Implication as inclusion and the causal asymmetry

by Daniel Saudek

April 2021

Abstract: How does causation in the physical world relate to implication in logic? This article presents implication as fundamentally a relation of inclusion between propositions. Given this, it is argued that an event cannot "causally imply" another, also given the laws of nature. Then, by applying the notion of inclusion to physical objects, a relation "possible with respect to" is developed, which generates a partial order on sets of such objects and is independent of time. Based on this, it is shown that changes of physical objects in time (at any rate, a great many of them) imply, and thus counterfactually depend on, what we call "causes"—an asymmetric dependence which is robust despite the perspectival nature of the concept of "cause".

Keywords: implication, causal asymmetry, inclusion, counterfactual dependence, perspectivalism

That there is an analogy between logical implication and the relation between cause and effect has been noticed since antiquity. Despite this, the precise relationship between the two has never, I think, been fully understood. It is clear that material implication does not adequately capture the cause-effect relation, since it holds also between causally unrelated propositions. The same is true for strict implication, understood as the combination of material implication with the necessity operator, which holds, for example, between the propositions "2+2 = 5" and "it is raining" (whether or not it really is raining). A widespread and influential idea, however, is that it is natural law which guarantees that effects are implied by their causes. This idea—which in the following will be called "natural law implication"—has been spelt out, in various ways, by McTaggart (1915), Lewis (1973, esp. 563-4 and 559), Weizsäcker (2002, 86-88), and in the deductive-nomological model of Hempel and Oppenheim (1948), to name but a few. It seems to me to be one of the odd quirks in the history of philosophy that natural law implication has often been used to explicate the logically *converse* notion of counterfactual dependence of an effect on its cause (e.g. Lewis 1973, 560-1), a confusion present already in Hume's thought (cf. the remark in Menzies and Beebee 2019, section 1).

In what follows, I will argue that a theoretically and empirically adequate account of the relationship between causation and implication can be given by viewing implication as a relation of inclusion. That is, the simple notion of being "in" will serve as the needed bridge between logic and causation (on this notion, cf. Strumia 2012). To that end, I will first introduce this view of implication. This will then be used to investigate whether an event can "causally imply" another, and in particular, whether natural law implication is tenable. Finally, it will be asked whether, and in what sense, events can be said to depend counterfactually on causes.

Some notes on terminology before we move on:

I will use collections, denoted by square brackets *[...]*, rather than sets. Collections work like sets, with two important differences: 1. Only a collection of two or more entities is an entity in its own right, whereas a singleton collection is identical to its constituent, i.e. $[x] = x$, for any x. 2. There is a relation "included in" which, unlike the set-theoretical "element of", is transitive. Thus, given, say, the collection $M := [x, [y, z]]$, both x and $[y, z]$ are included in M, as also are y and z, whereas $[x, y]$ is not. In addition, given a collection C, a "sub-collection" will be a collection C' of objects which are included in C. For example, $[x, y]$ is a sub-collection of M.

The notions of collection and inclusion can be extended, in particular, to propositions: By point (1), any atomic proposition is included in itself. In addition, any conjunction $Q := q_1 \wedge \ldots \wedge q_n$ is a collection of propositions, which can therefore also be written as $[q_1, \ldots, q_n]$.

Finally, let x be a concrete particular, such as a physical object, or an event. Then, "$\underline{x}$" will denote "x exists". This device will allow switching easily from the level of concrete particulars to that of propositions in order to use the tools of propositional logic. For example, "x implies y" is not a well-formed expression (concrete particulars cannot imply one another), but we can write "$\underline{x}$ implies $\underline{y}$".

1. Implication as inclusion

There is an obvious connection between inclusion and implication via the notion of sub-collections (or alternatively, subsets): for any sub-collection G of a collection F, and for any x, if x is in G, then it is in F. I will go beyond this and assume that implication *just is* a relation of inclusion between propositions. This leads to the fundamental principle:

FP: Any proposition implies all and only the propositions included in it.

