Domains, Discourses and Non-Reducibility
A Structural Exploration After Davidson

(Winner of Chadwick Prize 2026)

Mike Arnautov

This essay begins with a very brief summary of Donald Davidson's anomalous monism and abstracts from it a domain-independent structural pattern. It examines parallel instantiations of this structure in other domains, using them to clarify the commitments of anomalous monism and to reconsider some standard objections. The aim is not to defend anomalous monism, nor to assess Davidson's arguments for anomalism as such, but to explore the explanatory structure his position brings into view – including how non-reducibility, which may have more than one structural source, can arise even in structurally related formal systems that would seem least hospitable to it. Once made explicit, that structure may throw light back on anomalous monism itself.

In his seminal essay 'Mental Events*'* (1970), outlining anomalous monism, Davidson sets out three premises:

1. Some mental events interact causally with physical events.
2. Events related as cause and effect fall under strict laws.[1]
3. There are no strict laws on the basis of which mental events can be predicted and explained.

Davidson argues that taken together his three premises do not lead to a contradiction but instead necessitate the identity of mental and physical events, thereby turning the relatively modest physicalism of the second premise into full monist physicalism.

The argument hinges on two key insights.

1. It is necessary to distinguish between events of the world as they are, independent of any description, and the linguistic resources we use to refer to such events. I shall use a variant of Davidson's own terminology for events as such: *events-in-extension*.
2. Any law that can be articulated (a) is necessarily articulated within a language and (b) must employ generalisations – whether properties, predicates, or other classificatory devices – if it is to be more than a mere (possibly infinite) list of special cases. Hence events can be said to fall under laws only under a description, using a particular vocabulary.

What is striking about Davidson's argument is how little of its structure depends on anything specific to the mental. Apart from the third premise, defended at length in his Section II, the argument makes no appeal to consciousness, subjectivity, or

[1] Davidson stipulates deterministic law before disclaiming the necessity for determinism in his footnote 3.

intentionality as such. Instead, it turns on a distinction between events-in-extension and the vocabularies used to describe them, together with constraints on the articulation of laws. This suggests that anomalous monism may be understood not merely as a doctrine about mind and body, but as an instance of a more general structural relationship between a single underlying domain and multiple descriptive practices.

For brevity, I shall use ‘discourse’ to refer to an extended mode of description or classification – spoken, written, formal, diagrammatic, or otherwise representational – that supports systematic description and explanation by introducing characteristic predicates (used broadly: any discourse-relative classificatory distinction), criteria of individuation, inferential norms, and explanatory aims. Nothing in this usage implies that a discourse constitutes its own ontology or determines what exists.

Our concern is with possible nontranslatability between discourses so characterised. Nontranslatability here is not merely linguistic: it extends to any principled mismatch between the distinctions made available by different modes of presentation, including non-linguistic ones such as visual or diagrammatic representation.

The structure implicit in Davidson’s argument can be characterised like this:

1. There is a base domain consisting of entities or events related to one another by principled transitions. These transitions may be formal, deterministic, or probabilistic, but they constitute the domain’s internal structure.
2. Two or more descriptive discourses organise this base domain by introducing predicates in accordance with their respective explanatory aims and constraints, with no presumption as to the nature or relationship of these discourses.
3. The predicates employed within a given discourse cannot be assumed, without argument, to correspond directly to properties of the events in the base domain; they function as discourse-relative ways of describing, classifying, and explaining those events.

The instantiations of this structure that will be of interest here also satisfy two additional conditions:

4. The fact that different discourses concern the same underlying tokens does not entail the coextensionality or nomological intertranslatability of their predicates.
5. At least one discourse is maximally aligned, predicate by predicate, with the principled transitions obtaining within the base domain.[2]

In Davidson’s case, the base domain consists of events-in-extension standing in causal relations. The relevant discourses are, on the one hand, physical description,

[2] Where such alignment is aspirational and the discourse is pragmatically largely successful – as physical science arguably is – it can become tempting to mistake its descriptive resources for a direct inventory of the base domain.

which aims to capture causal regularities under strict laws, and on the other hand, mental description, whose predicates are governed by normative and rational constraints. Davidson's anomalism concerns the impossibility of formulating strict laws connecting predicates drawn from these two discourses, while his monism secures their common ontological domain. The abstract structure described above simply makes explicit a pattern of reasoning not confined to the mental case as such. It is not meant to replace Davidson's own defence of the anomalism of the mental; it explores a possible structural source of discursive autonomy or anomaly quite separate from rational normativity.

