Research Article

Teleology, backward causation, and the nature of concepts. A study in non-locality of reason

Marcin Poręba[1]

1. University of Warsaw, Poland

The paper analyses the traditional concept of teleology, as well as its modern descendant, the concept of function (as used in the context of so-called functional explanations), against the background of such notions as purposive action, concepts, causality, time, and space-time. The author distinguishes several meanings of teleology and shows that their dialectics reveal their dependence on the concept of backward causation. The classical approach to backward causation, due famously to Michael Dummett, according to which it is a relation between items such as macroscopic things, events, actions, etc., is rejected in favour of the view that future causes should be conceptualized in probabilistic terms. The paper lays special stress on the issue of concepts and their proper treatment as nonlocal entities, as opposed to their understanding as wholly present at dimensionless points in space-time. Using this approach, the author argues for the following disjunction: When trying to account for teleology and purposive action, we must either deeply reconsider the traditional, local view of concepts, or we must take backward causation seriously. It is of the nature of disjunction that finally both alternatives may turn out to be true.

1. Introduction

In this paper, I am trying to elaborate and defend a couple of claims about teleology, causality, concepts, and time. At first, they may seem unrelated, but it is part of my task to show that they are as tightly connected as possible, making up but different aspects of one and the same theoretical movement whose nature should become clear in what follows.

The claims I want to advance are the following.

 doi.org/10.32388/BIW86R

1. The concept of teleology is a mixture of three ideas: (a) teleology as presupposing purposive action, (b) teleology as a reflective principle, and (c) teleology as involving backward causation.
2. Functional explanation is a species of reflective-teleological thinking.
3. Backward causation is an idea essentially involved in any construal of teleology within the broadly deterministic framework of early-modern and modern philosophy and science, backward causation being a variant of deterministic causality: Things determined by their future appear as not determined only because they are indeed not determined *by their past*.
4. On the traditional view of concepts (or functions taken in intension) as local, i.e. as wholly present at dimensionless points in space and time, concepts are a prominent example of things not determined by their past.
5. On a nonlocal view of concepts (on which they can be instantiated only in regions of space and time), we can make good sense of teleology and at least leave room for backward causation.

2. Teleology

Intuitively speaking, teleology is the idea that, besides the purposiveness inherent in intentional actions of humans and (arguably) some non-human animals, there are a range of phenomena in nature and in history that suggest the operation of some plan or design behind them. For a philosophical analysis, I propose to distinguish three senses of teleology, whose dialectics should be helpful in developing what I believe to be a more adequate idea of teleology. They are: (a) teleology understood as grounded in purposive action, (b) teleology as a reflective principle, and (c) teleology as involving backward causation. Although almost every working philosophical concept of teleology is a mixture of these three ideas, for purposes of a more abstract discussion, it will be useful to keep them possibly apart.

a. *Telos* as purpose. This is perhaps the most natural way of thinking about teleology: to reduce ends to purposes understood as features of intentional action. The logic behind this concept of teleology is fairly simple: One starts from the seemingly obvious fact that at least some animals (including humans) typically act with a purpose 'in mind', and in the next step one extends this structure of purposive action to a broader class of phenomena that seem to suggest some purpose behind them. A striking feature of this account of teleology, be it teleology of nature or of history, is that it locates the strictly teleological element in the intentionality of the designer, while the processes designed to achieve the intended purpose are in principle taken to be causal in the

sense of efficient causality[1]. The marks of being teleologically organized, seemingly abounding in innumerable artefacts of animal, human, or godly craftsmanship, are typically not due to any intentionality immediately intervening to produce them, but rather to the operation of basically nonintentional factors, even if these are ultimately grounded in some intentionally conceived plan and actions immediately issuing therefrom[2]. Obviously, on this account of teleology, the whole mystery is relegated to the intentionality of the designer, which, taken for granted and assumed in this kind of teleological explanation, itself cannot be explained in its terms.

b. Teleology as a reflective principle ("as if" teleology). It seems to have been the solution of choice since the 18th century, flourished in many 19th century accounts of biological evolution or history, and is still live in the form of so-called functional explanations (in biology, social and cognitive sciences, and philosophy of mind). Its assumption is that there are no final causes in the sense (a), or at least that we do not have any reliable cognitive access to them. However, we need to adopt what might be called a teleological attitude, analogous to D. Dennett's 'intentional stance' if we want to make sense of a broad class of natural, social, and psychological phenomena that otherwise would appear unintelligible. Kant, whose contribution to this understanding of teleology is by far the most important in philosophical terms, attributed it to the so-called reflective teleological power of judgement, as opposed to the determinative power of judgement, responsible, e.g. for ordinary causal explanations. The reason why a teleological judgement cannot possibly be determinative lies chiefly in the fact that, once we drop the idea of an underlying intentionality of a designer, it refers to an end as what explains (or better, makes sense of) the events or processes to be explained. This obviously means that its *explanans* (an end) lie in the *future* of its *explanandum* (the events or processes made sense of as leading to this end), which is incompatible with the explanation being causal, at least in the Kantian sense of causality[3].