This extremely simple interpretation of implication is consonant with its etymological sense of "enfolding". I will use the symbol "→" for implication thus understood. The advantage of this interpretation is that no unrelated statements are connected by →, a motif shared by relevance logic (see e.g. Andersen and Belnap 1975, ch. 1, §3; Ferguson 2017, ch. 1), and connexive logic (see e.g. McCall 1966). But doesn't FP limit implication to abstract relations of inclusion between propositions, thereby making it too strong to account for causal connections? Before addressing this point (in sections 2 and 3), I will first spell out what this interpretation means, without however attempting a complete characterization of it:

First, any atomic proposition implies only itself. Any conjunction $Q := q_1 \wedge \ldots \wedge q_m$ (where the q_i's are atomic propositions) will be taken to be a proposition only if it is non-contradictory, i.e. for all i and j, $q_i \neq \neg q_j$, since I assume (for broadly Aristotelian reasons) that otherwise it is not only analytically false, but also meaningless. Any composite proposition of the form $P = p_1 \wedge \ldots \wedge p_n$ implies only and all of the 2^n combinations of p_i's. Thus, "I know it is raining" → "it is raining", because the former is a composite proposition including the latter, plus certain epistemic conditions. Similarly, for a composite object $S = [s_1, \ldots, s_n]$, where the s_i's are constituents, $\underline{S} \rightarrow \underline{s_i}$, for any i.

It is straightforwardly seen that "→" has the following properties: A truth never leads to a falsehood; a false composite proposition can, but need not, imply a true proposition; what is known as Aristotle's thesis $\neg(\neg p \rightarrow p)$ is satisfied; so also is Boethius' thesis, i.e. given $(p \rightarrow q)$, it is impossible that $(p \rightarrow \neg q)$.[1] We can, in addition, add to our inventory a "blocking-off" operator |...| defined by the property that, for any p, $|p|$ means "only p is true", that is, p is then taken to represent the "totality of facts" (Wittgenstein 1995, 1.1) which obtain. Then, analytically (and trivially), $|p| \rightarrow \neg q$ for any q distinct from p, where "distinct from" means "neither identical with, nor included in".

Second, despite what has been said so far, we are very often justified in making claims of the form "$p \rightarrow q$" also for q distinct from p. The reason is simply that we can assume background knowledge which need not be stated explicitly and which together with p implies q. For example, in the claim "you can't live on Mars, there's no oxygen there", it is assumed that some living beings need oxygen, and that "you" is such a being. Thus, FP is not violated by claims of this type.

Third, "→" admits addition: for any P and any q distinct from P, if p is a proposition included in P, $P \rightarrow p \vee q$. This would at first sight seem to fly in the face of what has been said so far, and it has been objected that such explosion of the content of a proposition leads to any two propositions sharing some content (see e.g. Gemes 1994; Ferguson 2017, 4-8; cf. Stepanov 2004, 1). To this, I answer that addition must be admitted, simply because $p \vee q$ is of course $\neg(\neg p \wedge \neg q)$, which is satisfied given P. Nor does this lead to an undesirable explosion: only $p \vee q$, but not q, is implied by P, so that we cannot deduce from the truth of P whether q is true or false. Thus, while the closure of P under "$\vee$" explodes, the truth-values deducible from P are well-behaved. The symbols "$\vee q$" therefore do not add content to p any more than the addition of "+ 0" changes an algebraic expression (all that "$\vee q$" does is to produce in the hearer a mental representation of q's subject matter, but it asserts nothing whatsoever about that subject matter). But if there is no extra content, it is legitimate to view $p \vee q$ as included in P.

Fourth, let F be a variable denoting a real intrinsic property, I a collection containing indices for these properties written as i, and J a sub-collection of I whose members are written as j. Then, if $S := [x{:}\cap_i F_i x]$ and $S' := [y{:}\cap_j F_j y]$, for all z, the proposition $\cap_i F_i z$, by which z is in S, includes the proposition $\cap_j F_j z$, by which z is in S'. Thus, for example, "Fred is a thrush" includes "Fred is a bird", since "thrush" includes "bird" in its definition. In this way, a relation of inclusion between property collections translates into a relation of inclusion, and hence of implication, between propositions.

