Explaining Behaviour by Formal Causes

An Account of an Overlooked Model of Explanation and Some Notes Towards a Reassessment of Three Philosophical Problems

Emil Brådd

Pro gradu-avhandling i filosofi

Handledare: Hugo Strandberg

Fakultetet för humaniora,

psykologi och teologi

Åbo Akademi

2023

ÅBO AKADEMI – FAKULTETEN FÖR HUMANIORA, PSYKOLOGI OCH TEOLOGI

Abstrakt för avhandling pro gradu

Ämne: Filosofi	
Författare: Emil Brådd	
Arbetets titel: Explaining Behaviour by Formal Causes: An Account of an Overlooked Model of Explanation and Some Notes Towards a Reassessment of Three Philosophical Problems	
Handledare: Hugo Strandberg	Handledare:

Abstrakt:

This thesis examines ways in which an awareness of the logical difference between distinct causal models of explanation can further and transform philosophers' understanding of problems related to causation, such as the problem of determinism. Since such problems arise in connection with, and logically depend on, specific causal explanatory models, any two models that differ in significant respects from one another will raise different problems. What this entails is that philosophers' assessments of the scope, validity, and importance of problems related to causation need be informed by a proper awareness of the variety of explanatory models encompassed by our causal language and by an adequate understanding of the relationship of dependence that holds between causal problems and models of explanation. This thesis articulates these insights and puts them into practice in an attempt to open new perspectives on three longstanding problems related to causation: the problem of determinism, the problem of physicalist reductionism, and the problem of whether reason-citing explanations of action are a type of causal explanation or not. These issues arise in connection with *efficient-causal* accounts of human nature and activity, and the concept of causation that has shaped the discussion of these issues has therefore, as a rule, restricted causation to mean generative relations between antecedent and subsequent events. Here attention is drawn to a different type of causal model that we use to explain certain kinds of mental activity and behaviour – a *formal-causal* model that logically differs from the efficient-causal models that philosophers have tended to focus on – and the abovementioned problems are reformulated and reassessed from the viewpoint of this model. This involves, among other things, reconnecting the overly intellectualized problem of determinism with our lived experience and identifying misguided intuitions that might partly explain the irresolvable nature of the debate over the causal status of reason-explanations.

Nyckelord: formal-causal explanation, explanation of behaviour, causal determinism, physicalist reductionism, action theory, descriptive complementarity, philosophy of mind, Jonathan Lear, Ludwig Wittgenstein	
Datum: 19.2.2023	Sidoantal: 89

Contents

Introduction

Background and Aim

The one idea of Wittgenstein's that has the most far-reaching implications for the practice of philosophy is no doubt the idea that philosophical problems arise from the language and concepts by and about which we do our philosophizing. This radical idea takes us to the heart of Wittgenstein's anti-metaphysical conception of philosophy, while also being the basic premise of his thinking that inclines him, according to some commentators, towards a form of idealism. Our problems in philosophy have their source, not in a mind-independent "reality," but in the language and concepts by which we interpret and engage with our social and natural world. What this entails, among other things, is that "good" philosophical work cannot be done without a proper understanding and appreciation of those areas of our language that shape the problems we are grappling with. My thesis examines some ways in which linguistic awareness of this sort can transform our understanding of some perennial problems related to causal explanation. Like Wittgenstein, I believe some of the worries philosophers have had about causal accounts of human nature and activity have arisen from an excessive focus on one or a few strands of our causal language, and that by acknowledging the existence of other strands of our causal language we can begin to allay those worries. In saying this, I am touching upon two related points that I make in my thesis: that philosophical problems related to causation arise from, and logically depend on, specific causal models of explanation, and that causal models that logically diverge in significant respects from one another will give rise to different philosophical problems. These points should be of interest to philosophers dealing with issues related to causal explanation, because they entail that the scope and generality of such problems depend on the degree of logical divergence among the causal models that make up our causal language and, further, that our ability to make sober assessments of the weight, validity and importance of such problems depends on our awareness of the diversity of our causal language. In developing these ideas, I also try to demonstrate the usefulness of a "methodology" I employ throughout the paper, one that involves approaching philosophical problems that are conceptually linked to specific models of representation from the viewpoint of other (conceptually unrelated) models and, by doing so, to challenge received, congealed understandings of those problems. In the context of my thesis, this involves approaching philosophical problems that arise in connection with "efficient-causal" explanations of human activity from the viewpoint of a causal model that conceptualizes causation, *not* as a generative

relation between antecedent and subsequent events, *but* as a patterning influence that shapes people's experience and perception of their environment. Having outlined the main features of this explanatory model, which I call, after Jonathan Lear, a "formal-causal" model, I try to present a fresh view of some longstanding issues related to casual explanation by approaching them from the viewpoint of this model. Due to limitations of space, I have confined my attention to three problems: the problem of determinism, the problem of physicalist reductionism, and the problem of whether reason-citing explanations are a type of causal explanation or not.

Structure

This thesis consists of five chapters. In the *first chapter*, I consider some ways in which an increased awareness of the diversity of our causal language can transform our understanding of philosophical problems related to causal explanation. The discussion is based on Ludwig Wittgenstein's posthumously published collection of remarks, titled "Cause and Effect: Intuitive Awareness". Here I also describe the problems of determinism and physicalist reductionism and problematize some assumptions that make these problems seem more urgent than they perhaps are. In the *second chapter*, I introduce the formal-causal model of explanation as it is characterized in Jonathan Lear's *Wisdom Won from Illness* and distinguish it from Aristoteles' version of this explanatory model. In the *third chapter*, I develop my own account of the formal-causal model of explanation, give examples, and illustrate the type of unfreedom that characterizes behaviour of the sort that we explain by referring to formal causes. In the *fourth chapter*, I explain why the problems of determinism and physical reductionism do not arise in linguistic contexts where behaviour is explained by formal causes and try to reconnect the problem of determinism – an overly intellectualized problem – with our lived experience by reformulating it from the viewpoint of the formal-causal model outlined in the previous chapters. Finally, in the *fifth chapter* I use the conceptual resources furnished by my account of formal-causal explanation to identify some misguided intuitions that may partly explain the irresolvable nature of the debate over whether reason-explanations are a type of causal explanation or not.

1. Wittgenstein's Investigation of Our Various Uses of "Cause"

About halfway through the posthumously published collection of Wittgenstein's remarks titled "Cause and Effect: Intuitive Awareness", the following passage makes its appearance:

> We may say: we are so exclusively preoccupied by contemplating a yardstick that we can't allow our gaze to *rest* on certain phenomena or patterns. We are used, as it were, to 'dismissing' these as irrational, as corresponding to a low state of intelligence, etc. The yardstick rivets our attention and keeps distracting us from these phenomena, as it were making us look beyond.—Suppose a certain style of building or behaviour captivates us to such an extent that we can't focus our attention *directly* on another one, but can only glance at it obliquely.[1]

The passage is a comment addressed to Wittgenstein's imagined interlocutor, who objects to the idea that the word "cause" has more than one meaning. The context in which it occurs is a sequence of remarks that begins with two examples of two distinct models of causal explanation, or causal "prototypes," by reference to which the meaning of "cause" can be explained – two examples Wittgenstein gives to his interlocutor, who responds with the objection: "But aren't these cases both of the same kind […]" The passage is thus intended to perform at least two functions: to draw the reader's attention to a tendency among philosophers to focus on one meaning of "cause" at the cost of overlooking, or downplaying the relevance of, the word's other meanings, and to suggest that such a one-sided focus can lead to some sort of trouble when we philosophize on issues related to causation. The trouble seems to arise from our consequent inability of making balanced assessments of the weight and importance of the questions we ask about causation. When the diversity of the meaning of "cause" is being overlooked, the questions we ask (and the problems we run into in our attempts to answer our questions) will likely be influenced by tacit assumptions that our restricted concept of cause will have prompted us to make – questionable assumptions that we may be wholly or partly unaware of. This is a general problem that ails any philosophical inquiry – the problem of bringing our assumptions to explicit awareness – but my interest here is limited to cases where our inquiry concerns issues related to causation, and more specifically, cases where we are dealing with philosophical problems associated with causal accounts of human nature and activity, such as the problems of causal determinism and physicalist reductionism. These problems may come to appear more necessary, inevitable, and pressing than they are, when considered in the light of a restricted concept of cause. The problem of determinism, to give an

[1] 1993a: 389. From here onwards, Wittgenstein's "Cause and Effect" is referred to as CE. The remarks in CE were written between the end of September and the end of October 1937 and published after Wittgenstein's death by Rush Rhees, one of Wittgenstein's literary executors.

example, is easily mistaken for a *general problem* that arises wherever causal language is being used, even though some meanings of "cause" might be logically incompatible with and, thus, resistant to the "problem." As we shall see, causal determinism, as a metaphysical doctrine, requires for its coherence a specific meaning of "cause" that goes with causal explanations of the "nomological" variety, which are, after all, only one type of causal explanation among others.[2] In "Cause and Effect," Wittgenstein identifies four other causal models, or prototypes, by reference to which the meaning of "cause" can be explained, each of which can be applied to a variety of natural and social phenomena that link up with various social practices, institutions, conventions, and areas of language. This chapter provides an overview of Wittgenstein's five causal prototypes and suggest some ways in which the picture he paints of our causal language can transform our understanding of philosophical problems related to causal accounts of human nature and activity. The main idea I want to convey here is that different causal models of explanation raise and lend support to different philosophical problems, and that problems associated with one causal model may not appear as problems when viewed from the perspective of another model. The discussion begins with an overview of Wittgenstein's five causal prototypes and from there proceeds to a quick rundown of what I consider to be the most important lessons to be taken away from Wittgenstein's remarks on causation.

1.1 Wittgenstein's Five Prototypes of Causation

In this section I describe and give examples of Wittgenstein's five prototypes of causation, which I will refer to as "causation by impact," "tracing the cause," "following the mechanism," "reacting to a cause," and "nomological causation."

Prototype #1: Causation by impact. This is likely to be the first prototype that comes to mind when we are asked to think of causation: the physical impact of two things colliding with one another. The paradigm example would be a billiard ball collision, where a stationary ball is set in motion by a collision with a moving ball; but it is not hard to think of other examples: a bird crashing into a window, a meteor hitting the moon, a cannonball tearing through the walls of a fortress, and so on. In a passage where Wittgenstein considers this prototype, he says: "in mechanics we are inclined to explain by this. If a thing has been explained by impact, it has

[2] It is the nomological model of explanation that Wittgenstein alludes to in the above-quoted passage by "The yardstick [that] rivets our attention".

been *explained*. Finding what strikes."[3] The point Wittgenstein is making here is that, in paradigmatic cases involving causation by impact, explanation comes to an end with the identification of the thing that caused an effect by striking something, for example, when the cause of the loud thud – a bird – is found lying stunned in the grass below the window.

Prototype #2: Tracing the cause. About this second prototype, Wittgenstein says: "We also speak of 'tracing' the cause; a simple case would be, say, following a string to see who is pulling at it".[4] In the context of explaining causal phenomena of this type, the person, the animal or machine that is found to be pulling the string, rope, chain, or what have you, is what we would call "the cause." This highly variable prototype seems to cover a wide range of activities which, despite not always involving traction, can be described in terms of "tracing the cause." I am thinking of acts and activities such as the following: the act of following the smell of grilling meat wafting through the park; the rescue team's operation of following the rising smoke to the crash site; or even the vengeful farmer tracking the fox that killed the hens in the chicken coop. These are "intermediate cases" that differ from Wittgenstein's string-example but can be seen as extensions of this prototype.[5]

Prototype #3: Following the mechanism. This protype involves observation of the kinds of chain-reactions that occur inside and in between the parts of mechanisms and automata of various types: "When I turn this wheel, then *this* wheel turns and the lever will strike the bell."[6] Here one's thoughts quite naturally drift to the synchronized sequences of operations of clockworks, but the range of objects that lend themselves to this type of observation is quite diverse: bicycles, pullies, revolvers, and so on. There are videos on Youtube that demonstrate what happens when you operate a door handle: by rotating the handle you rotate the spindle, which rotates the cam drive units that push the transmission plate, which then retracts the latch.[7] Wittgenstein suggests that the concept of "causal nexus" – the "mysterious link" between cause and effect that philosophers have puzzled over for centuries – has its roots in this prototype of causation: "It is from this, by the way, that we have the idea of a '*causal nexus*'. The idea that cause is not mere sequence, but is a *connexion*. But the connexion is a string or cogwheels."[8]

[3] CE 410
[4] CE 387
[5] For remarks on "intermediate cases," see Wittgenstein's *Philosophical Investigations* § 122 and "Philosophy", p. 175.
[6] CE 410
[7] Owen 2016
[8] CE 410

Prototype #4: Reacting to a cause. About this fourth prototype Wittgenstein says, "*one* root of the cause-effect language-game is to be found here, in our looking out for a cause."[9] This remark draws our attention to the link between our "primitive" reactions and our language of cause and effect. Wittgenstein is suggesting that our causal language, or one branch of it, depends on our pre-reflective reactions to things that impinge on us from our surroundings.[10] To illustrate, imagine someone getting hit from above by bird droppings, a person who "instinctively look from what has been hit to what has hit it," from the shit stain on his shoulder to the culprit in the sky. Here "Calling something 'the cause' is like pointing and saying '*He's* to blame!'"[11] If we were completely indifferent to our outward appearance and to all matters hygienic, or if our sense of touch were cruder and we would not notice when lightweight objects, like bird droppings, hit us, we wouldn't say some of the things we are inclined to say in situations like these. We wouldn't point to the gull in the sky and call it "the cause" as there wouldn't be anything for us to react to, no effect for the bird to have caused. Thus, this strand of our causal language depends on our unthinking reactions to things that affect us and are perceived as meaningful. Furthermore, this prototype links up our causal language with our practices of assigning blame and holding others responsible for the damage they cause us. Calling someone "the cause" implies that the effect is undesirable and something we wish to get rid of.[12] This is not to say that all causal events that instantiate this prototype necessarily involve reactions to undesirable effects caused by someone or something whom we hold to be deserving of blame. We do not reproach our co-player when the frisbee hits us because we aren't paying attention to the game. Also, the degree to which our reactions to things impinging on us from without are under our conscious control may vary from case to case. I may not be able stop my knee from jerking when the doctor gives it a tap; my reflexes are causally determined and beyond my control. However, I might learn to control my impulse to slap mosquitos by conscious training. So, to sum up, this prototype covers a wide variety of reactions to a wide variety of causes. One could perhaps extend it to include temporally extended "reactions" to things like drugs and chemical agents.

Prototype #5: Nomological Causation. This prototype lies at the root of the problem of determinism. In the literature it is called "nomological" (or "nomothetic") causation as it

[9] CE 373

[10] This comes very close to Wittgenstein's remarks on pain-language in *Philosophical Investigations*, where he suggests that our linguistic expressions of pain ("It hurts!", "I'm in pain!" etc.) have been grafted onto and extend our prelinguistic reactions to pain (moaning, crying, grimacing). See § 244.

[11] CE 373

[12] CE 373

construes causation in terms of lawful connections between causes and their effects.[13] According to this prototype, causation is a relation between two kinds of event that stand in a temporally ordered relationship of dependence, so that whenever the first event occurs the second always follows.[14] David Hume is the philosopher who gave this prototype its classic formulation: "We may define a cause to be *an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second."*[15] Causation of this type, then, is not a matter of one-off processes in nature, but involves recurring phenomena, where two kinds of event are "conjoined" in a relation of temporal succession. This is why causal statements of the nomological variety always go beyond bare particulars and attempt to say something of general validity, for instance, about mosquitoes that carry Plasmodium parasites and what happens when they bite human beings: malaria. Having written into it the assumption that causal knowledge is acquired through observation and experiment, the nomological prototype is associated with natural science. However, our use of nomological explanations is by no means limited to scientific contexts. Wittgenstein gives an example where a layman – a farmer or shepherd of goats – finds use for the nomological paradigm: "He has noticed that, since his goats have been grazing on that slope, they give less milk. He shakes his head, asks 'Why?'—and then makes some experiments. He finds that such and such a fodder is the cause of the phenomenon."[16] Other phenomena that lend themselves to explanation by the nomological model include events as diverse as the following: the rising and falling of the tide (caused by the gravitational attraction of the sun and the moon on our oceans), chemical reactions involving the mixing of liquids and transitions from one chemical state to another, and even the starting of a fire by holding a magnifying glass to the sun and focusing the light rays on combustible material. The high applicability of the nomological model in part explains why it is so tempting for philosophers to subsume all cases of causation under this prototype.

1.2 Lessons to Draw from Wittgenstein's Remarks on Causation

The principal value of Wittgenstein's remarks on causation lies, I think, in two areas. Firstly, in the reminder they give of the obvious but sometimes forgotten fact that causal models of

[13] "Nomos" is Greek for "law".
[14] This conception of causation is often accompanied by the idea that causal phenomena are governed by the laws of nature. See Tanney 2009: 95; Carroll & Markosian, 2010: 25.
[15] Hume 1935 (1748): 79; italics in original.
[16] CE 389

explanation are, *not* representations of a mind-independent reality, but linguistic constructs developed by human beings motivated by human interests, needs, and concerns.[17] This is an important reminder in a scientistic culture where it is often assumed that our explanatory models, and our language more generally, should aspire to capture the "essential" qualities of phenomena as they occur in nature, "objectively," and not as they appear in our "mistaken" everyday experiences ("the table in front of you is *really* not a table at all but a swarm of atoms"; "the pain you feel when you burn your finger is *really* a highly complex neuronal process"; "our feelings of pleasure are caused by the release of oxytocin"; etc.).[18] This assumption prevents us from gaining clarity about the nature of the phenomena we want to understand and explain, causally or otherwise, by suppressing awareness of the perspectival character of any explanation.

The second benefit we can gain from studying Wittgenstein's remarks on causation is an increase in our awareness of the tacit assumptions embedded in our causal language. The realization that there is more than just one model of causal explanation prompts us to inquire into the origins of, the purposes served by, the values and ideals inherent in, the ideas and attitudes connected with, our various causal models: "Why do we have all these models to begin with? why don't we use the same model across the board in all contexts of

[17] By using the term "linguistic construct," I am not implying that our causal language is an arbitrary system of symbols that we have "invented" to satisfy our needs and concerns, a system that could be radically altered without radically transforming the overall cultural context it forms part of. One can probably say without exaggeration that our form of life would be *inconceivably* different if we lacked our causal language. Lagerspetz explores the connection between the concept of "life-form" and our causal language in a recent paper titled "Wittgenstein's Forms of Life: A Tool of Perspicuous Representation." Nor do I mean to characterize our causal language as a predominantly intellectual phenomenon, as a symbolic structure we use in the service of speech and thought and is largely independent of our basic biological make-up. As we saw in the overview of Wittgenstein's causal prototypes, certain parts of our causal language depend on our "primitive reactions," which, in turn, depend on the biological constitution of our species. My use of the term "linguistic construct" is motivated by two things: (1) my wish to accentuate the contrast between the view of our causal language outlined in this chapter and the metaphysical view that accompanies the doctrine of causal determinism; and (2) my wish to indicate that the causal models I'm mainly interested in here – the nomological and the formal-causal models – have entered our everyday-language from the vocabulary of modern empirical science and depth-psychology, that is, from fields of knowledge that can, in certain contexts, be aptly described in constructivist terms.

[18] The philosophical position associated with this assumption is called "metaphysical realism." In *The Blue Book*, Wittgenstein discusses the "craving for generality" that may prompt philosophers to adopt this position: "Our craving for generality has another main source: our preoccupation with the method of science. I mean the method of reducing the explanation of natural phenomena to the smallest possible number of primitive laws…Philosophers constantly see the method of science before their eyes, and are irresistibly tempted to ask and answer questions in the way science does. This tendency is the real source of metaphysics, and leads the philosopher into complete darkness." (1958: 18)

explanation?"[19] An awareness of the diversity of our causal language also prompts the realization that philosophical problems related to causation are *always* connected with some specific causal model or models, and not necessarily with others. For the purposes of my investigation, this realization is important because it alerts us to the possibility that the worries philosophers have had concerning the use of causal explanation in the study of human nature and activity can perhaps only be reasonably entertained in regard to some particular causal model(s), and not in regard to causal explanations in general. In saying this, I am mainly thinking of the worry that causal explanations imply some form of determinism, and secondly, the worry that causal explanations reduce human nature and activity to biological phenomena. The former worry, or the kind of metaphysical speculation that raises the worry, is illustrated in Melden's classic investigation of free action:

> …whatever does happen, happens necessarily as it does, for given the conditions of its occurrence, the happening is causally necessary. Trace the causal antecedents of my conduct and my character back into the past as far as one pleases, to the conditions of my birth and my training, what happens now when I act as I do must happen in precisely the way in which it does. Hence I am no more responsible for what I am and do today than I am for the causal conditions of my birth, the training I received and the character I have, than I am for the fact that my father married my mother.[20]

This worry has occupied the attention of most of the major philosophers since the birth of modern empirical science in the seventeenth century. The core idea here is that agency drops out of the picture and morality is undermined if we look at human actions in causal terms. As was pointed out earlier, it is the nomological model that gives rise to this worry; but when the other causal models are overlooked and "causal explanation" comes to stand for nomological explanation, deterministic worries can begin to announce themselves indiscriminately wherever human activity is considered in causal terms.

The second worry I mentioned above, about the physicalist reductionism associated with causal explanation, is based on the assumption that proponents and practitioners of causal outlooks – neuroscientists, behaviouristic and evolutionary psychologists, etc. – reduce human beings and activities to biological phenomena. According to these critics, such causal outlooks define the concept of human being in relation to our organismic nature, as a biological, or even a mechanistic concept, instead of treating it as an open concept that evades final definition. Here,

[19] And conversely, the assumption that there is only one (legitimate) model of explanation tends to stifle such inquiries and is likely to prompt philosophers to develop theories about the necessity of our one and only model: "Our model for explaining natural phenomena arises from and reflects reality in its true existence."

[20] Melden 1961: 4-5. For a more recent (and perhaps the most influential) portrayal of the problem of determinism, see Peter van Inwagen's "The Consequence Argument".

for instance, are Harré and Secord criticizing such causal-reductionistic views of humans, of people:

> Psychologists seem to prefer to call the people they are studying "organisms" or "subjects." They think of their subjects as mechanisms that, like less complex physical objects, respond to the push and pull of forces exerted by the experimenter or the environment. In the classical exposition of this point of view an organism is regarded as being subjected to a certain *stimulus* situation, and it *responds* to it in a predictable manner.[21]

Continuing along these lines, a philosopher might argue something like this: "the concept of human being that emerges from behaviouristic and other casual-reductionistic viewpoints reduces human beings to their biological nature and thus fail to capture the characteristic that distinguishes us from other animals: our openness toward the "world" and ourselves[22] – an openness that enables us to define and to fashion ourselves (within certain limits) into the kind of beings we want to become. This is an essential part of human life – the activity of asking questions like "who am I?" and "what should my relation to the world be like?" and to organize one's life according to the answers one gives to such questions. Accordingly, a concept of human being that recognizes this aspect of human life must be one that changes based on our conception of the kind of beings we are and our hopes and visions of the beings we want to become in the future. This semantic openness of the concept of human being is overlooked by causal outlooks that define the concept in relation to our biological nature, and thus, these outlooks fail to capture the part of our nature that makes us human."

A similar argument can be made against causal explanations of human activity: "Causal explanations take too little account of the self-conscious understanding people have of their activities and their reasons and motives for engaging in them. The meaning of actions and their moral significance depends on the way in which the agent sees her action and the world in which she performs it. For the agent, it is this internal understanding that drives and structures her action; and for the observer trying to explain the agent's action from a third-person perspective, it is this internal understanding her explanation should try to capture. Causal outlooks, however, ignore this understanding and explain intentional actions and motivated behaviours from an external perspective, in terms of their physiological and neurobiological correlates. For example, a neuroscientist might characterize, say, an act of self-defence against

[21] Harré & Secord 1972: 27-32; italics in original

[22] The idea of our existential situation being characterized by an "openness" towards the world derives from early 20th century biologist Jakob von Uexküll, who describes humans as being more tightly embedded in a narrower and more closed world ("Umwelt") than other animals. See Jakob von Uexküll, *A Foray into the Worlds of Animals and Humans*.

a thug in the street in terms of neurophysiological processes and, by characterizing it so, makes it indistinguishable from other acts that differ with regard to the attitude and understanding they embody but correlate with sufficiently similar neurochemical processes, such as acts of fighting in combat sports. So causal outlooks iron out differences between kinds of activity by characterizing them in terms of physical features and processes that they have in common. This erasure of distinctions is problematic if we want to be able to distinguish intelligent behaviours that correlate with certain physiological processes from other kinds of intelligent and non-intelligent behaviours that correlate with similar physiological processes. But even more disturbing is the consequence this erasure of distinctions has for the moral side of human agency. For insofar as the assignment of responsibility to agents is done based on how agents see what they are doing, causal explanations, by ignoring this internal aspect of action, limit our ability to assign responsibility to agents for the acts they perform and their consequences."

Above I have presented three sets of objections that can be raised against causal accounts of human life and activity – one that takes issue with the deterministic implications of causal explanations, and two that take issue with the physicalist reductionism of causal outlooks and explanations. The first set of objections assumes that causal accounts rob humans of their autonomy and self-directedness by treating them as part of the "great causal network of nature;" and the other two assume that causal accounts fail to consider the internal understanding that partially constitutes human nature and activity by reducing human beings to biological phenomena. To what extent are these assumptions valid? And what can we say for and against these objections based on Wittgenstein's discussion of causation?