An obvious problem for the reflective-teleological paradigm can be put in the form of a question: What is the status of the concept of purposive action by analogy with which reflective-teleological explanations try to make sense of phenomena not related to purposive action? Is it a different (constitutive, determinative) concept of teleology, or just another application of the same reflective-teleological model? The latter seems hardly tenable: On such a reading, for example, 'the heart is so organized as if its purpose was to circulate blood'[4] boils down to something like: 'the heart is so organized as if it were something that works *as if* its purpose was

to circulate blood'. This brings us nowhere in terms of explanation, since in the obviously resulting infinite regress, there can never appear anything whose nature is really to act on a purpose, and not merely *as if* on a purpose. So, it seems that teleology as a reflective principle depends on teleology constitutively understood. To put it simply, if we want teleology as a metaphor, we have to be able to make sense of the literal meaning of purposive action.

c. As will be shown in the following, the idea of teleology as involving backward causation results from a dialectic to which the concept of purpose and purposive action inevitably leads. For the sake of this introductory typology, let us then outline only a general idea of teleology as backward causality. Purposive action, to which the concept of teleology refers – either literally or metaphorically – can be generally defined as an orderly behaviour whose orderliness can be causally explained in terms of a preceding mental activity as essentially involving a representation of some future states of affairs. Therefore, it seems that the only causality that comes into play in this context is forward causality whose relata are some future-oriented mental representations on the one hand and some features of the resulting behaviour on the other. Certainly our desires, decisions, and plans lie in the past (or at any rate, their beginnings lie in the past) of our actions aimed at their realization. To speak of backward causation in this context could only mean that some aspects of our behaviour, perhaps together with its circumstances and its outcomes, causally influence our past decisions. Before trying to make sense of what at first might seem a weird idea, let me observe that at least it does justice to a very primordial thought behind the more 'digestible' versions of teleology, namely that final causes essentially work backwards in time. A final cause is not any kind of intention or plan prior to the action or process meant to bring about some end. It is this end itself, insofar as it informs and directs the actions and processes leading to it. Since the end obviously lies in the future with respect to the measures leading to it, it seems that we have a clear case of the future informing the past.

3. A lesson from Kant

To many an insightful reader of the *Critique of the Power of Judgement* it is rather obvious that the rules of reflective teleological thinking outlined there also perfectly apply to the kind of reflection Kant engages in in the first critique in developing the core ideas of his transcendental philosophy. It is clear that he everywhere makes the working assumption that the cognitive faculties cannot be

understood in purely causal terms and that we only are able to make sense of their workings if we ascribe a kind of goal or purpose to them. Kant typically describes this goal as 'synthetic unity' (of consciousness, of apperception, of cognition, of representations, of various manifolds – different but closely related determinations pointing probably to different aspects of what Kant believed to be a unitary process). This synthetic unity – let us define it rather weakly[5] as an ordering of a given manifold (as defining a topology over it) – cannot possibly precede the process of synthesis whose effect it is. Therefore, an explanation of this process in terms of synthetic unity is necessarily a non-causal explanation. And since it is an explanation in terms of the effect, it is a species of teleological explanation. However, we have to keep in mind that synthesis is not a purposive action, it belongs to an utterly different category than e.g. going to the university to deliver a lecture. Whether it is conscious or not, it is certainly not intentional. Therefore, the only kind of teleology admissible in Kantian terms as an explanation of the synthesis seems to be the 'as if' teleology of the reflective kind studied in the third critique.

Let us have a look at an especially striking example of the reflective-teleological mode of explanation in the *Critique of Pure Reason*. It is taken from the "Deduction of the Pure Concepts of the Understanding" in the second edition:

> "The *I think* must *be able* to accompany all my representations, for otherwise something would be represented in me that could not be thought at all, which is as much as to say that the representation would either be impossible or else at least would be nothing for me. That representation that can be given prior to all thinking is called *intuition*. Thus all manifold of intuition has a necessary relation to the *I think* in the same subject in which this manifold is to be encountered. But this representation is an act of *spontaneity*, i.e., it cannot be regarded as belonging to sensibility. I call it the *pure apperception*, in order to distinguish it from the empirical one... (...).
>
> ... this thoroughgoing identity of the apperception of a manifold given in intuition contains a synthesis of the representations, and is possible only through the consciousness of this synthesis. For the empirical consciousness that accompanies different representations is by itself dispersed and without relation to the identity of the subject. The latter relation therefore does not yet come about by my accompanying each representation with consciousness, but rather by my *adding* one representation to the other and being conscious of their synthesis."[6]

No matter how we evaluate the conclusiveness of Kant's argument as an argument for the possibility of *a priori* synthetic cognition that essentially involves pure concepts of understanding, it obviously engages teleological reasoning in that the synthesis is viewed *as if* its purpose was to make the unity of apperception possible. It by no means explains the fact *that* the synthesis takes place (otherwise, it would be constitutive (determinative), and a teleological *and* determinative reasoning would necessarily be transcendent). It instead assumes the unity of consciousness as an 'end' to which the synthesis of representations serves as a 'means', with both the end and the means to this end appearing as contingent in terms of the essentially deterministic explanatory framework developed in the first critique.

Against the background of his theory of sensory perception, of concepts and judgements, of empirical and pure apperception, and of the synthesis underlying all of them, Kant was able to develop (in the *Groundwork of the Metaphysics of Morals* and the *Critique of Practical Reason*) a picture of rationality whose core idea is the ability to act on a representation of an end together with a conception of means leading to it. This is precisely the idea of purposive action, which serves as a model for a reflective-teleological explanation whose logic Kant develops in the third critique, and which he tacitly applies in the first. Therefore, it seems that in order to justify his conception of a rational being as acting on representations of ends and means, Kant uses modes of explanation presupposing the very idea of rationality to be justified. He needs them in order to account for how a creature is able to have representations and judgements to be acted on in the first place, while at the same time he refers to the concept of purposive action as to a concept in its own right, as if it were already justified philosophically. The resulting circularity is not easily removable. The rest of this paper is dedicated to showing that the dialectics to which the concepts pertaining to teleology give rise leads to the following disjunction: either we must admit some form of backward causation, or else we have to profoundly rethink our concepts pertaining to rationality. Needless to say, at the end both disjuncts may turn out true. Exploring this third possibility will bring us to some issues concerning the relation of reason to space and time.