2. Is sufficient causation possible?

According to sufficient causation, there are distinct concrete particulars c and e—conceived of as "cause" and "effect", respectively—such that $\underline{c} \Rightarrow \underline{e}$, where $\Rightarrow$ denotes *some* type of implication. c and e may be two events, or states of affairs, or objects. The obvious problem with this picture is that prevention of e can never be ruled out, in which case we get $\underline{c} \wedge \neg\underline{e}$, thereby ruining the implication, whatever its type. Simple as this point is, it is in my view still often not enough appreciated how far-reaching it is (cf. Anscombe 1971, 147). For example,

[1] By contrast, neither thesis holds if contradictory antecedents are allowed, as is easily seen. Cf. on this point Priest (2008, 178-9).

even death as the consequence of beheading—seemingly an unproblematic case of the type "$\underline{c} \Rightarrow \underline{e}$" (cf. McTaggart 1915, *passim*; Hume 2000, VIII, 1, 19)—can in principle be prevented (cf. Shewmon 2007).

Behind the problem of prevention in the empirical world is, I submit, a logical gap: c and e are supposed to be distinct particulars. But if implication is inclusion, it cannot be the case that $\underline{c}$ implies $\underline{e}$, because it does not include $\underline{e}$ (cf. Wittgenstein 1995, 5.135–5.1361). This is true also if c is, à la Mackie (1974), an entire causal condition, rather than just a little local particular. Similar reasoning also shows why natural law implication is, strictly speaking, impossible: Let $\mathcal{L}$ be a true conjunction including laws of nature, either currently known ones (such as Maxwell's equations, or the rules of quantum chromodynamics), or even "ultimate" laws of nature, if such there be. Natural law implication would then give $\underline{c} \wedge \mathcal{L} \Rightarrow \underline{e}$. But $\mathcal{L}$ does not include $\underline{e}$ any more than $\underline{c}$ does, and so the left-hand side cannot imply the right-hand side. Natural law therefore cannot fill the gap between distinct concrete particulars.

3. Counterfactual dependence

Events are changes of entities as their substrate. They are not isolated happenings, but rather happen "to something". This, at any rate, applies either to all events, or at least to a great many of them (cf. Esfeld 2011, 35). Insofar as it does, events counterfactually depend on something in some sense "other" than their substrate, a dependence which will now be derived on the basis of implication as inclusion. I will limit myself to doing this for relatively macroscopic physical objects with classical, not quantum-mechanical, identity conditions (cf. Lowe 2003, 78)—say, bacteria, rocks, or galaxies—in order to arrive at a simple model.

Consider, first, an entity a whose identity criteria are given by the collection $[c_1, \ldots ,c_n]$: any entity x is a if and only if it satisfies each c_i. The c_i's can refer, for example, to material constituents, sortal criteria (e.g. "is a mammal"), or qualitative properties. Of course, it is notoriously difficult to specify *which* criteria apply for a given entity. But for what follows, it matters only that *some* apply—an assumption which must be made in both everyday life and scientific practice, since otherwise it would not be possible to describe the evolution of self-identical objects in spacetime. We can now add to a specifications $d_1, \ldots ,d_m$, obtaining $[c_1, \ldots ,c_n , d_1, \ldots ,d_m]$, where the d_j's again refer to intrinsic properties of some sort (e.g. "is green") or to constituents (e.g. "contains iron"). Entity $[c_1, \ldots ,c_n , d_1, \ldots ,d_m]$ satisfies the identity criteria for a. Thus, a comes with a basic modality, that is, a potential to be in different states, where a "state" of a is any x satisfying $[c_1, \ldots ,c_n]$, with or without some specification. This basic modality is simply a consequence of a's definition. Given this, it is easy to see that the Leibniz principle holds—indiscernibles are identical—but not its converse.