By reminding us of the various uses of "cause," Wittgenstein enables us to approach questions concerning causation without feeling too encumbered by the worry of determinism. If there are ways of thinking and talking of causes and effects that do not imply law-governed regularities, there might be ways of characterizing behaviour causally that do not eliminate human agency. And we have in fact already seen examples of such nondeterministic characterizations in the section on the prototype I called "reacting to a cause." The examples we considered there, of getting hit by bird shit and slapping mosquitos that land on one's legs, are examples of causal explanations that do not require general laws.[23] What those examples also remind us of is the kind of agency involved in our efforts to regulate our instinctive and habitual reactions by

[23] The "crucial lesson" to be drawn from Wittgenstein's discussion of causation is according to Glock that "the basic cases of causal explanations are those of singular causation, which do not require general laws." Glock 2014: 34.

means of various cognitive and inhibitory functions we possess as human beings: our power to imagine and to foresee things, our power to engage in self-reflection and to change our outlook and attitude to things, our capacity to choose between action-possibilities, and so on. Unlike my knee-jerk reflex, which is causally determined and beyond my control, my impulse to slap mosquitos is susceptible to conscious control of various kinds.

Based on these observations, there seems to be scope for arguing that at least one of Wittgenstein's causal prototypes does not imply determinism in any straightforward manner. But what about the other protypes besides "reacting to a cause" and "nomological causation"? Do they involve some form of determinism? Some of the examples I gave of "causation by impact" and "following the mechanism" do seem to have deterministic undertones to them. There is a clear and straightforward sense in which the operations of clockworks and meteors hitting the moon are causally determined phenomena, is there not? Aren't these precisely the kind of phenomena one would expect to find described in 18th century textbooks on Newtonian physics? Indeed, they are – but perhaps even these descriptions do not quite saddle us with determinism. In regard to the prototype of "following the mechanism," Wittgenstein reminds us of the possibility of malfunction and mechanical breakdown:

> We talk as if these parts could only move in this way, as if they could not do anything else. Is this how it is? Do we forget the possibility of their bending, breaking off, melting, and so on? Yes; in many cases we don't think of that at all. We use a machine, or a picture of a machine, as a symbol of a particular mode of operation.[24]

Here Wittgenstein calls attention to one sense in which the mechanical motions of clocks, bicycles, pullies, and other such mechanisms, are *not* causally predetermined processes: even if objects of this kind are in their normal functioning predetermined by the laws of mechanics, physics, etc., those laws will not prevent these objects from breaking down. That is to say, the logical form of "following the mechanism" is deterministic, but reality does not always follow the logic of our explanatory models. And as for the determinism implied by causal descriptions that instantiate "causation by impact," the sniper's bullet might get deflected by a falling brick.

Even so, these reminders won't appear very persuasive for someone impressed by the arguments for determinism. A stronger case that can be made against determinism based on Wittgenstein's remarks on causation and the relationship between language and the world was

[24] *Philosophical Investigations* § 193. See also § 613.

hinted at at the beginning of this section.[25] Recall what I said about causal models of explanation being linguistic constructs that humans have developed based on their interests, needs, and concerns. In saying that, I was implying that our causal language and the aspects of the world that we organize, investigate, and engage with by means of our causal language do not "mirror" one another in such a way that we have reason to assume that the features that characterize the former have counterparts in qualitatively identical features in the latter. Our language is not a single homogeneous structure that stands in a relationship of one-to-one correspondence with "reality" but is a collection of substructures ("language-games") that stand in a variety of different kinds of relationships (functional, representational, etc.) to different aspects of the social and natural world. Hence our deterministic models of explanation, such as the nomological model, do not entail a deterministic universe. We can think and talk about the world deterministically without the world's being deterministic.

Wittgenstein's discussion of causation can then be taken to show that the charge of determinism can only be levelled against specific causal prototypes, and not against causal explanations in general. And even when levelled against those prototypes only, the charge might be motivated in part, Wittgenstein's writings suggest, by an unfounded assumption about the relationship between our language and the "world" – the assumption that our language mirrors, or should mirror, the structure of the universe. Wittgenstein challenges the idea that our casual language reflects "real" causal processes in the universe in at least two ways: by suggesting that the causal prototypes that carry deterministic implications are no more privileged or legitimate than the other prototypes that do *not* carry such implications, but only serve different purposes and meet different interests and concerns; and by demonstrating that the relationship between our causal language and the world is not of any single kind, as the determinist might assume ("our causal language *represents* real casual processes in the world"), but is of different kinds, depending on which strand of our causal language we are dealing with.

As to the validity of the assumption that causal explanations entail some form of physicalist reductionism, I will limit my discussion here to a brief comment about the causal prototype "reacting to a cause," since I explore this issue at some length in Chapter Four.[26] The observation I want to make here is that this prototype lends itself to the description of human behaviour of kinds that presuppose the agent's possession and exercise of various mental

[25] For an interesting discussion of Wittgenstein's understanding of our language and the world, see Hertzberg's paper, "Very General Facts of Nature".
[26] See section 4.4, especially pp. 49-55.

capacities. For instance, if I say, "it was the sight of the stranger that caused the child to hide behind its mother's skirt," I am not explaining the child's behaviour by a physical cause. The stranger is of course a physical being; but my explanation does not characterize the child's reaction as a (mere) physical reaction. Rather, my explanation presupposes that the child experienced this physical being in a discriminatory, intelligent way – *as* a stranger, *as* a scary-looking person, or whatever – and thus imputes to child certain fears, beliefs, and other such mental attributes. The cause of the child's reaction is the stranger *experienced in a certain way*, and in that sense, a "mental cause."[27] Thus, among Wittgenstein's causal prototypes, there is at least one that cannot be accused of physical reductionism.

1.3 Five Prototypes – Are Those All?

In the preceding section, I indicated some ways in which Wittgenstein's discussion of causation can transform our understanding of certain problems associated with causal accounts of human nature and behaviour. The main idea I have sought to convey is that the weight and importance we ascribe to such problems depend on the assumptions we have about causation and causal language, assumptions that lead, when unsound and misguided, to a distorted view of the problems we are dealing with. Two such misguided assumptions that Wittgenstein's discussion of causation particularly helps to raise awareness of are (1) the assumption that our problems concerning causation pertain to causal explanation in general (as opposed to specific causal prototypes), and (2) the assumption that true causal explanations will mirror "real" causal processes in a mind-independent reality. These assumptions are liable to make philosophical problems concerning causation, such as the problems of determinism and physical reductionism, seem more urgent and pressing than they perhaps are. Wittgenstein helps us realize that we needn't worry about these problems in contexts where human activity is characterized according to causal prototypes the conceptual features of which disagree with, or do not lend intuitive purchase to, these problems, such as the prototype of "reacting to a cause." Indeed, we may even come to regard these problems as trivial upon realizing, as Wittgenstein's discussion suggests, that our use of causal models of explanation does not entail our commitment to any metaphysical theses about the nature of "reality." These are, to my mind,

[27] For interesting discussions of such "mental causes," see §§ 9-11 in G.E.M. Anscombe's *Intention*. Of course, it might be argued that explanations of behaviour that refer to mental causes belong to a category of their own, to a causal prototype distinct from the one I've called "reacting to a cause." That may be so. What matters for my argument is just that there is *a* causal prototype that explains behaviour by mental causes.

the two most significant and philosophically liberating insights a philosopher of causation can draw from Wittgenstein's writings.

But in addition to raising our awareness about the assumptions we carry with us into philosophical discussions of causation, Wittgenstein's remarks can, by sensitizing us to the nuances and subtleties of our causal language, lead to discoveries of causal prototypes that might have escaped Wittgenstein's attention when working on the material that would eventually become "Cause and Effect: Intuitive Awareness." The five prototypes he considers there are, after all, not intended as a comprehensive list of the causal models that shape our talking and thinking of causation. Wittgenstein's primary aim in "Cause and Effect" appears to have been to identify uses of "cause" that might help philosophers steeped in the empirical (Humean) tradition to overcome the temptation to reduce causation to the nomological model. As Hacker observes, "other prototypes could be added, and further refinements within each of Wittgenstein's paradigms are possible."[28] I have suggested some such "refinements" in my overview of Wittgenstein's five causal prototypes, and now I would like to propose an addition to his list.

All five of Wittgenstein's prototypes conceptualize causal phenomena in terms of antecedent events that lead to, or condition, the emergence of subsequent events. Such prototypes can be regarded as subtypes of a broader category of causation, the common feature of which is, to use Anscombe's formula, "the derivativeness of an effect from its causes." Surprisingly, when Anscombe proposes this formula, she intends it as a definition of causation in general, not as a definition of certain types of causation: "causality consists in the derivativeness of an effect from its causes. This is the core, the common feature, of causality in its various kinds. Effects derive from, arise out of, come of, their causes."[29] The reason why the general applicability Anscombe claims for her definition may strike one as surprising is that Anscombe was deeply influenced by a philosopher who would have accepted her formula, if at all, as applicable to one type of causation and one only. The philosopher I am referring to is of course Aristoteles, who divides causation into *four* types, only one of which lends itself readily to Anscombe's definition. This one type is often translated into English as "efficient causation," and is accompanied in Aristoteles' theory by three other types, designated as "material," "formal," and "final" causation. From among Aristoteles' four types of causation, it is the "formal" type I would like to single out for special attention and add to Wittgenstein's list of prototypes. Or

[28] Hacker 2000:56
[29] Anscombe 1993: 91-2

to be more precise: the causal model of explanation I would like to add to Wittgenstein's list is sufficiently similar to the model Aristoteles calls "formal-causal explanation" for us to call it by the same name. I will describe the main features of this model of explanation in the next two chapters, where I also specify the ways in which it diverges from Aristoteles' conception of formal-causal explanation. At this point of the discussion, I only want to observe that we use this causal model for explaining certain kinds of mental activity and behaviour and, further, that explanations of this type exhibit conceptual features that elude Anscombe's definition of causality. Thus, the suggestion I am making and will have to substantiate later entails that our causal talk about the mental life and behaviour of people extends beyond talk about efficient causes that contribute to the generation of people's activities. Some causal explanations of human activity are, according to my view, better understood as instances of formal-causal explanation than as instances of efficient-causal explanation. The kinds of activity we typically explain by means of this formal-causal model are characterized by their subjection to some form of causal influence exerted by some nonrational mental mechanism or disposition of the agent. I will give several examples of this type of causation in the following chapters, but as a preliminary illustration the kinds of causal phenomena I am interested in, the reader can imagine, let us say, a person with an unconscious tendency to experience her social environment as cold, distant, and unhelpful. Imagine a person who cannot help viewing her social context under an uncaring, unsympathetic aspect, someone who expects nothing but insensitive, stony-hearted responses from those with whom she lives and interacts, someone who interprets everything that happens to her as a consequence of people's lack of interest and concern for her fate and well-being. Now, in our attempts to understand the behaviour of such a person we will on occasions be forced to interpret what she does in the light of her tendency to experience the world as cold, uncaring, etc., if we want to understand the significance of her actions and doings. For example, if we want to understand why our imagined person did not ask for help in a situation where doing so would have been the natural course of action, say, when moving into a new apartment, we might have to take into account the mental tendency described above if we want to understand the significance of what she does and did not do. What I shall argue in the following chapters is not that our conception of such an unconscious influence is of a causal nature – I take that for granted – but that it is, more specifically, an influence of a formal-causal character. I shall argue that explanations that explain people's behaviour by reference to such mental tendencies are best understood as formal-causal explanations, since they conceptualize the influence of the relevant mental tendency by which they explain the behaviour in question as a formative influence. A formative influence of this

kind differs in significant respects from the generative influence we conceptualize by means of efficient-causal models of explanation; so I shall argue in the following chapters. My main purpose, however, is not to persuade the reader that this is the best way of understanding explanations of this type, but to examine the shape some philosophical problems related to causal explanation assume *if* we approach them from a perspective informed by my account of such explanations. The question that guides my investigation is this: What happens to our philosophical problems related to causation, and what happens to the concepts connected with these problems, if we approach them armed with the distinction between efficient-causal and formal-causal explanation? The discussions initiated in this chapter, about whether and how the problems of determinism and physical reductionism appear from the point of view of different causal prototypes, will continue in the following chapters, but the focus will shift from Wittgenstein's five causal prototypes to the formal-causal model of explanation introduced in this paragraph. In addition to the problems of determinism and physical reductionism, I will also consider the problem of whether reason-citing explanations explain actions causally or not from the perspective of my account of formal-causal explanation.

1.4 Summary

In this chapter, I have overviewed Wittgenstein's five causal prototypes and considered some ways in which his discussion of causation can further our understanding of philosophical problems related to causal explanation. The view of causal language that emerges from Wittgenstein's discussion provides the basis for a sober reassessment of the worries philosophers have had about causal accounts of human nature and activity, most importantly the worries about determinism and physicalist reductionism. By showing that philosophical problems related to causal explanation are always relative to specific causal models, and by exposing the lack of grounds for the assumption that causal language automatically and by itself commits the language-user to some metaphysical theses about the nature of the universe, Wittgenstein allows philosophers to recognize such problems for what they are, in their true scope and stripped of their metaphysical content. Another reason to consider Wittgenstein's remarks on causation is that they help us become aware of the tacit assumptions we bring into philosophical discussions about issues related to causation by prompting inquiries which an essentialist understanding of causal language is likely to suppress or leave unexplored, such as inquiries regarding the interests and purposes served by different strands of our causal

language. The chapter ended with the introduction of a "formal-causal" model of explanation that has been widely recognized by psychologists but is largely neglected in the philosophical literature on issues related to causal explanation. It is this model that will occupy my attention in the remainder of this thesis, beginning in the next chapter, where I consider Jonathan Lear's discussion of the kind of mental causes we refer to when explaining behaviour by formal causes.

2. Jonathan Lear and Formal Causes

In his latest book, *Wisdom Won from Illness*, Jonathan Lear makes a critical remark about what he regards as a pernicious tendency among psychoanalysts:

> …something that went wrong in the history of psychoanalysis: the assumption that the only scientifically respectable causes in the explanation and treatment of human beings are efficient causes. Efficient causes are those antecedent states of affairs—in this case, mental states—that are sufficient to bring about the state of affairs that needs to be explained. It is this image of scientific explanation that is responsible for so much of the controversy—as vituperative as it was fruitless—that plagued the discipline throughout the last century. The fact is that if we want to understand human action and the role of self-consciousness in human action as well as the repetitions, distortions, and disruptions of unconscious fantasy, we need a rigorous understanding of the spontaneity of the human mind as well as the role of final cause in human action. Empiricism, the traditional model of scientific explanation, concerned as it is with efficient-causal explanation of a distinct object of inquiry, is incapable of providing insight into either area.[30]

According to Lear, there has been a tendency among psychoanalysts to assume that the analysand's mental activity is to be conceptualized principally in terms of efficient causes – as antecedent mental states (say, unconscious wishes) bringing about subsequent mental states (say, conscious fantasies) – and this tendency is a manifestation of a general cultural bias concerning the model of explanation that scientists are supposed to rely on in their search for understanding. According to this bias, scientists, as empirical researchers, are supposed to use efficient-causal explanation as their guiding model when designing investigations and interpreting data and results – a model of explanation which, on Lear's view, is incapable of providing insight into "the spontaneity of the human mind as well as the role of final cause in human action." But what are these areas of research that Lear is referring to, and what types of explanation do we need to gain insight into them? Before answering these questions, I need to make a brief detour into Aristotelian causality.

According to Aristotle, there are four types of causes we can refer to in answering the question "Why?": to material, formal, efficient, and to final causes. These are four different ways of explaining the nature of things – four types of causal explanation. We can explain the nature of a dining table, to use a typical example, by describing the matter it is made of (wood), by defining its form (four legs and a board), by identifying the agent that brought it about (a carpenter), and/or by referring to its purpose (to enable dining). To look for these causes is to engage in an Aristotelian causal investigation. And, as the cumbersome "and/or" indicates,

[30] Lear 2017: 156

these four types of causes are not mutually exclusive, but complementary: to bring a causal investigation to a satisfying conclusion, one might have to identify and come to understand all four causes of the object of one's investigation.

This simplified description of Aristoteles' theory of causation gives us a preliminary idea of the notions of *form* and *final cause*, which we need in order to understand what Lear is getting at in the quotation at the beginning.[31] The form of something is that which makes it into the thing it is, its essence, its inner principle of change and rest; and the final cause of something is that for the sake of which it is done, the end-state, or telos, towards which it is directed. What makes these concepts relevant for understanding the passage above, apart from Lear's explicit use of the term "final cause," is the allusion contained in the passage to the changes in scientific theory and practice that overthrew these concepts in the sixteenth and seventeenth centuries. In his first book, *Aristotle: the desire to understand*, Lear discusses some of these changes, among other things, the change in attitudes towards teleological explanations: "Since the seventeenth century teleological explanations have been in disrepute."[32] One of the reasons why teleological explanations fell into disrepute was that they came to be regarded as too metaphysical. The new breed of scientists that came after Galileo and Newton grew uneasy about the use of teleological explanation in the study of nature as they feared that teleological approaches assume the existence of a divine creator who has given the universe its divine purpose. And the concept of form, too, became problematic, as it refers to the kind of unobservable properties that modern science sought to reduce to simpler physical processes that were explainable in terms of efficient causes:

> Since the seventeenth century Western science has moved steadily away from conceiving forms as part of the basic fabric of the universe. It is thought that if we understand all the properties of the matter we will see form as emerging from these properties.[33]

It is to these and to other related changes in our ideas about what counts and what doesn't count as scientific explanation that Lear is alluding in the passage quoted at the beginning. Lear is expressing regret for the ways in which our excessive focus on efficient causes – our inheritance from the forefathers of empiricism – impedes advances in certain areas of scientific research. According to Lear, a one-sided focus on efficient causes precludes us from gaining

[31] For Aristotle's discussion of causation, see *Physics* II, especially II.3-9 194b16-195b30, and *Metaphysics* V.2, 1013a24-35. For Lear's in-depth discussion of Aristotelian causation, see Chapter 2 of his *Aristotle: the desire to understand*, especially pp. 26-42.
[32] Lear 1988: 40
[33] Lear 1988: 20

insight into certain aspects of the human mind, and, within the narrower context of psychoanalysis, prevents advances from being made in psychoanalytic technique. To remedy these problems, Lear attempts to import the Aristotelian concepts of form and final cause into psychoanalysis. The concept of final cause, which I won't discuss in my thesis, is of relevance in the psychoanalytic setting, according to Lear, as it allows the analyst to conceptualize the hoped-for end-state, or telos, of the psychoanalytic treatment, and, based on her conception of this telos, to choose between different approaches at significance moments in analysis.[34] And the notions of form and formal cause, which lie at the core of my inquiry, can on Lear's view help us understand what he describes as "the spontaneity of the human mind." Interestingly, however, when Lear uses the concepts of form and formal cause to characterize relevant kinds of mental activity and behaviour – activity and behaviour that manifest "the spontaneity of the human mind" – it is unclear whether and to what extent the content these concepts acquire in Lear's use corresponds to the content they have in Aristoteles' philosophy.

The concept of formal cause I am interested in was brought to my attention by Lear's *Wisdom Won from Illness*. There, in Chapters One and Two, we are presented with two case reports in which the problems suffered by two of Lear's patients – Lear is a trained psychoanalyst – are conceptualized in terms of formal causes. In the first report, we are acquainted with Ms. A, a person described as "inhabiting a disappointing world." "No matter what happened to her," Lear writes, "she would interpret it under an aura of disappointment." For example,

> if something she wanted occurred—getting promoted at work, asked out on a date by someone who interested her—she would diminish it: "The boss only promoted me because he wanted to promote my colleague, and he was too embarrassed not to include me" or "He invited me out because he got turned down by the person her really wanted to date." [35]

So here is a person with a tendency to experience the events of life as disappointing, a tendency which Lear describes in Freudian terms, as a kind of compulsion to repeat:

> We are, of course, familiar with the idea of unconscious repetition, but in calling the unconscious *timeless*, Freud asks us to envision what the repetitions are about. Each of the individual disappointments—over and over again—supports a structure of repetition. But the structure of repetition itself expresses a timeless thought: *that life shall be disappointing*. The thought functions as though it is an injunction, and its temporality is different from the familiar narratives of conscious life. Instead of a historical narrative of past, present, and future using familiar tensed verbs—"When I was a baby my mother wasn't there for me, now the boss at work lets me down"—the injunction hangs over all

[34] Lear 2017: 156-8
[35] Lear 2017: 18-9

> narratives, informing them with a timeless quality of disappointment. In this way, whatever the particular conscious narrative, a primordial structure of disappointment is timelessly held in place.[36]

Lear conceptualizes Ms. A's tendency to experience life as disappointing as a kind of mental structure, the essence of which he defines by identifying its main ideational component: the thought that life shall be disappointing.[37] This passage is followed by a brief description of how people who suffer from such "structures of disappointment" can become immune to "countervailing evidence" (i.e., evidence that speaks against viewing the world as disappointing), after which Lear introduces the notion of formal cause.

> This gives us a plausible way to understand the "psychic determinism" of unconscious mental life. The point ought not to be that there will always be a hidden, antecedent mental cause determining the will—how could we ever know that?—but that **disappointment functions as a formal cause, casting an aura over the events that do occur and providing them with a misleading and unhappy-making interpretation**. We cannot know with confidence what the chain of efficient causes has been, from past to present to future. But we can have confidence for thinking that ***whatever*** **happens, and** ***however*** **it comes about, there will be a tendency to incorporate it into an interpretive frame in which a sense of disappointment rules**.[38]

What Lear suggests here, in effect, is a new way of understanding mental structures that dispose us to view certain types of situations in certain ways, and, by the same token, suggests a new way of understanding explanations of mental activity and behaviour that refer to such structures: we can think of such mental structures as formal causes that have a distortive influence on people's lives and experiences, and we can think of explanations of mental activity and behaviour that refer to such structures as formal-causal explanations. My inquiries in the following chapters are premised on these suggestions.

In the second clinical vignette that I want to quote at some length, we are acquainted with Mr. B, a patient of Lear's whose personal freedom is restricted by a tendency to hesitate when choices and opportunities come his way, someone who envies other people who, from his point of view, seem better capable of exploring the possibilities of life. The vignette begins with a dream reported by Mr. B at the beginning of a session with his analyst, with Lear.

---

[36] Lear 2017: 19; italics in original

[37] Lear discusses such (Aristotelian) definitions in a passage where he considers the form of natural organisms: "Aristotle thinks that order is ultimately intelligible: it is that which is realized over and over again in natural organisms, it is that which a single definition can capture as the essence of these organisms, it is that which the mind can apprehend" (1988: 29).

[38] Lear 2017: 20; italics in original, boldface mine.

