

## Making Something Happen – Where Causation and Agency Meet

*Geert Keil*

The *reasons vs. causes* debate reached its peak about 40 years ago. Hempel and Dray had debated the nature of historical explanation and the broader issue of whether explanations that cite an agent's reasons are causal or not. Melden, Peters, Winch, Kenny and Anscombe had contributed their anticausal conceptions. The neo-Wittgensteinians seemed to be winning the day when in 1963 Donald Davidson published his seminal paper "Actions, Reasons, and Causes". Davidson's paper devastated the Wittgensteinian camp. It contained, among other things, a powerful attack on the logical connection argument. Davidson argued that the existence of a logical or conceptual connection between descriptions can never eliminate a causal relation, which holds between events *simpliciter*, not between events under certain descriptions.

Davidson maintained that in a way, reasons can be causes. When somebody acts for a certain reason, his intentional attitudes, or rather changes in his attitudes, cause his bodily movements. Davidson also argued that rationalization is a species of causal explanation. For the definition of action, he argued that intentional actions are bodily movements caused in the right way by beliefs and desires that rationalize them. Davidson's paper paved the way for causal theories of action, which superseded neo-Wittgensteinian analyses in the following decades. The causal theory was rapidly adopted by Alvin Goldman, David Armstrong, Paul Churchland, Myles Brand and many others, entering the mainstream and dominating the philosophy of action to this very day.

In 1971 Georg Henrik von Wright published his book *Explanation and Understanding*. The second chapter did not deal with agency, but with causation. It developed a new account of causation, the *interventionist* or experimentalist account. Focusing on causation, von Wright remedied a major shortcoming of the reasons vs. causes debate. The concept of causality, and the nature of the causal relation, received little attention in this debate, a fact that holds true for both camps. Mostly it was simply taken for granted that, as Hempel had declared, "causal ex-

planation is a special type of deductive-nomological explanation”.<sup>1</sup> One camp then aligned intentional explanations with D-N explanations, while the other camp insisted on their disparity. So strictly speaking, the label “reasons/causes debate” was a misnomer. The controversy dealt primarily with the question as to whether intentional explanations can take the form of D-N explanations, while the notion of causation, and the metaphysics of the causal relation, were left obscured.

With von Wright’s new approach, the situation changed. Von Wright was primarily concerned with causation, but his approach contained an implicit attack on the causal theory of action as well. His core idea was that the notion of causality is intimately linked with, or even derived from, the notion of intentionally making something happen. Other philosophers, even Hume, had considered such a connection before, but often just to reject this view, regarding it as a kind of myth belonging to the infancy of the human mind. Von Wright took the idea seriously. He submitted the analysis that  $p$  is the cause of  $q$  if and only if by doing  $p$  we could bring about  $q$ .<sup>2</sup>

The causal theory of action was also concerned with the relation between causation and agency, to which its name bears witness. The causal theory of action holds that actions are bodily movements with a certain causal history. This is why von Wright’s account constituted a momentous challenge to the causal theory: it reversed the direction of conceptual dependency between both notions. Davidson and his followers tried to define what an intentional action is by using the notion of causation. The causal condition which the causal theory sets is part of the definition of “doing something intentionally”. Von Wright claimed that the conceptual dependency is the other way round. He used the notions of doing, and bringing about, to explain what causal relations are. So, instead of a causal theory of action, he advocated an *agency theory of causation*, as it may be dubbed.

It is remarkable how seldom this clash of opinions about conceptual primacy is reflected in the literature. There are few exceptions: Fred Stoutland noticed the conflict, and he published a number of papers in

---

1 Hempel, C. G.: *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York/London 1965, p. 300.

2 von Wright, G. H.: *Explanation and Understanding*, Ithaca, New York 1971, p. 70.

which he compared Davidson's and von Wright's views.<sup>3</sup> Von Wright's book *Explanation and Understanding* was widely read and discussed in the seventies, especially in Europe. But it strikes me that especially in North America, where the causal theory of action became the orthodoxy of the day, von Wright's challenge went largely unnoticed. Even Davidson did not seem to take it seriously. He nowhere takes notice of the interventionist theory of causation, while he does discuss von Wright's earlier book *Norm and Action*. As is well-known, Davidson favoured an alternative account of causation, based on "the principle of the nomological character of causality", as he somewhat clumsily called it, or, later and less clumsily, "the cause-law thesis".<sup>4</sup> Davidson's firm adherence to a nomological theory of causality may explain why he did not take much interest in alternative accounts.