Consider now the simple case of a physical object a with the following identity criterion: any x is a if and only if it includes $s_1 , \ldots , s_n$, where the s_i's are material constituents. Suppose now that two states of a exist: the collections $[s_1 , \ldots , s_n]$ and $[s_1 , \ldots , s_n, t]$, for some material constituent t. No time order of these two states is assumed, that is, the former may exist before the latter, or vice versa. The existence of two states of an entity, independently of their time order, will in what follows be called a "proto-change", whereas a "change" has a definite time order. We now obtain the following simple syllogism, where again underlining symbolizes "exists":

1. $\underline{[s_1 , ... , s_n]}$.
2. $\underline{[s_1 , ... , s_n, t]}$.
3. $\underline{t}$. (from 2, using “→”)
4. $|\underline{[s_1 , ... , s_n]}| \rightarrow \neg\underline{t}$. (using the “blocking-off” operator)
5. $\neg\, |\underline{[s_1 , ... , s_n]}|$. (3 and 4, modus tollens)

In step (4), if we augment the left-hand side by any entity u which does not include t, writing $|\underline{[s_1 , ... , s_n]}, \underline{u}|$, we likewise get $\neg\underline{t}$. Therefore, given (1) and (2), there must be a collection T including t, whether it does so properly or improperly. In the latter case, $T = [t] = t$. T cannot always be included in a, since otherwise, a’s state $[s_1 , ... , s_n]$ could not exist, and must therefore, at least in some of its states, be distinct from a. But given this, we can now extend the meaning of “→” itself, and write: $\underline{[s_1 , ... , s_n]} \wedge \underline{[s_1 , ... , s_n, t]} \rightarrow \underline{T}$: t is included in $T \wedge T \neq a$.

I will call T an “aition” of t. το αἴτιον is the Greek word for “cause”, but in its etymological sense, an αἴτιον of x is simply something which has x as its part (αἴσα) (Gemoll and Vretska, 2006, “αἴτιος”). It is in this latter sense that I use “aition” here and in what follows. In the case considered above, the proto-change of an object depends counterfactually on a suitable aition.

In which other cases is there similar counterfactual dependence? To explore this, I will first define: An entity y will be called “possible with respect to” an entity x if $\neg(|\underline{x}| \rightarrow \neg\underline{y})$, and “impossible with respect to” x otherwise. The condition $\neg(|\underline{x}| \rightarrow \neg\underline{y})$, in turn, will hold if and only if y can be obtained given that, and only that, which is included in x. Thus, in the above example, $[s_1 , ... , s_n, t]$ is impossible with respect to $[s_1 , ... , s_n]$, but not conversely. When two states of an object exist such that at least one is impossible with respect to the other, this depends counterfactually on an aition of that which *per se* differentiates the two states. Given this, some proto-changes depend counterfactually on something distinct from the object in question, and some do not. Consider the following examples:

Proto-change of an object in its rest mass, total energy, and momentum depends counterfactually on a corresponding aition distinct from the object. This is because these three are real quantities, and on the Dedekind construction, any two real numbers x and y are sets such that, if $x > y$, then any element of y is also one of x, but not conversely (see e.g. Holmes 2012, 94-96). Thus, on our terminology, everything in y is included in x, but not conversely. Then, given an object a having two states with rest masses m^* and $m^{\sim}$ such that $m^* > m^{\sim}$, $|\underline{m^{\sim}}| \rightarrow \neg\underline{m^*}$, but not conversely, so that m^* is not possible with respect to $m^{\sim}$, whereas $m^{\sim}$ is possible with respect to m^*. This proto-change therefore implies that there is a distinct aition of the mass difference. And analogously for a’s total energy in a given frame of reference. As for momentum, if a has two momentum states p^* and $p^{\sim}$ such that in one frame of reference $p^* > p^{\sim}$, there is always a coordinate transformation which reverses the inequality. Hence, *each* state can legitimately be viewed as “greater than” the other, so that now, *neither* momentum state is possible with respect to the other. The existence of the two states therefore implies an aition of the momentum difference.