> I'm watching cars drive through red lights. They keep on going through. I have a series of reactions. They're getting away with something. They shouldn't be doing that. I feel angry; why didn't I do that? Could I get away with it? What stands out is: they got away with it. There's no accident, no police siren.[39]

Having related the dream, Mr. B observes,

> Maybe that's why I'm having dreams, seeing other cars racing through red lights. I'm not doing that. It makes me mad. I'm the one who stops for the green light. I even stop for the yellow light. Other people don't and they get away with it.[40]

Then Mr. B begins to associate to the dream, and

> to all sorts of examples, throughout his life, in which he had stopped himself at green lights—countless opportunities that he blocked, though he was hitherto unaware of doing this. Consciously he spent his time envious and angry with other people who could go through red lights—get away with things—and not get caught. He was unaware that he had spent his life—over and over again—wary and hesitating in front of green lights. This is an example of what Freud called the "timelessness of the unconscious." It is a timelessness *of form*. Consciously, history and human development proceed apace; but unconsciously, Mr. B was on the lookout for events and opportunities that could be interpreted as "green lights," and he enveloped them in an aura of suspicion. It was the manner in which he inhibited himself from potential successes. From one perspective, this looks like repetition; the same thing happens over and over again. But if we think of the unconscious as actively imposing a form on life's passing events—creating a structure of hesitation before "green lights" and envy at others who get to go through "red lights"—we can see that this structure is being timelessly maintained […] It is clear that the problem Mr. B faced was not this or that unconscious wish or fear, but a principle of mental functioning that shaped his life. How is one to intervene effectively in such well-organized mental activity—activity that seems to flow of its own accord, outside the reach of self-conscious reason? How could psychoanalysis undo such a formal organizing principle of mental life? [41]

The mental structure Lear discusses here is different from that which he discusses in the first case report, but the way in which he conceptualizes it is the same. Ms. A's "structure of disappointment" and Mr. B's "structure of hesitation" are both conceptualized as formal causes that operate on the lives and experiences of these individuals in systematic, specifiable ways. This is interesting not only because it opens a new way of thinking of certain kinds of mental complexes and dispositions, but also because it gives us a new way of thinking of explanations of mental activity and behaviour that refer to such mental structures. If it is plausible and profitable to think of certain kinds of tendencies to experience life in particular ways as formal causes, it might be plausible and profitable, too, to think of explanations of behaviour that refer

[39] Lear 2017: 40
[40] Lear 2017: 40
[41] Lear 2017: 41; italics in original

to such tendencies ("Ms. A reacted as she did because of her tendency to experience life as disappointing") as formal-causal explanations. This is the assumption that underlies my discussions in the following chapters. Lear is clearly right in pointing out that explanations of this kind are significantly different from the kinds of explanations we normally think of as causal explanations, that is, different from efficient-causal explanations. And though it might seem as though Lear were proposing an entirely new model of explanation, the formal-causal model he discusses here is actually already part of our everyday mental discourse. Or so I shall argue. The Freudian elaborations of Aristoteles' concept of formal cause in the above-quoted passages are Lear's conceptual innovations; but the non-theoretical, bare-bones model of explanation that his innovations build upon is an established part of our mental language and folk-psychology. Furthermore, our everyday-use of this explanatory model does not presuppose our commitment to Aristoteles' metaphysics about form or (in any non-trivial sense) to Freud's theory of the mind.[42]

According to Lear, Aristotle is committed to at least two metaphysical theses concerning forms: the thesis that "Forms must occupy a fundamental ontological position: they are among the basic things that are,"[43] and the thesis that "the primary source of change is form."[44] Neither thesis is implicated in our everyday use of the kind of formal-causal explanations I'm concerned with here. And indeed, it is even questionable whether these theses are implicated in Lear's own use of formal-causal explanations and the concepts of form and formal cause that his explanations are based upon. As far as I can see, Lear's concept of form does not stand in a referential relationship to a "real item in the world," but rather functions as a conceptual tool that enables us to see connections between certain patterns of thought, feeling, and behaviour of his patients. Nor do Lear's concepts of form and formal cause in the passages quoted above seem to involve dynamic "inner principles of change," but rather involve static, conservative compulsions to repeat certain pre-established cognitive and affective patterns. In Lear's use the concepts of form and formal cause are stripped of their metaphysical components. And while Lear's use *is* informed by Freud's theory of the mind, we, as non-psychologists, can use these concepts, and use formal-causal explanations, without familiarity

[42] The qualification about non-trivial senses takes into account the following (trivial) sense in which our use of such formal-causal explanations *does* presuppose our committed to Freud's theory of the mind: insofar as Freud's theory has been assimilated into "folk psychology," our use of formal-causal explanations that presuppose such assimilated parts of Freud will commit us to (a popularised version of) Freud's theory of mind.
[43] Lear 1988: 20
[44] Lear 1988: 35. For discussions of these two metaphysical theses about form in the philosophy of Aristotle, see for instance pp. 20, 22, 26, 28, 35, 40 in Lear's *Aristotle: the desire to understand.*

with Freud's principles of unconscious mental functioning.[45] We can give formal-causal explanations – "Margaret's experience of sexual abuse in childhood made her see an element of abuse in the most harmless of physical gestures between adults and children" – without knowing a thing about the timeless forces of the id.

2.1 Summary

In this chapter, I have considered Lear's suggestion that it can be philosophically clarifying to think of mental structures that have a distortive influence on people's perception and experience as formal causes. In making this suggestion, Lear is, in effect, suggesting the possibility of reconceptualizing explanations of behaviour that refer to such mental structures as formal-causal explanations. This explanatory model itself is not new – explanations that refer to causally significant mental complexes of the agent have a firmly established role in our ordinary mental discourse – but the novelty of Lear's account lies in his illuminating redescription of such explanations as "formal-causal." Folk psychology has assimilated this explanatory model from the technical language of Freudian and post-Freudian psychology to such an extent that we, as laymen, can use it without awareness of the Freudian underpinnings that support Lear's more theoretical usage. In the next chapter, I will look at some features that distinguish formal-causal explanations from efficient-causal ones and will illustrate the kind of unfreedom suffered by people whose behaviour lends itself to explanation by formal causes. Taken together, this chapter and the next provide the background for the investigations in the final two chapters, where I consider the problems of determinism, physicalist reductionism, and the causal status of reason-explanation, from the viewpoint of the formal-causal model of explanation I have begun to outline above.

[45] The parts where Lear discusses the "timelessness" of mental structures refer to one of the four principles that govern unconscious activity according to Freud: "*exemption from mutual contradiction, primary process* (mobility of cathexes), *timelessness*, and *replacement of external by psychical reality*". By describing unconscious processes as "timeless," Freud means that "they are not ordered temporally, are not altered by the passage of time; they have no reference to time at all. Reference to time is bound up, once again, with the work of the system *Cs.* [i.e., consciousness]" (Freud 1915: 187; italics in original).

3. Two Life-Stories and an Account of Formal-Causal Explanation

In this chapter, I seek to elaborate the concept of formal cause introduced in the previous chapter, to illustrate the kind of unfreedom suffered by people who are being acted upon by such causes, and to explicate the logical features of the formal-causal model of explanation that are most relevant for my purposes in the later chapters. The chapter divides into three sections and is based on two made-up examples, or "life-stories." The *first section* begins with a description of a teenager's experience of witnessing a drowning accident and the mental problems brought on by his experience. After the description there follows a commentary where I distinguish two causal perspectives we oscillate between when considering cases of mental disorder and characterize the formal-causal conception of mental illness that emerges from one of the perspectives. The *second section* begins with a description of a somewhat troubled individual who suffers from a fear of abandonment that prevents her from building enduring relationships with others. After the description there follows a commentary where I develop the concept of mental illness outlined in the first section, and illustrate the restricted personal freedom suffered by such troubled individuals. (The purpose of illustrating this kind of restricted freedom is to provide the foundation and background for a later discussion in which I connect the metaphysical problem of determinism with certain moral-existential problems familiar to many people, problems which might be best understood in formal-causal terms.) And the third and *final section* summarises and develops the main points made in the previous sections. The aim there is to give a concise account of the logical features of formal-causal explanations that are most relevant for my purposes in the later chapters.

3.1 The Story of Tom

Tom and Wesley were enjoying their holidays at a seaside resort when the tragedy unfolded. The boys had broken away from their families and were basking in the sun at a secluded section of the beach when Wesley decided to go for a swim. Tom was too comfortable to give up his spot in the sand and stayed behind and watched his friend wade in and swim out toward the buoy that marked the edge of the swimming area. On reaching the buoy, however, Wesley did not turn around as Tom had expected, but kept swimming further and further out until he could barely be seen from where Tom was lying. Then came a faint cry for help and Wesley's head went under. By the time his lifeless body was pulled out of the water, Tom's childhood friend was beyond rescue.

The psychological trauma caused by Wesley's drowning would stay with Tom far into adulthood, causing anxiety and feelings of guilt, intrusive memories, jumpiness, and an irrational fear of the sea. The first few years after the incident were the worst. Any large body

of water would revive the feelings of guilt that surrounded Wesley's death. Even a photograph of a beach could send Tom's heart racing and make him feel dizzy. Having recover from the first shock of Wesley's death, Tom would regain some sense of control over his life, but at the cost of his personal freedom. Excessive precautions had to be taken to keep a safe distance between Tom and large bodies of water. Any activities that entail contact with the sea had to be avoided. Beach vacations and trips to the coast were out of the question. Even swimming pools and bathtubs took on an ominous aspect and became a source of anxiety. All of this took a major toll on Tom's well-being and family-life, causing much grief and frustration, not least for Tom himself, who was aware of the strain his problems were putting on his friends and family.

This account illustrates two points of view we can adopt when describing and explaining mental disorders. The first point of view, which I will call the "etiological perspective," can give us insight into the past events and conditions that contributed to the development of a mental disorder, whereas the second point of view, which I will call the "symptomological perspective," helps us to understand the symptoms experienced by the individual suffering from the disorder. The first paragraph takes an etiological perspective and construes Tom's traumatic disorder in relation to the incident that led to its onset, whereas the second paragraph takes a symptomological perspective and construes Tom's disorder in terms of the influence it came to have on his life and his functioning. (The etiological conception of Tom's trauma can further be thought of as an instance of the causal prototype of Wittgenstein's that I called "reacting to a cause." But compared with the motoric responses I gave as examples of this prototype, such as the reflexive kick of someone having his knee-jerk reflex checked at the doctor's office, Tom's reaction to Wesley's death is temporally more drawn out and involves his exercise of cognitive-affective capacities that play little or no part in reflexive motoric responses.) What has happened in Tom's case is this: Tom has experienced a traumatizing event that has caused him to develop a mental disorder, a disorder which, having been caused, itself comes to act as a cause on Tom's experience and behaviour. These two phases of mental illness – the phase of emergence and the phase of manifestation – call for different kinds of causal story that feature different kinds of causal phenomena. The first kind of story is about past events and conditions and the subsequent events and conditions they have contributed to bringing about. In Tom's case, it is a story about the onset of his post-traumatic stress disorder, triggered by his experience of his friend's drowning. The second kind of story, which can be temporally scattered over the past and the present, is about pathological mental dispositions and the dysfunctional thoughts, feelings, and behaviours they prompt the ailing person to engage in. In Tom's case, it is a story about a set of mental dispositions that prompt him to engage in the cognitive, affective, and behavioural patterns that compose his traumatic

disorder. This second story makes Tom's mental illness appear, not as a reaction to something in the past, but as a force operating in the present. Here Tom's mental disorder is in the giving rather than in the receiving end of a "causal process," and the causal efficacy of his trauma seems independent of the event that led to its onset in the past. This conception of mental illness finds eloquent expression in an oft-quoted simile by Irish novelist John Banville: "The past beats inside me like a second heart."[46] When the protagonist of *The Sea*, Max Morden, makes this comparison, he seems to be saying something like this: "The past is not over and done with but is part of my thinking, feeling and acting in the present." In other words, Morden is thinking of his personal history, not as a series of extinguished events and conditions in a superseded past that contributed to the evolution of the person he has become, but as a piece of living history that he carries within him and continues to influence his acting and being in the present moment.[47] This formal influence of Morden's past endows his surroundings with personal, idiosyncratic meanings, and in that sense enriches his experience; but it also extends to his style of attention, to his moods and his surges of emotion, to his fears and his fantasies and other aspects of his mental life. This conception of the past is in key respects identical with the conception of mental disorder that emerges from the symptomological description of Tom's mental trauma above. From Morden's simile we get the idea of a past acting as a formal cause on the present in the same way as Tom's traumatic disorder – also a residue of the past – acts as formal cause on his experience and functioning in his present.

3.2 The Story of Kaitlin

Kaitlin is a 40-year-old woman with a history of troubled relationships. Her problems in love are due to a fear of desertion that prevents her from developing deep and lasting relationships with men. To avoid the experience of being abandoned by someone she loves, Kaitlin never permits herself to become emotionally invested in the men she is dating. She only gets involved with men who aren't looking to settle down, and her relationships always end before love has a chance to begin. Kaitlin systematically avoids men who are capable of intimate arrangements while being drawn to young guys who shun companionship and commitment, typically to young careerists of the creative-intellectual type: academics, musicians, interior designers, and the like. If a lover tries to turn their relationship into something more serious, Kaitlin gets anxious and finds a quick exit. And if a candidate for a long-term relationship comes along and shows interest, Kaitlin becomes vaguely avoidant and explains her evasive behaviour to herself by finding imagined fault with her suitor: "too possessive," "too bourgeois," "too boring." Her friends suspect something is wrong, but when they ask Kaitlin about her avoidant behaviour,

[46] Banville 2005: 13

[47] The psychoanalyst Hans Loewald has written about "the past which the patient carries within him as his living history." (Loewald 1972: 144)

she tells them not to worry: "I'm happy with the way things are going – I just don't want anything serious at the moment." And truly, Kaitlin does seem cheerful when she is with others – a happy, outgoing personality – but when alone, she feels anxious at times without knowing why.

Kaitlin's fear of abandonment has its roots in her relationship to her parents. When she was 10, her childhood ended abruptly when her father lost his job at the local steel factory and took to drinking. Before long, the father was a confirmed alcoholic, and the life of the family changed profoundly for the worse. To make ends meet, Kaitlin's mother had to take a second job and, because of her long hours, was never at home. Kaitlin suddenly found herself left to her own devices as her parents were always away – one working and the other drinking at the neighbourhood pub. This was a major blow for a girl who throughout childhood had received so much attention from her parents. Kaitlin became depressed and began to isolate herself from others. She couldn't trust people anymore, and love took on a sinister new meaning for her. Things started to look up when the family had to move to a smaller apartment in a neighbouring community and Kaitlin had to change schools. The move seemed a calamity at the time but turned out to be a blessing in disguise, as it forced Kaitlin to meet new friends who would help her to recover her trust in people. Kaitlin made a new start in life and her future was looking brighter. She was about to begin her sophomore year at university when the second disaster struck. Without explaining why, Kaitlin's boyfriend of ten months suddenly broke up and refused any contact with her. Kaitlin was devastated. She had been living in the belief that their relationship was going wonderfully, only to have the rug pulled from under her feet. To be dumped in such a callous manner was extremely painful, but Kaitlin was able to carry on her studies and would eventually graduate and start working as an architect.

Despite her fears and occasional anxiety, it would seem inapt, I think, to describe Kaitlin's problems as "symptoms of mental disorder." She has had a particularly hard time growing up, it is true; but there is something "normal" about her predicament too. I am not alluding to her problems with intimacy, of course, but more generally to the influence her past continues to have on her life in the present. Kaitlin's story is an example of the many ways in which our early attachments, dependencies, losses, and betrayals may influence our relationships later in life. I have portrayed her as a person who is largely unaware of the ways in which her fear of being abandoned by those she loves skews her perspective on herself and others, a person unaware of the ways in which her fear of abandonment alienates her from her emotional needs and prevents her from building deep and lasting relationships with others, especially with men. In a similar way as large bodies of water took on a sinister meaning for Tom after his traumatic experience of witnessing his friend's drowning, men who seek closeness and intimacy in their relationships with women acquire a threatening significance to Kaitlin because of her experiences of neglect and abandonment in childhood and early adulthood. But unlike Tom, who was aware of his fear – of its cause, of its effects, and its irrational nature – Kaitlin, who is unable to explain her avoidant behaviour with men, gives us reason to doubt whether, or to

what extent, she is consciously aware of her fears. Following Lear, we may think of Kaitlin's inability to account for her avoidant behaviour as a symptom of lacking "intrapsychic integrity": the conscious part of Kaitlin's mind that rationalizes and finds excuses for her behaviour is in conflict, and lacks the ability to communicate, with the unconscious part of her mind that prompts her to act as she does.[48] This intrapsychic tension and lack of integrity not only affects Kaitlin's ability to explain her conduct with "eligible suitors" but also prevents her from gaining control over the compulsions that are holding her back in life and love. The first step towards easing Kaitlin's fears would involve raising her fears to her conscious awareness so that she could begin to fix the problems they are creating for her. To overcome her fears in anything like a complete manner, however, other procedures might be needed as well. Without going into a general discussion of the therapeutic methods and procedures by which people can overcome irrational fears and other anxiety-related problems, I want to describe very briefly and schematically three such procedures that someone like Kaitlin might undertake to overcome her fears. The point of describing these procedures is to give us some idea of the measures a person whose relationships are unconsciously being shaped by a destructive "mental structure" (a formal cause) might need to take to gain more freedom in her life and relationships. The notions of freedom and unfreedom that emerge from the following descriptions are contrasted in the next chapter with the notions of freedom and unfreedom that go with the metaphysical doctrine of determinism.

(1) *Gaining conscious access to her fears.* The first thing for Kaitlin to do would be to raise her fears to the level of self-reflective awareness, not just in the sense of gaining access to knowledge and information about relevant "unconscious processes" at work in

[48] For an Aristotelian take on intrapsychic integration, see pp. 23-6 in Lear's *Wisdom Won from Illness*, and Chapter 2 in its entirety. I should note, however, that the idea of psychic integration is not an innovation due to Lear but a concept that underpins the theory and methodology of psychoanalysis from early on. Indeed, it's hardly surprising that a conception of the mind that conceives of mental ill-health in terms of psychic divisions, partitions, and subpersonalities, conceives mental health as a condition in which the subparts of the mind form a relatively coherent, integrated whole. For exemplary passages about psychic integration in the Freudian corpus, consider, for instance, the following: "In the process of a child's development into a mature adult there is more and more extensive integration of his personality, a co-ordination of the separate instinctual impulses and purposive trends which have grown up in him independently of one another" (S.E. 18: 79-80). And another passage: "The ego is an organization. It is based on the maintenance of free intercourse and of the possibility of reciprocal influence between all its parts. Its desexualized energy still shows traces of its origin in its impulsion to bind together and unify, and this necessity to synthesize grows stronger in proportion as the strength of the of the ego increases. It is therefore only natural that the ego should try to prevent symptoms from remaining isolated and alien by using every possible method to bind them to itself in one way or another, and to incorporate them into its organization by means of those bonds" (S.E. 20: 98). Freud's follower and Lear's "mentor," Hans Loewald, also writes about psychic integration: "…ego development is a process of increasingly higher integration and differentiation of the psychic apparatus and does not stop at any given point except in neurosis and psychosis" (1960: 224).

herself ("My therapist has told me that I'm unconsciously afraid of…"), but also in the sense of developing her ability to reflect on herself and her personal history from the viewpoint of her fears in an emotionally laden manner.[49] This procedure might require involvement in various therapeutic activities. Kaitlin might need to (*a*) recall and consider the events and conditions in the past that contributed to her fears, (*b*) to locate the larger constellation of ideas, beliefs, desires, passions, memories, fantasies, and other such elements of her mental life that form part of her fears, and (*c*) to learn to identify her fearful tendencies and to build the self-control needed to refrain from acting on them.

(2) *Breaking destructive patterns of thought, feeling, and behaviour, and exploring new ways of acting and relating to others*. Kaitlin would need to loosen the dispositional patterns her fears have prompted her to develop over the years and to challenge the maladaptive, irrational beliefs, wishes, passions, desires, fantasies, and other such attitudes, that underlie and support them. For example, if Kaitlin's avoidant behaviour is supported by an unconscious belief that any man whom she might get involved with is bound to abandon her no matter what she does, Kaitlin might need to make herself more receptive to "contradictory evidence" that can challenge her unfounded belief; or if her conduct is the result of her wish never to be hurt again, she might need to call this wish into question by reflecting on whether the risk of being abandoned by someone she cares for is a good reason to avoid all intimacy in life. Besides facing up to her fearful tendencies and challenging the unfounded beliefs, wishes, desires, and other attitudes that support them, Kaitlin would need to explore alternative ways of acting and relating to others, new ways of acting and being that she can endorse based on sincere reflection on what is good, right, desirable, and worth pursuing in life. In contrast to the old and destructive patterns that Kaitlin had developed and maintained over the years in an unreflective manner, these new ways of acting and relating should have substantive grounds and considerations speaking in their favour. While her old patterns were supported by a set of interlocking desires, wishes, and beliefs that Kaitlin, on reflection, should feel compelled to reject as unreasonable or inappropriate, these new patterns should reflect a more clear-eyed and considered view of her situation in life and the moral questions arising from her situation.

(3) *Cultivating new ways of acting and relating to others until they become second nature*. To overcome her fear in anything like a complete manner, Kaitlin would need to "re-condition" her disposition to think, feel, and act in fearful ways, so that her automatic reactions to situations that previously elicited fearful responses gradually come to express her more considered view of things. By consciously refraining from acting on impulses of fear and systematically repeating non-fearful ways acting and doing things, Kaitlin should be able to replace her old patterns of interaction with new ones that

[49] David Finkelstein, drawing on Freud, claims that "it is a defining characteristic of our unconscious mental states that we lack the ability to express them merely by self-ascribing them. Like all mental states, the unconscious ones may be expressed in our behavior. But what's distinctive about unconscious mental states is that we are unable to express them by self-ascribing them. If Harry unconsciously believes that he's unlovable, he might express this belief in any number of ways. But not by saying (or thinking), 'I believe that I'm unlovable.' Finkelstein 2008: 119-20. For a discussion by Freud on the relationship between language and the distinction between conscious/preconscious and unconscious states, see his 1915 article, "The Unconscious", especially section VII.

> harmonize with and reinforce the beliefs and desires she has arrived at through self-examination and moral reflection on how to live her life.[50]

So the process of overcoming her fears might require Kaitlin's engagement in various therapeutic activities that might give her more freedom in her interpersonal relating and open up new possibilities in life.[51] This effort to reorganize her mental life can be thought of as an attempt to reconcile and facilitate communication between the conscious and unconscious parts of her mind – a process of intrapsychic integration. Under this construal, the overcoming of her fears would involve an effort to resolve the conflicted relationship between different parts of her mind and of bringing them into better accord with one another.

Another way to think of this therapeutic endeavour – one that makes fewer assumptions about the nature and operations of the mind – is to consider it as an attempt to bring about a shift in Kaitlin's attitude to her fears, a shift that would allow her to gradually correct her fearful ways of thinking, acting, and relating to others. On this view, the struggle to face up to her fears would require Kaitlin's recognition, not only of her fears, but of the possibility of countering her fears in a purposeful and deliberate manner – a recognition by which Kaitlin would come to regard her fears as changeable parts of her personality that she herself is responsible for. What this recognition would amount to is a sort of "identification" with her fears. Kaitlin would come to "identify with her fears" in three senses: (1) in the sense of recognizing that her fears are part of her character and thus fall within the scope of her responsibility; (2) in the sense of becoming able to reflect on her life (past and present) from the viewpoint of the beliefs, desires, and other attitudes that form part of her fears, while simultaneously being aware of these attitudes not being founded on truth and fact, but rather being irrational, self-destructive, or in other ways "inappropriate"; and (3) in the sense of realizing that her fears are sensitive to her reflection on relevant questions of a moral nature, such as questions about the role that love should play in the good life, questions about what is reasonable to be afraid of in close relationships with others, questions about how to deal with irrational fears in a responsible manner, etc.. So the overcoming of Kaitlin's fears would require an attitudinal shift on Kaitlin's

[50] These therapeutic activities are supposed to represent a combination of the kind of procedures a contemporary psychodynamic therapist might use – someone whose practice is informed by a mix of psychoanalytic and cognitive-behavioural theory – in the treatment of patients with irrational fears and anxiety-related problems. For some core texts in the Freudian corpus about the treatment of anxiety disorders, see *Studies on Hysteria*, *A Case of Hysteria*, *Little Hans*, *The Rat Man*, and *Inhibitions, Symptoms, and Anxiety.* For a classic cognitivist treatment of the topic, see *Cognitive Therapy and the Emotional Disorders*, by Aaron T. Beck.

[51] Lear writes, "…psychoanalysis has a final cause—however open-ended, indeterminate, continuing, and active it may be: there is something that the psychoanalytic process aims to promote. Freedom is the final cause of psychoanalysis. Freedom is the kind of health that psychoanalysis aims to facilitate" (2017: 150)

part that would allow her to identify with her fears and make them answerable to the ideas and beliefs (about life, love, and things in general) she would arrive at through reflection on moral questions such as the ones above. In the course of this shift, Kaitlin's "state of alienation" – a state in which Kaitlin is unable to recognize her fears for what they are and experiences them (insofar as she is aware of them at all) as independent, alien forces beyond her control – would give way to a "state of identification" in which Kaitlin would relate to her fears (or what is left of them at this point of the therapeutic process), not as empirical facts about her mental constitution, but as changeable parts of the person she is and would like to become.[52]

In the previous paragraph I invoked the concepts of alienation and identification as means for understanding the efforts by which people like Kaitlin can overcome their irrational fears and anxiety-related conditions. I could end the discussion of these concepts here, but since they are related to another distinction that will bear on the discussion of determinism in the next chapter – the distinction between "involuntary" and "voluntary" actions – it might be worth the while to explore them a bit further and try to bring out their connection to the latter distinction. I will do so by considering some examples of involuntary and voluntary phenomena, beginning with movements of the body and proceeding to involuntary mental phenomena.

Here are two bodily movements that most of us are familiar with, the first of which is involuntary, and the second is voluntary: the reflexive jerk of a patient's leg when the doctor taps his knee with a hammer, and the kicking motion of a player shooting a penalty shot in a game of football. What are the relevant differences that distinguish these two movements of the body – relevant, that is, with reference to the concepts of alienation and identification? Here are three differences that seem relevant: Unlike the penalty shot, the reflexive movement is *beyond the patient's control* and is *lacking in understanding* and *outward expression*. The execution of the movement involves no conceptual processes on the part of the patient, and the reflex, viewed from the doctor's perspective, expresses neither the patient's attitude to, nor his understanding of, the event that triggered the reflex or the overall context of the situation at the doctor's office. The person having his reflexes checked cannot intelligibly identify with the jerky movements of his legs as a football player can identify with the kicking motion involved

[52] This paragraph draws on Richard Moran's paper, "Frankfurt on Identification." See especially pp. 146-147. The distinction between "active-engaging" and "passive-alienating" ways of relating to aspects of one's mental life can also be found in Freud's writings. In "Remembering, Repeating and Working Through", to give an example, Freud says that the patient's illness "must no longer seem to him contemptible, but must become an enemy worthy of his mettle, *a piece of his personality*, which has solid ground for its existence and out of which things of value for his future life have to be derived" (1914: 152; my italics).

in a penalty shot that she is about to shoot in an important game: a shot which the player can pour her *concentration* into, which can reveal her *command* of the ball – the product of years of *practice* – to those attending the game, a shot which the player can feel *responsible* for, can make her feel *proud* or plunge her into *despair*.