The remainder of the essay considers other instantiations of this structure, noting what they show about discursive autonomy under conditions of formal or causal closure.

The structure outlined above has at least two well-known instantiations beyond the mental case. The first may appear unexpected. I begin with a familiar episode from the philosophy of mathematics. The reason for this choice is methodological: arithmetic provides a domain in which the relevant distinctions can be made with exceptional precision, free from empirical complication.

Consider two discourses: Peano Arithmetic (PA) and its familiar meta-theory. Both concern the same arithmetical sentences about the natural numbers, but they apply different predicates to them. PA, which is an axiomatisation of arithmetic, introduces a syntactic predicate of provability, whereas the meta-theory applies a semantic predicate of truth.[3] These two predicates (provability in PA and truth in meta-PA) are so tightly correlated that it was long assumed that they must coincide as a matter of principle. Indeed, the very point of formalising arithmetic was often taken to be the capture of arithmetical truth by purely syntactic means. However, Gödel's incompleteness theorem shows that this expectation cannot be fulfilled: for any sufficiently strong, consistent formal system such as PA, there are arithmetical sentences that are true but not provable within the system. Gödel's 'sentence G' is true yet unprovable.

The significance of this example lies in the extremity of the conditions under which non-coextensionality arises. PA and its meta-theory are related by a precise and systematic correspondence: provability in PA is arithmetically definable, and the standard soundness results establish a close alignment between syntactic derivability and semantic truth. Indeed, for vast classes of arithmetical sentences the two predicates do coincide. Yet Gödel's result shows that this alignment cannot be global. Even in this minimalist, completely regimented setting – formal, exact, and free of empirical complication – the identity of domain and the strength of the correspondence do not force nomological intertranslatability between the two discourses.

---

[3] In the standard model of arithmetic. There are other models of arithmetic, but PA's standard model – the natural numbers with their usual properties – is what is relevant here.

This gap is not epistemic. The axiomatic character of PA provides a precise formal specification of what counts as a consequence of its axioms, yet the principled divergence between provability and truth remains. Access to the full formal description does not eliminate the autonomy of the semantic predicate.

Analogous features can also be observed in a very different domain, namely in simple rule-based dynamic systems, which instantiate the same basic structure while bringing additional features into view.

The instance in question is Conway's Game of Life (GoL). In brief, GoL is a discrete-time cellular automaton defined on a two-dimensions square grid/lattice of cells, each of which is either 'live' or 'dead' at any given time step. Each cell has eight neighbours – four orthogonal and four diagonal ones – and the evolution of the grid is governed by two simple, purely local rules:

1. A live cell at time T remains alive at time T+1 if and only if it has exactly two or three live neighbours at time T.

2. A dead cell at time T becomes live at time T+1 if and only if it has exactly three live neighbours at time T.

These rules are strictly local and purely numerical. They depend only on the number of live neighbours and are entirely agnostic as to which neighbouring cells contribute to that count. They have no resources for referring to distances beyond immediate adjacency at time T, or to direction. Nor do they retain any memory of cell states at any time prior to T.

Here the base domain consists of a necessarily finite actual implementation of the Game of Life, whether pencil-and-paper or a computer program with a visual display. The cell discourse – positional descriptions of live and dead cells, with the GoL rules as its law component – is perfectly base-aligned: it fully specifies how any such implementation must evolve from one time step to the next. It may be tempting to see the GoL rules as constituting the base domain, but the rules are merely a description of how any implementation must behave. The causal work takes place in the implementation.