On the other hand, consider an object composed of several sub-objects, such as $A := [U, V, W]$. Here, let U be in a state $u := [u_1, \ldots, u_l]$, V in a state $v := [v_1, \ldots, v_m]$, and W in a state $w := [w_1, \ldots, w_n]$, where the indexed objects are material constituents and l, m, and n are natural numbers. Consider, now, any entity obtained by recombination of constituents included in A, such as $u' := [u_1, \ldots, u_m, v_k]$, where $1 \leq k \leq m$. We find that $\neg(|\underline{A}|\rightarrow\neg\underline{u'})$, so that u' is possible with respect to A. Of course, u' is *not* possible with respect to u. Thus, the existence of u' depends counterfactually on an object distinct from U, but not on an object distinct from A. But *A-with-u* and *A-with-u'* are different states of A. Thus, for a composite object, not all proto-changes depend counterfactually on something distinct from it. This means, in particular, that an object can have two qualitatively different states which are possible with respect to each other. After all, the recombination of A's constituents can result in a difference in qualities such as colour, opacity, or conductivity. Hence also, quantitative measures of such qualities may differ for two states which are possible with respect to each other. For example, *A-with-u* and *A-with-u'* may have different electrical conductivities. The case of such quantitative measures is therefore different from those of mass, total energy, and momentum considered previously.

Also, two entities may be impossible with respect to each other. For example, this is true of $[s_1, \ldots, s_n, t]$ and $[s_1, \ldots, s_n, u]$ (where $t \neq u$), since each contains a consituent which is not in the other.

In sum, for any two non-identical entities x and y, y may be possible with respect to x, or vice versa, or both, or neither. The relation "possible with respect to" therefore generates a partial order on any set of entities. To illustrate: a complex carbon chain is possible with respect to the early Solar System 5 billion years ago (the former can be obtained from what is in the latter, with no momenta or energy from outside the Solar System requried), but not vice versa. The Milky Way and Andromeda galaxies are not possible with respect to each other (given only one, you cannot obtain the other). A cat rearranging its limbs as it falls in mid-air has different states which are possible with respect to each other.

We can now analyse the change of a physical object A in time: as with proto-change, A has two states, but in addition, there is now a distinction between an earlier and a later state. Also, let "cause of x" have its "naïve", everyday meaning: something "from which" x originates as its "source" (cf. Suárez 1965, XII, 2, 4). Do changes require causes, just as proto-changes require suitable aitia?

From the above considerations, given a change of A there is some a included properly or improperly in A having two states a^* and $a^\sim$ such that at least one is not possible with respect to the other. Let $a^\sim$ be that state. We now get two cases: 1. a^* is before $a^\sim$, as is the case, for example, when a acquires a constituent c. The existence of the two states depends counterfactually on an aition of c. Given the assumed time order, this aition is one from which a acquires c, and which can therefore be viewed as a cause of a's change. 2. $a^\sim$ is before a^*, as occurs when a loses c. Then, the aition of c acquires c from a, and cannot be identified with a cause of a's change. However, in this case, there is now a sub-object of a itself which acquires momentum in its own rest frame. But this depends counterfactually on the existence of an aition of this momentum from which the sub-object acquires it, i.e. on a cause of it.

These examples illustrate how the notion of “cause” is linked, via the “from-to” distinction, to our temporal perspective (cf. Price 1996; Price 2007): reverse the direction of time and, at least in many cases, also the cause-effect relation will be reversed. However, it was seen in cases 1 and 2 that the change of *a* does indeed depend counterfactually on what we would call a “cause” in everyday language. This counterfactual dependence is therefore not a perspectival effect. Thus, we are justified in concluding, at least for the cases considered, that the change of a physical object implies another object as its cause. Hence, events, *qua* such changes, imply causes.

4. Conclusion

Implication as inclusion can, given the above, be used to understand not only relations between propositions and collections *in abstracto*, but also causation in the physical world: one event cannot “causally imply” another, but since there is counterfactual dependence of proto-changes on suitable aitia independently of time, there is also, in time, one of changes on causes. Causal asymmetry is therefore, I submit, much simpler than is often thought: we do not need to take a detour via global laws or regularities (Humeanism), or even regularities over sets of possible worlds (Lewisian counterfactual theories) in order to account for it. We need only to look, locally, at what is included in concrete objects, and what is not.