4. Functional explanation

The above remarks regarding Kant's concept of teleology and his own use of teleological modes of reasoning can be easily generalized, not least because most of the accounts of teleology that followed Kant are indebted to him overtly or tacitly. As an example, let us consider functional explanations,

widespread in biology, psychology, social, and cognitive sciences, and in philosophy of mind. Functional explanation consists in assigning a 'function' to its explanandum (an adaptation, an organ, a certain physiological characteristic, a social institution, a mental property like belief, desire, emotion, etc.) in order to make it better understandable and in a sense more clearly visible in the overwhelming tangle of structures and processes we face when we think of natural, social, and psychological phenomena. Functions smell better than ends and are certainly more digestible in our secular era. Yet, on closer inspection, functional explanations reveal a logic similar to Kant's reflective teleological judgements and, accordingly, the same type of dialectics leading to a reconsideration of the nature of concepts and rationality. We shall see, however, that the concept of function featuring in this nowadays more fashionable form of teleology gives us some clues to these issues, which the more archaic concept of an end does not.

Here is a list of typical formulae for assigning functions.

1. The function of the heart is to circulate blood;
2. The f. of blood is to bring oxygen and nutrients to all parts of the body;
3. The f. of marriage is to provide security to women and to increase men’s power;
4. The f. of religion is group cohesion;
5. The f. of pain is to warn the individual of danger;
6. The f. of anxiety is to warn others about danger;
7. The f. of a belief is to participate in inferences that lead to the satisfaction of desires;
8. The f. of a belief is to aim at the truth.

Observe that in these and similar formulations, the 'function' can assume two different meanings. First, it can mean as much as the 'role' of something (like the heart, or blood, or family, or belief, etc.) in the context of some broader structure or process (living organism, society, culture, mental life, etc.), i.e., the service something does to a whole, the contribution it makes to the overall working of some complex system. Part of a description in terms of function thus understood is frequently the reference to a specific process through which something performs its function (e.g. with the heart it is blood circulation). However, such a reference is sometimes lacking, probably due to the difficulty of specifying any clear causal link between the state performing the function and the contribution it makes – typically it happens with function assignments for mental states[7]. As a rule, when assigning a function in terms of a role or contribution, one takes for granted the meaning or the role of the

larger structure to which the part or process in question belongs, even if in some more general context also this larger structure can be assigned some functional role with respect to a still larger system (e.g. an organism can be considered a functional part of a population, and this in turn as part of an ecosystem).

Second, a function can mean a certain kind of relation, characterized by uniqueness: every element of the domain is associated with only one element of the range. To be sure, 'one element of the range' typically means an ordered n-tuple of items of different types whose combination defines the structure of a value of the function in question. For purposes of functional explanation, functions are typically taken 'in intension', that is, as rules or abstract procedures that convert their arguments into their values. For example, the belief that it is raining can be defined as a functional property corresponding to a function whose arguments are, for example, certain perceptory states, and whose values are some other beliefs (like 'the streets will be wet tonight') combined with some behavioural dispositions (to stay at home, to take an umbrella when going out, etc.). One of the merits of this abstract-intensional approach is that it allows multiple realizations of one and the same function by physically or even metaphysically[8] different processes and setups. An extensional treatment of functions as sets of ordered pairs would render multiple realization more problematic due to the reference to a concrete domain and range consisting of objects or events whose nature would have to be specified in advance.

On the face of it, both uses of the function concept are vastly different. So different that one can even suspect that it is only due to a historical coincidence that they bear the same name. In fact, this impression is completely wrong. What mostly misleads us as to the relationship of both concepts is the fact that the second is typically couched in abstract, quasimathematical terms and that it completely leaves out the spatiotemporal character of phenomena to be explained functionally. However, when it comes to the question of what function a given empirical process realizes, it immediately becomes clear that ascribing a function to a system necessarily involves assumptions concerning its future states. This can happen in two ways, weaker and stronger. The weaker version is when it is *us* who make these assumptions concerning the future of a system; a function attribution is then clearly a variant of reflective-teleological judgement: we describe a system as if its intended goal was to produce a certain output. The stronger version consists in attributing to the system itself a kind of intentionality – an intention to yield certain values in the future; in this variant, the system itself *knows* which of its possible future states are in accord with the function it implements. It is especially

easy to take this stronger position with respect to artificial systems such as computers. There it seems that one literally has a function built into the system, because programs are written in terms of functions, and one can think that once installed on a machine or loaded into the code segment of the memory, an algorithm becomes 'internalized' by the system and constitutes its 'intentionality'. In fact, the sole intentionality that comes into play in this case is the old good one of the engineer or the programmer. It would require an independent argument to show that there is more to the 'intentionality' of a computer than simply realizing the intentions of its designer and its users, not less than in the case of a screwdriver to show that it has the intention to drive screws.

So what we in fact observe is rather a perfect analogy between both uses of the function concept with respect to their teleological character: both are overtly or tacitly teleological, both admit of stronger and weaker interpretations depending on how literally we ascribe intentionality to a system or process we want to explain. The only difference consists in the considerably higher level of abstraction characteristic of the concept of function as an operation that takes some arguments and returns some values. The abstract formulation has two merits. One is the already mentioned merit of lending support to the anti-reductionist accounts of functional roles based on multiple realizations. The other is that the abstract formulation provides a clear link to the issue of concepts, which must be addressed if we want to explain the literal meaning of intentionality and purposive action behind reflective teleological models in which concepts such as end or purpose are used in a metaphorical, 'as-if' manner. But before we begin this task, we must introduce one more element into the picture.

5. Backward causation

It seems that in order to understand the behaviour of a system in terms of its purpose, its role, or its function at a given moment or interval it is necessary to refer to some *future* states of this system. This observation follows clearly from the meaning of terms such as “purpose” or “function”: a purpose is always something belonging to the future of an action having this purpose. Otherwise, it would make no sense to say of an action that it *failed* to achieve its purpose, since this involves (a) that the action has taken place and (b) that the state of affairs being its purpose did not occur. Similarly, with a function: no matter whether we define the function extensionally, as a set of ordered pairs, or intensionally, as an operation (eg, a certain computation), it essentially involves a reference to its value for a given argument[9]. And obviously, when applied to the behaviour of a system, where function is realized by a process evolving in time, the value referred to lies necessarily in the future

with respect to the process realizing the function in question. This reasoning can be generalized from purely temporal to spatio-temporal relations. In effect one can say that in order to understand the behaviour of a system at some point or in some region in space-time, it is necessary to refer to its future understood spatiotemporally.

