence that no counterfactually robust pattern/causal relatum, including the modal characteristics that yield its causal profile, can be identical to any given microstate of the causal nexus. Instead, it is identical to that microstate *plus* the counterfactual buffer zone around it, encompassing trajectories which could have but did not

occur, but whose occurrence would not have altered the causal profile of the pattern thus tokened. The point plus the volume around it is required to represent a counterfactually robust relatum without altering the causal profile.

This is relevant for a wide range of other philosophical discussions, for instance Davidson's Cause-Law Thesis.[29] His view relies on the assumption that a single token can be redescribed indefinitely in a variety of different vocabularies while remaining identically the same token. If my claims here are correct, it will be the exception rather than the rule that a redescription in a different vocabulary can reproduce precisely the same volume. Such redescription will usually alter at least some part of the causal profile of the relatum, and thus, by Davidson's own characterization of causation as extensional, not pick out the same token.

Counterfactual robustness and using causal profiles to individuate relata also precludes the possibility of even formulating Kim's causal exclusion problem.[30] If we were to collapse each counterfactually robust pattern-token into just the microstate that happened to have occurred, we would be changing its causal profile dramatically - it would just be a different relatum with a different causal profile. Using a volume in phase space allows us to fully convey all the modal characteristics about what could have been different while retaining the same relatum. In order to have adequate expressive power to describe the richness of causal structure in the world, most causal relata must be represented with volumes rather than points in phase space.

This lays the ground work to connect the metaphysics to the interventionist semantic. Variables, in Woodward's account of interventionism, collect these tokens into sets by their identification criteria. Each pattern has a generic causal profile closely related to but not identical with the causal profile of any one of its instantiations. This generic variable causal profile is rather like an averaged version of the causal profiles of each token. It is by dint of the tokens having the causal profiles that they do that a variable has the causal profile it does. This illuminates the way by which variable **A** might genuinely cause variable **B**, even while a particular token of A did not cause any B; it is also compatible with some particular token C causing a particular token E, even though for variables, **C** does not cause **E**.

The contrastive character of interventionist variables, where variables may take value x rather than y or z, for instance, set the boundaries for the counterfactual robustness zone –

[29] Donald Davidson, "Laws and Cause," *Dialectica*, xlix, 2-4 (June 1995): 263-280.
[30] See Jaegwon Kim, *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation* (Cambridge, MA: The MIT Press, 2000).

taking value x needs to be distinguished from value y, but x itself need not be distinguished into any further fine-grained subvolumes. Put differently, the variables and values for the variables provide the groupings for points in the space into meaningful volumes. They show better and worse ways to 'chunk' that space based on its causal behaviour. The boundaries of the counterfactual robustness zone thus capture counterfactuals in a way that can be tied to contemporary interventionism.

There is also a point carried over from thermodynamics, one with deep significance for thinking about causation and determination, but which is currently rather under-theorized. It doesn't really matter, in the way one would naively expect of a determinate world, which exact microstate actually occurs. This is not merely an epistemic point about the inaccessibility of knowing which microstate any system is in at any given moment. It is an ontological point: many elements of the actual microstate for most causal relata are both part of that very token occurrence, while also being boundedly but genuinely causally irrelevant. The bounded causal irrelevance of elements in the instantiation of a causal relata is striking for the way in which higher level causal relata both are clearly instantiated by physical systems comprised of states of the microphysical nexus, while also giving a kind of autonomy from the details of those microphysical edges and nodes that are the tokenings of the pattern. The bounded causal irrelevance of all microphysical elements, taken together, yields the boundedness of the autonomy of the pattern token from the exact microstate of the nexus.

Thus, the pixel-level mechanistic causal processes support but do not exhaust the counterfactuals of arbitrarily higher-level causal patterns. Counterfactuals are not mysteriously emergent from those pixels of the causal nexus, but also cannot be simply reduced to those basic edges and nodes in the causal nexus. The modal features of causal relata, their counterfactual robustness, means that the lowest level of the causal nexus will not be a sufficient replacement for higher level causal profiles. The modal properties of higher-level relata is thus explicable in terms of the characteristics of the volume in phase space, as the parts of the microphysical causal nexus that actually instantiated given higher-level relata, plus the counterfactual robustness zones around those points defined by the range of other values that the relevant portions of the causal nexus could have taken to instantiate the very same higher-level relata. The way in which counterfactual robustness of pattern instantiation arises from pure pixel level processes in the causal nexus, even deterministic ones, unifies physical processes and counterfactuals into a single stereoscopic view.