### The Interventionist Account of Causation

Let me briefly review some features of von Wright's interventionist account. In *Explanation and Understanding*, von Wright makes the following observation:

In the idea of putting systems in motion the notions of action and of causation meet. [...] It is natural to speak of the causes of phenomena as factors which 'produce' or 'bring about' their effects.<sup>5</sup>

- 
- 3 Stoutland, F.: "The Causal Theory of Action", in: Juha Manninen and Raimo Tuomela (eds.), *Essays on Explanation and Understanding. Studies in the Foundations of Humanities and Social Sciences*, Dordrecht 1976, pp. 271-304. Stoutland, F.: "Philosophy of Action: Davidson, von Wright, and the Debate over Causation", in: Fløistad, G. (ed.), *Contemporary Philosophy. A New Survey*, The Hague 1982, Vol. 3, pp. 45-72. Stoutland, F.: "Davidson on Intentional Behavior", in: LePore, E. and McLaughlin, B. P. (eds.), *Actions and Events. Perspectives on the Philosophy of Donald Davidson*, Oxford/New York 1985, pp. 44-59.
- 4 Davidson, D.: "Mental Events" (1970), in his *Essays on Actions and Events*, Oxford 1980, pp. 207-225. Davidson, D.: "Laws and Cause", *Dialectica* 49 (1995), pp. 263-279.
- 5 von Wright, G. H.: *Explanation and Understanding*, Ithaca, New York 1971, p. 64.

This lead-in lays down one desideratum of an adequate account of causation, namely, to explain the *asymmetry* of the causal relation.<sup>6</sup> The notion of bringing something about is clearly asymmetrical, in a way that does not seem to rely simply on the asymmetry of temporal succession.

Von Wright's appeal to the notion of bringing about finds a parallel in a remark of John Searle's. In his book *Intentionality*, Searle states:

there are not two kinds of causation, regularity causation and Intentional causation. There is just one kind of causation and that is efficient causation; causation is a matter of some things making other things happen.<sup>7</sup>

It is no accident that Searle speaks of "efficient causation". The notion of making something happen has an affinity to Aristotle's *causa efficiens*.<sup>8</sup>

Now obviously the development, or reconstruction, of a general notion of efficient causation can succeed only if causal efficacy is not confined to human agents. Even if human agents should be the only beings who can bring about something intentionally, intentional actions must not be the only instances of the cause-effect relation. Von Wright surely was tempted to go over to *agent causation*, but he resisted that temptation. The agent causality model, as advocated by Chisholm, gives an account of the relation between causation and agency that differs both from interventionism and the causal theory of action. Chisholm invokes, or invents, a second *species* of causation in addition to event causation, namely, agent causation, or, "causality through freedom", as fellow agent causalist Kant called it. This species of the causal relation is supposed to hold literally between agents and events. Adherents to agent causation include Richard Taylor, Randolph Clarke, Timothy O'Connor and Jonathan Lowe.

Let me briefly mention what I consider to be the decisive *ontological* objection to agent causation. Using a bit of jargon, agents or persons are *continuants*, i.e. Aristotelian substances that persist in time, while

---

6 Cf. *Ibid.*, p. 42.

7 Searle, J. R.: *Intentionality. An Essay in the Philosophy of Mind*, Cambridge, Mass. 1983, p. 135.

8 Recently, James Woodward used the phrase "Making Things Happen" as a heading for a general theory of causal explanation: *Making Things Happen. A Theory of Causal Explanation*, Oxford 2003.

actions and events are *occurrences*, i.e., dated happenings. (This terminology goes back to W. E. Johnson.) Now Charlie Broad raised the following objection against agent causation:

How could an event possibly be determined to happen at a certain date if its total cause contained no factor to which the notion of date has any application? And how can the notion of date have any application to anything that is not an event?<sup>9</sup>

One might object that it is an exaggeration to claim that the notion of date “has no application” to persons or Aristotelian substances. The problem is rather that these entities, being continuants, simply last or live too long. The acting person existed before her action took place, and she will live on afterwards. For this reason, she cannot be, in a literal sense, the triggering cause of any occurrence. What demands explanation is why this action took place at a certain moment in time, and citing the agent as the cause cannot answer that question. Arguably, that question cannot be answered without reference to something dated, to a particular occurrence, e.g., to some mental event inside the person, perhaps to a decision of hers or a volition. This is Broad’s *datedness objection* against agent causation. It is a very powerful objection, and as far as I can tell is still without a convincing reply.

Surely there are semantic connections between agent-causal statements and event-causal statements, but it seems impossible for entities that persist but do not *occur* to be causes in a literal sense. Most endorsers of agent causality do not take seriously the challenge that lies in promoting continuants to causes. Only Kant bit the bullet. He developed an account of “timeless agency”, holding that the initiation of a free act is not part of the temporal order at all. The transcendental subject, he claims, belongs to the noumenal world, while bodily movements belong to the empirical world. Virtually no one has found the view credible that noumenal selves can cause empirical events in a literal sense.