But reference to future states in order to understand the behaviour of a system can mean two clearly different things. First, it can mean dependence of our concepts concerning the behaviour of a system on some assumptions concerning its future states. This seems uncontroversial, since it in no way implies that the future is somehow already present and affects what happens now. Typical examples of such dependence are concepts used in reflective-teleological judgements, describing a system as if it were designed with some end or purpose in view. Part of functional explanations can also be interpreted after this reflective model. Second, it can mean that reference to future states of a system is indispensable for an *explanation* of its present functioning, which roughly translates to saying: "The system S is in state s_1 at t_1, *because* at a later moment $t_2>t_1$ it will be in state s_2". If this is to be interpreted as a species of causal explanation, then the causality involved in it is clearly a backward one, a causality working backward in time.

Before I elaborate a bit on the idea of backward causation, a remark is in order: Even on a purely intuitive understanding of backward causation, it stands out that it does not quite overlap with the idea of purposive action being informed by its purpose, understood as something belonging to the future of the action. For one thing, the failure of such an action to achieve its purpose, that is, the non-occurrence of its purpose (or even, like in Greek tragedy, the occurrence of something contrary to it) by no means implies that the action had no purpose, or that its purpose was different from what the agent had believed before the actual outcome occurred[10]. So, if purposive action is to be informed by its (intended) purpose, then, so it seems, rather not causally, on the pain of admitting non-existent future events as causes of existing ones. Less controversial is the consistency of backward causation with reflective-teleological explanations, functional ones included. Certainly, if the heart of an animal stops pumping blood, it immediately becomes problematic whether it still exemplifies the function of a heart. Generalizing a bit, we could say that the non-occurrence of the effect with respect to which a function is defined essentially changes the status of processes that were supposed to lead to it. *If* it were to be interpreted as a case of backward causation, then there would be at least no obstacle of the kind mentioned above with respect to purposive action, since a function that does not yield the desired output simply is a different function, if any, than the one that does.

But are there any independent reasons for believing in backward causation in the context of teleology and purposive action? To address this question, let us first look at what is involved in the concept of backward causation. For a general model, we refer to Lewis's counterfactual analysis of causation. Sure, the counterfactual analysis also applies backward in time in cases where a sufficiently reliable causal link works forwards. For example, one can say that if a gun had not fired, the trigger would not have been pulled, which means that the gun that did not fire is a reliable indicator of not pulling the trigger. In such cases, a counterfactual clause working backwards clearly presupposes a 'normal', forward-working causal chain. In order to have a clear case of backward causality, we would have to make sure that the respective counterfactual clause is not backed by any forward-working causal link. Obviously, in terms of counterfactual analysis, it would amount to saying that (1) there is a true counterfactual clause whose consequent refers to an earlier event than its antecedent, and (2) there is no chain of counterfactually dependent events leading from the earlier event to the later. Informally speaking, what we need to make sense of is that an event e_2 at t_2 could have occurred even if an event e_1 at t_1 (t_1 earlier than t_2) had not occurred, while e_1 would not have occurred if e_2 had not occurred. In terms of possible worlds: (1) both e_1 and e_2 occurred at the actual world, (2) among the closest non-e_1-worlds there are worlds at which e_2 does occur, (3) all the closest non-e_2-worlds are also non-e_1-worlds.

Before applying the above remarks to teleology and purposive action, let us have a look at what backward causation implies for our idea of time and temporal order. The fundamental conception behind the majority of modern and contemporary accounts of time and causality says that causes of a given event must belong to its past, i.e. must lie in its past light cone. Since this is frequently used to define the orientation of space-time with respect to past and future (people like Kant took simpler route and spoke of causality as defining the temporal order), it seems almost an analytic truth that causality never works backward in time, or, from the more complex and more interesting spatiotemporal perspective, that no event can be causally influenced from outside its past. Therefore, to save the idea of backward causation from inconsistency, we have to assume that the spatio-temporal order is not defined by causality (in accordance with the principle that the past is where the causes belong) but perhaps by some more abstract principle, say, of a topological character. The order thus defined could be better described as orientation: for a simple, purely temporal order it means that from any point one can look in two directions, of which one can be called 'the past' and the other 'the future', but these are meant as mere labels whose only purpose is to distinguish both directions from

each other. What matters is only that these directions be always distinguishable and that if we start to travel in one, we cannot possibly get back from the other. Against the background of temporal or spatiotemporal order thus abstractly defined, it is of course an open question whether the causes of a given event must belong to its past. And this perfectly suffices to make sense of backward causation as not altogether absurd.

Now, what sort of reasons could persuade us that there might be actual backward causation working in teleology and purposive action? We already know at least these two things: (1) What causally informs the purposive action rather cannot be the state of affairs that makes up its intended purpose: there are cases in which the action failed to achieve its purpose and, nevertheless, *was* the action having this purpose. However, it is still thinkable that some more complex future configurations of states of affairs, not necessarily including the desired outcome, inform the present action, making it, e.g., have such and such purpose. (2) On a reflective-teleological model the nonoccurrence of a specific outcome leads to a change in qualification of the 'responsible' structure or process with respect to its function. But it is rather a change on the level of conceptualization of what is going on than in the processes themselves, of which we rather tend to think that they are sensitive to the conventional, forward type of causality. And the said change of qualification in its turn is rather posterior to the non-occurrence of the expected/desired outcome, so that it can be safely regarded as caused in the forward sense by its own past.