Despite the extreme simplicity of the rules, they support a rich taxonomy of stable and mobile configurations referred to as patterns. Of these, only one need concern us here: the glider – a configuration of five live cells which recurs every four time steps, displaced one cell diagonally. The glider therefore exhibits persistent, directed motion across the grid. This behaviour has been widely exploited in Game of Life 'engineering projects' , including the construction of logic gates, memory devices, and ultimately a functioning Turing machine realised entirely as a GoL pattern.[4] Yet even granted a state vocabulary in which the glider configuration is definable, its

[4] See e.g. Wikipedia, *'Conway's Game of Life', version date 13 May 2026:* https://en.wikipedia.org/w/index.php?title=Conway%27s_Game_of_Life&oldid=1354003362.

identity over time is not: at cell level nothing moves – five cells die, five others are born.

The rules lack the expressive resources required to classify that recurring configuration as a persistent mobile object. This limitation, again, is principled rather than epistemic. Unlike the arithmetical case, where the predicates of the two discourses were largely coextensive, the cell discourse cannot do the classificatory work of 'is a glider': it yields only a disjunction of static templates, not stages of a single persisting object.

Gliders therefore cast a particularly clear light on what is often taken to be the most counter-intuitive aspect of anomalous monism: how can elements of the base domain, described under one discourse, be considered to be causally efficacious without constituting a separate causal chain and thereby vindicating Jaegwon Kim's causal exclusion argument (1989)? The first level answer, given by Davidson himself, is that what is causally efficacious are events *simpliciter*, events of the GoL instantiation (1993). There is just one causal chain and two distinct descriptions of it. Hence Kim's revised, explanatory variant of the causal exclusion argument (1993). But the GoL example demonstrates an interesting additional factor.

The patterns vocabulary does not and cannot offer a rival causal chain, but it offers the only means of referring to perfectly real organisations of the base domain as persisting, re-identifiable objects. The cell discourse specifies states and transitions cell by cell, but possesses no transtemporal identity for configurations: it cannot say that the same pattern has moved.

We could, of course, enrich the cell discourse so that 'glider' figures as one of its kinds, but doing so would be question-begging – an enrichment motivated only by the explanatory aims of the pattern discourse. To dismiss gliders as mere epiphenomena is therefore dialectically hollow. It simply forfeits the only vocabulary capable of describing how the implementation's causal operations organise into the recurrent structures that actually make the system's pattern-level behaviour tractable and explicable.

The pattern discourse provides indispensable vocabulary for tracking causal processes, yet involves no downward causation. The systemic constraints of particular cell configurations do not modify the rules or introduce new causal powers; they merely fix how those rules operate in a given case. Nothing mysterious is added, and nothing at the base level is violated. Nevertheless, in GoL explanatory privilege belongs to the discourse capturing the regularities of interest, not the one that is most closely – in this case perfectly – aligned with base domain regularities.

This represents a striking inversion of a standard reductionist assumption. In the present case, the causal closure of the base domain not only permits the explanatory role of the pattern discourse – given the expressive limitations of the

rules level, it demands it. There is simply no other level at which the regularity of the glider's behaviour can be referred to at all.

This also exposes a limitation in a common background assumption of explanatory exclusion arguments. The argument assumes that the discourse most closely aligned with the base domain – physical discourse in the case of anomalous monism, or the cell discourse in the case of GoL – is necessarily explanatorily privileged, and any other discourse therefore reducible to it. GoL shows this assumption cannot be taken for granted. Even where the base domain is causally closed and the base-aligned discourse descriptively complete, it may lack the classificatory resources required for explanation, forcing appeal to a higher-level discourse instead. The explanatory exclusion argument therefore cannot simply be applied to anomalous monism; its central privileging assumption itself requires justification.

There is one further lesson to be drawn from this instance of the general structure. As we have seen, the pattern discourse in GoL is clearly autonomous. It also supports genuine regularities of its own: the behaviour of gliders and of more complex 'spaceship' patterns is highly predictable and the well-known 'speed of light' limit on their motion is easily established.

However, because GoL can instantiate a Turing machine, it is capable of universal computation. Although any particular computation occupies only a finite region, any complete account of its pattern-level behaviour would therefore have to become an open-ended catalogue of special cases. In any case, pattern types, unlike their implementations, are grid-independent, so the pattern taxonomy is unbounded by its very definition. Thus, on Davidson's conception of lawhood, where a list of special cases does not amount to a law, GoL's pattern discourse is more than merely autonomous, though not fully anomalous. "Weakly anomalous" is therefore an appropriate characterisation.