Bibliography

Andersen, A. R., Belnap, N. D. (1975). *Entailment: The Logic of Relevance and Necessity*, vol. I. Princeton and London: Princeton University Press.

Anscombe, G. E. M. (1971). Causality and Determination (inaugural lecture at Cambridge University). Reprinted in M. Tooley (ed., 1999), *Laws of Nature, Causation, and Supervenience* (pp. 283-297). New York and London: Garland Publishing.

Esfeld, M. (2011). *Einführung in die Naturphilosophie,* 2nd edn. Darmstadt: Wissenschaftliche Buchgesellschaft.

Ferguson, T. M. (2017). *The Proscriptive Principle and Logics of Analytic Implication.* CUNY Academic Works. https://academicworks.cuny.edu/gc_etds/1882

Gemes, K. (1994). A new theory of content I: Basic content. *Journal of Philosophical Logic* 23, 595–620. https://link.springer.com/article/10.1007/BF01052779

Gemoll, W., Vretska, K. (2006). *Griechisch-deutsches Schul- und Handwörterbuch,* 10th edn. Munich: Oldenbourg.

Hempel, C. G., Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science* 15(2), 135-175.

Holmes, M. R. (2012). *Elementary Set Theory with a Universal Set.* Cahiers du Centre de Logique (10), Université catholique de Louvain, Département de Philosophie. https://randall-holmes.github.io/head.pdf

Hume, D. (2000). *An Enquiry concerning Human Understanding,* edited by T. L. Beauchamp. Oxford: Oxford University Press.

Lewis, D. (1973). Causation. *The Journal of Philosophy* 70(17), 556-567. Reprinted in M. Tooley (ed., 1999), *Laws of Nature, Causation, and Supervenience* (pp. 178-189). New York and London: Garland Publishing.

Lowe, E. J. (2003). Individuation. In M. J. Loux (ed.), *The Oxford Handbook of Metaphysics* (pp. 75-95). Oxford: Oxford University Press.

Mackie, J. L. (1974). *The Cement of the Universe: A Study of Causation.* Oxford: Oxford University Press.

McCall, S. (1966). Connexive implication. *Journal of Symbolic Logic* 31(3), 415-433.

McTaggart, J. E. (1915). The Meaning of Causality. *Mind* 24(95), 326-344.

Menzies, P. Beebee, H. Counterfactual theories of causation. *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), Edward N. Zalta (ed.).

Price, H. (1996). *Time's Arrow and Archimedes' Point: New Directions for the Physics of Time.* Oxford University Press.

Price, H. (2007). Causal Perspectivalism. In H. Price, R. Corry (eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited* (pp. 250-292). Oxford: Oxford University Press.

Priest, G. (2008). *An Introduction to Non-Classical Logic*, 2nd edn. Cambridge & New York: Cambridge University Press.

Shewmon, D. A. (2007). Mental Disconnect: 'Physiological Decapitation' as a Heuristic for Understanding 'Brain Death'. In Pontifical Academy of Sciences Scripta Varia 110: *The Signs of Death* (pp. 292 - 333). Vatican City.

Stepanov, A. (2004). Towards a theory of causal implication. http://stepanovpapers.com/TOWARDS%20A%20THEORY%20OF%20CAUSAL%20IMPLICATION.pdf

Strumia, A. (2012). *The Problem of Foundations: An Adventurous Navigation from Sets to Entities, from Gödel to Thomas Aquinas.* Siena: Cantagalli.

Suárez, F. (1965). *Disputationes Metaphysicae*, edited by C. Berton, reprinted at Hildesheim by G. Olms Verlagsbuchhandlung.

Weizsäcker, C. F. von (2002). *Aufbau der Physik.* Carl Hanser Verlag.

Wittgenstein, L. (1995). *Tractatus Logico-Philosophicus*. In *Werkausgabe*, 2nd edn., vol. I. Frankfurt am Main: Suhrkamp.

All cited online sources were retrieved in April 2021.