I have mentioned some ways in which one cannot identify with certain reflexive movements of one's body. But what about involuntary mental phenomena? In what sense can people who suffer from, say, intrusive thoughts and mental images be "alienated from," and "incapable of identifying with," those parts of their mental lives? Imagine intrusive thoughts of a violent or a sexual nature, such as thoughts about jumping from tall buildings, about using knives to harm others, about having sex in public places, and so on. To suffer from thoughts and images of this sort is to be subject to ideas – unwanted, distressing, uncontrollable ideas – one experiences as incompatible with one's values and conscious desires, with one's moral convictions and sense of self. In contrast to the alienation that may accompany movements of the body that bypass one's autonomous agency completely, this type of alienation involves phenomena that originate from semi-independent mental structures that override one's moral judgments and rationality.[53]

So much for now about alienation and identification. I will return to these concepts in the next chapter, but now I want to end this section by addressing a misunderstanding which the language I have used here might give rise to. The misunderstanding concerns the notion of mental health implied by my descriptions of pathological, and borderline pathological, mental activities and symptomatic behaviours. The vocabulary I have used to describe various aspects of mental problems, such as the words "involuntary," "unconscious" and "irrational," might suggest an inclination on my part to think of mental health as a total absence of these qualities, as a condition in which the individual has perfected her capacity for self-governance and self-

[53] I have borrowed the term "semi-independent structure" from Donald Davidson, who argued that the Freudian mind "contains a number of semi-independent structures," which are "characterized by mental attributes like thoughts, desires, and memories" (Davidson 1982: 290). In Davidson's use, the term "semi-independent structure" means something like "sub-personality," or "mental complex." The definition Laplanche and Pontalis give of "complex" in *The Language of Psychoanalysis* is in line with Davidson's characterization: "Organised group of ideas and memories of great affective force which are either partly or totally unconscious. Complexes are constituted on the basis of the interpersonal relationships of childhood history; they may serve to structure all levels of the psyche: emotions, attitudes, adapted behaviour." It is particularly behaviour influenced by enduring, habitual, encompassing, complex mental structures of this kind I am interested in in my thesis. To explain behaviour by reference to a mental complex of this kind is, on my account, to give a formal-causal explanation. Having said that, I also see no reason why the account of formal-causal explanation I provide here couldn't be applied, in its current or an adjusted form, to dispositional explanations that refer to "simple" mental attributes, such as, beliefs, desires, memories, inhibitions, etc.

determination and any activity she engages in issues from her autonomous rational agency. Accordingly, the pursuit of mental health would involve Kaitlin in an effort, not so much to gain conscious access to her fears, but to become Conscious and Rational in some absolute sense, perhaps to harness her unconscious mind and gain perfect control over herself. This is not how I understand mental health. What my discussion of the procedures by which Kaitlin might overcome her fears is meant to suggest is that therapeutic processes of this sort are more a matter of replacing destructive unconscious impulses and tendencies by new and less destructive ones than a matter of abolishing unconscious mental activity. None of my formulations are meant to imply the possibility of unifying the mind into an internally consistent, voluntaristic, reflective agency that can fashion itself and its components however it pleases. The concepts of intrapsychic integration and alienation/identification were introduced here as tools for understanding the conflicted relations that can exist between different parts of the mind – between cognitions and affects, between the conscious and the unconscious – parts which can in various ways and to varying degree be brought into accord and communication with one another. The concept of intrapsychic integration presupposes that communication between different parts of the mind can be disrupted and blocked as well as facilitated and restored, but not that the parts can be united so that they form a single unified entity. And similarly, the concept-pair alienation/identification carries no assumption of a possible ideal state in which the individual has "become one" with her mental life in its entirety. Rather, if these concepts presuppose anything regarding the pursuit of mental health it is this: that there can be no end to a person's efforts to integrate her mind and identify with beliefs, desires, wishes, memories, and other mental contents and attitudes she finds painful and at variance with her self-conception.

3.3 Summary and Additional Remarks

In the discussions above, I have chiefly tried to do three things: (1) to distinguish formal-causal explanations of behaviour from efficient-causal explanations; (2) to draw out some of the key features of the conception of mental disorder that emerges from formal-causal explanations; and (3) to outline the notions of freedom and unfreedom that go with this conception of mental disorder by describing some therapeutic procedures by which people who suffer from fear and anxiety-related issues might try to overcome their problems. I opened the discussion by distinguishing two perspectives we oscillate between when we consider mental disorders: the

etiological and the symptomological perspectives. The *etiological perspective*, I explained, gives us a view of the past circumstances and events that contributed to the formation of a person's mental disorder, whereas the *symptomological perspective* affords us a view of the symptoms suffered by the afflicted person. These are two distinct ways of conceiving mental illness. When viewed from the etiological perspective, mental illnesses appear as *reactions* to the events and conditions of the past that contributed to their formation, and causal explanations articulated from this perspective take the form of *efficient-causal explanations*. Such explanations give mental disorders their meaning by linking them to the past conditions and events that contributed to bringing them about. When viewed from the symptomological perspective, by contrast, mental disorders appear as *mental structures* that exert a formative influence on the afflicted person's perception, experience, and functioning. The causal explanations we articulate from this perspective take the form of *formal-causal explanations* and give mental disorders their meaning by linking the dysfunctional patterns of thought, feeling, and behaviour that together compose the mental illness. These two types of explanation, efficient and formal-causal, though logically distinct, are not mutually exclusive or in any way oppositional. The relationship between them is rather one of complementarity and relative self-containedness. That they complement one another shows in our tendency to combine efficient-causal and formal-causal explanations when describing and explaining mental disorders. A clinical vignette in a psychological study, to give an example, would typically include explanations of both sorts. And their relative self-containedness, in turn, shows in our ability to form some understanding of mental disorders based on one type of explanation without having access to explanations of the other type.[54] Without arguing the point, I also suggested the possibility of conceiving mental dispositions of a less pathological nature along similar lines, as formal causes that pattern our perception and experience of the world in ways that we ourselves may not be aware of.

But if the etiological and symptomological perspectives are concerned with different types of causes – with efficient causes and formal causes, respectively – what kinds of causal processes or interactions, more precisely, do these causes partake in? What are the distinguishing features of the kinds of causation that these two types of causes are involved in? Based on my

[54] Having said that, it is true that in contexts where mental disorders are considered wholly from one perspective to the exclusion of the other, the omitted perspective is often familiar to the interlocutors from before. For instance, if Tom's friends were to have an extended discussion of Tom's trauma wholly from the symptomalogical perspective, their ability to understand each other and keep the discussion going without considering the developmental story of Tom's trauma – i.e., without shifting to the etiological perspective – would be due to their prior familiarity with the developmental story of his trauma.

discussions of Tom and Kaitlin's cases, the key distinguishing features might be described as follows. The symptomological perspective, which views mental activities and behaviour in reference to their "structural determinants," construes causation as a *subjection to a patterning influence*, whereas the etiological perspective, which views mental activities and behaviour in reference to their "historical determinants," construes causation as a *subjection to a generative influence*. The former conception – causation as a patterning influence that *shapes* human experience and activity – involves a disposition on the part of the agent to view the world in a certain way and to repeat certain pre-established pattern(s) of thought, feeling and/or behaviour, while the latter conception – causation as a generative influence that conditions the *emergence* of human experience and activity – involves a motivational effect generated by a particular event, state of affairs, or process (be it neuro-chemical or socio-economic in nature) that produces an impression and/or spurs the agent to some form of activity. The comparison that comes to mind when thinking of such patterning influences is with the organizing power of the Kantian mind. Analogously to the way in which Kant's faculties of "intuition" and "understanding" organize the data of sense according to an array of transcendental "categories," mental structures acting as formal causes organize the world of experience by investing objects and events with idiosyncratic meanings, by making certain spatial locations and encounters with people evocative of the past, and by otherwise charging the world of experience with significance and emotional valence. The comparison with Kant however only goes as far as *patterns of experience* are involved and fails to capture the *activity-patterns* that accompany them.[55] These two aspects of formal causation – the patterning of experience and the patterning of activity – are related to one another functionally and developmentally, in the sense that they operate in unison and tend to emerge, develop, and expire together.[56] And a second limitation to the comparison with Kant is that his categories are supposed to be *universal*, *unchangeable*, and *spread out over human consciousness* in its entirety,[57] whereas the patterning influence of mental disorders (conceived as formal causes) is *subjective*,

[55] Kant's categories, which impose a certain structure on the objects of experience, do not uphold or support any particular cognitive, affective or behavioural patterns, but rather condition the possibility of our having *any* such patterns at all, in the sense that they condition the possibility of our experiencing the objects that such patterns are directed at *as objects*.

[56] In Tom's case, these two aspects of formal causation concerned his experiential and activity patterns with regard to large bodies of water; and the relationship of dependence between experiential and activity patterns is reflected in the fact that Tom is likely to have abandoned the cognitive and behavioural patterns (the activity patterns) he had developed in relation to large bodies of water had those bodies of water lost the threatening character they assumed in his experience due to the patterning influence of his mental disorder.

[57] For example, the spatial and temporal forms of objects of experience are, according to Kant, the same for everyone, are unalterable, and are supposed to hold good in all states of consciousness.

changeable, and *more local* in the sense that the influence can be limited to particular objects of experience.[58]

These are the principal points I want to make about formal-causal explanation, formal causes, and formal causation. My treatment of these matters has, admittedly, been brief and schematic, but a more comprehensive descriptive account is beyond the scope of this study. My primary interest is, after all, not so much with formal-causal explanation as such as with the changes in the philosophical landscape that occur when we approach certain problems concerning causal explanation armed with the distinction between efficient-causal and formal-causal explanation. I will add some detail to my account of formal-causal explanation in the chapters that follow, but my main objective is to open new perspectives on old problems by approaching them in the light of the understanding gained in the foregoing discussions. In the next chapter, the focus will be on the problems of determinism and physicalist reductionism – two problems that are often thought to plague causal accounts of human nature and activity. And the final chapter is a discussion of the question of whether reason-citing explanations of actions and doings are a type of causal explanation or not.

[58] Here I am drawing on Freud's discussion of the distinction between reaction-formations and general dispositions. See Freud 1926: 158.

4. Determinism and Physicalist Reductionism from the Viewpoint of Formal-Causal Explanations

The condition of being unfree is often discussed among philosophers in terms of "predetermination." A philosopher might say, for instance, that someone is "predetermined to act in a certain way and in no other," meaning that that someone's future action is in some strong sense decided in advance. The timing has been set and the agent has been chosen – everything has been prepared down to the least detail – all that remains is for the action to play itself out according to the laws of nature. This manner of speaking obviously comes with some metaphysical baggage, some of which I will unpack in a moment. But there is also a neutral way of using this metaphysically suggestive term that brings into view certain aspects of the unfreedom involved in formally caused mental activity and behaviour. One can speak of unfreedom in terms of "predetermination" without thereby implying any metaphysical doctrines about the deterministic nature of the universe. I want to devote the opening section of this chapter to a brief discussion of this innocuous use of "predetermination," mostly to provide myself with a common denominator that will facilitate comparisons between the distinct forms of unfreedom that go with causal determinism, on the one hand, and formal-causal explanations of human activity on the other hand. Having discussed this sense of "predetermination," I will pass on to considering the question of why formal-causal explanations do not lend themselves to metaphysical speculation about the fundamental nature of the universe as readily as efficient-causal explanations do. The first two sections of this chapter are discussions of these topics, that is, discussions of the notion of "predetermination" that goes with formal-causal explanations and of the incompatibility of formal-causal explanations with the metaphysics of causal determinism. In the last two sections I turn to the notion of freedom that goes with formal-causal explanations and to the objections against casual outlooks on human life presented in Chapter 1. My aim there is to explain why formal-causal explanations of mental phenomena and behaviours are resistant to those objections.

4.1 Two Kinds of Unfreedom

To get a preliminary sense of the notion of unfreedom that goes with formal-causal explanations of mental activity and behaviour, let us begin by considering some examples.

> "Tom's trauma led him to avoid the seaside while travelling in Spain."

> "Kaitlin's fear of intimacy has kept her emotionally isolated from her lovers."
>
> "Robert's obsession about burning down the house by accident caused him to doubt whether the stove had been properly turned off."

These explanatory statements are about individuals who cannot help experiencing certain types of situations in certain ways and acting accordingly. That is what they have in common – the common plight of these people – the sort of unfreedom they are suffering from. Indeed, from the explanations above these individuals may come across as compulsively driven to the point that someone versed in philosophical jargon might feel inclined to characterize their condition in terms of "predetermination." Tom, who is bound to stay at a safe distance from large bodies of water, and Kaitlin, who dumps her lovers if they start developing real feelings for her, and Robert, who is overcome by anxiety and checks the stove several times when leaving the house – these are troubled individuals whose personal freedom has been restricted in ways that might prompt one to think of their lives and behaviours as being "predetermined" in one sense of the word. This is not the standard sense in the literature on determinism, however, and the idea of unfreedom evoked by descriptions of people whose lives and activities are "predetermined" in this sense differs from the idea of unfreedom we associate with the standard philosophical usage. The idea of unfreedom suggested by the nonstandard sense in which I am using the word "predetermined" is familiar enough to most people from various kinds of moral-existential questions we ask ourselves at various moments in life, such as questions concerning the influence our past experiences might have on our choices and aspirations in the present. The idea of unfreedom we associate with the standard philosophical usage, by contrast, is a theoretical construct that arises, not from the moral-existential concerns that pervade our experienced lives, but from metaphysical speculation on the fundamental nature of the universe. To give an idea of this kind of speculation, I will quote a passage from Carroll and Markosian's textbook on metaphysics, where they invite you, the reader, to

> consider some action of yours, and focus on the fact that (if Determinism is true) it was physically necessary a million years ago that you would perform that action. It begins to look like you couldn't be responsible for that action, since, on our supposition, it was "pre-determined" by past conditions and the laws of nature.[59]

The speculative kind of unfreedom that Carroll and Markosian are considering in this passage involves a state in which every event in the history of the universe is necessitated by antecedent events and conditions, together with the laws of nature. The future is conceived as "completely

[59] Carroll & Markosian 2010: 51.

determined" by the past, in the sense that "for any time, *t*, given the way the world is at *t*, and given the laws of nature, there is only one way that things could continue."[60] Clearly this unfreedom is of a different order from the one I had in mind when using the word "predetermined" in connection with the explanatory statements at the beginning of this section. In describing the behaviours of Tom, Kaitlin, and Robert as "predetermined," I was using the word to refer to an internal psychological compulsion, whereas Carroll and Markosian are using the word to refer to an external metaphysical constraint. My use of "predetermined" was an invitation to reflect on the restricting effects that deeply ingrained patterns of thought, feeling and behaviour can have on our personal freedom, whereas Carroll and Markosian are inviting us to consider the course of a law-governed universe. As unrelated as these two types of unfreedom might seem at first glance, it is not implausible to suppose that the metaphysical variant finds experiential support from the non-metaphysical variant. It is not unlikely, that is, that the metaphysical problem of determinism – a theoretical, intellectual problem – is prompted by, and finds a degree of legitimacy, from the ways in which we conceive and experience our agential freedom at certain moments in life, as attenuated and constrained. If this is so, and the metaphysical problem of determinism has an experiential dimension that gives it its meaning and urgency, then the "problem" cannot be adequately solved by metaphysical speculation alone. The existential problem of how to transcend closed patterns of experience will not be resolved by any intellectually satisfying articulation of the fundamental character of the universe, but needs to be recognized and understood for what it is – an ongoing difficulty of life that admits of no final solution. It is one of my hopes that my thesis will stimulate further thinking on this experiential dimension of determinism, as well as critical assessments of the suggestion that this dimension can be understood in formal-casual terms.

4.2 Reasons Why Formal-Causal Explanations of Behaviour Do Not Encourage Deterministic Thinking

In the previous section, I contrasted the unfreedom expressed in behaviours considered under formal-causal descriptions with the unfreedom we associate with the metaphysical doctrine of determinism. We saw the difference between these two kinds of unfreedom reflected in the

[60] Carroll & Markosian 2010: 50. Peter van Inwagen, in "The Consequence Argument", expresses the same thought as follows: "Determinism says that the past (the past at any given instant, a complete specification of the universe at any given instant in the past) and the laws of nature together determine everything, that they leave no open possibilities whatever" (2008: 453-4).

difference in meaning between the standard philosophical use of "predetermined" and the meaning this word took on when applied to behaviour conceived in formal-causal terms. Since the standard philosophical use is associated with a deterministic conception of unfreedom, a pressure to reconceptualize predetermination was created when the word was applied to behaviours that evoke a non-deterministic conception of unfreedom. Furthermore, since the unfreedom of determinism has conceptual links to certain metaphysical ideas that are unconnected to the unfreedom expressed in behaviour that lends itself to formal-casual explanation, such behaviour fails to prompt those ideas and the sort of metaphysical speculation in which they normally occur. This point is related to the fact that the metaphysical ideas I have in mind depend for their application on certain logical features that characterize efficient-causal explanations but are missing from the logic of formal-causal explanation. Consider, for instance, the metaphysical notion of endless causal chains extending backwards in time to infinity. The reason why this notion does not chime with formal-causal explanations is that it construes the relation between cause and effect according to the efficient-causal model, as a generative relation between antecedent and consequent events, whereas formal-causal explanations construe the relation between cause and effect as a formative relation between a mental structure and the activities and experiences it shapes.[61]

Another metaphysical idea that goes awkwardly with formal-causal explanations is the idea that causal phenomena are "necessitated" by natural laws. This idea receives support from efficient-causal statements of the nomological variety, which, as the reader might recall from Chapter 1, take the logical form of universal propositions: "given an event of kind A, an event of kind B follows"; "given an event of kind B, it will have come out of an event of kind A." It is precisely because they are phrased as generalizations that nomological statements raise the question of what it is that necessitates the relationship between events related as cause and effect. This question, which does not arise in connection with single-case causation, can be understood in two ways, both of which anticipate deterministic responses. If understood as an ontological question about the universe, the deterministic response might look something like this: "The unfailing succession of events of kind A and kind B is necessitated by natural laws." And if understood as an epistemological question about our knowledge of the universe, the deterministic response might look something like this: "Judgments concerning the causal

[61] One might also wonder whether the concepts of causal chain and mental structure are compatible, and if so, how they relate to one another. Does it make sense to think of a set of mental dispositions as a sequence of temporally restricted events?

relations between events of kind A and events of kind B are underpinned by the laws of nature." The main thing here is not the difference between the ontological and the epistemological sense of this question, however, but the fact that the question (in both of its senses) springs up in linguistic contexts where efficient-causal statements of the nomological variety are concerned. Where formal-causal statements are involved, this question simply does not arise, since formal-causal statements do not take the form of universal propositions. Unlike nomological statements, formal-causal ones can have explanatory merit regardless of whether the phenomena they explain or describe give us reason to regard them as manifestations of exceptionless regularities. That is to say, the truth of formal-causal statements about mental phenomena and behaviour does not depend on their universal validity. Take this formal-causal statement as an example: "Tom's phobic fear of the ocean made him lose self-control when the airplane flew over the coast of Spain." The truth of this assertion does not depend on whether Tom will react in the same way on all relevantly similar future occasions. If Tom manages to stay calm on later occasions of seeing the Spanish coast from an airplane, his state of calm on those occasions will not automatically invalidate this explanatory statement, or demand that it be qualified in some way. If this statement about Tom is true, it will remain true in the future, no matter how Tom reacts on future occasions. The reasons for this is that formal-causal statements, unlike nomological ones, allow for single-case causation.

So there are certain metaphysical ideas associated with causal determinism that find no purchase where formal-causal statements are concerned, and this partly explains why formal-causal statements do not encourage deterministic thinking. However, there are other reasons besides this why formal-causal statements do not steer our thoughts in the direction of determinism. I would like to consider two of those reasons briefly before passing on to the notion of freedom that coemerges and forms a concept-pair with the notion of unfreedom that accompanies formal-causal explanations.

In contexts where human behaviour is viewed and explained causally, the temptation to draw deterministic conclusions about the agent's lack of autonomy seems to increase as the causes by which the explaining is done become temporally and qualitatively more remote, distinct, and different in nature, from the *mental powers* by which humans have traditionally been thought to exercise self-governance. In Western culture, the foremost of these include the *will* and the capacities for *self-consciousness* and for *reason*. Thus, we are under more conceptual pressure to assign responsibility to an agent for a piece of destructive behaviour in contexts of explanation where our causal statements make reference to *affective states* that purportedly

prompted the agent's behaviour, or to "triggering events" that occurred *shortly before the behaviour took place*, than we are under pressure to assign responsibility in contexts where our statements refer to *waves of neural excitation in the agent's brain* as the cause of the behaviour, or to exciting causes *in the distant and irretrievably lost past*. Accordingly, if the results of a neuroscientific experiment were to convince us that someone's behaviour was triggered by the person's brain before she consciously "decided" which course of action to take – *that* would seem harder to square with the agent's status as a morally responsible being than would a causal explanation that appeals to the agent's irrational feelings of guilt as the cause of her behaviour.[62] So, to say it again: in contexts where human behaviour is explained causally, the more remote, distinct and different in nature, the determining cause is from the mental powers we regard as integral parts of our human freedom, the greater the temptation to draw deterministic conclusions based on the explanation. Having stated these preliminaries, I can add two more reasons why formal-causal explanations do not encourage deterministic thinking:

1) The causal influence of mental structures and dispositions (conceived as formal causes) that shape people's experiences and behaviour does not emanate from a distant and irretrievably lost past but is concurrent with the experiencing and acting. Hence, the causal influence appears to be "within the temporal reach" of the agent's self-governing powers; there is no "temporal gulf" to separate the self-governing powers from the source of the causal influence, but the self-determination and the causal influence are, as it were, occurring at the same instant.

2) The causal influence of mental structures and dispositions (conceived as formal causes) is not exerted from outside of the agent's self-governing powers, but shape the agent's willing and reasoning *from within*, in a similar way as Kant's categories shape the Kantian mind from within. This makes it unclear whether such causes are sufficiently external and distinct from the self-governing powers they shape for their influence to involve anything like determinism. If the cause of an action turns out to be a constitutive part of the agent's self-governing powers, does it make sense to think of the agent as being moved by something other than herself? A corollary of this is that formal causes and the self-governing powers they shape belong to the same (psychic) "realm of existence," suggesting that the formative influence might work in the opposite direction too, from the self-governing powers to the formal causes – a possibility that contradicts determinism.

4.3 Freedom as the Acquired Capacity to Adjust Unwanted Mental Structures and Dispositions

[62] The neuroscientist Benjamin Libet famously argued that our volitional actions are determined by neuronal activities in the brain that we are unaware of. See Libet 1985.

In the discussion of Kaitlin's life story, I tried to describe a form of agential freedom that is absent from symptomatic mental acts and behaviours of the sort that lend themselves to formal-causal explanation. This freedom was outlined in brief discussions of some therapeutic procedures that people who suffer from fears and anxieties like Kaitlin's might engage in in order to alleviate their problems and to open new horizons in their personal development. In Kaitlin's case, these procedures, if carried through properly, should result in increased self-knowledge and an enhanced set of mental capacities by means of which she could loosen or undo the restrictive effects her fear of abandonment had come to have on her personal freedom: she would develop a better understanding of her relationship problems and their psychological basis; she would gain more control over her fear-induced patterns of thought, feeling and behaviour, she would cultivate new habits and dispositions that would reflect her more considered views on how she ought to live her life;[63] and so on.

I conceptualized the overall therapeutic process involved in such journeys to freedom by means of concepts borrowed from Jonathan Lear and Richard Moran. Following Lear, I described such therapeutic undertakings as processes of "intrapsychic integration" in which a person brings the conscious and unconscious parts of her mind into better accord and makes accessible to change parts of her mental life that have split off from her conscious personality and taken on a life of their own.[64] In Kaitlin's case, this involved an integrative process through which her fear of abandonment would be integrated into her conscious experience and recognized as a part of her personality.

In Moran's terminology, I described such therapeutic undertakings as attitudinal shifts whereby a person comes to relate to some part of her mental life that she has become "alienated from" in a more personalized and responsible manner, comes to "identify with" that part of her character, and thereby might find the motivation needed to drive an appropriate process of change. In Kaitlin's case, this would involve a move from an alienated condition in which she experiences her fear of abandonment (insofar as she is conscious of it all) as an empirical fact about her mental constitution – a fact that she has little or no responsibility to change – towards

[63] On the account I've given here, what marks the difference between "good" and "bad" mental dispositions is largely the individual's self-reflective awareness of the pattern in question: Can she express and reflect on it feelingly? is the pattern susceptible to conscious influence and modification? does it harmonize with individual's ideas of what is good, right, desirable, and worth pursuing in life? can the individual identify with the disposition, or does she feel alienated from it? is she passively repeating it or actively adjusting it (if it needs adjustment)? how rigid is it? does the individual relate to it as an empirical fact about herself or as a character trait answerable to moral demands and assessment? Etc.

[64] Lear 2017: 45.

a condition of identification in which she relates to her fear as a character issue that she herself has unconsciously developed and maintained over the years.