IV. Putting it all together: the information-theoretic account of causation

These metaphysics plus representational tools provide the materials to support the interventionist semantics of Woodward and causal modelling approach of Spirtes, Glymour, and Scheines, and others. Information theory can provide a wealth of additional methodological tools for discovering and modelling causal structure, with only such that only one final step is then needed. In order to directly apply information theory to those volumes in phase space, they must be partitioned with a probability distribution over the partition. Information-theoretic tools can be directly applied to these probability distributions. This can then be used to calculate quantities such as mutual information between two volumes, mutual entropy, joint information, and more. These partitions can be made at finer or coarser grains, which allows modellers to find grainings that maximize or minimize informational connectedness between volumes. Quantities that measure other aspects of causation than relationship(s) between two individual volumes can also be used: the causal gradient of a whole region of the nexus can be measured, and the rate and/or 'flow' from one region of phase space to another. This is the real pay-off of the metaphysical view, where the rubber hits the road. The details laying out such applications and the methodological opportunities afforded by this approach, are the second part of this project, to be further explored in a future paper. Recall the core metaphysical view, now with additional emphasis: causation *is* a *set* of information-theoretic relations between patterns instantiated in the rich causal nexus. The pluralism of what has been characterized as different construals of causation (for instance, Hitchcock, "Of Humean Bondage") can be unified by defining different facets of causation using different informational measures.

An informal explanation may be helpful. In intuitive terms, partitioning a volume in phase space means dividing it into smaller subvolumes, to any degree of coarse or fine grainedness. The same volume can be divided into a large number of very small subvolumes, or a smaller number of larger subvolumes; the subvolumes can be divided so that they are all of equal volume, or so that the system is likely to spend an equal amount of time in each, or in a host of other ways. There isn't a single 'right' way to partition a volume. A choice of partition is largely driven by standard modelling considerations about the kind of system being modelled, the goal of the model, etc. There are ample guidelines for partitioning in statistical mechanics and thermodynamics.

Again informally, to put a probability distribution over the partition means we assign a probability to each subvolume in the partition such that the total probability for the whole volume is equal to 1. There are a lot of ways to partition and add probability distributions, but there will be nothing special about this – it is merely is a tool for representing various features of different specific systems, and already well-discussed in modelling literature. As a reminder, the volume over which we put the partition and distribution is the counterfactual robustness zone representing causal relata as patterns instantiated in the causal nexus.

With causal relata thus represented as appropriately delineated volumes in phase space, partitioned with a probability distribution, the material is ready for the application of information theoretic techniques. We can apply measures like Kullback-Leibler distance, or joint entropy, or mutual information, etc., between such volumes in phase space. Those equations are designed to be applied to probability distributions over partitions, and the work up to here was to find a way to put causation into such a form that informational quantities can be used to measure causal relationships. The informational relationships between the distributions are causal because the volumes themselves represent causal patterns in the nexus.

This approach addresses the shortcomings with Salmon's original view by providing a precise way to determine the right "grain size" for maximizing stability, proportionality, and specificity in representing complex, multi-level causal systems.[31] The level at which one describes a system corresponds quite closely to the grain of the partition over the counterfactual robustness volumes. By varying the grain of partitions over two candidate causal relata, we vary the amount of mutual information between them. We can use this fine-tuning to discover proportional and maximally specific causal relationships by looking for partitions that maximize the mutual information between causal relata. The level at which, for instance, mutual information is maximized is the 'right' level at which to describe the relata in order to effectively represent their degree of causal connectivity. Maximizing informational connectivity is a non-ad-hoc way to identify a privileged grain size for particular systems. With such a non-arbitrary choice of the level at which to characterize causal relata, we can then give very precise answers as to the stability of the causal relationship in question, and we can assess the specificity of our causal relata in terms of the value of the mutual information thus achieved. This opens up a huge new range of modelling opportunities, and allows for precisification and justification of

[31] See, for instance, James Woodward, "Causation in Biology: Stability, Specificity, and the Choice of Levels of Explanation," *Biology & Philosophy*, xxv, 3 (June 2010): 287-318.

existing work where such levels are already treated as privileged but without systematic justification for doing so beyond the fact that it works.

This sounds very modelling-oriented and rather un-metaphysical, but it is key to recognize how such methodological consequences follow directly from certain metaphysical claims that are 'baked into' the representational format being deployed. There are core metaphysical commitments about causation that have a sui generis role to play in modelling causal systems. These assumptions need not be specifically invoked during more focused or practical aspects of engaging in causal discovery and modelling. The metaphysical commitments are required for the whole view of causation to hang together coherently; this is part of the old-fashioned natural philosophical approach of this this paper.