Looking back at Davidson's justification of the third premise in Section II of 'Mental Events' , we can now observe that it relies entirely on global features of the mental discourse: its holism, its constitutive normativity, and it being subject to standards of rational interpretation. Davidson does not address the possibility of more tightly scoped local lawhood connecting mental and physical descriptions. The GoL example shows that this omission is not a weakness.

Even if the contentious argument of Section II establishes only weak anomalism of the mental, this strengthens Davidson's overall schema rather than undermining it. Weak anomalism is entirely compatible with the existence of locally law-like correlations, and thus protects anomalous monism against the prospect of any future neurobiological discoveries establishing limited regularities between mental and physical descriptions.

Once the abstract structure implicit in anomalous monism is made explicit, the structural intelligibility of mental anomaly follows fairly directly. Where multiple discourses organise a common base domain, and where predicate/property correspondence is not to be assumed, there is no antecedent reason to expect global nomological alignment between them. Davidson's own ambition was, of course, stronger: to show that the mental is in fact anomalous. The present discussion does not attempt to establish that conclusion, but aims to clarify how and why anomalism is structurally intelligible within the framework he proposes.

The examples considered here show that non-reducibility can arise even in systems that appear least hospitable to it: formal, fully specified, and causally closed.[5] Even a discourse whose predicates track a domain's underlying structure as closely as is in principle possible, need not possess the expressive resources required to recover every genuine regularity that other coherent discourses over that domain can articulate. Thus familiar exclusionist worries need not arise within this framework. This does not entail that the same form of non-reducibility applies to the mental, but it does establish such a structural possibility.

The arithmetical and GoL cases occupy opposite extremes within the space of such discursive structures. In the Gödel case, the two discourses are exceptionally closely aligned, so that their principled divergence appears only at the limit. By contrast, in the GoL case the cell discourse and the pattern discourse lack any comparable predicate alignment. Neither case is offered as a direct analogue of anomalous monism. If Davidson is right, the mental case, as an instantiation of the same structure, would lie somewhere between these extremes.

Taken together, these considerations suggest that non-reductive physicalism need not be treated as a last resort, but may instead be regarded as a coherent and viable position with the clear advantage of avoiding the ontological inflation of dualism without imposing the explanatory austerity of reductionism. Any nontranslatability of phenomenal, mental, or rationalising descriptions into physical ones may be an inescapable limitation of descriptive practice, not a fact of ontological multiplicity.

---

[5] It should be noted that comparable structures also arise in more complex settings, such as the relationship between hardware and software, the role of effective theories in physics, or the non-uniqueness of smooth structures on a common topological manifold. Pursuing such cases lies beyond the scope of the present essay.

Bibliography

Davidson, Donald. ‘Mental Events.’ In *Experience and Theory*, edited by Lawrence Foster and J. W. Swanson, 79–101. Amherst: University of Massachusetts Press, 1970.

Davidson, Donald. ‘Radical Interpretation.’ *Dialectica* 27 (1973): 313–328.

Davidson, Donald. ‘Thinking Causes.’ In *Mental Causation*, edited by John Heil and Alfred Mele, 3–17. Oxford: Clarendon Press, 1993.

Gardner, Martin. ‘Mathematical Games: The Fantastic Combinations of John Conway’s New Solitaire Game ‘Life’.’ *Scientific American*, October 1970.

Kim, Jaegwon. ‘The Myth of Nonreductive Materialism.’ *Proceedings and Addresses of the American Philosophical Association* 63, no. 3 (1989): 31–47.

Kim, Jaegwon. ‘Explanatory Exclusion and the Problem of Mental Causation.’ In *Supervenience and Mind: Selected Philosophical Essays*, 239–260. Cambridge: Cambridge University Press, 1993.

Nagel, Ernest, and James R. Newman. *Gödel’s Proof*. Revised edition. New York: New York University Press, 2001.