

Forthcoming in *Acta Philosophica Fennica* (a special volume in honour of G.H. von Wright's Centennial Anniversary)

PANU RAATIKAINEN

Explanation and Understanding Revisited

1. Introduction

Georg Henrik von Wright's book *Explanation and Understanding* (1971) is a modern classic in analytic hermeneutics, and in the philosophy of the social sciences and humanities in general. In this work, von Wright argues against naturalism, or methodological monism—that is, the idea that both the natural sciences and the social sciences follow broadly the same general scientific method and aim to achieve causal explanations. Against this view, von Wright argues that the social sciences are qualitatively different from the natural sciences: according to his view, the natural sciences aim at causal explanations, whereas the purpose of the social sciences is to understand their subjects. In support of this conviction, von Wright also offers a version of the so-called logical connection argument.

Von Wright distinguishes two traditions in the history of science and philosophy, which he calls the Aristotelian and the Galilean traditions; these involve two different views of explanation: causal-mechanistic explanation and finalistic explanation. Von Wright views causal explanation along the lines of the traditional covering law model of explanation. Finalist explanations, by contrast, lean on intentions, goals, or purposes. Causal explanation, according to von Wright, is characteristic of the natural sciences, while the latter is distinctive of the social sciences.

Von Wright argues that human action cannot be explained causally in accordance with the covering law model, but must be understood as intentional. As a model of intentional understanding, he presents instead “practical syllogism”, which puts actions into a kind of “logical connection” with beliefs and desires.

Von Wright also aims to demonstrate that intentional understanding, characteristic of the social sciences, is essentially different from that of causal explanation: in a causal relationship, cause and effect are conceptually independent of each other, while in a practical syllogism the action is—so the argument seeks to demonstrate—in a strong conceptual (“logical”) dependency between intentions and beliefs and desires.

2. The key ideas of *Explanation and Understanding*

Following a powerful tradition, von Wright assumes the traditional view that causal explanation in the natural sciences is aptly characterized by the covering law model, also known as “the subsumption theory of explanation” (or, with reference to its most famous advocate, “Hempel's model”. Hempel in fact formulated two variants: “the deductive-nomological model” and “the inductive-statistical model”; see Hempel 1965). According to it, a phenomenon or an event is explained by subsuming it under general laws. Schematically, such explanations have the following form:

L1, L2 . . . Ln	general laws
C1, C2 . . . Cn	initial, specific conditions

E	the <i>explanandum</i> —an event to be explained

Causal explanation, so understood, is thus a form of inference: the event to be explained is inferred from the initial conditions and one or more general laws; in their light, the *explanandum* ceases to be surprising. According to Hempel, scientific explanations are always nomological—in other words, they always involve at least one law. Hempel famously argued that this model also covers explanation in history (Hempel 1942); this was von Wright’s point of departure.

In the background of this model of explanation lurks the regularity theory of causation, popular among empiricist philosophers ever since Hume. According to the regularity theory, roughly, *c* causes *e* if and only if all events of type *C* (i.e., events that are like *c*) are regularly followed by events of type *E* (i.e., events like *e*).

In addition, von Wright presented in his work an original picture on causation: a version of the manipulability theory of causation (similar ideas had been advocated already by Collingwood (1940) and Gasking (1955)). That is, von Wright suggested that we should analyze causation in terms of the concept of manipulability:¹

p is a cause relative to *q*, and *q* an effect relative to *p*, if and only if by doing *p* we could bring about *q* or by suppressing *p* we could remove *q* or prevent it from happening (von Wright 1971, p. 70).

On the one hand, von Wright thus ties scientific explanation—at least in the natural sciences—closely to the covering law model, and thus implicitly to the regularity view of causation. On the other hand, von Wright also defends a specific manipulability theory of causation. This raises the question of whether von Wright clearly saw the difference between these two views of causation (more of this below). In any case, the presence of these two theories in von Wright’s thought creates some kind of tension.

Be that as it may, von Wright presents, as an alternative to the covering law model in the social sciences, what he calls “practical syllogism”, a scheme for understanding human actions. It is an inference which has the following form:²

X intends to bring about *P*.
X considers that he cannot bring about *P* unless he does *A*.