So, it would be far too simple to say that it is the purpose, the desired state of affairs that causally informs the action or process leading to it. The temptation to rely on this simplistic model is due probably to the fact that when theorizing about teleology, purposiveness, intentionality, and related matters, we are chiefly interested in successful cases, i.e., cases in which the action or process indeed produces the desired outcome (otherwise, it seems, there would be no point in making so much fuss about teleology and intentionality). The unsuccessful cases appear in this context as a kind of anomaly, a rather marginal phenomenon that can be explained away (eg due to a wrong initial conceptualization[11]) or left for future treatment, whereas the basic theory, which is a theory of successful action and of well-functioning systems, may remain essentially unchanged. To some extent, it resembles the situation in perception theory, where the dominant orientation is towards the so-called veridical perception and where the typical explanatory pattern has the form: "I perceive that *p*, *because p*". The temptation to think that the default explanation of a perceptual belief whose content is *p* is the state of affairs signified by *p* is so strong that it leads to a completely nonchalant

attitude towards the innumerable cases in which one has a perception with content *p* while there is nothing in the vicinity even remotely corresponding to *p*. So my suggestion is that in both cases – in the less problematic case of perception being informed by features of the environment and in the more problematic one of purposive action or functioning of a system being informed by its future – any minimally adequate explanatory model of what is going on in these cases must be considerably more complex than the dominant simplistic view.

What is required of such a model is, among others, that the distinguished cases (veridical perception, successful purposive action and well-functioning of a system) be considered as elements of some broader class of situations to which also unsuccessful cases (e.g. illusory perceptions and hallucinations, failed purposive actions and malfunctions of systems) belong, and that the explanations of both kinds of case overlap to an extent sufficient to account for, e.g., the ability of illusions to deceive us, or for the fact that we sometimes undertake actions whose intended purpose is impossible or even self-contradictory. *Provided that* an explanation of some aspects of purposive action and functioning of systems in terms of their future causes makes any sense in the first place, and taking perception as our reference point, we could say that in some cases of intentional action its future causes can produce an illusion of there being a state of affairs in the future which, accordingly, becomes the intended or desired purpose. Since obviously it is not the desired state itself that might have contributed to the illusion, because as it happens there is nothing of the sort in the future, it must have been some other future parameters that made it look as if the desired outcome was there, awaiting. What I mean is a mechanism analogous to the one behind the illusion that, when standing on the seashore, we can see a clear line, some 4 kilometres away, where the sea meets the sky. When we say that this perception is produced by something in a causal way, this is certainly true, but it is clearly not the horizon line that is causally responsible for the perception quasi of a line. The appropriate causal story would have to be roughly about the curvature of the Earth, the light and the laws of its propagation and refraction, as well as about some aspects of the functioning of perception. Certainly not about the line, so clearly, it seems, visible out there. However, any satisfactory account of this kind should explain why the perception is quasi one of a line out there. Analogously, for purposive action, the putative causal story about its future should provide, among others, an explanation of its intended goal as something represented, even if, in fact, the action is doomed to failure.

Let us go a little further and ask which parameters of the future might be causally responsible for some characters of teleology and purposive action. First, we should observe that all standard candidates for causes (objects, states of affairs, events, facts, properties or instantiations of properties) are inappropriate with respect to backward causality, since their very conception presupposes forward causality[12]. We can say that objects, events, and properties of our everyday experience are fundamental aspects of reality considered as lying in the past. It should be observed that the reality thus understood cannot be possibly considered as a single one. This is a rather obvious consequence of the fact that the past is relative to the point or region in space-time: different spatiotemporal locations have different pasts, and that means, strictly speaking, different realities. Reality as consisting of objects, events and their properties on the one hand and the idea of causality as operating forwards in time on the other are parts of one and the same package. Therefore, if we want to make sense of the idea of backward causality, we have to look for suitable candidates for causes not among future objects and events because, in a sense by definition, there are no such things. But are there other options?

Indeed, there are and they are of two kinds. First, we can descend below the level of middle-sized objects of everyday experience to which the traditional idea of causality as essentially a forward working mechanism primarily applies. According to a hundred years old wisdom what we find there is quantum systems and processes whose nature and laws drastically differ from almost everything we are used to at the classical level. Perhaps the most obvious divergence from the classical idea of forward causality are the so-called non-local correlations, which on some interpretations involve causal transactions between space-like separated events. However, such correlations do not necessarily imply backward causality in the sense of causality working backwards from within the future, if the future is to be understood in terms of the forward light cone of an event[13]. Non-local correlations seem to imply only that what happens at a given point in space-time can be causally influenced by something *from outside* its past, not necessarily from its future strictly understood. What might suggest such an influence from the future are, e.g., the phenomena of light propagation in different media. It has long been known that the index of light refraction at the boundary of two transparent media follows Fermat's principle according to which light travels along a path that minimizes time, depending on the different speeds of light propagation in different media, such as air and water or layers of air with different temperatures[14]. Before the advent of quantum mechanics and quantum electrodynamics it was difficult to account for these phenomena in causal terms. It looked as

if light knew in advance at what speed it was going to travel after crossing the boundary, but this would precisely mean that it knew its future. In terms of classical, forward causality, this amounts to a miracle. Quantum mechanics, with its utterly different conception of time and causality, allows for a better understanding of this 'miracle'. In some interpretations, we can say that at the moment when light passes the boundary, it has already travelled along its future path (even perhaps along all its possible future paths), and this future 'experience' can somehow causally influence what happens now at the boundary. What is important is that this (admittedly hypothetic and highly theoretical) story is about what happens at the quantum level. At the classical level we only have forward causality, and that is precisely why certain classically identifiable phenomena like light refraction must remain utterly unexplained in classical terms. As a last remark, let us observe that the distinction between classical and quantum levels is far from obvious. I rather think that the 'classical level' is a kind of appearance in a sense similar to the Kantian 'world of appearances'. What underlies them is always quantum processes, no matter whether they work forward or backward in time. The classical processes that we project onto the past are at bottom quantum processes, the only difference being that we conceptualize them in terms of middle-sized objects, their properties and events involving them. I stress this conceptualization aspect since I believe that, e.g. what Kant termed sensory intuition, has nothing classical to itself – no 'collapse of the wave function' occurs when I, for example, observe something to be red. The only classical thing in our experience is what features in stories we tell with the help of our concepts. As to these stories and concepts themselves, considered sub specie of their realization, they are equally quantum processes.