Von Wright always steered clear of the difficulties of promoting agents to causes. He holds that when an agent makes something happen, i.e., when he brings about *q* by doing *p*, it is the *occurrence* of *p* that causes *q*:

---

9 Broad, C. D.: *Ethics and the History of Philosophy*, London 1952, p. 215.

I am anxious to *separate* agency from causation. Causal relations exist between natural events, not between agents and events. When by *doing p* we *bring about q*, it is the *happening* of *p* which *causes q* to come. And *p* has this effect quite independently of whether it happens as a result of action or not.<sup>10</sup>

So, far from subscribing to agent causation, von Wright conceives of the interventionist account as a superior analysis of ordinary event causation.

Now, how does he arrive at his view that *p* is the cause of *q* if and only if by doing *p* we could bring about *q*? Putting the details of his view aside, let us recall the vital role that *the confirmation of causal laws* plays in his account. Von Wright starts with a familiar criticism of the empiricist regularity analysis: mere regularities might be accidental. They do not reflect the alleged necessity of causal connections, and they do not support counterfactuals. Causal *laws*, being nomic truths, would support counterfactuals, but where to get them from?

According to Wittgenstein's *Tractatus*, the world is "everything that is the case". Now in causal reasoning, we make claims about what *would have been* the case under different conditions, and perhaps even about what *must* be the case. But how do we establish such modal claims? How does one, for example, 'verify' the counterfactual conditional that if *p* had not been the case, when in fact it was the case, *q* would not have been the case either? Counterfactual cases cannot be observed in nature, after all. Von Wright is well aware of this fact:

It is logically impossible to verify on any single occasion when *p* was (is) not there, what would have been the case, had *p* been there.<sup>11</sup>

In order to 'verify' the counterfactual statement we ought somehow to make the actual and the non-actual 'change place'. How can this be done? Literally this can of course not be done at all.<sup>12</sup>

Now, fortunately we are not mere bystanders who merely observe what actually happens. We have the capacity, as von Wright puts it,

---

10 von Wright, G. H.: *Causality and Determinism*, New York/London 1974, p. 49.

11 von Wright, G. H.: *Explanation and Understanding*, Ithaca, New York 1971, p. 71.

12 von Wright, G. H.: "On Causal Knowledge", in his *Truth, Knowledge, and Modality* (= *Philosophical Papers III*), Oxford 1984, pp. 86-95, p. 92.

to interfere with the course of the world, thereby making true something which would not otherwise (i.e., had it not been for this interference) come to be true of the world at that stage of history.<sup>13</sup>

It is important to note that descriptions of ordinary actions already have such counterfactual implications. For example, when we say that somebody opened a window, we imply that the window would have remained closed without his interference. Generally speaking: “Productive action is the producing of a change on an occasion when no change would otherwise have occurred”.<sup>14</sup>

According to von Wright, the test procedure for causal counterfactuals is this: “doing something and noticing that a certain thing follows; refraining from doing and noticing that the same thing does not follow”.<sup>15</sup> This procedure, which is, incidentally, reminiscent of Mill’s “method of difference”, “is as ‘near’ as we can come to the verification of a counterfactual conditional”.<sup>16</sup> Von Wright hastens to add that this

does not mean that causal laws, nomic connections, can be ‘conclusively verified’. But it means that their confirmation is not a mere matter of repeated lucky observations. [...] One could say that we can be as certain of the truth of causal laws as we can be of our abilities to do, and bring about, things.<sup>17</sup>

Let me stop here to take stock.

- Von Wright rejects the regularity account of causality, for familiar reasons. Mere regularities are too weak, because they might be accidental.
- So he invokes causal laws, which support counterfactuals. He seems to accept what Davidson calls the principle of the nomological character of causality.

---

13 von Wright, G. H.: *Causality and Determinism*, New York/London 1974, p. 39.

14 von Wright, G. H.: “Replies”, in Juha Manninen and Raimo Tuomela (eds.), *Essays on Explanation and Understanding. Studies in the Foundations of Humanities and Social Sciences*, Dordrecht 1976, pp. 371-413, p. 376.

15 von Wright, G. H.: *Causality and Determinism*, New York/London 1974, p. 45.

16 von Wright, G. H.: *Explanation and Understanding*, Ithaca, New York 1971, p. 72.

17 Ibid., p. 73.

- The nomological analysis, however, cannot be the whole story about causal relations, since it leaves open the question as to how to confirm causal laws.
- This is where agency comes into play. Our ability to interfere with the course of events, and to make the actual and the non-actual ‘change place’ hypothetically, enables us to confirm causal laws.

The picture that emerges is this: interventionism does not reject the cause-law thesis, rather it fills a gap in the nomological account. Nomic truths, it is said, support counterfactuals, but the very notion of a counterfactual condition is indigestible by empiricist analyses. What *would* have been the case is not part of the observable world. Interventionism demonstrates where science has to go beyond mere observation, and has to take agency and experimentation seriously.

### **Four Objections to Interventionism**

A number of objections to interventionism have been raised. These objections are mainly directed against von Wright’s analysis that  $p$  is the cause of  $q$  if and only if by doing  $p$  we could bring about  $q$ .