Therefore, *X* sets herself to do *A*.

¹ The use of “could” in this analysis shows that von Wright thought of causation in broadly counterfactual terms.

² This is only a very rough formulation. Von Wright put forward a number on refinements in his (1971) and in later publications (cf. Kusch 2003). Such further details are, though, quite irrelevant for our present discussion.

According to von Wright, practical syllogism provides “the sciences of man” with “an explanation model in its own right which is a definite alternative to the subsumption-theoretic covering law model” (von Wright 1971, p. 27). He continues:

Broadly speaking, what the subsumption-theoretic model is to causal explanation and explanation in the natural sciences, the practical syllogism is to teleological explanation and explanation in history and the social sciences (*ibid*).

Von Wright denied that an intention (for example, *X*'s intention to bring about *P*) could ever be the cause of the behavior at stake; in other words, the relationship between them is not, in his view, causal. He points out, first, that philosophers, particularly after Hume, have distinguished between cause and effect, on the one hand, and (logical or conceptual) ground and consequence, on the other hand; cause and effect are logically independent of each other. Accordingly, von Wright calls a causal relationship which satisfies this requirement of logical independence between its relata “Humean” (this is a much more general idea than the regularity theory of causation, often associated with Hume) (von Wright 1971, p. 93). More exactly, then, he considers the question of whether intention or will could be a *Humean cause* of an action (*ibid*).

Von Wright contends that *if* the relation between intention and behavior is causal, there must be a general law (non-logical nomic connection) involved (von Wright 1971, p. 97)—and practical syllogism would be only a disguised form of nomological-deductive explanation in accordance with the covering law model (von Wright 1971, p. 98). However, he notes that the premises of practical syllogism contain explicitly no laws. Furthermore, he argues that no laws are presupposed in the background of the inference—for example, laws that would connect specific types of intentions to the behavior or action of a particular type.

Von Wright further argued that the conclusion of practical syllogism cannot be verified in any other way except verifying the premises, and *vice versa*—and that their relationship is therefore conceptual and not (Humean) causal (von Wright 1971, pp. 94-, 107-). This is his special version of the logical connection argument.³

3. Recent developments in the philosophy of science

Since the publication of von Wright's classic work over four decades ago, the terrain of the philosophy of science has changed in many ways. As we have seen, von Wright bound causal explanation in science tightly to the covering law model. This was indeed the received view at the time. In the philosophy of science today, by contrast, it is practically universally agreed that the regularity theory of causation is implausible, and that the covering law model is also utterly untenable as an account of explanation in the natural sciences, and of causal explanation in general.

First, the regularity theory is not able to distinguish true causal dependencies and mere accidental correlations. For example, a sudden drop in the reading of a barometer is regularly

³ Other versions of the logical connection argument were presented by Anscombe (1957), Winch (1958), Melden (1961), and Taylor (1964), for example.

succeeded by the occurrence of a storm. However, it does not follow—*pace* the regularity theory—that the barometric reading caused the storm; rather, a drop in atmospheric pressure caused both the barometric reading and the storm.⁴

Second, even in many natural sciences other than physics (in the biological and medical sciences, for example), there are hardly any genuine laws, but they still give explanations all the time—explanations which are apparently causal. Von Wright, however, assumes that if the connection between the two events or states of affairs (e.g., between intention and action) is causal, it must necessarily involve a general law.

Traditionally, the covering law model of explanation has been motivated by the positivist-empiricist conception of causation: at the level of observation, we can only know that certain types of observable events occur regularly after certain observable events of a second type. Postulating some deeper “necessary” causal connection between the events is considered to be illegitimate metaphysics. This is the regularity view of causation.