These commitments are few but required to make the theoretical package work. They are shaped by a deeply pragmatist approach to causal metaphysics. First, there is the idea of a pattern. Which patterns we want to use will be a small-o ontological question; there are intriguing questions about how patterns must fit together in order to adequately cover a given range of phenomena, for instance, and to cohere and inferentially connect in the right sorts of ways. There is a capital-O Ontological commitment, though, to the idea of a pattern as what causal relata *are*. There may be multiple different ontologies of patterns developed to deploy in different circumstances, but still just one Ontology of patterns.

Second, there is the idea of a causal nexus. Talking about the causal nexus just is a way of talking about the actual world, namely, talking about its causal bits. There is some causal nexus, separating shadows from that which casts shadows; the question of the precise details of this nexus is an empirical one. But before we can answer that question, we must have a prior understanding of what it would take to be the *causal* nexus. We can recognize this commitment as metaphysical rather than empirical or ontic because we can easily consider alternative causal nexus options and because, while the question is open, we do have a good sense of what we are looking for to answer that question.[32] The idea of a causal nexus, thus, is not an empirical question tied to only our world, but is something we must already have at least largely in hand

[32] For instance, consider the ways in which we can recognize genuine causal nexi in fictional worlds. The Force is a recognizably law-based supplement to a fictional world's causal nexus that is otherwise similar to the one in our own world; Mrs. Weasley is bound by magical conservation laws to prepare food and vanish it, in order to conjure it back, since prepared food cannot be conjured from nowhere.

before being able to identify the fine-grained details of the causal nexus in our actual world. This is akin to what Hitchcock has called a Carnapian explication project.[33]

Finally, there is the metaphysical commitment that causal relationships are *informational* relationships in the nexus between causal patterns. This is not a small commitment; it implicitly involves the commitment to the existence of a range of informational relationships any of which may track different aspects of the way in which we deploy causally rich terminology.

Causal pluralism is thus understood in the array of different kinds of informational relationships that all fall under this treatment, in addition to the pluralism of pattern ontologies.

While I have emphasized that information theory can be applied to anything, and that the results are not thereby causal, it is also important to clarify that information itself, even applied to causation, is not itself some kind of extra physical quantity. Information is a tool for very precisely describing features of the world, but is not the features thereby described. It lacks intrinsic physical content. Thus, the claim here is not that the world itself just is information, nor is the claim that causation just is information, as if information were some kind of pure physical stuff out of which the world could be made. Timpson challenges the slogan that "Information is physical,"[34] and this paper is in line with his broad approach. Information itself is nothing, in the ontological sense; what *is* is the causal nexus and patterns instantiated in it, which are informationally structured, but where the information itself is a structure *of* something else, not a reified extra substance. To draw on a kind of Neo-Aristotelian analogy, the causal nexus is the substratum; patterns, including their informational connections, are the forms. Pattern-tokens in the nexus are broadly similar to primary substances, and patterns themselves, as definitional or conceptual objects, to secondary substances.

Even though I rely on Sider's notion of fundamentality,[35] the view of structure that comes out of this approach requires the categorical rejection of Sider's "knee-jerk realism," according to which there must be One and Only One Right Way to carve up the world. There can be multiple pattern ontologies, and no further answer about which one is the 'real' one. This account also thereby undermines the sharpness of Sider's distinction between substantive and conceptual disputes. As Dennett describes them, something counts as a real pattern if betting

[33] Hitchcock, Christopher, "Events and times: a case study in means-ends metaphysics," *Philosophical studies* 160, no. 1 (2012): 79-96.
[34] Christopher G. Timpson, *Quantum Information Theory and the Foundations of Quantum Mechanics* (New York: Oxford University Press, 2008).
[35] Sider, *Writing the Book of the World.*

on it over the long term results in winning over not betting at all. Analogously, causal patterns are real if we can intervene on them, even weakly, to change causally downstream patterns. Disputes about concepts are also disputes about which will help us 'win' in the long run, making the substantive and conceptual disputes two perspectives on the same question. There may be multiple sets of patterns that work for a given chunk of the nexus, with no further meaningful or non-ad-hoc answer as to which is the 'real' way to carve that part up. This approach thus shares Dennett's pragmatist orientation.