(a) The first objection says that the analysis has an air of *circularity*, since the notion of bringing about itself has causal connotations.

(b) The second objection, which we may dub the *sceptical* objection, says that we simply do not possess a reliable procedure for the verification of the counterfactuals in question. It can always happen that our certainty as to having brought something about was deceptive, because the effect was due to some external cause. This can happen even with basic actions, as recent psychological research on the so-called “control illusions” has demonstrated. It just does not seem to be the case that agents have infallible first-person knowledge of their basic actions, let alone of the further effects of those actions.

(c) The third objection argues that von Wright’s analysis covers only *general* causal statements. This shortcoming is rarely mentioned in the literature, so let me explain. It is clear that the interventionist analysis of “ $p$  causes  $q$ ” has to be given in generic terms, in order to license causal judgements about cases that are beyond our reach. Von Wright clearly concedes that non-actions also have causes and effects: “Causation

operates throughout the universe – also in spatial and temporal regions forever inaccessible to man”.<sup>18</sup> So, the interventionist analysis must allow for extrapolation to such cases. But von Wright underrates the price one has to pay for the possibility of extrapolation. The price is simply that the truth conditions for singular causal statements are not provided by interventionism. Being told that an event belongs to the kind of event that can or could be brought about by a certain means does not tell us what in fact caused the event on any particular occasion. One can easily imagine a given case where  $p$  was not the cause of an instance of  $q$ , even if generally, doing  $p$  might be an effective means of bringing about  $q$ . Von Wright’s generic analysis leaves this possibility open. Recipes, even good recipes, do not always work.

In my view, there is a grain of truth in each of these objections, and more than just a grain in the third. But in order not to lose sight of my original concern, which is to describe the challenge that interventionism poses to the causal theory of action, I shall make an attempt to locate more precisely the role that agency plays in the notion of causation. For that purpose, I shall put the three objections aside and bring up a fourth, which is closely related to the third.

(d) Von Wright holds that “we can be as certain of the truth of causal laws as we can be of our abilities to do, and bring about, things”, as quoted above. Now my question is this: why not simply cut short the detour through causal laws and their confirmation, and turn to counterfactual claims about causal sequences directly?

Leaving causal laws aside has at least two advantages. First, it might well turn out that no causal laws are available anyway, that is, no exceptionless generalizations that cover cause-effect pairs are true. Nancy Cartwright’s slogan “the laws of physics lie” is certainly an exaggeration, because her arguments apply only to laws of a certain kind. Many laws of physics describe relations between universals, and such laws cannot have counterinstances, strictly speaking, because they make no claims about instances in the first place. *Causal* laws, however, do describe the temporal behaviour of physical systems, and such laws can be falsified by counter-instances.<sup>19</sup> In fact, no one has ever presented a

---

18 Ibid.

19 The distinction between laws of succession and laws of coexistence goes back to Mill (cf. Mill, J. S.: *A System of Logic* [= *Collected Works* VII/VIII], Toronto

strict causal law of the kind Davidson requires.<sup>20</sup> I cannot substantiate this bold claim here, but if I could, my conclusion would be thus: the good news is that only *some* laws of physics lie, the bad news is that these are precisely those laws which the champions of the cause-law thesis invoke, and that they need, namely strict laws of succession that cover cause-effect pairs.<sup>21</sup> Physical science is doing quite well without strict laws of succession. The laws that physicists are rightly proud of are different in character. The only ones who are left empty-handed are the philosophical champions of the cause-law thesis.

The second advantage of avoiding the detour through causal laws is that the task of providing truth conditions for singular causal statements is facilitated. Laws not only support or license counterfactual judgments, every singular causal statement does so too. Von Wright is well aware of this fact, and accordingly, he sometimes weakens the conceptual link between causation, agency and the confirmation of causal laws. Consider the following passage:

I do not believe in the existence of a ‘causal glue’ over and above the fact that, if p causes q, then, if p had been when in fact it was not, q would have been too. The ‘causal glue’ is the truth of the counterfactual statement, one could perhaps say.<sup>22</sup>

In this passage, von Wright plainly subscribes to the *counterfactual* account of causation. According to this view, if an event caused another event, the counterfactual dependence between cause and effect holds regardless of whether we have or could have brought about the effect, and regardless of whether a strict causal law exists that covers the case.

---

1974 [1843], Bk. III, Ch. V, § 1). Mill also makes it clear that “unless there had been laws of succession in our premises, there could have been no truths of succession in our conclusions” (ibid.).

20 “If *c* and *e* are related as cause and effect, there exist descriptions of *c* and *e* that instantiate a strict law.” Davidson, D: “Replies”, in R. Stoecker (ed.), *Reflecting Davidson*, Berlin/New York 1993, pp. 312-3.

21 For the details, see Keil, G.: “How the *Ceteris Paribus* Laws of Physics Lie”, in: Jan Faye/Uwe Scheffler/Max Urchs (eds.): *Nature’s Principles*, Berlin/Heidelberg/New York (Springer) 2005, pp. 167-200.