As noted above, however, von Wright presents also his own, quite different view of causation. The manipulability theory of causation, as formulated by him and some others (Collingwood, Gasking), did not as such gain much popularity at the time. First, this theory of causation was commonly seen as too anthropocentric, that is, too dependent on the human agent: It seems to entail that causation could occur only where a human agent can manipulate the initial conditions. However, there are causal relations, for example, between some distant celestial bodies or between some past events of history, even if they are virtually beyond the reach of human impact. Second, the manipulability theory has also been considered to be circular, as it seems to presuppose causal notions.⁵

More recently, the so-called “interventionist” theory of causation has become increasingly popular in the philosophy of science. It has been developed especially by Woodward (1997, 2000, 2003), although related ideas have been put forward, for example, by Pearl (2000) and Spirtes et al. (2000). An account of scientific explanation based on it is also a promising candidate for replacing the covering law model as the standard model of causal explanation.

What is interesting in our present context is that this theory is a direct heir and a variant of the earlier manipulability theory of von Wright and others. The core concept of this theory is “intervention”. Heuristically, one may think of interventions as manipulations that might be carried out by a human agent in an idealized experiment. Nevertheless, the approach is not in any problematic sense anthropocentric, as completely natural hypothetical interventions (i.e., interventions that do not involve human agents) are possible, and

⁴ This is a problem for the simple regularity theory of causation. Both the later advocates of the covering law model such as Hempel (see e.g. Hempel 1965, 338-9) and von Wright seem to have been aware of such problems to some extent (see, e.g., von Wright 1971, p. 15). Refining the notion of law in terms of counterfactuals was considered (von Wright 1971, pp. 21-22). There was not, though, yet a satisfactory analysis of counterfactuals in turn available. Moreover, it is not obvious how well a counterfactual account of laws coheres with the regularity theory of causation and the general empiricist view behind it. Finally, neither Hempel nor von Wright ever seriously questioned the requirement that causal explanation must be necessarily based on laws (cf. the second objection that follows).

⁵ See Woodward 2013. How fair such accusations are in von Wright’s case is open to debate. He seems to have been aware of such problems, and at least attempted to circumvent them. This is even more clear in *Causality and Determinism*, where he points out that “causation is ontically independent of agency” (von Wright 1974, pp. 49-50).

intervention can be defined in purely causal terms, without any reference to human activity. That a causal vocabulary is openly presupposed means that the interventionist theory does not even aim to give a reductive analysis of causation. Still, this does not make the approach viciously circular: “*c* causes *e*” is explicated with the help of other causal relations and correlational information.

The interventionist theory is a well-developed version of the more general counterfactual approach to causation (cf. Menzies 2014). Counterfactual accounts are based on different “what-if-things-had-been-different” considerations. Causal relations are analyzed by examining certain contrary-to-facts (“counterfactual”) alternative possible scenarios. David Lewis (1973) in particular has been an influential figure here.

Assume thus that *c* has occurred first, and then *e* has occurred. When exactly is it warranted to conclude that *c* caused *e*? Under what circumstances it is correct to say that the relation between these events is really causal and not merely accidental correlation? The basic idea of the counterfactual theories of causation is that the meaning of causal claims can be explained in terms of conditionals of the form:

If *c* had not occurred, *e* would not have occurred.

Such conditional statements, in which the antecedent is contrary to the facts (assuming that the alleged cause *c* actually occurred), are referred to as “counterfactual conditionals”, or simply as “counterfactuals”. Roughly, then, the relation at stake is causal (i.e., *c* is a cause of *e*) if and only if the above counterfactual conditional holds.⁶

As was already noted, the interventionist theory is a version of the counterfactual theories of causation. According to it, whether a relation is causal can be evaluated with the help of counterfactuals which have to do with the outcomes of hypothetical interventions. Such counterfactuals are called “active counterfactuals.” These are such that their antecedents are made true by an intervention, and have the form:⁷

If *c*’s occurrence had been changed by an intervention into *c*’s absence, then *e*’s occurrence would have changed into *e*’s absence.

Note in particular that no laws (covering *c* and *e*) are required in the counterfactual and interventionist approaches for causation.

One way of motivating the interventionist approach is to ask the following questions: What is the point of our having a notion of causation (in contrast to, say, a mere notion of correlation) at all? What role or function does this concept play in our lives? Why do we care to distinguish between causal and merely correlational relationships? (cf. Woodward 2003, p. 28) According the interventionist approach, the answer is that such knowledge of genuine causal relationships is, sometimes, practical and applicable: by manipulating the

⁶ Put this simply, the view faces immediate problems (the “pre-emption” cases). Lewis (1973) was already aware of them, and presented a more sophisticated theory; cf. Menzies 2014.