Taken together, one might say of the notion of freedom that emerges from these observations that it concerns our capacity to reorganize ourselves as individuals by means of cognitive resources that we can acquire and apply to unwanted components of our character. This freedom is the acquired capacity to make infantile, impulsive, split-off, and repressed parts of our character answerable to knowledge and reason. Such adjustments are of course not just a matter of changing our "inner" mental tendencies, but are just as importantly a matter of adjusting the aspects of the "outer" world that we experience as "coloured" by those tendencies, for instance, the idiosyncratic meanings objects and events have for us as individuals, or the action-possibilities we are responsive to in our environment, or the general character of the world of our experience.[65]

4.4 Reasons Why the Charges of Determinism and Physicalist Reductionism Miss the Mark

Now I have reached a point in my exposition of formal-causal explanations from which I am able to assess whether, or how far, the objections raised at the start of the paper against causal explanations of human nature and activity apply to the formal-causal model. The most important of these was the objection that causal outlooks undermine morality. This challenge comes in a variety of forms, but the core idea is that agency drops out of the picture if our mental life depends on causal forces that lie beyond our conscious control. So, translated to the case of formal-causal explanation, the criticism would be that moral agency is eliminated when human lives and behaviours are considered under formal-causal descriptions. Is this objection justified? According to the account I have put forward, the answer is No for various reasons, the most important of which I will review in what follows.

As I noted earlier, causal determinism, as a philosophical position, relocates agency from the self-governing powers of the agent to physical factors that lie outside her influence and control, for instance, from the agent's free will to neuronal processes in her brain, and regards those external factors as causal determinants in the absence of which thoughts and actions that appear

[65] For an example of a change in the general character of the world our experience, imagine the change involved when the benign world of a person who enjoys mental health turns hostile and uncontrollable at the onset of a psychotic disorder.

free to the agent had not been carried out. I also remarked that deterministic outlooks conceive of the influence of such factors in generative terms, as a generative influence that brings about, or contributes to bringing about, the agent's activities and behaviours. Formal-causal explanations, by contrast, explain human activity by reference to mental structures and dispositions that form part of and shape the agent's self-governing powers, for instance, by reference to unconscious fears that shape the agent's reasoning, and regards those internal factors as causal determinants in the absence of which certain thoughts and actions of the agent had assumed a more or less different shape. The influence of such psychological factors is conceived of in formative terms, as a formative influence that shapes the agent's activities, responses, and behaviours according to prototypes laid down in the agent's past. Such internal factors do not remove the agent's self-governing powers from their "generative" role in the performance of actions but subject them to a formative influence. As formal causes, such factors are neither distinct from nor beyond the influence of the agent's self-governing powers, but partly make up and determine the functioning of those powers, while being susceptible to the agent's efforts to change the character and extent of their influence. This last point – that the agent can change and overcome the influence of psychic structures acting as formal causes – is another reason why formal-causal explanations do not undermine morality: insofar as these psychological factors by reference to which formal-causal explanations explain human activities are, or can be made, accessible to change, it is "reasonable" to hold people responsible for the measures they adopt, or fail to adopt, in order to change such factors in cases where changes seem appropriate.

So, to sum up, formal-causal explanations cannot be said to eliminate human agency for three reasons: (1) the causes they appeal to are too closely intertwined with the agent's self-governing powers for us to consider their influence as flowing from an external source; (2) the causal influence of the kind of nonrational factors by which formal-causal explanations explain human behaviour is of a patterning, rather than of a generative, nature; and (3) it is a commonly acknowledged fact that humans have the capacity to adjust and overcome the causal influence of the kind of psychological factors by reference to which formal-causal explanations explain actions and behaviour. But in addition to these, there was a fourth reason why formal-causal explanations do not eliminate human agency that I devoted some attention to in the section 4.2. There I observed that the logical form of nomological statements ("given an event of kind *B*, it will have come out of an event of kind *A*") lends support to certain metaphysical ideas about the nature of the universe, one of which was the idea that causes "necessitate" their effects.

Philosophers impressed by this idea take the "constant conjunction" of *A*s and *B*s to indicate the existence of a metaphysical necessity that lawfully governs the causal relationship of these types of events. So, on this view, the logical form of nomological statements is not accurately represented by the formula cited above, but should be reformulated as follows: "Given an event of kind *B*, it will *of necessity* have come out of an event of kind *A*." But this formula, if correct, seems to pose a threat to our agential freedom. For if causal connections are governed by necessity, then true explanations of human activities that appeal to such connections will eliminate the voluntary character of our actions: "One cannot coherently think of a person's behaviour as necessitated by laws while at the same time considering it an expression of the agent's free will." –That was the worry that arises from nomological statements. But as I pointed out earlier, this problem simply does not arise where formal-causal explanations are concerned, because the misconception it rests upon – that causal connections involve some sort of necessity – finds no purchase in formal-causal explanations. Why? Because explaining human behaviour in terms of formal causes does not involve making universal claims about causal connections between types of behaviour and the types of causes they derive from. Formal-causal explanations are not exceptionless generalizations about law-governed regularities but, at their most ambitious, are probabilistic generalizations about contingent and conditional regularities that can be observed in the lives of individuals, and at their least ambitious, are explanations of single-case causal phenomena. Hence there is no justification for saying that formal-causal explanations eliminate human agency by explaining voluntary actions in term of necessary connections, because the notion of causal necessity is incompatible with the logic of formal-causal explanations.

The second critique against causal outlooks that I presented in the first chapter involves the claims that proponents and practitioners of such outlooks, by approaching their subject-matter from an external perspective, reduce human beings to their biological nature and human activities to their physiological correlates. The essence of the first claim is that causal outlooks seek to define the concept of *human being* in relation to our organismic nature, as a biological concept, instead of treating it as an open concept that evades final definition. The concept *human being* evades final definition, according to the critic, because its content changes based on our conception of the kind of beings we are and our hopes and visions of the beings we wish to become in the future. In other words, the meaning of "human being" is not wholly determined by our biological nature but is partly open and determinable by our self-understanding and self-fashioning. And the second claim, which I lay out in more detail on

pages 10-11, is that causal explanations fail to take into account the internal understanding that partially constitutes human actions and behaviour, and thus reduce actions and behaviour to mere physical phenomena. Can these lines of criticism be turned against the formal-causal model of explanation? Are formal-causal explanations reductive, or guilty of ignoring or taking too little account of the internal understanding embodied in action? I will use a slightly modified version of a formal-causal statement I considered earlier in this chapter as a point of reference for a discussion of why these criticisms do not apply to the formal-causal model. The statement I have in mind is about a person suffering from OCD and his compulsion to check the stove, and the modified version of this statement that I shall consider reads as follows:

> "Seized by his obsessive worry about burning down the house, Robert had to turn around and go back three times to make sure the stove had been properly turned off."

It is true, of course, that we can describe and explain compulsive behaviours in terms of chemical imbalances, brain abnormalities, and other such biochemical and neurophysiological factors, without heeding the understanding and ideational content that enters their composition. One such explanation, proposed some years ago by a team of scientists, involves the claim that OCD-behaviours are caused by overactivation of a signalling pathway in the amygdala region of the brain.[66] In everyday contexts of explanation, however, we tend to be more interested in the beliefs, desires, wishes, fears, and other such conceptually shaped mental states and attitudes that give compulsive behaviours their meaning for the agent than in their neurobiological correlates. The formal-causal statement above is an example of this everyday sort of explanation. As an explanatory statement, it explains Robert's compulsion to check the stove, not in terms of neurobiological processes, but in terms of Robert's obsessive worry about burning down the house – a worry that presupposes familiarity with various social practices, customs and institutions that form part of the cultural context of the worry. Robert couldn't possibly worry about starting a fire by forgetting to turn off the stove unless he knew what stoves are and how they function, unless he knew about things that can go wrong if one leaves the house with the stove on, unless he had some idea of the responsibilities that might befall one if an accident should happen, of the financial burden, the guilt and the shame, and so on and so forth.[67] Much of this knowledge will be of a tacit nature and seldom if ever be voiced

[66] See Ullrich, M., et al. 2018.

[67] Robert's obsession about burning down the house need not be the only, or even the predominant, feature of his condition. It may be only one component in a complex of symptoms that might include *non-ideational* features as well (e.g., tics, twitches, and other such "meaningless" movements and vocalizations; panic attacks and feelings of anxiety; stomach issues and other psychosomatic ailments for which there is no medical explanation), features that "embody understanding" in a much more restricted sense, if at all, than do his

or reflected on. But it nonetheless informs Robert's behaviour and is in principle available to his articulable awareness. Robert can become aware of this knowledge and its conceptual components, similarly as he can become aware of the network of ideas in his mind that his worry about burning down the house is connected to. These are the ideational ingredients of Robert's compulsion to check the stove: the conceptual prerequisites needed for worrying about accidentally burning down the house and for taking the precautionary measures Robert takes to prevent a house fire from occurring, on the one hand, and the ideas in Robert's mind that his worry is connected to on the other hand. The former components are essentially *public* and have their source in the language, the social practices, and other normative structures that form part of the cultural context of Robert's worry, while the latter – though dependent on the same public normative structures – are *personal* and have their source in Robert's previous experiences and personal history. These two sets of ideational components together compose the understanding embodied in Robert's stove-checking expeditions. Moreover, it is these ideational components, in combination with the aim-driven character of Robert's activity, that allow us to characterize his stove-checking as a set of "actions" rather than as mere "behaviour." The reason for calling attention to the distinction between actions and behaviour here is that we find it most natural to conceptualize Robert's stove-checking in terms of its ideational content when we regard it as a series of particular actions. When we regard his activity as an instance of OCD-behaviour, by contrast, we are more likely to ignore the ideational content embodied in the activity and attempt to understand it in more general terms, as neuroscientists and biologists do when making universal claims about types of behaviour. Conceived of as actions, Robert's stove-checking is carried out for the sake of achieving a goal and, as such, requires teleological explanation. Biologists of course use teleological explanation as well to explain intentional (human and non-human) animal behaviours; but biologists characterize the ends served by animal behaviours in different terms than those we typically use to characterize the ends served by human actions. Biologists describe the ends pursued by animals by reference to instincts and the biologically adaptive functions of organisms and species, whereas human actions serve ends that we typically describe by

conceptually structured symptoms. Such symptoms can only be said to "embody understanding" in the sense that (i) they originate from problems in living (insofar as Robert's condition is caused by the strains and stresses of life, rather than by neurobiological factors), (ii) they give expression to problems in living, and (iii) they are susceptible to influence by attempts on Robert's part to intelligently (as opposed to medically) deal with those problems in living. For another example of non-ideational symptoms, the reader can imagine a person living abroad and suffering from stomach cramps and vomiting due to her unconscious guilt over not being able to care for a parent diagnosed with cancer. For a clinical case-history of such a person, see Marilyn Wedge's "The Mystery of Psychosomatic Symptoms".

reference to beliefs, desires, wishes, and other such attitudes which involve the exercise of conceptual capacities on the part of the agent. Rather than describing the migratory movements of salmon in terms of beliefs and desires that motivate salmon to travel as they do, a marine biologists would describe the movements of salmon in terms of instinctive responses to natural events, such as aging and the change of seasons. This is not to deny that the migratory movements of salmon are of an intentional character, but only that their intentionality is of the same kind as the intentionality we ascribe to human actions. The latter are "intentional" in a stronger sense than the former. We mean something else by "intentional" when using the word in connection to human actions than we do when using it in connection to aim-driven behaviours of non-articulate creatures. And this, I think, applies to ill-judged human actions just as much as it applies to well-reasoned ones. Provided that the performance of an action involves certain conceptual and linguistic abilities on the part of the agent, we think of it as "intentional" in this stronger sense of the word, no matter how poorly or well-reasoned the action happens to be. These abilities include, above all, the agent's ability to see what she does and the world in which she does it under certain descriptions, and the ability to give reasons for one's actions and doings. Robert, as I have described him here, has these abilities, whereas non-human animals have them, and are able to acquire them, to a lesser degree or not at all. Either way, an animal's possession of these abilities is not a criterion of application of "intentional" where the weaker sense of the word is concerned. A chimpanzee does not have to give reasons for its use of a blade of grass to fish termites from a mound for us to be warranted to describe its termite-fishing as "intentional" in the weaker sense of the word.[68]

So the agent's possession of certain conceptual and linguistic abilities is a criterion for ascriptions of strong intentionality of the sort we associate with reasoned (human) actions, one of which was the agent's ability to give reasons for her actions. However, now the reader might wonder what I mean by "reason" more exactly. Furthermore, what is involved in the act of giving reasons for one's actions? I have devoted a substantial part of the next chapter to discussions of reasons, reason-giving, and reason-explanations, so my treatment of these issues here is limited to minimal observations needed to keep the discussion going.

[68] The contested question is whether great apes, such as chimps and bonobos, can acquire the relevant conceptual and linguistic abilities to such an extent that their behaviours can be characterized as "purposive" in the stronger sense of the word. For interesting discussions of these and other related issues, see Pär Segerdahls *Djuren i kulturen* and his co-authored *Kanzi's Primal Language*, with William Fields and Sue Savage-Rumbaugh.

The question I will consider here is this: What might Robert say that would count as reasons if someone were to ask him why he turned around and went back three times to make sure the stove was turned off? Let me begin by stating the obvious. The possibility of the stove being on is a reason to go back since the act of doing so might prevent an accidental house fire from taking place – an event that any person would recognize as an evil. Thus, if someone were to ask Robert upon his return from the *first* stove-checking expedition why he turned around and went back home, the following straightforward response would count as a reason: "I had to turn around when I realized that I had forgotten to check the stove." But what about the second and third time around? What sort of factors might Robert mention that would count as reasons for going back twice more? Well, one thing that is clear is that the factors he would mention need not *justify* his actions to qualify as "reasons." For after all, just as bad art is still art, bad reasons are still reasons.[69] And yet there must be *some* criterion by which to distinguish things Robert might say that would count as reasons from others that would not. The conceptual features of reasons is a topic I discuss in the next chapter, so I will limit myself here to a rough suggestion as to what the criterion might be: Robert's reasons for going back twice more to inspect the stove must refer to some factors which in certain situations, according to common logic, speak in favour of checking the stove. So, according to this criterion, the factors Robert mentions in following two explanations, offered by Robert upon his return from the second and the third stove-checking expedition, would count as reasons for going back: "I had to go back a second time because I wasn't paying proper attention the first time around"; and the final explanation: "I had to go back one last time when I realized that I had only checked the light that indicates whether the stove is on and had forgotten to feel the plates with my hand." As I pointed out earlier, if the factors Robert mentions in these explanations – his inattentiveness and his omission to feel the plates – carry little justifying force, that does not automatically preclude them from being reasons. If Robert is damaging the reputation of the company he is working for by always being late, his failure to pay attention while doing the first stove-inspection may not justify his decision to go back a second time and arrive even more late for work. But even so, his inattentiveness may count as a reason for going back to check the stove. From the perspective of someone whose mental life is dominated by a worry about burning

[69] It is one thing to assess the justifying force of reasons – which can only be done on a case-by-case basis – and another to assess whether some factor can be referred to as a reason at all. The fact that we can and often do justify people's actions, our own and those of others, by giving reasons might lead to the false assumption that *all* reason-giving serves a justifying function. In a similar way as claims to knowledge ("I know that…", "it is certain that…") are always open to assessment, reason-explanations are always open to assessment, but not all reason-explanations are given or asked for with this possibility in mind.

down the house, it may seem reasonable to turn around and check the stove a second time if one was not paying attention the first time around, and it may seem a reason to go back a third time if one forgot to feel the plates during the previous inspection.

At this point the reader might be wondering about my use of the vocabulary of rationality. Am I seriously suggesting that it is *reasonable* to feel an irresistible urge to go back and check the stove one final time after already having done so twice? Am I saying that it is reasonable to carry out actions which will have the effect of further entrenching the agent's destructive mental patterns and habits? Am I claiming that actions which may seem reasonable from the restricted point of view of a certain worry of the agent can be described as "reasonable" even if they clash with the agent's overall values, desires, and view of things? No, I do not wish to endorse these views. My aim has been to awaken an appreciation of the different ways in which we can use reason-explanations and the vocabulary of rationality to characterize less-than-fully-reasonable actions. Robert's stove-checking acts, affectively charged and performed for "bad" reasons, are examples of such actions. Considered in the light of reasonable ideas about how one ought to live one's life, Robert's stove-checking seems glaringly unreasonable. In fact, one way of describing his agential perspective is in terms of corrupted rationality. Robert's capacity for reason has been compromised by his compulsive worry about burning down the house so that considerations which others would dismiss as far-fetched or unlikely have come to seem imperative to him. To give an example of this, the reader might imagine Robert responding to the question of why he closed all the windows when leaving for work by saying, "I had to close the windows in case lightning should strike through and set the house on fire while I'm away. The weather is supposed to stay sunny, but a forecast is a forecast, not a guarantee." For a compulsive worrier like Robert, the tiniest risk might seem unacceptably high. So even if it is true, as I maintained earlier, that Robert's compulsive actions presuppose his possession of certain conceptual and linguistic abilities, it is also true that his use of those abilities, especially the ability to give reasons for his actions, make his compulsions seem more rational than they are. For Robert is, after all, acting on reasons that arise from his compulsive worry. His judgement has been compromised by his worry about burning down the house in a similar way as Tom and Kaitlin's judgment had been compromised by their early traumas and fears. In Tom's case, it was his phobia of large bodies of water that had skewed his perspective, and in Kaitlin's, her fear of abandonment. So the suggestion I am developing here, and will continue to develop in the next chapter, is that we use formal-causal explanation to explain some of the things that can go wrong in people's reasoning when they fail to think and act as

rational agents.[70] Such formal-causal explanations are particularly relevant for explaining overtly irrational behaviour, but their application extends, I think, to less clear-cut cases involving seemingly rational thoughts and actions that are formally shaped by nonrational structures of the agent's psyche.

However, now I owe the reader an explanation of how the considerations above relate to the original question posed at the outset of this section – the question of whether the formal-causal model of explanation is susceptible to the criticisms that might be levelled at causal outlooks that ignore the internal understanding embodied in human action and reduce human beings and activity to biophysical processes. The relevance of these considerations for this question lies in the relationship I have tried to tease out between formal-causal explanations of actions, on the one hand, and reason-explanations of actions that refer to the agent's reasons for acting, on the other hand. I have tried show that reasons and formal causes operate at the same level of explanation, make use of the same register of language, and thus stand and fall together with respect to these criticisms. One cannot consistently blame formal-causal explanations for ignoring or downplaying the significance of the internal understanding of action without blaming reason-explanations for the very same thing. And since it would be absurd to blame reason-explanations for ignoring the internal understanding of action, seeing that reason-explanations are specifically concerned with this understanding, one cannot blame formal-causal explanations on this score either. Or to put it another way: formal-causal explanations are immune from the criticisms above because the mental dispositions they refer to – obsessive worries, unconscious fears, fantasies, and what have you – belong to the same vocabulary of mind as the beliefs, desires, and intentions we impute to agents by explaining their actions by reference to reasons for acting. If I explain someone's decision to take an umbrella when she goes out by referring to the reason that it might start raining soon, then I also impute to that person the belief that it might start raining while she is out and the desire not to get wet. And similarly, if I explain someone's act of refraining from cheating by referring to her love for her partner, then I also impute to that person the intention to stay faithfully committed to the person

[70] This is in line with Lear's discussion of formal causes in *Wisdom Won from Illness*: "Freud's point is that a person's capacity for reason can be pervaded by unconscious, nonrational, mental forces [of the kind that Lear conceptualizes as formal causes]; and when that happens, reason can be pervasively distorted by a nonrational form of thinking" (39).

she loves.[71] Our beliefs, desires and intentions give us reasons to act in particular ways. The belief that it is about to start raining soon gives one reason to take an umbrella, and the intention to stay committed to the person one loves gives one reason not to cheat if the occasion arises. This much should be uncontroversial. However, a more controversial idea that I am developing here is that causes *too* generate reasons for acting in particular ways in particular situations. This is the idea I was getting at when I said that Robert's compulsive worry – the formal cause of his stove-checking – gives him reason to act as he does. The worry Robert has at the forefront of his mind shapes his way of seeing things and, in doings so, gives him reasons for action.

Another point I made in the discussions above was that compulsive worries, and other such nonrational mental dispositions that can act on us as formal causes, presuppose our possession of various kinds of understanding in the form of conceptual and linguistic abilities. Robert could not possibly worry about burning down the house unless he knew various things about stoves and the normative practices that shape the culture he lives in, things he could not possibly know without having certain conceptual and linguistic capacities. This is another reason why the criticisms above do not apply to formal-casual explanations. In contrast to the neurochemical processes and biologically adaptive functions that neuroscientists and biologists refer to in their explanations of human behaviour, the mental structures and dispositions referred to in formal-causal explanations have ideational content that presupposes the agent's possession of various conceptual and linguistic abilities.

So the criticism that causal explanations ignore or downplay the significance of the internal understanding embodied in human activity does not apply to formal-causal explanations, because formal causes are *about* something, are mental causes with ideational content, and whatever explanatory force they have they owe to this ideational content.

As to the criticism that formal-causal explanations reduce the meaning of "human being" to our biological nature, there is not much to say. If there is some particular idea of human being that attaches to formal-causal explanations, it can only be one that embraces our cognitive, affective, and behavioural dimensions in their entirety – an idea that far surpasses our biological nature.

[71] Here I am drawing on Donald Davidson's account of reasons for action. For Davidson, to explain an action in terms of reasons is to attribute to the agent a "pro-attitude" that consists of beliefs and desires that speak in favour of the action. See Davison's "Actions, Reasons, and Causes", p. 25.

4.5 Summary

In this chapter, I have positioned the formal-causal model of explanation in relation to some philosophical problems associated with causal outlooks on human life and activity. I have explained why formal-causal explanations do not encourage deterministic thinking and why they escape the charges of determinism and physicalist reductionism. Formal-causal explanations do not tempt us in the direction of determinism as they are logically incompatible with the metaphysical ideas that are wedded to determinism. The attenuated autonomy of agents whose activities are considered under formal-causal descriptions can be thought of as a form of predetermination – the agent is "predetermined" to think, feel, and act in certain ways in certain kinds of situations – without evoking the idea of endless causal chains in a law-governed universe. Formal-causal explanations cannot be said to eliminate human agency because (1) the causes they appeal to are too closely intertwined with the agent's self-governing powers for us to consider their influence as flowing from an external source; (2) the causal influence of the kind of nonrational factors by which formal-causal explanations explain human behaviour is of a patterning, rather than of a generative, nature; (3) it is a commonly acknowledged fact that humans have the capacity to adjust and overcome the causal influence of the kind of psychological factors by reference to which formal-causal explanations explain actions and behaviour; and (4) formal-causal explanations admit of single-case causation. The reason why formal-causal explanations do not involve physicalist reductionism is that the kinds of mechanisms and dispositions they refer to – obsessive worries, unconscious fears, etc. – belong to the vocabulary of mind rather than to the vocabulary of physical forces and matter. I have also discussed some cases in which reason-explanations fail to account for actions and behaviour and stand in need of complementation by formal-causal explanations. Such cases, as we shall see in the next chapter, offers us a new perspective on the debate in the philosophy of action of whether reason-explanations explain actions causally or not.

5. An Examination of the Significance of Formal-Causal Explanations for the Debate Over the Causal Status of Reason-Explanations

As observers trying to understand the activities of others we oscillate between causal and teleological orientations. Some actions and doings we try to understand in terms of events and conditions in the past that are likely to have influenced the agent's conduct in significant ways, and others we make sense of in terms of motives and reasons that seem to have guided and motivated the behaviour. My aim in this chapter is to gain new insight into some longstanding philosophical problems concerning the relationship between these two orientations by using the conceptual resources provided by the account of formal-causal explanation outlined in the previous chapters. I am particularly interested in the problem of whether reason-citing explanations of actions are a type of causal explanation (in the sense that Wittgenstein's five causal prototypes are types of causal explanation) or whether they belong to, or compose, an altogether different category of explanation. Another closely related but not identical problem I discuss here is whether the reasons for which actions are performed are also the causes of those actions. If I go swimming to cool myself down, to give an example, is my reason for going swimming – my desire to cool down – also the cause of my going swimming? If we combine these two problems in a certain way, we get the central question I address in this chapter: Why is it that a philosopher who acknowledges the conceptual difference between reasons and causes might nonetheless be inclined to think of reason-explanation as a type of causal explanation? This inclination of thought, which has been discussed by Glock (2014) and Queloz (2017), might lead a philosopher to reject the view that agents' reasons for acting are the causes of their actions, while asserting that there are reason-explanations that are best regarded as a type of causal explanation. I have found this inclination in myself when I have considered certain ways in which we explain certain kinds of behaviour (which I shall specify later), which partly explains my eagerness to discuss such an abstruse issue of little general significance. My main reasons for pursuing this topic, however, are my hunch that the intuitions lying behind the abovementioned inclination might go some way towards explaining the continuing appeal of causal accounts of reasons and reason-explanation (despite the strong arguments against such accounts) among contemporary philosophers, and secondly, my hope that the account of formal-causal explanation outlined in these pages can help us to identify and evaluate those intuitions.

5.1 Three Arguments Against the View That Agents' Reasons for Acting Are the Causes of Their Actions

The main concern of this chapter is to identify and assess some intuitions that might incline philosophers to think of reason-explanation as a type of causal explanation. Before approaching those intuitions, however, I must say something about the distinction between reasons and causes and explain why some philosophers have found it misguided to identify the former with the latter. I too find it misguided to identify, or otherwise to assimilate, reasons with causes, and the investigation I conduct in this chapter presupposes that one cannot do so without bending or breaking the rules of our discursive practices (I will explain what this means in a moment). Since my understanding of the conceptual difference between reasons and causes is largely based on the philosophy of Wittgenstein and his followers, I will begin by presenting what I consider to be their most important arguments against the view that agents' reasons for acting are the causes of their actions. The arguments are three in number and concern certain features of our discourse of giving and asking for reasons. The first argument involves the claim that acts of reasoning and reasoned actions, unlike causal processes and caused actions, are subject to rules. The second argument involves the claim that reasons, but not causes, have a justifying function in our practices of giving and asking for reasons. And the third argument highlights certain first- and third-person asymmetries that characterize our talk about reasons but are absent from our talk about causes.