22 von Wright, G. H.: “Replies”, in Juha Manninen and Raimo Tuomela (eds.), *Essays on Explanation and Understanding. Studies in the Foundations of Humanities and Social Sciences*, Dordrecht 1976, pp. 371-413, p. 384.

The counterfactual account is weaker than both the interventionist and the nomological account. The question that suggests itself is: *What do we miss if we replace interventionism proper with the simple counterfactual analysis?*

## The Counterfactual Analysis of Event Causation

On the counterfactual analysis of event causation, as it was worked out by David Lewis, the singular causal statement “*c* caused *e*” is true if and only if,

- *c* and *e* are individual events that both occurred, and
- if *c* had not occurred, *e* would not have occurred either.<sup>23</sup>

One may feel inclined to add the third condition that *c* must have preceded *e*, but according to Lewis, “backtracking” counterfactuals should not be affirmed anyway, so that the temporal condition is redundant.

Yet there is a further condition needed, a clause that fixes certain circumstances under which the counterfactual dependence holds. Without such a clause, it becomes difficult to assess the truth-value of the counterfactuals in question. Lewis himself invokes a relation of overall similarity between possible worlds. To my mind, this move is a characteristic example of explaining the obscure by appeal to the more obscure. But perhaps we can do better. Causal counterfactuals, which deal with events that actually occurred, can be supplemented with a *ceteris paribus* clause. In the philosophy of science there is much controversy about the exact content of the *ceteris paribus* clause, but in this case, the clause fixes the circumstances that *actually pertained* when the causing event occurred. If, in counterfactual reasoning, we were free to change the prehistory of the causing event, the counterfactual in question would no longer have a determinate truth value. For reasons I have explained

---

23 “If *c* and *e* are two actual events such that *e* would not have occurred without *c*, then *c* is the cause of *e*” (Lewis, D.: “Causation” [1973], in his *Philosophical Papers*, Vol. II, New York/Oxford 1986, pp. 159-213, p. 167). I neglect Lewis’ distinction between causation and causal dependence, because I do not share the sole motivation for this distinction: his assumption that the causal relation is transitive.

elsewhere, the *ceteris paribus* (hereafter cp) clause needed must have an *indexical* content, so that the hedged counterfactual reads: “If *c* had not occurred, *and if everything else had been as it was when c actually occurred*, *e* would not have occurred”.<sup>24</sup>

The indexical nature of this cp clause, incidentally, precludes its being used to restore the truth of causal *laws* in the face of counterinstances, for laws must not contain singular terms referring to particular objects, locations or times. Hence, if cp clauses have an indexical content, there can be no such things as cp laws. Indexical cp clauses need singular propositions as their habitat, so that they cannot be used to defend the cause-law thesis. But they can do something better for the theory of causality, as I just pointed out. They are perfectly intelligible if combined with singular counterfactuals.

Another advantage of the counterfactual analysis is that counterfactual dependence may cut across the mental/physical distinction. There are many true mixed counterfactuals with physical antecedents and mental consequents, and vice versa. Perception provides examples of the first sort: “If the alarm clock had not rung, I would not have acquired the belief that it was time to get up”. This is a counterfactual with a physical antecedent and a mental consequent. In the case of intentional action, the counterfactual dependence runs the other way round: “If I had not decided to raise my arm, my arm would not have gone up”. Or, if you prefer the belief/desire version, as Davidson does: “Had the agent not desired that *p* and believed that *q*, the action would not have occurred”. Note that no special reference to mental causation is made here. Of course, the antecedent of such conditionals describes something mental, but this does not mean that there is a special *kind* of mental causation. This fits well with Searle’s observation: our common sense picture does not contain two kinds of causation. Causation is a matter of things making other things happen, regardless of whether there are mental events in the causal chain – or events that can have a mental description, as Davidson would put it.<sup>25</sup> So, the counterfactual account leaves room

---

24 See Keil, G.: “How the *Ceteris Paribus* Laws of Physics Lie”, in Jan Faye/Uwe Scheffler/Max Urchs (eds.): *Nature’s Principles*, Berlin/Heidelberg/New York (Springer) 2005, pp. 167-200.

25 Von Wright, in his later years, does not even commit himself to the existence of mental causation: “[H]is intention (volition) in doing this does not cause the bodily movements in which his actions may be said to ‘consist’. These move-

for mental causation without making a fuss about it. And that is a good thing.

A further advantage of the counterfactual account is that it represents our epistemic situation better than the regularity theory or the nomological theory. As Searle puts it:

I am much more confident of the truth of my original causal statement and the corresponding causal counterfactual than I am about the existence of any universal regularities that would cover the case.<sup>26</sup>

In sum, I submit the view that the detour through causal laws is an aberration, because the counterfactual analysis already yields the correct truth conditions for singular causal statements. I am fully aware that substantiating this claim would require a far more rigorous analysis than space permits in this paper. Among other things, one would have to go into a discussion of alleged counterexamples to the counterfactual analysis, specifically, cases of pre-emption and overdetermination.