⁷ This is an extremely rough formulation and omits many niceties of the theory. In reality, the interventionist theory of causation has been developed into a sophisticated theory. In any case, the interventionist setting helps to avoid many of the problems that the standard counterfactual theory faces; see, for example, Menzies 2014, §4.3.

cause we can influence the effect. If there is a real causal relationship between *A* and *B*, manipulating *A* is a way to change *B*; mere correlation between *C* and *D*, on the other hand, just disappears if one attempts to affect *D* by manipulating *C*. (Obviously, our knowledge of causal relationships and our interest in them need not be restricted only to applicable causal relations; it can certainly be purely theoretical and based on curiosity. Not all science is applied science.) Thus, we can try to find a cure for AIDS, suppress poverty, or prevent eutrophication of the Baltic Sea on the basis of knowledge about the causal relationships associated with them.

Real causal relationships can, in favorable circumstances, be distinguished from accidental correlations experimentally by manipulating the initial conditions (the putative causes) and investigating whether this has consequences on the effects (surely, this is often in practice impossible). The interventionist theory of causation thus emphasizes the close connection between causal thinking and experimental research (and manipulation and control).⁸

4. Causal explanations and the social sciences again

In recent years, several philosophers have argued that from the point of view of the counterfactual theories of causation and causal explanation, it is unproblematic to think that, for example, in history and the social sciences, one could present and will present causal explanations (see, for example, Lebow 2007, 2010; Nolan 2013; Raatikainen 2011; Reiss 2009).

From this perspective, human actions, singular historical events, and such may also well be explained causally, for example, by citing the mental states—such as desires and beliefs, intentions and reasons—of the relevant human agents as causes. For example, the event that a King declared war may be explained causally, for its part, by the King's (perhaps false) belief that the neighboring country was planning an invasion. The belief can well be viewed as a cause of the declaration of war and subsequent battles: if the King had not had that belief, he would not have declared war.

Accordingly, Lebow (2010) examines historical explanations of the outbreak of the First World War from the perspective of the counterfactual theory of causation.⁹ Reiss (2009) discusses a few interesting historical episodes in some detail from the counterfactualist perspective. I have myself presented (Raatikainen 2011) a fairly detailed case study on historical research concerning the so-called Club War, the 1596 peasant uprising in an area of what is now Finland, which has attracted much interest among Finnish historians. I argue that the interpretations of the historians can be naturally analyzed in accordance with the counterfactual and in particular the interventionist theory of causation, and it is natural to take them as presenting causal explanations.

Although the specific beliefs, desires, intentions, and motives that guided the main historical agents are internal mental states, even they can very well be the causes of actions

⁸ The interventionist theory of causation is also beginning to play an important role in the philosophy of mind and the issue of mental causation; see, for example, Raatikainen 2010; Woodward 2008.

⁹ As is happens, von Wright reflected on the very same example; he obviously attempted to show that the explanation is not causal. From the counterfactualist perspective, though, it is difficult to see why it could not be.

and events, and causally relevant factors from this point of view: if a counterfactual intervention would have changed the relevant beliefs (or desires) of these agents, for example, the course of events would have been different.

This suggests that even if the social sciences perhaps study a specific human aspect of reality which cannot be reduced to the reality studied by the natural sciences, the social sciences can nevertheless be viewed as providing causal explanations in fundamentally the same way as the natural sciences do. Therefore, a proper understanding of these issues in the philosophy of science removes a key reason to oppose methodological monism, or naturalism.

In sum, if causation is understood—in accordance with the widespread consensus in philosophy nowadays—as counterfactual dependence and not as necessarily based on laws, there is no real obstacle to taking the various explanations in the social sciences as normal causal explanations. The internal states of mind or motives of the human agents may also be causes.