Thus, construing causal relata as patterns means that two claims, which are often taken to be in tension, hold. The first is that there is a very real distinction between relationships by which we can do things[36] and relationships by which we can make predictions but which cannot be used for intervention. The second is that there is no unique causal structure in any particular part of the world (or, in the world as a whole, but we'll focus on the more limited claim here). There really are better or worse ways to pick out patterns, and there are patterns that, as much as one might want them to be, simply are not instantiated in some given spatiotemporal section of the causal nexus. Nonuniqueness does not imply that anything works as well as anything else.

It is worth clarifying that there can be relationships between patterns that are not *causal* relationships; patterns themselves may stand in mathematical, or compositional, etc., relations. Once instantiated, the relationship between those patterns *in the causal nexus* is causal. There are counterfactuals one can evaluate regarding a variety of relationships between patterns, but they are not in and of themselves causal relationships. They are proto-causal, or causal in potentia. The causal nexus is thus labelled causal not because the 'pixel' level alone is genuinely causal. It is labelled causal because that which is causal is ultimately required to be instantiated, or minimally, instantiatable, in that nexus. Such relata are not exhausted by that instantiation – they may have modal features that never become actualized in the nexus, but which nevertheless shape their causal profiles. There can be patterns that are coherently defined such that it is an empirical question if they ever actually are instantiated. But they if they never are actually instantiated, they are causal relata in an attenuated way, causes 'in name only'. There is a clear requirement of actuality for causation, from which individual patterns may deviate, but which renders the entire account actual in character.

[36] See, among others, Nancy Cartwright, "Causal Laws and Effective Strategies," *Nous*, xiii, 4 (November 1979): 419-437, Woodward, *Making Things Happen*.

Further on this point, the lowest causal level might not be the lowest physical level; it is possible that there are lower physical levels that are non-causal. It is part of the metaphysical commitment to the very idea of a nexus that there is some lowest physical causal level analogous to pixels. The precise character of that nexus is left to physics, but cannot be left *only* to physics, in that extra-physics explication of what it is that a causal nexus could be is required.[37] It might be that the conserved quantity account of physical causation has enough empirical trouble that it is eventually discarded. Comparatively little depends on this. The key thing is that physics provides the material for this lowest physical causal nexus level, segueing the formerly philosophical question into a more tractable empirical question. Deferring to physics on the causal nexus, and updating our views of the precise nature of the causal nexus in light of developments in physics, can be accommodated with little to no change in many higher level patterns.

At this point, there is still an enormous amount of interesting philosophical work to be done, but it is no longer quite *metaphysical* work. It is methodological articulation and application work. The nature of the informational relationships that can exist between these volumes, and the range of causal relationships in causal systems across the sciences especially, can be investigated at all levels of abstraction, size, and/or organization independently of the smallest details of the nexus itself.

V. Laplace's Pattern

One objection is commonly raised at this stage. The broad concern is that, really, the causal work is still being done at the microphysical level. This concern has been raised about Woodward's account, as well, where it is relegated to being a convenient way to talk about higher-level variables but where the 'real' causal story remains microphysical. Here it takes the form of a worry that the pixels of the nexus are still the only 'really' causal part of the view. We can do clever things with patterns, but any causal 'oomph' displayed by a higher-level pattern just comes from the pixels of the rich causal nexus. I will call this general intuition appealed to

[37] The question of causation by connection versus disconnection does not map onto any meaningful distinctions within this account. Some causal relata will be connected in that way; some may not, but will still be causally related. That will depend on particular systems. The interesting questions relating to e.g. causal gradients, or divergence of a causal field in a particular part of the nexus, look muddier rather than clearer if we insist on the question "but are they connected or not?"

in this framework Laplace's Pattern. Laplace's Pattern just is keeping track of every conserved quantity in the nexus through propagation and exchange. It is the analogue to the bitmap of the pixellated Game of Life. The challenge can thus be put: if we have Laplace's Pattern, don't we thereby have all the causal oomph there is? What could be left for other patterns to do?

There are two ways in which this intuition goes wrong. One is that it constitutes a misunderstanding of the nature of a pattern ontology. The second is that it relies on an empirically inaccurate view of the very basic microphysical nexus itself. In both cases, the result of relying on Laplace's Pattern is that genuine causal structure of the world is dramatically underdescribed. Laplace's Pattern never says anything wrong about what causal structure there is - everything contained in that pattern is indeed part of the causal structure of the world. It never yields a false positive. Yet that pattern leaves out a great deal of genuine causal structure; the false negatives are monumental. Insofar as a theory needs have adequate resources to describe what there is, Laplace's Pattern is not adequate. I'll break this down into responses to the two ways the intuition goes astray.