Some philosophers argue that counterfactual dependencies must be *grounded* in something factual, since they are not ground-floor properties of the physical world. Well, they are grounded. The truth conditions for singular causal statements are not given *solely* by the counterfactual analysis. They include the physical properties of the substances involved. For example, claiming that the litmus tincture would not have turned red, had the acid not been poured into the test-tube, will fail to enlighten anyone who has witnessed the experiment. In science, we seldom ask for the triggering cause and leave it at that. Often the causing event is obvious enough, and we rather want to know *why the effect produced was F, rather than G*. In order to answer that question, we will have to mention the physical properties of the substances involved, further facts about their arrangement and the physical forces present. These properties and facts are what ground the counterfactual judgement.

---

ments were caused by nervous processes.” (von Wright, G. H.: *In the Shadow of Descartes. Essays in the Philosophy of Mind*, Dordrecht/Boston/London 1998, p. 109).

26 Searle, J. R.: *Intentionality. An Essay in the Philosophy of Mind*, Cambridge, Mass. 1983, p. 118.

## The Role of Free Agency

I promised above to make an attempt to locate more precisely the role that agency plays in the notion of causation. But now it seems that I have parted company with the interventionist analysis altogether, and that, in giving preference to the counterfactual account, nothing of von Wright's insights is preserved.

What exactly do these insights consist of? I have argued against making a detour through the confirmation of causal laws, but I have not yet addressed the deeper issue as to how we are capable of counterfactual reasoning in the first place. Von Wright correctly stresses that counterfactual dependencies cannot be observed in nature. The question emerges as to *how we come to have beliefs about what would have been the case?*

Imagine that Laplacean determinism were true. In a deterministic world any given series of events is governed by strict laws of succession, so that nothing could ever have been other than it actually was. There are no open possibilities whatsoever. What about applying counterfactual reasoning about an alternative series of events in such a world? To be sure, with counterfactual reasoning we do not commit ourselves to the claim that the past could have been different. We just say that *if* the past had been different, this difference would have propagated. But if universal determinism is true, the past *could not* have been different. That is, any divergence from actuality that we think up immediately calls for being antedated. Our world cannot have been different only at that point in time, the course of events must have diverged before. David Lewis describes it thus:

Had I raised my hand [when in fact I have not], [t]he course of events would have diverged from the actual course of events a little while before I raised my hand, and at the point of divergence there would have been a law-breaking event – a divergence miracle, as I have called it.<sup>27</sup>

Now if determinism is true, this pushing back in time of a divergence from actuality calls for iteration. It seems that if the world is deterministic, then “if a certain pebble had rolled at a moment when in fact it did

---

<sup>27</sup> Lewis, D.: “Are We Free to Break the Laws?” (1981), in his *Philosophical Papers*, Vol. II, New York/Oxford 1986, pp. 291-298, p. 294.

not roll, the entire previous history of the world would have had to be different”.<sup>28</sup> Alternatively, agents would have to be able to render the laws of nature false. This second alternative, though, seems even less attractive. Under determinism, our actions could only be up to us if we had the miraculous ability to change either the laws of nature or the past – this is the core of the “Consequence Argument” for the incompatibility of free will and determinism.<sup>29</sup>

Now if any divergence from actuality calls for being pushed back in time, we may well ask what the point of the demand is that the divergence miracle be predated at all. In the mouth of the determinist, the plea for predating the divergence is simply not sincere, his real conviction being that the divergence from actuality could not have taken place *at all*, not a minute ago, not yesterday, and not at the time of the Big Bang. It would have taken a miracle to make the divergence possible, but surely determinism does not admit miracles.

But if this is how matters lie, then counterfactual reasoning about what would have been the case under different conditions tends to lose its point. If we believed in Laplacean determinism, counterfactual judgements about alternative histories would be futile. We could not explain why we assign truth values to some of these judgements. If the “least overall departure from reality” that we have to consider, according to Lewis, always and indiscriminately requires a world that is different from ours in its entire history, then the whole possible worlds semantics could hardly be justified. Under determinism, it makes no sense to calculate a “least” overall departure from. I find no solution to this problem in Lewis’ work. Therefore I take the demand, inflicted by determinism, to reschedule any divergence from actuality back to the Big Bang, or perhaps beyond, to be a *reductio ad absurdum* of counterfactual reasoning about alternative histories. *A fortiori*, then, counterfactual judgements could not be used to *analyse* or elucidate causal judgements. Last but not least, it would be hard to explain how we could have acquired the ability for counterfactual reasoning in the first place.