5. What about the logical connection argument?

It is certainly natural to think of a person's beliefs and desires as being causally relevant to the person's behavior. But as we have noted, the so-called logical connection argument advocated by von Wright and many others denies the very meaningfulness of such a "causalist" view: It is argued that there is a tight analytical ("logical") connection between a person's action and her reasons for it. Hence, so the argument goes, the relation between the latter cannot be causal.

Davidson (1963), by contrast, put forward a response rather early on that has become quite influential. His argument begins with the observation that an event can be described (in a language) in many different ways; there may be conceptual dependence between particular descriptions of two events, for example, of a mental event and the subsequent action. The causal connection, though, holds between a pair of events in the world, and not between particular linguistic ways of describing them: "the truth of a causal statement depends on what events are described; its status as analytic or synthetic depends on how the events are described." Hence, Davidson argues, the logical connection argument fails.

Davidson's own view of causation was nevertheless quite naïve. He thought that causation is an absolute and simple relation between events. However, there are reasons to think that causal judgments are more relative than that: it has become increasingly popular to think that causal claims involve (even if often only implicitly) a contrastive class for both cause and effect, that is, they contrast alternatives to the putative cause and effect. Contrasts may be chosen differently, depending on one's interests and background assumptions (see, e.g., Hitchcock 1996). Some contrasts need to be fixed; otherwise, causal claims are not even unambiguous. Note that different choices of contrasts lead to different causal claims, some of which may be false, some true. Causal relation is thus also in a sense relative to descriptions.

Does this rebut Davidson's response? I do not think so. I contend that we must distinguish two different ways in which a causal judgment can be relative to descriptions. Even when particular contrast classes have been chosen and fixed, it may be still possible to pick up the cause and the effect by different descriptions. And at this second stage,

Davidson's key idea becomes relevant. In sum, I think that Davidson's argument is at least on the right track.

Accordingly, even transparently causal natural relations can be described as "logically" connected; for example:

Radiation sickness is caused by exposure to high amounts of ionizing radiation

The victim's passing away was brought about by her cause of death

Overexposure to the sun causes sun-stroke

But although in all these cases there is a conceptual connection between the relata, it would be preposterous to insist that the relation between them is not causal.

The logical connection argument admittedly has a certain intuitive appeal; I grant that some explanations in terms of reasons and intentions really are somewhat empty and trivial. But this only means that they do not satisfy certain further criteria of a *good* and *informative* explanation¹⁰—they may share the shortcomings of the notorious "*virtus dormativa*" explanation.¹¹ This does not mean that it is not a causal relation that is in question; some true causal statements just are not that informative.

6. Conclusions

In retrospect, we can conclude that the whole juxtaposition between explanation and understanding—so popular in the philosophy of the social sciences for so long—has been, above all, a confusion based on the outdated positivist view of causation and causal explanation.

Even in many natural sciences other than physics, such as the biological sciences, there are very few genuine laws, if any, and typical explanations in these sciences, which intuitively count as causal explanations, do not involve any laws. Consequently, it is hardly surprising that the social sciences do not have many explanations which involve laws either. The reasonable conclusion is that causation does not require laws—not that there are no causal explanations in the social sciences. The popular alternative view today is that causation only requires counterfactual dependencies.¹² It is very natural to see social scientific explanations as referring to such dependencies.

"Understanding" can, of course, mean a number of things in different contexts, but in the dispute at hand, it refers most commonly to understanding the behavior and action of

¹⁰ Ylikoski and Kuorikoski (2010), for example, suggest that different dimensions of the goodness of an explanation include non-sensitivity, cognitive salience, precision, factual accuracy, and degree of integration.

¹¹ In Moliere's play "*Le Malade Imaginaire*" (1673), a doctor explains opium's disposition to make people sleepy by its *virtus dormativa*, that is, "*dormative virtue*"—which is, of course, utterly empty and useless as an explanation.

¹² It is of course reasonable to think that counterfactuals are more generally grounded on various laws and regularities. The point here is simply that there need not to be any law connecting events of type *C* (events that are like *c*) to events of type *E* (events like *e*) for *c* to be a cause of *e*. The corresponding counterfactual may be often grounded on regularities other than the one between *c* and *e*.

human agents on the basis of internal motives and the reasons that guide them. Understanding in this sense is not, in the light of the above-mentioned advances in the theory of causation, the qualitatively different opposite of causal explanation—it is a kind of causal explanation.