The first response to show the inadequacy of Laplace's Pattern highlights the radicalness of pattern ontology. To treat the causal nexus as the only genuinely causal part of the story, and patterns as a mere way of keeping track of something, is to fail to have a pattern ontology. The lowest level of the nexus is not even the most causal, much less the only really causal, part of the world. Rather, it is degenerately causal, in the mathematical sense of degeneracy. A linear equation is a degenerate second order equation with a zero in front of the squared term. The lowest level in the causal nexus is a mathematically degenerate pattern. Just as pixels in the Game of Life have a degenerate pattern, that of the bit map, the 'pixel' level in the causal nexus is degenerately causal in that it counts as the most basic possible pattern. In both cases, this is a pattern. In both cases, it is the least efficient pattern, in that it involves the full bit map equivalent, and not the *only* pattern, since there are many others at design levels that can also be identified.

Not only is it not the only genuine pattern, it is also not privileged with respect to the other possible patterns. Anything that can be picked out as a pattern in the nexus has the same status as any other pattern. Any genuine pattern is equally real. This has the potentially counter-intuitive consequence that higher-level causes are just as real as lower-level causes, and that special kinds of causes such as intentions are, if they can be reliably picked out with a description and noise tolerance, just as legitimately causal as more straightforwardly scientific

ones. A surprising consequence of this commitment to pattern ontology is that it reorients our notion of fundamentalality from horizontal, with the smallest at the bottom and fundamentality decreasing as one goes up, to vertical, where fundamentality is more scale-free (and, thus, not exclusively microphysical). If one wants to select the smallest set of patterns such that any other pattern could be derived in some way from that set, or, if one wants to find the most fundamental patterns in the nexus, such a set will have to be vertically integrated. An incredibly common assumption about fundamentality, for instance in seeking it almost exclusively in quantum theories, is that the fundamental is horizontal, entirely at the smallest physical size scales and levels of organization. In a pattern ontology, the smallest set of patterns may need to include higher-level patterns that cannot be adequately derived from smaller size scale patterns. Just as there is a smallest world size that is yet large enough for a Turing machine to be buildable in it, there can be patterns that count as fundamental yet require at least a certain volume of phase space possibility to be instantiatable.[38]

The second response as to why Laplace's Pattern is not adequate highlights an overlooked and underappreciated part of Salmon's account, one he recognized but the consequences of which he did not fully explicate. It involves a breakdown in the analogy between pixels and the causal nexus. The bit map description in the Game of Life is also the uniquely most accurate for predicting future states, whereas Laplace's Pattern cannot be assumed to be the most accurate, let alone uniquely so, for the causal nexus. We often have a half-formulated intuition about the microphysical bits of the world that pictures them somewhere between miniature billiard balls colliding and tiny little pixels flashing on and off. But this is incorrect. The causal nexus has edges and nodes that are the propagation and exchange of conserved quantities, but the *nexus itself is not conserved*. There is no quantity "conserved quantities" that is meta-conserved. There are several equivalent ways to put this.

One way to put this point is that the nexus itself does not merely march forward relentlessly with all the little pixels in a row. The nexus itself can add or lose pixels in any given interaction. Each time there is a node where causal processes intersect and exchange conserved quantities, the *quantities* exchanged must be conserved, but the *nexus* itself might grow or shrink in terms of the number of edges it contains. Salmon discusses this possibility in *Causality and Explanation*, although the implications are somewhat obscured by his examples involving chickens and snakes. A standard causal interaction he labels an X, because there are two

[38] Life and rationality are two potential candidates for this.

processes that enter the interaction, and two (modified) processes that leave the interaction. But there are two further types of interactions. In a Y interaction, one process splits or fissions at an interaction-node into two processes. In other words, more processes leave the interaction than entered it. The third possibility he calls a λ (lambda) interaction: two causal processes enter the interaction node and one causal process leaves it. Conserved quantities are conserved across all these interactions, but the number of causal process-edges bearing those quantities can change at interaction-nodes.

This changes the bottom-level structure of the nexus, since the nexus is defined in terms of the edges and nodes. It's like your computer screen suddenly stretching out and adding more pixels in one area, and shrinking by losing pixels in another. It is not merely that the pixels change size relative to one another; in the causal nexus, it is that such 'pixels' can actually cease existing, or begin existing. It is not a 'flat' conserved web. The same amount of each conserved quantity can be distributed across ten edges, or across one. If one thinks of the little particles in deterministic rows like a miniature marching band, it turns out that band members, instead of merely moving across the field, also fission and fuse. Where there was one flute player there are now two piccolos; where two trumpet players collide there is now a single trombone. Conserved quantities are conserved, but the piccolos can now play two tunes whereas the flute could only play one.