(1) To act for reasons is to act according to rules. The concept of reason is related to mental acts that involve "reasoning," that is, related to procedures in which we think and form judgments logically. The mental act performed in such procedures can be described as a movement in the direction of a conclusion along the normative connections that constitute a rule-governed practice or activity. That is what it means to "reason one's way to a conclusion": to arrive at a conclusion in the form of an act or a proposition by taking one or more steps according to the shared rules (the norms, customs, standards, or what have you) of some practice or activity. The rules that determine such practices and activities can be as tight and specific as the laws of arithmetic, or they can be as loose and indeterminate as the "principles" according to which one fixes a date with someone by going through one's diary and finding a free time.[72] Any normatively structured practice or activity, any "language-game" to use Wittgenstein's dictum, can serve as a framework for reasoning. Physical processes, by contrast – earthquakes and other such processes that are causally rather than normatively determined – cannot intelligibly be characterized as involving "reasoning," or as taking place "according to reasons," since our explanations of such

[72] These are Wittgenstein's examples: "…a reason may be the way one arrives at a conclusion, e.g., when one multiplies 13 × 25. It is a calculation, and is the justification for the result 325. The reason for fixing a date might consist in a man's going through a game of checking his diary and finding a free time." Wittgenstein 1982: 5

processes do not explain their occurrence as the result of a mental act of a thinking subject, or as taking place according to the shared rules of some practice or activity. So the concept of cause, which we use to explain physical processes, lacks certain features that are central to the concept of reason and therefore cannot be equated with the latter.

(2) *The justifying function of reasons*. We can justify our acts of reasoning and reasoned actions by appealing to the shared rules we follow in carrying them out. For example, if I add 7 and 5 and get 12, I can justify the operation by appealing to the rules of arithmetic; or if I mix sugar rather than chili-powder in the doughnut dough I am making, I can justify my choice of ingredient by referring to the recipe I am using in my baking. The rules we invoke as our reasons for acting as we do have a justifying function in our discourse of giving and asking for reasons. The causes of our actions, by contrast, cannot be appealed to to justify our actions, because causes lie outside the normative realm where our justifying practices take place.[73] A reason can justify a conclusion by virtue of the normative connections (relations of entailment, exclusion, valid inference, etc.) that legitimate (and proscribe) transitions from certain reasons to certain conclusions, for instance, from the recognition that one is feeling groggy to the conclusion that "the drug is starting to take effect." But a cause (e.g., a neural process that conditions my trip to the marketplace) cannot justify the effect that it helps to bring about because the conditions out of which something emerges do not by the fact of conditioning the emergence of that something justify its emergence. To think otherwise is to confuse the origin of something with its nature or value.

(3) The first-person authority of the reason-giver. In our attempts to understand the behaviour of others from an observer's (third person) perspective, we make inferences about their aims and purposes based on observation. We see someone poking the ceiling with a broomstick and infer from his behaviour that he is trying to remove a bees' nest. Or we see someone pulling out items from the kitchen cupboard – a bag of flour, dark chocolate, a packet of butter, four eggs, sugar, cocoa powder – and infer that she is making a mud cake. Inferences of this type are hypothetical, inductive, and open to falsification. If we were to ask the person with the broomstick why he is messing with the bees, he might reply that he is trying to get some honey for his evening tea. In these respects, third-personal statements about people's reasons are similar to causal hypotheses: both are based on past observation and inductive reasoning, and both are open to falsification.[74] In the case of first-person statements about one's own reasons, however, these grammatical features do not apply. In explaining some action of mine by stating my reasons for doing it, I do not base my explanation on observation or inductive reasoning, and my explanation is not vulnerable to contradicting observations in the way that causal hypotheses are. The fact that it is *me* who is stating

[73] Waismann 1997: 123

[74] In *The Blue Book,* Wittgenstein contrasts first-personal statements about reasons and motives with statements about causal relations: 'The difference between the grammars of "reason" and "cause" is quite similar to that between the grammars of "motive" and "cause". Of the cause one can say that one can't *know* it but can only *conjecture* it. On the other hand one often says: "Surely *I* must know why I did it" talking of the *motive*. When I say: "we can only *conjecture* the cause but we *know* the motive" this statement will be seen later on to be a grammatical one. The "can" refers to a *logical* possibility' (p. 15; italics in original).

my reasons, rather than someone else making third-person statements about my reasons, matters in ways that the identity and the perspective of a speaker reporting contradicting observations on a causal hypothesis do not matter. Where my own reasons are concerned, I enjoy a certain "authority" and cannot go wrong in what I say – not in the way I can go wrong in making third-person statements about other people's reasons or in making causal hypotheses. This is not to deny that I am fallible in stating my reasons,[75] but to draw attention to certain grammatical discrepancies between first-person avowals of reasons, on the one hand, and third-person statements of reasons and causal hypotheses on the other hand. Among those discrepancies is the fact that we can become aware of our own reasons in ways that we cannot become aware of connections between causes and their effects, and the fact that we can question causal hypotheses in ways that we cannot question someone else's reason-avowals. In our language these discrepancies are reflected in our use of certain words and phrases in causal discourse that we would not use in first-person avowals of reasons ("discover", "to look for", "make hypotheses about", etc).[76] These discrepancies are related to differences between the procedures by which we become aware of our own reasons and those by which we discover causal relations and make conjectures about other people's reasons. Unlike the procedures by which we guess other people's reasons and make conjectures about causal relations, the procedure by which we become aware of our own reasons does not involve observation of any kind, neither "external" observation of our bodily movements nor "internal" observation of our inner goings-on. When I explain my actions by reference to reasons, I do not turn my gaze inwards, as it were, and report on what I find,[77] nor do I make inferences about my reasons based on the way that my body moves.[78] The procedure by which I become aware of my reasons is altogether different. Wittgenstein puts it simply: "On being asked for the reasons for a supposition, one calls them to mind."[79] I call my reason(s) to mind by asking myself which considerations did in fact weigh, or do weigh, or will weigh, or should weigh, the heaviest in my acting, thinking, deciding, feeling, or what have you, in a particular situation.

Overall then these arguments show that, and in what sense, the concept of reason differs from the concept of cause by drawing attention to some ways in which certain linguistic practices and conventions that play a role in determining the concepts differ from one another. Given this divergence between our reason-talk and causal talk, and in view of the "value" it affords

[75] For example, I can go wrong in stating my reasons in situations where I have more than one ground for doing something and am asked to single out the one that weighed the heaviest in my decision to carry out the action.
[76] In *The Blue Book* Wittgenstein says: "The proposition that your action has such and such a cause, is a hypothesis. The hypothesis is well-founded if one has had a number of experiences which, roughly speaking, agree in showing that your action is the regular sequel of certain conditions which we then call causes of the action. In order to know the reason which you had for making a certain statement, for acting in a particular way, etc., no number of agreeing experiences is necessary, and the statement of your reason is not a hypothesis" (p. 15).
[77] *Philosophical Investigations* §§ 290, 683
[78] Cf. *Philosophical Investigations* §246
[79] *Philosophical Investigations* § 475. This two-sentence remark ends with the question: "Does the same thing happen here as when one considers what may have been the causes of an event?"

us as language-users – among other things, the value of having concepts by which we can discriminate intentionally performed actions and natural processes of change – it would be false and misguided to equate reasons with, or subsume them under, the concept of cause. Reasons are, according to the arguments above, in significant respects different from causes, and the failure of philosophers to recognize the difference involves some form of unawareness or confusion regarding one or more of the discursive practices that give content to our concepts of cause and reason. Rather than offering new insight into our established conceptual scheme, philosophers arguing for the view that reasons are causes are, whether they realize it or not, acting as promoters of a process of conceptual revision through which two previously distinct concepts would merge into one: "Let us from now on treat *X*s and *Y*s as more alike than distinguishable." This is another way of formulating the point I made earlier, that philosophers arguing for the view that reasons are causes are "bending and breaking the rules of our discursive practices": they are bending and breaking the rules of our causal talk and reason-talk and, in effect, promoting a new conception of causes and reasons. And while conceptual revision might not be bad in itself, the revision being promoted here would obscure a distinction which our ability to make moral judgments depends upon – the distinction between caused and reasoned behaviour – and should thus be opposed on moral grounds.[80]

What is interesting, however, is that a philosopher persuaded by these (or other) arguments for the non-causal nature of reasons might nonetheless be inclined to regard certain reason-citing explanations as causal explanations. The ostensibly self-contradictory position such a philosopher might argue for would involve asserting that (*a*) there are reason-citing explanations that are best regarded as a type of causal explanation, while rejecting the claim that (*b*) the reasons for which agents act are the causes of their actions.[81] A more modest and

[80] To illustrate the relevance of this distinction for our moral judgment, imagine a school-shooter who murdered 16 people, Charles Whitman, whose autopsy revealed a brain tumour pressing on the region of his brain that conditions our ability to control our emotions and behavior (the amygdala). In determining Whitman's moral responsibility, would we not use the distinction between "caused" and "reasoned" behavior? For an interesting article about Whitman's case, see Micah Johnson's "How Responsible are Killers with Brain Damage?" (2018). https://www.scientificamerican.com/article/how-responsible-are-killers-with-brain-damage/

[81] For interesting discussions of some versions of this position, see Hans-Johann Glock (2014) and Matthieu Queloz (2017). Both philosophers distinguish between "causalism about reasons" and "causalism about intentional explanation" and suggest that a philosopher may coherently subscribe to the latter while rejecting the former. Here, to quote a passage that discusses the distinction, is Queloz in the conclusion of his paper: "Wittgenstein's arguments all bear primarily on causalism about reasons (CR), and not on causalism about intentional explanation (CE). Where he does speak about intentional explanation, he emphasizes that a key characteristic of intentional explanations which is alien to the causal explanations of physics is that they refer to rule-governed practices and institutions, and thus to the agents' local perspectives and idiosyncrasies […] Yet even granted this difference, it does not follow that intentional explanations cannot be a *sui generis form of* causal explanation."

less paradoxical version of this view would simply note that there are reason-citing explanations the non-causal character of which comes to appear questionable when we recognize the extent to which their explanatory value depends on our ability to causally explain the motivating force of the reason(s) by which they seek to explain the activity that needs explaining. Or to put it more simply: there are reason-citing explanations ("Mary did X because of Y") the explanatory force of which depends on our awareness of relevant causal factors to throw light on the significance that the reasons they appeal to have for the agents personally. Both versions of this view, the bolder and the more modest, raise doubts about the sharpness of the distinction between causal and reason-explanations and, thus, both versions of the view, or either, if correct, make a claim on the attention of scholars from philosophy and other disciplines engaged in enterprises in which the distinction occupies a prominent position, especially those who take the distinction to be of a sharp, categorical character.[82] What remains of this chapter is therefore a critical evaluation of these views. My aim is primarily to assess the plausibility of, and to specify the intuitions that motivate, these views, and secondarily, to demonstrate how an increased understanding of the formal-causal model of explanation might lead us to reformulate certain question philosophers have asked about the distinction between causal and reason-explanation. The strategy by which I attempt do this involves close examination of an imagined scenario in which one person is confronted with the task of explaining a piece of behaviour that appears to elude this important distinction.

5.2 The Complementarity of Causal and Reason-Explanation as a Source of Philosophical Confusion

The example I discuss in this section follows up the description of Kaitlin's traumatizing childhood experiences in Chapter Three, so I encourage the reader to go back and re-read the second paragraph of the section entitled "The Story of Kaitlin," on page 28, before moving on. Having done so, consider the following scene between Kaitlin and her recent friend, Linda, a scene which takes place during a film-night at the latter's house.

[82] Ethics and the philosophy of law are two branches of philosophy, besides action theory, where the distinction between causal and reason-explanation plays an important role. And speaking more generally, much of the research done in the humanities and social sciences takes some stance, explicitly or implicitly, as to which explanatory model is to be preferred when describing and explaining human actions and behaviour. Svend Brinkmann, to give an example, is someone from the humanities who has argued for the view that psychology, his discipline, should principally work with reason-citing models of explanation to describe and to explain its subject-matter. See Brinkmann's 2006 article "Mental Life in the Space of Reasons".

Kaitlin and her new friend Linda are watching a tear-jerker about the strains and hardships of working-class life during the Thatcher era. About three-quarters into the film, Kaitlin bursts out in tears over a scene in which one of the characters, a girl of ten whose parents have recently lost their jobs, finds her drunk father passed out on the floor when she comes home from school. Although the event is disturbing enough in its own right, the scene doesn't stand out as particularly sad or upsetting in the larger context of the film. Or so Linda thinks: "Why is Kaitlin crying so uncontrollably now when she wouldn't bat an eyelid at those heart-wrenching parts earlier on?" Linda is surprised and slightly alarmed by her friend's reaction. Not by its emotional type or tone – there is nothing surprising about a person's reaction of tearing up over an upsetting film-scene, after all – but by its excessive intensity. Kaitlin's reaction seems out of proportion to the scene that elicited it. "Why is Kaitlin crying so hard? is she alright? what should I do? how should I respond?" Linda grows more and more worried until she feels compelled to ask Kaitlin about her crying: "Kaitlin, dear, why are you crying so?"[83] *In between sobs, Kaitlin replies, "When that girl found her father passed out on the floor, it was so sad I had to cry."*

If Kaitlin's reply does little to help Linda understand her emotional reaction, it is because it leaves open the question of why the relevant scene made such a strong impression on her. Kaitlin's reply confirms what Linda already knew or assumed – that it was the scene with the drunken father that moved Kaitlin to tears – but sheds no light on the question of what it was that made the scene so powerful to her. Kaitlin has a reason to cry – I assume that the pain of a child who is forced to see a parent drunk to the point of unconsciousness counts as a reason to cry[84] – but her reason only goes halfway to answering Linda's question. The reply Linda receives fails to acknowledge her incomprehension as to what it was that made Kaitlin experience the film-scene as intensely as she did. Why did it affect her so deeply and throw her into such an agitated state? If Linda knew what we know about the disruption of Kaitlin's family life in childhood, she would notice the analogy between Kaitlin's experiences as a child and the experiences of the girl in the film and, thus, her reaction would lose much of its puzzling character. If Linda knew about Kaitlin's experiences connected to her father's unemployment and subsequent alcoholism, she would understand the charged significance the scene with the drunken father acquires for her personally. But what kind of knowledge, or awareness, is this

[83] I hope my wording here conveys that Linda's question is posed in a non-judgmental manner, in a genuine attempt to understand her friend's reaction, mostly to figure out how best to respond to her emotional state. In other words, the question is not intended as censure ("why are you crying like a baby? this scene isn't meant to be as sad as your crying makes it seem"). Nor is Linda asking Kaitlin about her crying in an attempt to assess the "appropriateness" of her emotional reaction ("did I miss something? is this scene *really* as sad as that?"). When we are puzzled by someone's reaction to something, our attempts to understand the reaction may require that we adopt a non-judgmental attitude toward the reacting person.

[84] The pain of a child who is forced to see a parent drunk to the point of unconsciousness may count as a reason for other reactions besides crying, for instance, for anger and indignation. However, the crucial point here is that we can evaluate the appropriateness of such reactions, whatever they are, and this evaluability speaks in favour of conceptualizing the scene that elicited the reaction as a reason rather than a cause.

that would help Linda understand Kaitlin's reaction better? It is, as far as I can tell, awareness of two kinds of causal factors: awareness of the mental disposition that influenced Katilin's experience of the scene with the father,[85] and awareness of the circumstances and events in the past that led Kaitlin to develop that disposition. It is the degree of awareness Linda has with regard to these causal factors that decides whether, or to what extent, she is able to understand Kaitlin's reaction, as well as her explanation of why she reacted as she did. For how else, if not through such causal awareness, could Linda even begin to give content to Kaitlin's description of the film-scene as "so sad"?

If what has been said so far seems reasonable, what are we to make of the fact that the explanatory force of Kaitlin's response to Linda's question depends on Linda's awareness of causal factors that influenced Kaitlin's experience of the film? Are we to conclude from this that Kaitlin's response must be some sort of causal explanation after all, despite having the form of a reason-explanation? No, it would be more appropriate to conclude that there are reason-citing accounts of behaviour that fail, or do not suffice, as "explanations," accounts that need supplementation in the form of descriptions of causal factors that influenced the behaviours that require explanation. According to the view I am outlining here, the inadequacy of such accounts can be traced to some measure of indeterminacy that surrounds the referential object of the reason – the rule, the event, the film-scene, or whatever – that is offered in explanation of the behaviour in question. Such accounts are inadequate insofar as they fail to determine the subjective meaning that the referential object of the reason has for the agent personally, the idiosyncratic meaning (and motivating force) that the object has acquired by becoming "linked" with some past (remembered) experience of the agent. What such linkages do is they carry over the character and/or intensity of the agent's past experience(s) to her present experience of the referential object of the reason and, thus, create a kind of correspondence between the agent's past and present.[86] In Kaitlin's case, the indeterminacy concerns the subjective meaning that the scene with the drunken father takes on for her

[85] By "mental disposition" I mean "composite of interconnected cognitive, affective, and behavioural patterns." As a general remark, I believe our mental vocabulary is best understood, not as words that refer to inner processes and entities, but as words that allow us to see connections between such interrelated patters of thought, feeling, and behaviour. For an apt example of this, consider what Freud has to says about jealousy in this passage: "It is easy to see that essentially it is compounded of grief, the pain caused by the thought of losing the loved object, and of the narcissistic wound, in so far as this is distinguishable from the other wound; further, of feelings of enmity against a successful rival, and of a greater or lesser amount of self-criticism which tries to hold the subject's own ego accountable for his loss" (Freud 1922: 223).

[86] Here I am drawing on the writings of Hans Loewald. For interesting discussions of mental mechanisms and dispositions that link past experiences with present ones, see the fourth chapter of *Sublimation*, and his articles, "On the Therapeutic Action of Psychoanalysis" and "Primary Process, Secondary Process, and Language".

personally (by becoming linked with her experiences involving her own alcoholic father) and raises two connected questions. The first question is: "What meaning has the scene with the drunken father for Kaitlin personally?" The second question: "Which mental structure, disposition, or other part of Kaitlin's mind, is responsible for investing the scene with the meaning it has for her?" This second question is asking for a specification of the formal cause of Kaitlin's experience of the relevant film-scene. Such specifications, when convincing, are informed by an adequate understanding of the circumstances and conditions that gave rise to the mental disposition that is being specified. This, however, – that mental dispositions are conceptually linked to the circumstances that contributed to their emergence – does *not* make the second question a question about efficient causes, at least not primarily so. As I observed earlier, questions about formal causes are distinct from questions about their origin.[87] In Kaitlin's case, the distinction concerns two different kinds of investigation: investigations that seek to determine the nature and dynamics of the mental disposition that influenced her experience of the relevant film-scene, and other investigations that seek to locate the events and circumstances that contributed to development of that disposition.

At this point of the discussion, I would like to interject some remarks about ambiguity which will allow me to approach the question that has occupied me in this section – the question of whether Kaitlin's response to Linda's inquiry can reasonably be regarded as a causal explanation – from another angle.

I think the question Linda poses to Kaitlin can help us distinguish different kinds of ambiguity that can be found in contexts where why-questions are asked and answered. Most importantly for my purposes, I think Linda's question ("why are you crying?") can help us distinguish why-questions the ambiguity of which is intended and goes all the way down to the state of incomprehension where the inquiry issues from other why-questions the ambiguity of which is an unintended by-product of the speaker's use of language with more than one meaning. It

[87] Our attempts to form a formal-causal understanding of someone's unconscious fear or other such mental disposition typically involves, not only an effort to understand the nature and symptoms of the disposition, but also an effort to understand the efficient causes that contributed to its emergence and to any significant developments it may have undergone since its initial formation. In the course of our such investigations, the disposition being conceptualized as a formal cause becomes increasingly self-contained with regard to its aetiology, in the sense that we, as interpreters of the disposition, become increasingly able to think of and discuss it without referring to the circumstances and conditions that contributed to its formation and development. As Nietzsche puts it in *Daybreak*: "the more insight we possess into an origin the less significant does the origin appear" (44). Having developed a formal-causal understanding of a mental disposition, we are able to consider it under two aspects: under its efficient causal aspect, as a reaction to certain things in the past, and under its formal-causal aspect, as a psychic agency that patterns the agent's life in the present.

would be crass and inappropriate to reduce the ambiguity of Linda's question to a consequence of the overlap in our ways of talking about causes and reasons, if it turned out to be an expression of her equal, undifferentiated interest in the reasons for *and* the causes of Kaitlin's emotional reaction; it would be crass, that is, to think of the ambiguity of Linda's question along these lines: "By formulating her question as she does, using a word that can serve both as a request for a reason and as a request for the cause, Linda inadvertently introduces a touch of ambiguity to her question." As Wittgenstein repeatedly noted, the fact that "why?" can serve both as a request for a reason and as a request for the cause is a fertile source of confusion.[88] However, this observation by no means implies that the ambiguity made possible by our double-use of "why?" is always, in any given linguistic context, of an unintended and misleading character.

Sometimes we use "why?" as a plain, unequivocal request for a reason why someone acted in a particular way and refuse to take causes for an answer, as when we ask our spouse, "why did you paint the kitchen blue when we decided to paint it yellow?" On other occasions and in other contexts, we use "why?" as a request specifically for the cause of something that has happened and, conversely, refuse to take reasons for an answer, as when we ask a tsunami scientist, "why did the tsunami in 2004 happen?" In these kinds of unambiguous situations, we may come to question the linguistic competence of our interlocutor, or her moral character, or her general understanding of the subject matter of our question, if she finds our plain, unequivocal "why?" open to more than one interpretation. In other situations, however, our interlocutor's sensitivity to actual ambiguity in the question we are asking might give us confidence in her judgment and understanding. The ambiguity might be deliberate on our part, or it might be of a less deliberate sort which we only become aware of while asking, or after having asked, our ambiguous question, an ambiguity that registers some degree of interpretive openness regarding the subject matter of our inquiry. There are likely to be other variants of this type of ambiguity, but the one that interests me here involves a particular kind of incomprehension on the part of the person who is asking the ambiguous why-question. This person (the inquirer) is puzzled by something someone does to the point of asking her (the agent) why she did it, but since her puzzlement can be partly allayed both by responses that

[88] Wittgenstein repeatedly comes back to the ambiguity that results from the overlap of our ways of talking about causes and reasons: "…it is frequently asked "How do you know it?" There is an ambiguity here between reasons and causes.' (1982: 28); "…another confusion sets in, that between reason and cause. One is led into this confusion by the ambiguous use of the word 'why'" (1958: 15). See also Wittgenstein 1982: 5 & 39 and 1966: 21.

refer to reasons and by ones that refer to the causes of the behaviour, the inquirer is open and receptive to both types of explanatory response. The inquirer is dismissive of neither reason-citing responses nor of cause-specifying ones, because the motivational basis of the behaviour that puzzles her might include, or is likely to include, both rational and causal elements. The behaviour might be purposive and guided by normative constraints (rules, norms, conventions, and other such socio-cultural prescriptions) of the kind that we express in terms of reasons, while being also motivated by a mental mechanism or disposition of a causal character, such as a depressive disposition that prompts pessimistic thoughts and interpretations of social events and the agent's future.[89] In such cases, the significance of behaviour might be best conveyed by explanations that combine reasons with causes, and a person trying to understand such behaviour might reasonably approach the agent with questions that express her equal interest in the reasons for and the causes of her behaviour.

Having distinguished intended from unintended ambiguity, I can now reframe the question regarding the status of Kaitlin's response to Linda's question – causal explanation or not? – to see whether an alternative formulation might affect the answer I feel compelled to give. Here, then, is the old question in a new form: If it is true, as I have suggested, that a speaker can express her equal interest in the reasons for and causes of an agent's behaviour by posing an ambiguous why-question, might it be possible for the agent to respond to such a question in a correspondingly ambiguous manner, with a because-statement that somehow expresses by the same words at the same time both the reasons for and the causes of her behaviour? This formulation, I must admit, turned out queerer than I expected, and I am not sure if it even makes sense. However, if it does, I am strongly inclined to answer in the negative: No, one cannot express both the reasons for and the causes of one's behaviour at the same time by the same

[89] Aaron Beck describes depressive dispositions, or "schemas" as he calls them, as follows: "The characteristics of depression can be viewed as expressions of an underlying shift in the depressed patient's cognitive organization. Because of the dominance of certain cognitive schemas, he tends to regard himself, his experiences, and his future in a negative way. These negative concepts are apparent in the way the patient systematically misconstrues his experiences and in the content of his ruminations. Specifically, he regards himself as a 'loser.' First, he believes that he has lost something of substantial value, such as a personal relationship, or that he has failed to achieve what he considers an important objective. Second, he expects the outcome of any activity he undertakes to be negative. Therefore, he is not motivated to set goals and, in fact, avoids engaging in 'constructive' activities. Furthermore, he expects his entire future to be deficient in satisfactions, achievements. Third, he sees himself as a 'loser' in the vernacular sense; he is inferior, inept, lacking in worth, awkward, and socially undesirable" (1976: 264)

words. Without a positive answer to this queer question, we still lack a reason to regard Kaitlin's response to Linda's question as a causal explanation. Even if Linda's question were best understood as an expression of her undifferentiated interest in the overall significance of Kaitlin's behaviour, this consideration should not lead us to revise our earlier assessment of the noncausal status of Kaitlin's response.