The second option is in a way the opposite of the first. It consists in invoking a special kind of entity which is typically believed to be indifferent with respect to the distinction of past and future, the abstract objects. The abstract objects like numbers, sets, functions, propositions, etc. are arguably time-independent, or better space-time independent. They spread so to speak over the whole of space-time: even if, for example, the proposition "Epaminondas was fatally wounded in the battle of Mantinea" is about a particular historical event located at a certain point in space-time, it is true everywhere, at every point, since it could be truly asserted anywhere at any time. However, this time- and spacelessness of abstracts has a price, and this is their lack of causal powers, be they powers to affect or to be affected[15]. That Epaminondas died had certainly some causes and effects, but the truth of the proposition signifying this fact lies beyond any causal chains, no matter whether classical or quantum, forward or backward. Precisely because they are everywhere, the abstracts cannot be

causally active. Causes active at every point in space-time would make no difference, and making difference is precisely what we need causes for. Philosophers have devised several remedies for this causal inertia of the abstracts, all of them consisting of making them to some extent relative with respect to space and time. One of these remedies was to give the abstracts a semi-concrete body in the form of laws, like the laws of physics. Obviously, laws are meant in this context neither as theoretical formulae nor as mere statistical regularities, but as principles that are truly operative in nature and "governing" it. Although obscure, this idea gives some meaning to the abstracts being causally relevant. It is also clear why this meaning involves a spatio-temporal relativization of the abstracts: a law operates only at points and in regions where certain conditions are met, and it is through these conditions that it exerts its influence upon what happens at these points or regions. However, it is far from clear whether the interpretation of abstracts as laws allows for backward causation. It certainly depends on the kind of laws and on the interpretation we give them. To the extent, for example, that a law of nature can be given the form of a variational principle, it has at least the air of something 'taking account' of the future to operate at present.

Another way of thinking of the abstracts as participants in causal order is to relate them to concepts to which I now turn.

6. Concepts and Space-Time

Concepts can be understood in two ways, objective and subjective. Taken objectively, concepts are abstract entities whose nature can be approximated as operations (functions taken 'in intension'). Subjectively concepts are features of thought and action, present in humans and some non-human animals and responsible for such characters of their behaviour as orderliness, purposiveness, and rule-governedness, an important aspect of which is the applicability of normative concepts to thought and action (linguistic behaviour included)[16]. Let us illustrate the objective-subjective distinction with a fairly simple example: the concept 'addition of two natural numbers'. Objectively understood, adding natural numbers is the simplest arithmetic operation that yields all and only ordered triples of the kind <5, 7, 12>. Therefore, the concept 'addition' is time- and spaceless, as well as completely independent of anyone's ability to understand it. Of course, the faith in there being such concepts is not obligatory on a philosopher, but at least this much should be admitted: their conception is not self-contradictory and not *obviously* obscure. Subjectively understood, the concept of 'addition' is something possessed by a creature who has mastered a suitable portion of arithmetics, so that it is a

sufficiently reliable computer of sums of natural numbers. Tentatively, we can say that concepts understood subjectively are certain dispositions (behavioural, mental, neurophysiological, spiritual, etc.) captured by counterfactual clauses of the kind: "If the person O had been given 5 and 7 as input at t, she would have returned 12 as output".

An immediate remark is in order. Due to something very similar to Kant's transcendental illusion, counterfactual clauses such as these tend to be interpreted as expressing objective concepts. On such a reading, the above conditional is taken to imply that, had O returned something different than 12, for example 10, she would have betrayed the lack of understanding of the (objective) concept of 'addition'. But the true import of the conditional is different: It rather expresses a condition of O's conforming to a *subjective* concept of addition, namely, to the one possessed by someone who utters this conditional. When attributing concepts to others (or to ourselves), we never judge their behaviour from a quasi-godly perspective, defined as involving possession of objective concepts. We judge from the perspective of our understanding of concepts (so, strictly speaking, on my mere output, without taking into account the reactions of others, I cannot consistently refuse, e.g., the concept of addition to myself; the maximum I can do in this 'private' way is sometimes detect errors in my calculations and possibly correct them). Even if we imagined an omniscient God watching our calculations and saying 'This one has mastered addition, and this one not', this would be only a limiting case of judging on a subjective concept, namely the godly concept of addition. This is what Wittgenstein meant by his impressive remark: "Auch Gott kann Mathematisches nur durch Mathematik entscheiden"[17].

Concepts understood subjectively are not the same as our subjective grasp of objective concepts. My subjective concept of "addition", due to which I calculate 5 + 7 = 12, is not my subjective way of knowing about addition as an objective operation. What I know through my concept of addition is rather a set (arguably finite) of identities of the above kind, together with a couple of more general facts like that for all a and b, a + b = b + a. But to even think about the objective concept 'addition', it is not enough that I know (if imperfectly) how to add numbers. What I need are new concepts, like 'operation', 'function', 'function in intension', 'set', 'relation' etc., which in their turn are absolutely superfluous for adding numbers. As to the relationship between objective and subjective concepts, for lack of better categories (may future developments of abstract philosophy provide for them), we can think of it in terms of instantiation or realization: one and the same objective concept has realizations in thoughts and actions of different individuals and groups. The realization on the level of groups is

especially important since it is responsible for our sharing of the same or at least overlapping subjective concepts.