But we do make causal judgments, and we do reason counterfactually about the past. We judge that if the driver had not overlooked the

---

28 Bennett, J.: “Counterfactuals and Temporal Direction”, *Philosophical Review* 93 (1984), pp. 57-91, p. 68.

29 See van Inwagen, P.: *An Essay on Free Will*, Oxford 1983, p. 16 and 56.

other car, the crash would not have occurred. We even imprison people as a result of such reasoning. And when it comes to intentional action, we not only reason hypothetically that *if* the past had been different, such-and-such would have been the case, we actually affirm categorically *that* we could have done otherwise, good libertarians that we are in real life. In what Strawson calls our descriptive metaphysics, free actions are the paradigmatic divergence miracles that need not be predated. And we need such paradigmatic cases in order to make sense of counterfactual reasoning about the past.

So the picture I am trying to sketch is this: the counterfactual account gives the correct truth conditions for singular causal statements. It need not be relativized to human interventions, or to potential human interventions. Our belief in the truth of the counterfactual does not depend on whether we were, or could have been, involved as agents. Von Wright's analysis of "p causes q" is beyond remedy.

The counterfactual account yields truth conditions for causal statements, but it cannot by itself explain why counterfactual claims have truth values in the first place, nor why we are capable of counterfactual reasoning. This is where free agency comes into play. Causation is not directly linked with agency, but in a more roundabout way. And it is not linked conceptually with agency as such, but with libertarian freedom, with our ability to do otherwise in given circumstances.

Of course, none of this amounts to a *proof* of freedom. Kant is to be praised for his insight that no direct proof of freedom is possible, and that all we can do is check whether anything stands in the way of freedom, i.e., whether there is any reason for not taking the commitments of our ordinary conception of agents as choosing their actions seriously. Kant maintained that rational beings can only act "under the idea of freedom". Nonetheless, he thought of the empirical realm as being governed by deterministic laws. In order to resolve this tension, he took recourse in transcendental idealism. Though we have the "power of spontaneously beginning a series of [...] states",<sup>30</sup> we cannot exercise this causal power in the world of appearances. This twist is beyond me, for human actions take place precisely in the world of appearances. Those of us who do not find transcendental idealism credible will crave some elbow room in the empirical realm. If it is supposed to be literally

---

30 Kant, *Critique of Pure Reason*, B 476.

true that we are able to act thus or otherwise in given circumstances, then the world must contain open possibilities.

None of this amounts to a proof of the existence of alternative possibilities. It is up to physics or to metaphysics to find out whether our world contains open possibilities. But if we look at human agency from the inside, as it were, then we are not concerned with the *ratio essendi* of freedom, i.e., with the question as to whether or not the way of the world can ever bifurcate, but with its *ratio cognoscendi*. In performing an action, “under the idea of freedom”, as Kant puts it, believing that our arm would not have gone up had we not decided to raise it, we *become aware of* a range of alternative possibilities. We do not *create* these possibilities, however. They are real enough, but we would not be in a position to “see” these possibilities were we mere observers of regularities or irregularities. The picture that emerges is that our experience of choosing our actions is the *ratio cognoscendi* for the real possibility of alternative courses that the world can take.<sup>31</sup>

If we take our pre-philosophical commitment to libertarian freedom seriously, thinking differently about our abilities and opportunities is not an option for us. It seems to be a built-in feature of our common sense conception of agency that the world leaves room for our choices. In this respect, my result is still von Wrightian in spirit, since he assumes a *conceptual* link between agency and freedom:

[T]he concept of an action, the ascriptions of actions to an agent, belong to discourse in which ‘free will’ is taken for granted. [...] The ‘freedom’ or ‘free will’ of a man consists in the fact that he acts, one could say. [...] The ‘mystery’ of freedom, if there is one, is the ‘mystery’ of the fact that there are agents and actions.<sup>32</sup>

---

31 This sketch is developed in more detail in Keil, G.: *Handeln und Verursachen*, Frankfurt am Main 2000, especially pp. 388-95 and pp. 444-57.

32 von Wright, G. H.: “Freedom and Determination”, *Acta Philosophica Fennica* 31 (1980), pp. 5-88, p. 78-9.

## Causality in the Human Sciences

I wish to conclude with some remarks about the role of the notion of causality both in our common sense world view and in the scientific picture.

Both von Wright and Davidson had a feeling for the crucial function played by the notion of causality in our world-view. In Davidson's much-cited words:

The concept of cause is what holds together our picture of the universe, a picture that would otherwise disintegrate into a diptych of the mental and the physical.<sup>33</sup>

In order to appreciate this insight, one just has to think of the many mixed causal counterfactuals that we affirm. Common parlance does not care for the difference between mental and physical causes in such judgments.

Now contrary to the anti-causalist view in the *reasons vs. causes* debate, causation is not only central to our *scientific* world view, it is essential also to the understanding of intentional action, perception, memory and a host of other intentional phenomena. In the decades that followed the reasons vs. causes debate, many causal theories of intentional phenomena were developed.<sup>34</sup> As causal relations are essential to the holding of intentional relations between mind and world, the philosophy of mind and action is not well advised to surrender the notion of cause to the naturalistic camp. This is an insight shared by both Davidson and von Wright.