Kaitlin's response to Linda's question is *not* a causal explanation of her crying but a reason-citing statement that draws, or is likely to draw, Linda's attention to the need for elaborative description of the mental disposition(s) that causally contributed to Kaitlin's experience of the film-scene that brought her to tears. Kaitlin's response fails as an explanation insofar it leaves Linda puzzled as to why the scene with the drunken father made such a strong impression on her. An "explanation" is, after all, a statement or account that *explains* – an act of telling or showing that dissolves someone's state of puzzlement or incomprehension – and Kaitlin's reply fails to do so. At best, it might qualify as a "partial" or an "incomplete" explanation, as it only partly, if at all, reduces Linda's puzzled state of mind. It might seem to merit the description of "incomplete explanation" as it might make a positive contribution, even if only a slight one, to Linda's understanding of Kaitlin's emotional reaction by offering, or affirming, the reason for her crying – the sad aspect of the film-scene with the passed-out father – while leaving open the further question of why she experienced the film-scene as intensely as she did. This further question is, according to my view, a causal question about Kaitlin's mental life. Yet, the fact that Kaitlin's response leaves open a causal question does not transform it into a causal explanation. No, Kaitlin's response is and remains a reason-citing explanation, albeit an incomplete one. It is incomplete because, while not being altogether insubstantial, it fails to give Linda the causal details she would need to understand the significance of her friend's emotional reaction.

I have now reached a point in my inquiry from which I am able to evaluate the two arguments presented earlier (pp. 62-3), against philosophical views that take the distinction between causal and reason-citing explanations to be of a sharp, categorical character. The first and the bolder of the arguments maintained that there are reason-citing explanations that ought to be regarded as causal explanations, and the second and more modest argument (merely) expressed misgivings about exaggerating the sharpness of the distinction and overstating its applicability. The question that confronts me now is whether Kaitlin's reason-statement about her crying ("When that girl found her father passed out on the floor, it was so sad I had to cry") lends support to either of these arguments.

I have in fact already given my answer regarding the first and the bolder of the arguments. The fact that Kaitlin's reason-citing response might raise further questions about the causal determinants of her reaction does not transform it into a causal explanation. Despite the causal questions it raises, Kaitlin's response is a reason-citing explanation of her crying, insofar as it counts as an "explanation" at all. It cannot reasonably be regarded as a causal explanation because it contains conceptual features that are lacking from the concept of cause (and which are contained in the concept of reason), such as the feature of being evaluable in regard to its better or worseness,[90] and the feature of being stated by someone who occupies a privileged, first-person standpoint in the act of offering her explanation.

As for the second argument, I find it much harder to say anything definite as to whether the film-example supports it or not. Does Kaitlin's response to Linda's question blur the distinction between causal and reason-citing explanations and/or raise doubts about the scope of its applicability? My inclination is to answer both Yes and No. Yes, if I take the question as an empirical question about the "causal" and "rational" modes of thinking that converge in the act of interpretation by which Linda might try to understand Kaitlin's explanation of her crying; and No if I take it as a philosophical question about the relationship between Kaitlin's explanation, as a particular instance of language-use, and the distinction between causal and reason-citing explanations, as a conceptual difference that depends on our discourse of causes and reasons. Let me explain what I mean in more detail, beginning with the negative answer.

The distinction between causal and reason-citing explanations is – as all distinctions are – conceptual and based on our conversation and thinking of causes and reasons in various practices, activities, and situations. Were we to examine our discourse of causes and reasons in those various contexts in which it takes place, we would come upon instances of language-use where the distinction finds clear and straightforward application – instances involving assertions and statements such as the following: "My friend Alex bought a sauna-tent because they are cheaper than building an actual sauna"; "Cindy got a toothache because of the sudden drop in temperature." There is no question of which of these statements is a causal explanation and which a reason-citing one.[91] But there is a feature of these statements, or rather, a feature

[90] Why is it that Kaitlin's response enables Linda to assess the appropriateness of her reaction to the film? Because Kaitlin's reply reveals her implicit evaluation ("so sad") of the object of her feelings (the film-scene with the passed-out father) – a revelation that situates her reaction in the "normative dimension" where the better and worseness of human thought and behaviour can be assessed.

[91] For a more fanciful example of explanatory statements on which the distinction between causal and reason-explanations finds easy application, consider two explanations that Isaac Newton – the Newton of legend –

of our spontaneous way of reading and apprehending them, that I would like to highlight in this connection. Notice how these statements come across as *explanations* even when they are abstracted from the kinds of explanatory context in which they would be used, if ever, in ordinary everyday speech. Even when removed from their proper surroundings, these statements, free-floating and detached from their natural habitat, come across as "explanations," seem to satisfy our criteria for calling them so: they explain, or seem to explain, why and how something is as it is in a sufficiently conclusive manner; they seem to serve the purpose of explaining; and so on. The same can also be said about the differentiating features of the concepts of cause and reason that I outlined earlier in this chapter. Those features too – the justifying function of reasons, the asymmetry between first and third-person reason-statements, etc. – seem instantiated in these statements, free-floating as they are from the kinds of contexts in which they would be used in everyday discourse. Taken together, then, these observations suggest a kind of (circular) explanation of the ease with which the distinction of causal and reason-explanation applies to these particular statements: the ease of application is due to the clear and definite way in which these statements seem, on the face of it, to satisfy the criteria of application for the words that make up the compound terms by which we refer to the distinction that concerns us here: "cause," "reason," and "explanation." By inflecting and hyphenating these words appropriately, we get the compounds "causal explanation" and "reason-explanation."[92]

However, not all attempts to specify the causes and reasons of behaviour lend themselves to the distinction of causal and reason-explanations with the ease, or seeming ease, of these statements. There are cases in which it is unclear whether a statement about someone's acting or doing qualifies as an "explanation" of her behaviour, and, if we are led to conclude that it does so qualify, unclear, too, whether the statement is better regarded as a "causal" or a "reason-

might have given in response to the question, "How did you come up with the law of universal gravitation?" The first response, which takes the form of a causal explanation, is this: "The theory came to me when I was sitting in the orchard and suddenly an apple fell on my head". And the second response, the reason-citing one, goes something like this: "While I was observing apples in the orchard and wondering why they always fall straight to the ground, I realized that there must be a force of attraction of some sort that exists between any two bodies in the universe, which then led me to the idea that…etc…etc."

[92] My use of "seem" in this paragraph, annoying as it might be, is meant to indicate awareness of the problems associated with judging whether statements abstracted from their proper contexts satisfy certain criteria of application or not. If the context of use not only provides a setting for a statement, but also contributes to its meaning, how can one judge whether a statement abstracted from its context satisfies some criteria or not? Such abstraction casts doubt on the meaning of the statement and, hence, makes it difficult, if not impossible, to say whether the statement satisfies some criteria or not.

explanation" – or, indeed, whether it can be properly regarded as *either*. I think Kaitlin's response to Linda's question combines these uncertainties. It is unclear whether her response counts as an "explanation," as it does not explain her reaction to the film-scene in anything like a conclusive manner. And if one deems it appropriate to think of and describe her response as an "explanation," one might still hesitate to label it as a "causal" or a "reason-explanation." Why? Because to grasp an *explanation of Kaitlin's reaction* is also to grasp and understand *Kaitlin's reaction*, and a genuine attempt to understand her reaction will engage the person seeking understanding in a pursuit of both causal and normative questions – questions such as the following: "Has Kaitlin got a reason for crying?"; "What is it that gives the film-scene the charged significance it has for her?"; "What experiences has Kaitlin had that have led her to develop a disposition to react thus-and-so in such-and-such situations?" This is one reason, I think, why some philosophers might find it awkward to choose between the labels "causal explanation" and "reason-explanation" in cases of this sort. It might seem counterintuitive and misleading to think of Kaitlin's explanation in terms of these labels once it has been recognized, first, that the interpretation of her avowed reason for crying engages one in interpretation of the act of crying itself, and secondly, that causal considerations play a significant role in the interpretation of the act of crying (and thus also in the interpretation of her explanation). If the interpretation of her crying prompts us to interpret the reason we are offered for her crying in the light of certain causal considerations about her mental constitution, is it appropriate for us to think of Kaitlin's explanation as "noncausal"? Does it not rather show that there are cases of behaviour that go awkwardly with the distinction between causal and reason-explanation because of the mixed (causal and normative) considerations that go into our attempts to grasp their significance? These are, I think, legitimate questions to ask and ponder. However, in asking and pondering these questions, one must not mistakenly think or suppose that the usefulness of the distinction, let alone its intelligibility, presupposes our ability to apply it universally, without doubts or misgivings, in any context where there is behaviour that requires explanation. The existence of contexts of explanation in which the application of the distinction is doubtful or problematic does not by itself detract from the intelligibility or usefulness of the distinction in other contexts. This is why part of me is inclined to answer negatively to the questions raised earlier – the questions of whether Kaitlin's explanation of her crying blurs the distinction between causal and reason-explanation, and whether it raises doubts about the usefulness and useability of the distinction: Kaitlin's explanation does *not* blur the distinction, because its sharpness or blurriness is a conceptual issue independent of particular cases of

language-use; *nor* does Kaitlin's explanation raise doubts about the usefulness or useability of the distinction as long it holds good and finds application in other contexts of use.

There is, then, a clear sense in which Kaitlin's response to Linda's question does not blur the distinction between causal and reason-explanation or raise doubts about its usefulness or useability. However, I have also identified a sense in which Kaitlin's response might nonetheless be said to do these things. I have drawn attention to the causal questions that Kaitlin's explanation of her crying gives rise to and used the absence of answers to those questions to account for the weakness of her explanation. Without answers to those causal questions, I pointed out, Kaitlin's reason-explanation is likely to appear "enigmatic" and lacking in force to its recipient. This was explained in terms of a kind of indeterminacy of meaning that surrounds the reason Kaitlin gives for her crying. The referential object of her reason, I explained, is unclear and unspecified in the sense that the personal meaning it has for Kaitlin is unknown to Linda. There is a variance between the overt, public meaning of the scene with the passed-out father and the covert, subjective meaning it acquires for Kaitlin personally, a variance hinted at ("…it was so sad…") but left unresolved in Kaitlin's explanation. Linda cannot make sense of Kaitlin's reason for crying, because her experience of the scene with the passed-out father differs too much (in the direction of "appropriateness") from Kaitlin's experience of the scene for Linda to understand why it should elicit such uncontrollable crying. In Kaitlin's experience, the scene merges with her past in such a way and to such an extent that it becomes difficult to say what exactly it is that she is reacting to. But while the indefiniteness of the object of Kaitlin's emotions makes it difficult for Linda to make out Kaitlin's reaction, Kaitlin's explanation of her crying does little to eradicate the vagueness. Her explanation, if anything, exacerbates the vagueness by taking the form of a reason-explanation, and thereby – since reasons involve publicly available rules, norms, conventions – downplaying the variance between the public meaning of the film-scene and the personal meaning it takes on for Kaitlin. By singling out the scene as her reason for crying, Kaitlin suggests, in effect, that her crying was brought on by the scene as it is experienced by the average "reasonable" viewer, as an object of experience the significance of which corresponds (closely enough) to the overt public meaning of the scene. This is misleading because Kaitlin's experience is very much shaped by the personal past she, as a person suffering from trauma, carries within her and brings to bear on the objects of her experience. Indeed, the meaning and emotional valence the relevant film-scene takes on for Kaitlin (by becoming linked with her past) departs so widely from the significance the average viewer would attach to the scene that Kaitlin's act of putting forward

the scene as her reason comes to appear almost spurious, in that it characterizes her reaction as more "reasonable" than it is. As a reason for crying, the scene does not explain Kaitlin's reaction, because her reaction does not correspond to the general public's (and Linda's) idea of what an "appropriate" reaction to the scene would be like. The *type* of emotion Kaitlin's reaction instantiates (sad feelings, crying, etc.) is "appropriate" enough, but its *degree* of intensity is "excessive." Kaitlin's reaction is, Linda thinks, out of proportion to the scene with the passed-out father. Her reaction is over-great, and this over-greatness is what her explanation, more than anything else, fails to explain. Her reaction is not "reasonable" enough to be conclusively explained, or explainable, by reasons. For while the reason Kaitlin offers points to a normative connection that she evidently understands and is aware of, a connection that informs her reaction – the connection that holds between "sad events" and "crying" –, her reaction is in important ways influenced by a nonrational mental disposition that itself requires causal explanation. Such explanation, I have argued, comes in two varieties: formal-causal explanation that explains the nature and symptomatology of some part a person's mental constitution, and efficient-causal explanation that explains the formation and development of that part of the person's psyche. In Kaitlin's case, the formal-causal explanation would explain the nature and symptomatology of such things as Kaitlin's childhood trauma of being "abandoned" by her parents, her fear of rejection and intimacy, and other related mental dispositions and mechanisms that shape her thinking, feeling, and behaving; and the efficient-causal explanation would locate and characterize the past events and circumstances that contributed to the development of those mental dispositions and mechanisms. Both kinds of explanation are potentially relevant in cases of behaviour that are "reasonable enough" to lend themselves to reason-explanation,[93] but not enough for reason-explanations to bring out their "full" significance. In such cases, rational and causal modes of thinking may converge in the act of interpretation in ways that might make a philosopher hesitate to label the explanations that result from the interpretation as either causal or reason-explanations, especially if the difference between these forms of explanation is taken as being of categorical character. The reason why a philosopher might feel uncomfortable about such labelling is, I think, this: she assumes that the categorical status of the distinction between causal and reason-explanation requires as its legitimizing counterpart a form of "categorical thinking" that follows the logic of one type of explanation and one type alone, and then, proceeding on this assumption, she

[93] Here "reasonable enough" means, minimally, that the behaviour involves the agent's exercise of her conceptual capacities, and the agent makes connections between ideationally related elements, connections which are evaluable as to their better or worseness.

comes upon explanations of behaviour that require non-categorical interpretive measures and takes them as evidence of the non-categorical status of the distinction: "The distinction between causal and reason-explanation might not be of categorical character after all, since there appears to be instances of reason-explanation that require interpretive measures that employ causal thinking."

It was for the above considerations I said that part of me is inclined to give a positive answer to the question of whether Kaitlin's response to Linda's inquiry blurs the distinction between causal and reason-explanation: Kaitlin's response does seem to blur the distinction in the sense that rational and causal modes of thinking converge in the efforts by which its recipient is likely to interpret its meaning, and converge in such a way that it becomes hard to separate one mode of thinking from the other. The reason Kaitlin gives for her crying cannot be properly understood in this context without recourse to elaborative descriptions of the mental disposition(s) that causally contributed to her experience of the film-scene that brought her to tears. Insofar as Kaitlin's response to Linda's question counts as an "explanation," it is, specifically, a reason-explanation, but one that calls for supplementation in the form of description of the mental disposition(s) that influenced her experience of the referential object of her reason. For these reasons, Kaitlin's explanation does not lend much straightforward support to philosophical positions that seek to sharpen the distinction between causal and reason-explanation to the point of describing it as a "categorical distinction." What this example shows is that our ordinary ways of thinking and talking about the motivational basis of behaviour involves more shifting between forms of explanation than such philosophical positions might suggest.

Having now considered all the major issues I intended to discuss in this part of my paper, I am approaching the end of the present chapter. But there is one more task to be completed. Before presenting the film-example with Kaitlin, I stated as one of the aims of the subsequent discussion the demonstration of how an increased understanding of the formal-causal model of explanation might change our conception of, and approach to, certain questions philosophers have asked about the distinction between causal and reason-explanation. Now that subsequent discussion has drawn to a close without my having specified which philosophical questions I had in mind when making that statement. I would therefore like to end by briefly discussing those questions.

One of the things I have sought to illustrate is the complementary relationship that can be found between causal and reason-explanations in some interpretive contexts where genuine efforts are made to understand people's actions and doings. Kaitlin's emotional reaction was presented as an example of behaviour the significance of which can be brought into focus, *not* by any one explanation of one variety or another, *but* by a set of explanations of different types. Such partial explanations can together bring out the significance of certain kinds of behaviour in a more perceptive and psychologically plausible manner than any single explanation of any given sort could. Furthermore, this implies that different explanations need not conflict one another, despite being of different kinds, but can enrich, clarify, and expand one another in various ways. We saw this with Kaitlin's explanation of her crying. The reason she gave for her emotional reaction urged us to further interpret her crying in causal terms, by explaining her experience of the relevant film-scene as (formally) caused by her mental disposition to think, feel, and to act in certain ways in certain kinds of situations, and further, by trying to grasp her mental disposition as the (efficient-)causal outcome of her past experiences. So these different kinds of explanation we are mentally juggling with here stand in a relationship, *not* of competition ("the reason-explanation is more important than the causal explanations"), nor of mutual exclusion ("the reason-explanation contradicts the casual explanations"), nor of reducibility ("the reason-explanation is reducible to a causal explanation"), nor even of mutual independence ("the causal and reason-explanations account for the same phenomenon on different levels of explanation and, hence, neither compete with nor support one another"), *but* stand in a relationship of complementary dependence. Kaitlin's emotional reaction reveals its significance when considered from an inter-explanatory perspective, in terms of both reasons and causes, efficient and formal. This complementarity, as I hinted above, should make us suspicious of philosophical claims and doctrines that assume the relationship between causal and reason-explanation to be oppositional, reducible, mutually exclusive or mutually independent. So we have grounds to be suspicious, for instance, of claims that exclude the possibility of actions the significance of which require combined explanations that refer to both causes and reasons, claims such as the following: "There simply is no such thing as a causal explanation of action because to explain an action is to explain it rationally and to explain an event is to explain it causally."[94] Generalizations of this sort leave us with fewer resources to

[94] D'Oro 2012: 212. This statement occurs in the context of a discussion of the philosophical debate between 1960s non-reductivists – Dray, Melden, Anscombe, von Wright – who maintained that action-explanations are rational explanations of how agents ought to act in response to norms of instrumental reasoning, on the one hand, and Hempelian reductivists who claimed that action-explanations are a species of causal-nomological

interpret human activity by erasing distinctions between kinds of action that require different kinds of explanation. There are, of course, plenty of actions that we explain, and consider "fully" explainable, by reasons. However, the ubiquity of such actions gives no reason to conclude that *all* actions are explainable by reasons and reasons alone. Moreover, if dispositional explanations of the sort I have considered in this study have a (formal) causal aspect to them, as I have argued they do, then there are actions that we explain by reference to both reasons and causes. Imagine for instance someone who finds herself in a situation in which there is more than one course of action available to her, all of which find some support from reason: one possible course would satisfy some healthy desire of the agent (say her desire to learn more about classical music); another course of action would contribute to the well-being of her family in some specifiable way; and a third would benefit her career and professional life. After some self-reflection and consultation with friends and family, our agent decides in favour of the course of action the goal of which she takes herself to have the best reason to pursue in her situation in life: the course she expects to contribute to the well-being of her family. But now comes the twist in the story. Our agent also has an unconscious motive to adopt the course of action that she has the best reason to pursue: by taking that particular course she can avoid certain social situations she is unconsciously afraid of. In other words, our agent is self-consciously adopting the course of action she has the best reason to pursue, while being unconsciously inclined toward that course by a nonrational bias.[95] Now, I think many philosophers would consider in many a context an explanation that omits the agent's unconscious fear as insufficient or incomplete. Unconscious fears, desires, jealousies, and other such nonrational motives are obviously relevant to our attempts to understand people's actions and behaviour. But how do we think of and explain such motivating mental factors? I claim that we already in our non-theoretical, mundane, everyday inferences about each other's lives and behaviour regard such motivating factors as being of a causal nature. There is something

explanation on the other. The statement represents, *not* D'Oro's own position, but the non-reductivists side of the debate. D'Oro's own position, if I have understood her correctly, is that causal and reason-explanations are categorically distinct but mutually compatible, so that a physician's causal explanation of Alexander Litvinenko's death does not compete with a political historian's reason-explanation of Litvinenko's death (D'Oro's example). Causal explanations do not compete with reason-explanations, but neither do they enrich and complement the latter. On D'Oro's view, which equates "causal explanation" with "efficient-causal explanation," causal and reason-explanations stand in a relation of mutual independence to one another. My citation from D'Oro is not a critique of her own position *vis-à-vis* the philosophical debate in question.

[95] This example, I want to note, reconceptualizes and makes implicit use of Donald Davidson's distinction between "reasons for acting" and "the reason why we act". According to Davidson, the reason why someone acts in a particular way (but not the other reasons she may have for acting) is also the cause of her action. See Davidson's "Actions, Reasons, and Causes".

causal about having one's thinking, acting and relating to others, influenced by nonrational parts of one's mind, which lead one to gravitate toward certain kinds of people, activities, and social contexts rather than others. What I have proposed in these pages, following Jonathan Lear, is a way of bringing our theory closer to our common-sense intuitions. My suggestion is that we in philosophy start thinking of the contribution of nonrational mental dispositions to the significance of actions and behaviour in terms of formal causes, and further, that we conceptualize explanations of the influence of such factors as formal-causal explanations. However, doing this demands that we accommodate our concept of action to the implication of there being actions that are best explained by means of explanations that combine reasons with causes.

So a better understanding of the formal-causal model of explanation might persuade philosophers to reformulate some questions they have asked about the distinction between causal and reason-explanation, for instance, questions that exclude the possibility of actions that require explanatory statements couched in terms of both causes and reasons. Instead of asking "Are actions explained causally or rationally?" we might ask "Which action can be explained rationally?" or "Which features of actions can and which cannot be explained rationally?" Whatever form the new questions would take, they would harmonize with our new appreciation of the ways in which causal and reason-explanations can complement one another and together bring out the significance of certain actions and behaviour.

Besides enhancing our understanding of the ways in which causal and reason-explanations complement one another in some interpretive situations, an enriched understanding of formal-causal explanation gives us a new diagnosis of the inclination of some philosophers to regard reason-explanation as a type of causal explanation. As we have seen in the preceding discussions, this inclination is misguided as far as Kaitlin's explanation of her crying, and other relevantly similar reason-explanations (which I specify in the next paragraph), are concerned. If a statement that refers to a reason for acting in a certain way counts as an "explanation," then it counts as a "reason-explanation" too, no matter how much or how little complementation it requires in the form of causal explanation to sufficiently explain the behaviour in question. Whatever causal considerations the statement might occasion on our part are no reason for us to regard the statement as a causal explanation. Or so I have argued.

However, in addition to offering this argument as to why the inclination to regard certain reason-explanations as causal should be abandoned, I have also hinted at an explanation of why

someone might come to suffer from this unfounded inclination to begin with. The intuition that drives the inclination is motivated, I have suggested, by the ways in which causal and rational thinking may converge in our interpretive efforts in contexts of explanation where we are interested in the overall motivational basis of people's actions and doings. Such interpretive situations may involve cases of overly neurotic activities, where the agent's capacity for reason has clearly been compromised by some nonrational mental mechanism or disposition, such as an unconscious fear of something, and her reasoning thus only partly guides, if at all, the performance of the activity that needs explaining. But such situations may also involve less overly unreasonable activities where the agent's avowed reason for acting in a certain way actually goes some way towards explaining her behaviour, but we are nonetheless left pondering whether and how her mental dispositions and past experiences might have influenced her behaviour in some way. When confronted with the task of interpreting behaviour that falls into one of these categories – overly neurotic behaviour, or (borderline-)normal behaviour that is influenced by the agent's mental disposition(s) or past experience(s) in some nonpathological way(s) – our desire to understand and make sense of the behaviour might lead us to formulate two or more partial explanations that refer to both causes and reasons, partial explanations that clarify and expand the meaning of one another and together provide a sense of the overall significance of the behaviour. The interpretive effort we imagined Linda to engage in in relation to Kaitlin's emotional reaction was an example of this. Due to its lacking explanatory power, the reason-explanation Kaitlin offered to Linda for bursting out in tears urged Linda to interpret her friend's emotional reaction further by means of causal explanations that sought to identify the mental disposition that influenced her experience of the relevant film-scene, on the one hand, and to locate the past experiences that led her to develop that mental disposition on the other hand. These explanations complemented one another in various ways and together gave Linda and us (when we imaginatively placed ourselves in her situation) some sense of the significance of Kaitlin's emotional reaction. But this complementarity, as we saw, was also potentially misleading, as the temporal and semantic interrelations between the explanations we were mentally juggling with could easily lead us to confuse the conceptual features instantiated by one explanation with the conceptual features instantiated by another. In particular, the causal-logic, or causal character, of the supplementary explanations by which we sought to understand some aspects of Kaitlin's mental life could rub off on the reason-explanation she gave for her crying and make it seem somehow "causal." In other words, Kaitlin's reason-explanation could take on a (false) causal appearance because of the temporal and semantic relationships it formed in our thinking with the formal and efficient-causal

explanations by which we sought to make sense of her reason for crying. This analysis offers us a new perspective on the inclination to regard reason-explanations of certain kinds of behaviour as a type of causal explanation. According to the account I am proposing, those (neurotic and nonpathologically-influenced-by-the-past type of) behaviours may prompt us to interpret them in terms of both causes and reasons, by means of two or more partial explanations that are so closely connected to one another that their conceptual features might become indistinguishable and create the impression of a reason-explanation that has something "causal" about it. If this diagnosis is correct and describes a psycholinguistic phenomenon that affects philosophers (and other language-users) more broadly, then my account of formal-causal explanation has helped me to identify and explain an unfounded intuition that motivates some of the conceptual confusion that contributes to the appeal causal accounts of reason and reason-explanation continue to have among contemporary philosophers, even among thinkers who acknowledge the conceptual difference between causes and reasons.