What distinguishes subjective concepts from the objective ones whose realizations they are is, among others, their spatio-temporal relativization. Concepts thus understood are anchored in specific regions of space-time, roughly speaking those in which the respective thoughts and actions are enclosed. But concepts being anchored in a specific region can mean two essentially different things. It can mean either that a concept is pointwise instantiated, i.e. can operate at every point within the region, analogously e.g. to an electromagnetic field acting at every point within a certain region at which a charged particle can happen to be present. Or it can mean that a concept is a global property of the region without being instantiated at any specific point, being, so to speak, everywhere without being anywhere. The first meaning is the option of choice for those who consider concepts as a special kind of dispositions that – be they realized mentally, physically or spiritually – can be fully actualized at any moment at which a system (e.g., a person) possessing a given concept faces a task to the solution of which this concept seems appropriate. Obviously, this option is fully compatible with purely forward causality, since conceptual operations appear on this interpretation as actualizations of potentialities that are already fully present (typically due to an *earlier* process of learning) at the moment of their first full-fledged actualization. On the second interpretation, concepts cannot be identified with dispositions; otherwise they would be clearly pointwise instantiated (if I have a certain disposition during some period, I plainly have it at every moment within this period). But, according to the second interpretation, identifying concepts with such dispositional properties rests on a mistake: even if it is trivially true that in order, for example, to utter '12' when asked to add 5 and 7, I must have a disposition to do this, this disposition cannot be identified with my (subjective) concept of 'addition', since *whatever* there is to my disposition to utter '12' at that moment, it is fully compatible with any imaginable future behaviour, including behaviour that I would unhesitatingly disqualify as incorrect with respect to my concept of addition. So, what is essential for my being in possession at a given moment of the subjective concept addition is not only my past and present behaviour but also my behaviour in the future, and if I indeed *can* be said at a given moment to possess the concept, the fact signified by this statement depends not only on my past and present, but also on my future.

The line of reasoning just sketched is, of course, a variation on a familiar theme from Wittgenstein, known as the rule-following argument. It is not my purpose here to provide any independent reasons

for or against this argument. However, even if in itself not fully compelling, this argument clearly shows what it can mean that the present fact that I possess a concept (e.g. addition) depends on future facts[18]. But what kind of dependence could this mean? Once again, analogously to the ambiguity concerning teleology, it can be understood weakly or strongly. On the weak interpretation, it means that future facts can alter our judgments about what concepts now govern someone's behaviour. This is rather uncontroversial. The stronger interpretation can be put in the form of a backward counterfactual conditional: if some aspects of the future had been different, I wouldn't have possessed the concept (say, of addition) I now have. This in turn might suggest causal dependence, but certainly not from any future *facts*, since the level of facts, as we already observed, is the domain of forward causation. What lies in the future is mere probabilities, not probabilities in the guise of facts, which is how past probabilities appear to us. Accordingly, the relevant conditional might look as follows: *If the probability distribution on the outcomes belonging to the future of a certain piece of putatively intentional behaviour were in some respects different, the behaviour in question would not have displayed the concept it actually displays.* This cannot be interpreted as an instance of backward causation, understood as the influence of putative future facts upon what happens at present, since facts-on-facts influence is a paradigm limited to forward causation. So we either have to enrich our paradigm of causation to include such causal relata as probability distributions on the one hand and possessions of concepts on the other, or elaborate on our conception of a concept so as to allow concepts to extend over regions of space-time *always* including some portion of the future. Both policies are worth pursuing; perhaps even only when combined would they yield a satisfactory account of the 'space of reason'. Below, I briefly comment on the second option, leaving the first for another time, especially since it depends on some obviously extra-conceptual considerations, pertaining e.g. to the notion of causality in the context of quantum physics.

7. Teleology, Concepts, and Time

When describing a system in reflective-teleological terms, i.e. as behaving *as if* its purpose was to perform a certain function, we rely on our intuitions concerning the literal (determinative) meaning of teleology, referring primarily to purposive action and, more generally, to intentionality. But on closer inspection, this putatively literal meaning is itself in danger of slipping into the reflective and metaphorical, which would finally lead to the conclusion that concepts like purpose or intentionality are metaphorical through and through, having literally *no* literal meaning. As an alternative to this, I

developed a view on which it is indeed possible to save the literal meaning, but on a relatively high cost: we have to either admit of some sort of backward causality, or profoundly reconsider the received view of concepts. Very likely we have to do both.

Regarding concepts, the decisive move consists in abandoning what might be called the 'locality assumption', that concepts are items (most probably dispositional properties) that are fully instantiated at every point in the spatio-temporal region in which the respective conceptual competence is located (typically this is a segment of a person's life trajectory). Instead, I propose to treat concepts non-locally, i.e. as pertaining to such regions as wholes. Now, especially if we look at the matter from topological perspective and consider such regions in terms of open subsets of a larger topological space, we can even demand that every point within such a region has a neighbourhood containing, among others, points belonging to its future. Thus, to say of someone that she is in possession of a certain concept *c* at time *t* means that *t* belongs to a region to which *c* can be attributed as a global property. That is why future can in principle falsify any present concept ascription: not because future can anyhow change the past *after* it has already passed, but rather because any present concept ascription is at once a statement about the past and future of this present. It is clear that this account is independent of any sort of backward causation, and it is rather reminiscent of the idea of considering past and future as parts of one, at the bottom an indivisible whole (analogously to how quantum nonlocality can be interpreted in noncausal terms)[19].

That purposive action is essentially a conceptual achievement, perhaps does not need special advertisement. On the nonlocal treatment of concepts, concepts constitutive of a given piece of action pertain to a region extending into the future and covering the time at which the supposed outcome of the action will or will not take place. Some features of this future time (not necessarily the desired outcome, since this can fail to obtain without losing the intended purpose of the action) are constitutive of the fact that *this* concept was at work in forming the representation of the intended purpose of the action. So we can say that indeed the future is relevant for purposive action, which is precisely what was to be shown in order to give substance to teleology understood as not merely a reflective idea behind which a purely deterministic landscape based on forward causation may very well stand.