University of Tampere

References

- Anscombe, G.E.M. (1957), *Intention*, Basil Blackwell, Oxford.
- Davidson, D. (1963), "Actions, Reasons, and Causes", *Journal of Philosophy* 60, 685–700; reprinted in D. Davidson, *Essays on Actions and Events*, Clarendon Press, Oxford, 1980.
- Collingwood, R.G. (1940), *An Essay on Metaphysics*, Clarendon Press, Oxford.
- Gasking, D. (1955), "Causation and Recipes", *Mind* 64, 479–487.
- Hempel, C. (1942), "The Function of General Laws in History", *Journal of Philosophy* 39, 35–48; reprinted in Hempel 1965.
- Hempel, C. (1965), *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, Free Press, New York.
- Hitchcock, C. (1996), "The Role of Contrast in Causal and Explanatory Claims", *Synthese*, 107, 395–419.
- Kusch, M. (2003), "Explanation and Understanding: The Debate Over Von Wright's Philosophy of Action Revisited", in L. Haaparanta & I. Niiniluoto (eds.) *Analytic Philosophy in Finland, Poznan Studies in the Philosophy of the Sciences and the Humanities. 2003, Vol. 80*, pp. 327–353.
- Lebow, R.N. (2010), *Forbidden Fruit. Counterfactuals and International Relations*. Princeton U.P.
- Lebow, R.N. (2007), "Counterfactual Thought Experiments: a Necessary Teaching Tool", *History Teacher* 40, 153–176.
- Lewis, D. (1973), "Causation", *Journal of Philosophy* 70, 556–567.
- Melden, A.I. (1961), *Free Action*, Routledge and Kegan Paul, London.
- Menzies, P. (2014), "Counterfactual Theories of Causation", in E.N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), , URL = <http://plato.stanford.edu/archives/spr2014/entries/causation-counterfactual/>
- Nolan, D. (2013), "Why Historians (and Everyone Else) Should Care about Counterfactuals", *Philosophical Studies* 163, 317–335.
- Pearl, J. (2000), *Causality*, Cambridge University Press, New York.
- Raatikainen, P. (2010), "Causation, Exclusion, and the Special Sciences", *Erkenntnis* 73, 349–363.
- Raatikainen, P. (2011), "Teon teoria ja historiallinen selittäminen", *Tiede & Edistys* 4/2011, 291–300.
- Reiss, J. (2009), "Counterfactuals, Thought Experiments, and Singular Causal Analysis in History", *Philosophy of Science* 76: 712–723.
- Spirtes, P., C. Glymour and R. Scheines (2000), *Causation, Prediction, and Search*, 2nd ed., MIT Press, New York.
- Taylor, Ch. (1964), *The Explanation of Action*, Routledge and Kegan Paul, London.
- Winch, P. (1958), *The Idea of Social Science and its Relation to Philosophy*, Routledge and Kegan Paul, London.
- von Wright, G. H. (1971), *Explanation and Understanding*, Routledge and Kegan Paul, London.
- von Wright, G. H. (1974), *Causality and Determinism*, Columbia University Press, New York.
- Woodward, J. (1997), "Explanation, Invariance, and Intervention", in *PSA 1996*, volume 2, 26–41.
- Woodward, J. (2000), "Explanation and Invariance in the Special Sciences", *British Journal for the Philosophy of Science* 51, 197–254.
- Woodward, J. (2003), *Making Things Happen: A Theory of Causal Explanation*, Oxford University Press, New York.
- Woodward, J. (2008), "Mental causation and neural mechanisms", in *Being Reduced—New Essays on Reduction, Explanation and Causation*. Ed. J. Hohwy and J. Kallestrup. Oxford: Oxford University Press, 218–262.
- Woodward, J. (2013), "Causation and Manipulability", in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition), URL = <http://plato.stanford.edu/archives/win2013/entries/causation-mani/>.
- Ylikoski, P. and J. Kuorikoski (2010), "Dissecting Explanatory Power", *Philosophical Studies* 148, 201–219.