More technically, it means that the total volume of phase space is not conserved over time. There can be more of it, or less, over time, no matter how widely we draw the boundaries for our system. Including the entirety of the universe in the bounds, there will be changes in the total volume of phase space representing the nexus over time. Liouville's theorem does not generally hold.[39]

The fact that conserved quantities are not themselves conserved opens up further measurement options as well. Call it *causal amplification* when a given unit of a conserved quantity is distributed across more edges leaving an interaction than entering the interaction. Call it *causal dampening* when a given unit is distributed across fewer edges leaving an interaction than entering. In causal amplification, there is more possibility space for causal structure: the volume of phase space that can be occupied is larger. In causal dampening, there is less room. We can then treat the causal nexus like a field of sorts, much like a wind map gives a vivid picture of air flow despite being comprised of point measurements of wind speed and direction. This can be used

[39] Thanks to James Mattingly for raising this issue.

to find gradients, and areas of divergence and convergence. A causal source is a region in the nexus where more edges exit than enter; a causal sink is a region where fewer edges exit than enter. All of this provides methodological traction on means by which to express extremely precise claims about causation in the manner expected in the sciences.

VI. Conclusion

Before information theory can be used to generate methodological resources for investigating and representing causal systems, it must be clear to what information theory would be applied. A recognizably philosophical account of causation can be put in a form where the right kinds of volumes with partitions and probability distributions are generated for the direct application of information theory. This accounts lays the foundation for claims about information-theoretic causal connections.

With this theoretical foundation for what causation *is* and how counterfactuals and physical causal processes are to be identified from the empirical sciences, there is a unification of different strands of thought in the philosophical discussions of causation. This provides the metaphysical basis for the Woodwardian semantics of interventionist counterfactuals. The view here has significant implications for ways to find causation in the world, for how best to model a variety of causal systems, and provides a foundation for the existing and deeply influential work on causal search.

On the other hand, this view also leaves much work to be done. This accoun does not yet uniquely determine the application of various mathematical tools of information theory to causation, and there will be several distinct, incompatible applications, such that further considerations need to be marshalled to choose the most appropriate. That is further work for a further paper, however, and will involve a great deal of detailed empirical analysis that is of a different character than the analysis performed here.

The remaining questions are largely modelling questions, not metaphysical ones: they are more like the projects in which scientists in various fields engage, rather than the kinds of projects in which philosophers engage. Subsequent development of the application will continue the already-well-begun process of handing over traditional philosophical material to other departments, especially to statistics, computer science, etc. It is not a fully un-philosophical matter. But this account opens up a lot of straightforwardly empirical questions that can be addressed using the epistemological methods of the sciences. It is for this reason that it warrants

being called natural philosophy, marking the beginning of the end of causation as a specifically philosophical rather than scientific discipline.

H.K. Andersen

Simon Fraser University

Acknowledgements: This work has been developed over a long period of time, and as such owes a great deal to feedback over the years. Much thanks to audiences at the University of Pittsburgh, University of Pennsylvania, 2016 Philosophy of Science Association meeting, University of Victoria, 2014 Causality and Complexity in the Sciences workshop, University of British Columbia, Carnegie Mellon University, and the 2013 Pacific Division American Philosophical Association meeting, and to my Philosophy of Science students and Explanation seminar students for their patience and feedback. Thanks for helpful feedback to Kathleen Akins, Frederick Eberhardt, Steve Esser, Chris Hitchcock, Kareem Khalifa, Stefan Lukits, Samantha Kleinberg, Liam Lazenby, Roberta Millstein, Alexander Reutlinger, Joel Smith, Hao Tang, Imran Thobani, Michael Weisberg, Cory Wright, and Jim Woodward. Thanks also to Jim Bogen and Sandra Mitchell for supervision and discussion on the early development of this work. Particular thanks to Kathleen Creel, in discussion with whom core features of this view were developed and who provided feedback on several early drafts. I am deeply grateful to Endre Begby for extensive discussion, detailed comments, and invaluable editorial assistance. This project was partially supported by a grant from the Social Sciences and Humanities Research Council of Canada, and took place on unceded Coast Salish territory.