5.3 Summary

In this chapter, I have examined some philosophical problems concerning the distinction between causal and reason-explanation through a perspective informed by an awareness of the difference between efficient and formal causation to see whether an appreciation of this subdivision of causal explanation might lead me to reconsider our intuitions regarding those problems. My central focus has been on the question of whether reason-explanations explain people's actions and doings causally or not. I began the investigation as someone holding the view that reason-explanations do not explain their explananda causally, and at the end of the investigation, I find myself still holding that view. What has changed in the course of these pages, however, is my understanding of an inclination I also find in myself to regard certain reason-explanations as having something "causal" about them. The intuition that drives this inclination is motivated, I believe, by the ways in which causal and rational modes of thinking may converge in our interpretive efforts in contexts of explanation where we are interested in the overall motivational basis of people's actions and doings. In cases where our desire to understand someone's behaviour leads us to formulate two or more partial explanations that refer to both causes and reasons, the close temporal and sematic relations between the explanations in our thinking can lead us to confuse the conceptual features of one explanation

with the features of the other. The causal-logic of our causal explanations can spill over into our reason-explanations and create the impression of "causal" reason-explanations.

Besides considering such causal impressions that reason-explanations can give when accompanied by causal explanations, I have devoted a large portion of this chapter to examining the nature and relative sharpness of the distinction between causal and reason-explanation, and in doings so, have articulated some consideration that speak against philosophical views that treat the relationship between these explanatory models as oppositional or noncomplementary. For a philosopher to treat their relationship in these ways involves a failure to recognize the complementarity we find between causal and reason-explanations in interpretive situations of the kind that Linda found herself in in relation to Kaitlin's emotional reaction. What the example with Kaitlin showed was just how hard it can be to separate reason-explanations from causal ones when dealing with certain kinds of behaviour, and how certain behaviours reveal their significance only when apprehended in terms of both reasons and causes. This complementarity between causal and reason-explanation puts pressure on philosophical views that construe the distinction between causal and reason-explanation as a sharp distinction of type rather than a less sharp one of degree, especially views that emphasize the function of this distinction to distinguish between what is and what is not the product of the agent's will.

6. Conclusion

Our causal language for describing and explaining behaviour includes an explanatory model that differs in significant respects from the efficient-causal prototypes discussed by Wittgenstein in "Cause and Effect," a model that represents actions and doings in terms of enduring mental structures, such as childhood traumas, that exert a formative influence on the agent's thinking, feeling, and acting, by way of patterning her experience of the world and the people and things in it. Such explanations, I have found, can open new perspectives on philosophical problems that arise in connection with causal accounts of human nature and activity. I have illustrated this by examining three such problems from the viewpoint of the formal-causal model: the problem of determinism, the problem of physicalist reductionism, and the problem of whether reason-explanations are a type of causal explanation or not. My accounts of why these problems do not arise in connection with formal-causal explanations have also served to exemplify the kinds of insights one might gain by approaching "causal problems" from the viewpoint of extraneous, conceptually unrelated models of explanation, insights that can help philosophers make better-informed assessments of the scope, weight and validity of such problems.

7. Svensk sammanfattning

Denna pro gradu-avhandling undersöker sambandet mellan filosofiska frågor och problem förknippade med kausalitet och specifika kausala förklaringsmodeller. Fokuset ligger på frågor som uppstår då mänskligt handlande förklaras kausalt, i termer av orsak och verkan, exempelvis frågan om den fria viljan, och sambandet mellan dessa frågor och vissa typer av orsaksförklaringar. Vår orsaksdiskurs omfattar flera förklaringsmodeller som tillsammans med andra språkmedel (kausativa verb etc.) ger orsaksbegreppet dess mening, men den filosofiska litteraturen har tenderat att bortse från denna språkliga mångfald och överbetona vissa delar av orsaksdiskursen. Historiskt är det de naturvetenskapliga orsaksförklaringarna som dominerat det filosofiska tänkandet i den västerländska traditionen, och orsaksbegreppet har således ofta reduercats till dess innebörd i naturvetenskapliga sammanhang. David Hume, vars kausalitetsteori harmoniserar med den moderna empiriska naturvetenskapen som växte fram på 1600-talet, har varit speciellt inflytelserik inom den anglo-amerikanska filosofin. Fram till 1970-talet var den huemska (”nomologiska”) synen på kausalitet dominerande bland empriskt-analytiska filosofer. Enligt Hume handlar kausalitet om lagbundna förhållanden där en viss typ av händelse, A, alltid följs av en annan typ av händelse, B. Denna syn på kausalitet överrensstämmer med vissa typer av kausala påstående som görs inom de empiriska naturvetenskaperna men fångar inte vårt språk kring kausalitet i dess helhet. Det finns andra typer av kausala förklaringar som inte handlar om lagbundenheter mellan orsak och verkan, förklaringar som ofta har förbisetts i den moderna filosofin om kausalitet. I min avhandling vill jag lyfta fram sådana förbisedda kausala förklaringstyper. Mitt syfte är att utmana antaganden som ligger till grund för den filosofiska problematiken med att förklara mänskligt handlande kausalt, till exempel antagandet att kausala förklaringar reducerar mänskligt handlande till fysiologiska skeenden och leder till en form av determinism.

I avhandlingens första kapitel granskar jag Ludwig Wittgensteins anmärkningar om kausalitet i ”Cause and Effect: Intuitive Awareness”. I dessa anmärkningar påminner Wittgenstein läsaren om andra orsaksförklaringar än de ”nomologiska” förklaringarna som betonats i den moderna västerländska filosofin. Orsaksförklaringar av den nomologiska varianten utgör endast en del av vårt kausalitetsspråk, som innefattar åtminstone fyra andra typer av orsaksförklaringar, typer som väcker andra frågor än de som varit i fokus sedan Humes och Kants dagar. Wittgenstein hjälper oss att inse att de återkommande frågorna förknippade med kausalitet är mer eller mindre lokala frågor som uppstår i samband med särskilda typer av kausala satser och inte med andra. Vårt kausalitetsspråk inbegriper orsaksförklaringar av olika

slag, som ger upphov till olika frågeställningar och problem. En följd av dessa insikter om å ena sidan orsaksdiskursens mångfald och å andra sidan sambandet mellan särkilda orsaksförklaringstyper och särskilda frågor om kausalitet, är att filosofer som engagerar sig med frågor förknippade med kausalitet behöver vara medvetna om den språkliga mångfalden som karaktäriserar vår orsaksdiskurs för att kunna bedöma frågornas omfattning, betydelse och relevans. Om inte denna lingvistiska medvetenhet finns, finns det risk att man överskattar frågornas omfattning eller bygger ens resonemang på ogrundade antaganden som ligger gömda i frågornas formulering. Sådana missförstånd är vanliga i filosofiska diskussioner om problematiken med att förklara mänskligt handlande kausalt. Det är vanligt att frågan om determinism och den fria viljan uppfattas som allmäna frågor som berör alla typer av kausala förklaringar, trots att de är logiskt oförenliga med vissa typer av kausala satser. Och samma sorts generalisering kan också utmynna i det felaktiga påståendet att kausala förklaringar (oavsett typ) reducerar mänskliga handlingar till fysiologiska skeenden – ett påstående som bortser från orsaksförklaringar av mänskliga reaktioner som implicit tillskriver olika mentala (kognitiva, perceptiva, emotionella) egenskaper till en kausalt påverkad agent. Det är alltså vanligt att lokala problem som hör ihop med vissa typer av orsaksförklaringar – huvudsakligen med nomologiska förklaringar – uppfattas som allmänna problem med orsaksförklaringar och därför upplevs som större och mer angelägna än de egentligen är. Vad som ytterligare bidrar till felbedömningar och missförstånd av denna art är vissa metafysiska antaganden som åtföljer orsaksförklaringar av den nomologiska varianten, ogrundade antaganden om förhållandet mellan språk och verklighet och om verklighetens grundläggande natur.

Efter en inledande beskrivning av den här sortens felbedömningar och missförstånd som kan uppstå till följd av en bristande insikt om vårt kausalitetsspråk och dess mångfald introducerar jag en ytterligare typ av orsaksförklaring som förblivit ouppmärksammad i litteraturen om kausalitet. Till skillnad från de ”verksamkausala” förklaringarna man fokuserat på i den moderna filosofin om kausalitet – förklaringar som hänvisar till *generativa* relationer mellan orsak och verkan – hänvisar förklaringarna jag intresserar mig för till *formativa* relationer mellan orsak och verkan. Jag betecknar dessa förklaringar, efter Jonathan Lear, som ”formell-kausala” förklaringar. I mina exempel på sådana förklaringar är den angivna orsaken ofta ett barndomstrauma av något slag som påverkar den traumatiserade individens varsebliving och beteende på förutsägbara sätt. Två exempel jag diskuterar lyder så här: “Tom’s trauma led him to avoid the seaside while travelling in Spain”; “Kaitlin’s fear of intimacy has kept her emotionally isolated from her lovers.” Förklaringar av den här typen har en kausal karaktär,

men handlar inte om händelser som följer varandra i tid och rum, utan om djupt rotade mentala tendenser som regelmässigt påverkar vårt sätt att se och reagera på vår omgivning. Från och med kapitel 2 ligger avhandlingens fokus på den här typen av orsaksförklaringar. Jag vill med hjälp av en ökad förståelse för dessa förklaringar belysa den lokala karaktären av filosofiska problem förknippade med kausalitet och öppna upp nya perspektiv på problematiken med att förklarar mänskligt handlande kausalt.

Kapitel 2 och 3 innehåller mer utförliga beskrivningar av den formell-kausala förklaringsmodellen och ger bakgrund till diskussionerna i de påföljande kapitlen. Jag börjar med att redogöra för Aristoteles begrepp om ”formella orsaker,” eller snarar för Jonathan Lears freudianska anpassningar av detta begrepp. Lear menar att det är fruktbart och motiverat att se på mentala komplex som omedvetet styr vårt tanke- och känsloliv som formella orsaker. Ett exempel på sådana mentala komplex, eller ”mentala strukturer” som Lear kallar dem, är ett barndomstrauma som formar den traumatiserade personens liv och relationer, bland annat genom att disponera hen att tolka sociala situationer på ett visst sätt. Min avhandling tar fasta på Lears förslag och undersöker de filosofiska implikationerna av att konceptualisera denna typ av dispositionella förklaringar som ”formell-kausala” förklaringar.

I kapitel 3 vidareutvecklar jag det formell-kausala orsaksbegreppet som introducerades i kapitel 2, skiljer mellan verksamkausala- och formell-kausala förklaringar av mental ohälsa och skisserar frihets- och ofrihetsbegreppen som framträder ur formell-kausala förklaringar av vissa typer av mentala besvär. Diskussionen baserar sig på två påhittade historier om mental ohälsa. Den första historien handlar om en tonårspojke som utvecklar posttraumatiskt stressyndrom efter att ha varit med om en drunkingsolycka, och den andra handlar om en kvinna som lider av anknytningsproblem orsakade av bristande omsorg i barndomen. I en kommentar till den första livshistorien skiljer jag mellan verksamkausala- och formell-kauala förklaringar genom att kontrastera min beskrivning av *uppkomsten* av protagonistens mentala besvär med beskrivningar av *symptomen* som hans mentala besvär orsakar. För mina ändamål är den avgörande skillnaden mellan beskrivningar som konceptualiserar kausalitet som en *verkande kraft* och andra som konceptualiserar kausalitet som en *formgivande struktur*. I den första livshistorien utgår den verkande kraften från den traumatiserande händelsen som ger upphov till protagonistens mentala tillstånd, och den formgivande strukturen består av protagonistens minnen, rädslor och andra mentala dispositioner, som tillsammans utgör den subjektiva polen av hans besvär. (Senare i kapitlet jämför jag den här typen av formgivande strukturer med de transcendentala kategorierna som, enligt Immanuel Kant, formar den

upplevda världen.) Efter att ha beskrivit skillnaden mellan verksamkausla- och formell-kausala förklaringar av mental ohälsa riktas diskussionens fokus till frihets- och ofrihetsbegreppen som framträder ur formell-kausala förklaringar av mentala besvär. I kommentaren till den andra livshistorien försöker jag skissera upp konturerna till dessa begrepp genom att beskriva vissa ”terapeutiska åtegärder” som protagonisten kan vidta för att få kontroll över sina problem – åtgärder som kan tänkas hjälpa henne att ersätta sina destruktiva tanke- och beteendemönster med nya och mer funktionella mönster och mentala färdigheter. Syftet med dessa beskrivningar är att ge bakgrund till diskussionen om determinism och formell-kausala förklaringar i det påföljande kapitlet.

I det fjärde kapitlet återvänder jag till frågeställningarna som presenterats i det första kapitlet. Diskussionen kretsar kring frågan varför den filosofiska problematiken med att beskriva mänskligt handlande kausalt inte berör formell-kausala förklaringar. Här undersöker jag kopplingen mellan deterministiska ståndpunkter och vissa metafysiska idéer (om oändliga orsakskedjor och nödvändiga samaband mellan orsak och verkan) som är logiskt oförenliga med formell-kausala förklaringar, och jämför formell-kausala förklaringar av mänsklig aktivitet med verksamkausala förklaringar som utlokaliserar de motiverande orsakerna till externa faktorer utanför agenten själv. Jag diskuterar fyra begreppsliga egenskaper som förklarar varför formell-kausala förklaringar inte understöder deterministiska ståndpunkter: (1) formella orsaker har ett formgivande (snarare än generativt) inflytande på handlingarnas tillkomst och (2) är för intimt sammankopplade med de självreglerande mentala förmågorna vi förknippar med handlingsfrihet (självmedvetandet, rationalitet etc.) för att kunna uppfattas som externa och bortom agents kontroll. Dessutom (3) är det allmänt vedertaget att vi som människor besitter kognitiva förmågor som möjliggör för oss att påverka mentala dispositioner och mekanismer av det slag som formell-kausala förklaringar anger som orsaker, och (4) formell-kausala påståenden beskriver sannolika och enskilda orsakssamband (”singular causation”), i motsats till lagbunda händelseförlopp. På grund av dessa egenskaper uppstår inte frågan om determinism i samband med formell-kausala förklaringar av mänskligt beteende. Och eftersom formell-kausala förklaringar förklarar beteende i mentala termer uppstår inte heller frågan om fysikalistisk reduktionism.

Men även om formell-kausala förklaringar inte reducerar mänskligt handlande till ett fysiskt skeende eller hotar våra vardagliga uppfattningar och intuitioner om människans handlingsfrihet, betyder det inte att de är irrelevanta för dessa problem. Tvärtom menar jag att en ökad förståelse för formell-kausala förklaringar kan öppna upp nya perspektiv på den

filosofiska problematiken med att förklara mänskligt handlande kausalt. Beträffande frågan om determinism kan en sådan förståelse ge oss insikt om vissa filosofiska intuitioner som kanske delvis förklarar varför filosofer upplever denna teoretiska, metafysiska fråga som viktig och meningsfull. Formell-kausala förklaringar riktar vår uppmärksamhet mot olika typer av inre, mentala tvång som kan manifestera sig i destruktiva, dysfunktionella tanke- och beteendemönster av olika slag. I de flesta exemplen jag diskuterar i min avhandling tar detta inre tvång formen av ett barndomstrauma av något slag, ett trauma som på ett påtagligt sätt begränsar en persons frihet och relationer, men det kan också bestå av mindre patologiska mentala disposistioner och mekanismer som omedvetet formar vårt tanke- och känsloliv. Den här typen av inre tvång, som förstås bäst i formell-kausala termer, ger upphov till existentiella frågor som är nära besläktade med den metafysiska frågan om den fria viljan, och likheten mellan dessa frågor är, tror jag, en orsak till det oupphörliga intresset filosofer visar för diskussioner om determinism. Abstrakta, metafysiska frågor av typen ”Är alla mina känslor och handlingar bestämda av tidigare orsaker?” kan verka angelägna och relevanta om man brottats med existentiella frågor av typen ”Är det *här* vad jag faktiskt vill med mitt liv eller är det mina känslor från barndomen som spökar?” Den senare frågan, som många av oss grubblar över i vissa livsskeden, exempelvis då vi står inför viktiga beslut om karriär eller äktenskap, får den första mera abstrakta frågan att verka mer angelägen än den kanske är.

Liksom diskussionen om determinism, syftar det sista kapitlet till att kasta nytt ljus över en långvarig debatt om kausala förklaringar, nämligen debatten om huruvida rationella förklaringar av handlingar är en typ av kausal förklaring eller inte. Ska rationella förklaringar som förklarar handlingar i termer av *skäl* tillskrivas kausal status, eller utgör de en väsensskild kategori av förklaringar? Undersökningen i det sista kapitlet ifrågasätter värdet av denna fortskridande debatt genom att identifiera vissa filosofiska intuitioner som kanske bidrar till dess tillsynes olösliga karaktär. Här visar jag att rationella förklaringar kan få en kausal ”aura” över sig i språkliga sammanhang där de står i en kompletterande relation till formell-kausala förklaringar. När rationella förklaringar upplevs som otillräckliga, exempelvis då vi anar att en skenbart rationell handling också drivs av ett omedvetet motiv, kompletterar vi ofta med dispositionella förklaringar av en formell-kausal karaktär, vilket kan leda till att den rationella förklaringen börjar framstår som ”kausal.” Det sista kapitlet demonstrerar att en ökad förståelse för formell-kausala förklaringar kan hjälpa oss att identifiera ogrundade filosofiska intuitioner som denna illusion kan framkalla hos filosofer.

Bibliography

Anscombe, G. E. M. 2000 [1957]. *Intention*. Cambridge & London: Harvard University Press.

—— 1993 [1971]. "Causality and Determination". In: E. Sosa & M. Tooley (eds.). *Causation*. Oxford: Oxford University Press: 88-104.

Aristoteles. 1941. *Metaphysics*. Trans. W. D. Ross. In: R. McKeon (ed.) *The Basic Works of Aristoteles*. New York: Random House: 689-926.

—— 1941. *Physics*. Trans. R. P. Hardie & R. K. Gaye. In: R. McKeon (ed.). *The Basic Works of Aristoteles*. New York: Random House: 218-394.

Banville, J. 2005. *The Sea*. London: Pan Macmillan.

Beck, A. T. 1976. *Cognitive Therapy and the Emotional Disorders*. New York: International University Press.

Brinkmann, S. 2006. "Mental Life in the Space of Reasons". *Journal for the Theory of Social Behaviour* 36 (1): 1-16.

Carroll, J. W. & N. Markosian. 2010. *An Introduction to Philosophy*. Cambridge: Cambridge University Press.

Davidson, D. 2006 (1963). "Actions, Reasons, and Causes." In: E. Lepore & K. Ludwig (eds.). *The Essential Davidson*. Oxford: Oxford University Press: 23-36.

—— 1982. "Paradoxes of Irrationality". In: R. Wollheim & J. Hopkins (eds.). *Philosophical essays on Freud*. Cambridge: Cambridge University Press: 289-305.

D'Oro, G. 2012. "Reasons and Causes: The Philosophical Battle and the Meta-philosophical War". *Australasian Journal of Philosophy* 90 (11): 207-221.

Freud, S.. 2001. *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. J. Strachey et al. (eds. & trans.). London: Vintage.

—— 1914. "Remembering, Repeating and Working-Through". S.E. 12.

—— 1915. "The Unconscious". S.E. 14.

—— 1921. *Group Psychology and the Analysis of the Ego*. S.E. 18.

—— 1922. "Some Neurotic Mechanism in Jealousy, Paranoia and Homosexuality". S.E. 18.

—— 1926. *Inhibitions, Symptoms and Anxiety*. S.E. 20.

Finkelstein, D. H. 2003. *Expression and the Inner*. Cambridge & London: Harvard University Press.

Glock, H.-J. 2014. "Reasons for Action: Wittgensteinian and Davidsonian Perspectives in Historical and Meta-Philosophical Context". *Nordic Wittgenstein Review* 3 (1): 7-46.

Hacker, P. M. S.. 2000. *Wittgenstein: Mind and Will, Volume 4 of an Analytical Commentary*

on the Philosophical Investigations, Oxford: Blackwell.

Harré, R. & P. F. Secord. 1972. *The Explanation of Social Behaviour*. Oxford: Basil Blackwell.

Hertzberg, L. 2011. "Very General Facts of Nature". In: O. Kuusela & M. McGinn (eds.). *The Oxford Handbook of Wittgenstein*. Oxford: Oxford University Press: 351-372.

Hume, D. 1935 [1748]. *An Enquiry Concerning Human Understanding*. In: T. McCormack & M. W. Calkins (eds.). *An Enquiry Concerning Human Understanding and Selections from a Treatise of Human Nature*. Chicago: Open Court: 1-176.

Lagerspetz, O. 2020. "Wittgenstein's Forms of Life: A Tool of Perspicuous Representation". *Nordic Wittgenstein Review* 9 (1): 107-131.

Laplanche, J. & J.-B. Pontalis. 1973. *The Language of Psychoanalysis*. Trans. D. Nicholson-Smith New York: W. W. Norton.

Lear, J. 1988. *Aristoteles: the desire to understand*. Cambridge: Cambridge University Press.

—— 2017. *Wisdom Won from Illness*. Cambridge & London: Harvard University Press.

Libet, B. 1985. "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action". *The Behavioral and Brain Sciences* 8 (4): 529-566.

Loewald, H. 1980. *Papers on Psychoanalysis*. New Haven & London: Yale University Press.

—— 1960. "On the Therapeutic Action of Psychoanalysis". In: *Papers on Psychoanalysis*: 221-256.

—— 1972. "The Experience of Time". In: *Papers on Psychoanalysis*: 138-147.

—— 1978. "Primary Process, Secondary Process, and Language". In: *Papers on Psychoanalysis*: 178-206.

—— 1988. *Sublimation: Inquiries into Theoretical Psychoanalysis*. New Haven & London: Yale University Press.

Melden, A. I. 1961. *Free Action*. London: Routledge & Kegan Paul.

Moran, R. 2017. "Frankfurt on Identification: Ambiguities of activity in mental life". In: *The Philosophical Imagination*. New York: Oxford University Press: 136-157.

Nietzsche, F. 1997 (1881). *Daybreak: Thoughts on the prejudices of morality*. M. Clark & B. Leiter (eds.), R. J. Hollingdale (trans.). Cambridge: Cambridge University Press.

Owen, J. 2016. *How does a Door Handle work?* [video] https://www.youtube.com/watch?v=WX8NG0275R4 [20 September, 2022]

Queloz, M. 2017. "Two Orders of Things: Wittgenstein on Reasons and Causes". *Philosophy* 92 (3): 369-97.

Tanney, J. 2009. "Reasons as Non-Causal, Context-placing Explanations". In: C. Sandis (ed.). *New Essays on the Explanation of Action*. Basingstoke: Palgrave Macmillan: 94-111.

Ullrich, M, et al. 2018. “OCD-like behaviour is caused by dysfunction of thalamo-amygdala circuits and upregulated TrkB/ERK-MAPK signaling as a result of SPRED2 deficiency”. *Mol Psychiatry* 23: 444-458.

Waismann, F. 1997 (1965). *The Principles of Linguistic Philosophy*. R. Harré (ed.). London: Palgrave Macmillan.

van Inwagen, P. 2008. “The Consequence Argument”. In P. van Inwagen & D. W. Zimmermann (eds.). *Metaphysics: The Big Questions*. Oxford: Wiley-Blackwell: 450-456.

Wedge, M. 2019. “The Mystery of Psychosomatic Symptoms”. *Psychology Today*. Available: https://www.psychologytoday.com/us/blog/suffer-the-children/201911/the-mystery-psychosomatic-symptoms [2 October, 2022]

Wittgenstein, L. 1958. *The Blue and Brown Books*. Oxford: Blackwell.

—— 1966. *Lectures and Conversations on Aesthetics, Psychology and Religious Belief*. C. Barrett (ed.). Oxford: Blackwell.

—— 1981. *Zettel*, 2nd ed. G. E. M. Anscombe, G. H. von Wright (eds.), G. E. M. Anscombe (trans.). Malden & Oxford: Blackwell.

—— 1982. *Wittgenstein’s Lectures, Cambridge 1932-35, from the notes of Alice Ambrose and Margaret Macdonald*. A. Ambrose (ed.). Oxford: Basil Blackwell

—— 1993a. “Cause and Effect: Intuitive Awareness”. In: *Ludwig Wittgenstein: Philosophical Occasions 1912-1952*, J. Klagge & A. Nordmann (eds.). Indianapolis & Cambridge: Hackett: 370-405.

—— 1993b. “Philosophy”. In: *Ludwig Wittgenstein: Philosophical Occasions 1912-1952*, J. Klagge & A. Nordmann (eds.). Indianapolis & Cambridge: Hackett: 161-99.

—— 2009 (1953). *Philosophical Investigations*, 4th ed. G. E. M. Anscombe, P. M. S. Hacker, & J. Schulte (eds.). Malden & Oxford: Wiley-Blackwell.