Davidson has gone one step further. His official doctrine – the cause-law thesis – is that wherever there is a singular causal relation, there is a strict law that covers the case, whether we know that law or not. For the purposes of his “anomalous monism”, it suffices to know that such causal laws exist, whereas it is up to physical science to spell out those laws. But there is a strong counter current in Davidson's thinking about causality. In his later years, Davidson noticed a tension between the scientific search for detailed and accurate laws on the one

---

33 Davidson, D.: *Essays on Actions and Events*, Oxford 1980, p. xi.

34 For an overview, see Davis, S. (ed.): *Causal Theories of Mind: Action, Knowledge, Memory, Perception, and Reference*, Berlin/New York 1983.

hand, and the use of causal notions on the other. Bertrand Russell had referenced that tension before, in his famous paper “On the Notion of Cause”. Just like Russell, Davidson says: “[I]t is a sign of progress in a science that it rids itself of causal concepts”.<sup>35</sup> And again: “Unavoidable mention of causality is a cloak for ignorance; we must appeal to the notion of cause when we lack detailed and accurate laws”.<sup>36</sup>

Now, this insight into the prescientific character of the causal idiom gives rise to an exciting new perspective on the *reasons vs. causes* debate. I am thinking of the following remark of Davidson’s, where he pleads for a reversal of the received wisdom:

It is often thought that scientific explanations are causal, while explanations of actions and mental affairs are not. I think almost exactly the reverse is the case: ordinary explanations of action, perception, memory, and reasoning, as well as the attribution of thoughts, intentions, and desires, are riddled with causal concepts; whereas it is a sign of progress in a science that it rids itself of causal concepts.<sup>37</sup>

So, according to Davidson, the notion of cause, which becomes extinct in advanced physics, finds an asylum in a sphere where the Wittgensteinians deemed it particularly out of place: in the explanation of action and mental affairs. Davidson does not allow, however, that reversed perspective to have repercussions for his *theory* of causality. He sticks to the cause-law thesis, and he is not prepared to draw Russell’s moral, which can be encapsulated in the slogan “the more nomic, the less causal”. Davidson still calls the strict laws that he thinks make causal concepts superfluous, “causal laws”. This looks to be an unstable position.<sup>38</sup> The truth conditions for singular causal statements, i.e., the existence of strict laws, are such that they, if fulfilled, make the phenom-

---

35 Davidson, D.: “Representation and Interpretation”, in Said, K. A. M., Newton-Smith, W. H., Viale, R. and Wilkes, K. V. (eds.), *Modelling the Mind*, Oxford 1990, pp. 13-26, p. 23.

36 Davidson, D.: *Essays on Actions and Events*, Oxford 1980, p. 80.

37 Davidson, D.: “Representation and Interpretation”, in Said/Newton-Smith/Viale/Wilkes (eds.), *Modelling the Mind*, Oxford 1990, pp. 13-26, p. 22-3.

38 For a diagnosis of the tension in Davidson’s view, see Bieri, P.: “Mental Concepts: Causal because Anomalous”, in Stoecker, R. (ed.): *Reflecting Davidson. Donald Davidson Responding to an International Forum of Philosophers*, Berlin/New York 1993, pp. 305-310.

enon disappear. The literal truth of causal talk, according to this view, is preserved only by a veil of ignorance. Davidson would retort that knowledge of strict laws would only make causal *concepts* superfluous, not causal *relations*, which hold in nature between mind-independent events. Unfortunately, he is silent about the question of what causal relations essentially are. His repeated assertion that they are extensional relations applies only to singular causal *statements*, not to the relations themselves.

The tension in Davidson's position can be resolved if we take into account a fact that Russell pointed to, and which Davidson, curiously enough, overlooked; namely, that the strict and precise laws that physicists are rightly proud of, have no causal interpretation, because they are not even laws of succession.<sup>39</sup>

Still, Davidson is to be praised for having corrected the misconception that causality is essentially a scientific notion. The application of causal concepts is not at all confined to scientific purposes. Nor is it up to any scientific theory to determine the "nature" of the causal relation, or the truth conditions for causal statements. These are philosophical tasks *par excellence*. In carrying out these tasks, von Wright has made further progress than Davidson. Though von Wright's interventionist analysis of "p caused q" is too narrow, he made it clear that causal judgments turn on our ability to engage in counterfactual reasoning, and that in devising and assessing counterfactuals, we must go beyond the attitude of the bystander who merely observes what actually happens.

---

39 See Keil, G.: "How the *Ceteris Paribus* Laws of Physics Lie", in Jan Faye/Uwe Scheffler/Max Urchs (eds.): *Nature's Principles*, Berlin/Heidelberg/New York (Springer) 2005, pp. 167-200.