To what extent this conceptual relevance of the future is backed by its real causal pressure on the past is, as already said, a different and immensely more difficult question. However, let me observe one interesting point: The issues of conceptual and causal relevance of the future seem to be inextricably

linked in one fundamental respect. Our firm belief that the causes of a given state of the world must necessarily lie in its past owes its seeming obviousness to the fact that it concerns the middle-sized objects and events of our everyday experience in which our ordinary and a good deal of our philosophical understanding of reality, time and causality is rooted[20]. As happens, the conception of the world as consisting of such objects and events is itself a result of a conceptualization with essentially the same battery of concepts with the help of which our goals and aims are formed and which govern our purposive actions. So what seems to block our thinking of our agency as being causally informed by the future is first of all the fact that we think of it in terms of the past-oriented worldview of our everyday experience, shaped by the idea that strictly speaking we can perceive and cognize only what has already happened.

On the other hand, the conceptual relevance of the future, suggesting spatiotemporal non-locality of concepts, invites us to re-consider our familiar, past-oriented worldview. If concepts are spatiotemporally extended, if they are spread over regions ranging not only into the past but also into the future, then the idea that the future might be no less real than the past considerably gains plausibility. And this in turn opens up a perspective on backward causality as a factor shaping the reality on a par with the familiar forward species. We might be encouraged in this line of thought by the idea that it is concepts, not the middle-sized objects of everyday experience, that serve as a "gateway" through which the future informs the present and the past. Led by this kind of considerations, we could start to look for suitable physical realizations of backward causality.

Footnotes

[1] That is why Aristotle was able to consider the activity of e.g. an architect or a sculptor to be an instance of efficient, instead of final, causality, even if this activity itself is a clear example of intentional, purposive action.

[2] However, this requires two qualifications. First, on some teleological worldviews there is some place left for minor additional tampering on the part of the designer, meant to correct the slight deviations from the original project, to which it might come e.g. through a long operation of purely efficient-causal factors. A typical example of this additional, tampering intentionality is given by Newton whose God from time to time corrects the motion of heavenly bodies, gradually disturbed in their movement by the long operation of gravitational forces. However, on the majority of teleological accounts the

role of this additional tampering is kept to a minimum. Second, especially on some teleological accounts of history and (perhaps) evolutionary biology, part of the factors bringing about the intended goal (be it salvation, or the Prussian state monarchy, or communism, or development of a more advanced species etc.) are, like the behavior of individuals pursuing their particular, mostly egoistic ends, or striving for food, or looking for sexual partners, themselves intentional, even if the goal they produce in the long run is completely unintended by the actors themselves.

[3] We have to bear in mind that the core of Kantian treatment of causality is the idea that causal relations are indispensable for the asymmetry of past and future, constitutive of the direction of time (of time's being an orientable manifold).

[4] Or, to use contemporary teleological idiom, "The function of the heart is to circulate blood throughout the body".

[5] With stronger definitions my point would be the more compelling.

[6] I. Kant, Critique of Pure Reason, trans. P. Guyer, A.W. Wood, Cambridge University Press 1998, p. 246-247 (B 131-133).

[7] Probably due to our relatively poorer understanding of the relationships between mental processes specified in psychological terms and the neurophysiological mechanisms meant to realize them.

[8] E.g. one and the same mental function like that of a belief or desire can be realized, among others, by a biological system, an electronic device or even an immaterial soul – all these are metaphysically possible options.

[9] It might be perhaps not quite obvious for functions understood in intension, since e.g. the instruction how to add two natural numbers does not explicitly mention the result for any specific pair. But, on the other hand, a certain result, e.g. 237 for 123 + 114, is among the necessary conditions of the operation's being an instance of addition.

[10] To be sure, in the classical tragic vision the characters are superficially acting on their purposes, but essentially they are realizing a kind of 'program' of which this vision assumes that it never fails.

[11] Like when we say „Oh, then something else must have been her purpose" or "So it seems that we wrongly identified the function of this adaptation".

[12] This is the reason why I consider the classical account of backward causation proposed by Michael Dummett as fundamentally wrong (see his "Can an Effect Precede its Cause?", Proceedings of the Aristotelian Society, 28 (Supplement), pp. 27-44, and "Bringing about the Past", Philosophical Review, 73, pp. 338-359). As a paradigm of backward causation Dummett considered intentional

action aimed at bringing about some state of affairs in the past. However, his arguments trying to make sense of ordinary objects and events being good candidates for causes of states of affairs in their past were rather obscure. At most they showed how it is possible for someone to intend to influence the past, not how to make sense, e.g., of causal relevance of my prayer today for the survival of someone in a yesterday's plane crash. Unfortunately, the majority of discussions following Dummett concentrated on the same uninteresting paradigm of event causation at the macro-level (see e.g. M. Black, "Why Cannot an Effect Precede its Cause?", Analysis, 16, pp. 49-58; P. Forrest, "Backward Causation in Defence of Free Will", Mind, 374, pp. 210-217; J.H. Schmidt, "Newcomb's Paradox Realized with Backward Causation", British Journal for the Philosophy of Science, 49, pp. 67-87; R. Swinburne, "Time and Causation", American Philosophical Quarterly, 51, pp. 233-245). The problem with this paradigm is that it inevitably projects our everyday conception of things and events onto the future, to which it cannot be sensibly applied.

[13] However, on some interpretations of quantum entanglement and non-locality this is precisely what is going on. See e.g. J.G. Cramer, "Generalized absorber theory and the Einstein-Podolsky-Rosen paradox", Physical Reviev D, 22, pp. 362-376, and "The transactional interpretation of quantum mechanics", Review of Modern Physics, 58, pp. 647-688.

[14] This is of course a grossly simplified picture. Strictly speaking, the behaviour of light follows the so-called variational principle: it takes the path that – generally speaking – either minimizes or maximizes time. But for our purposes the simpler treatment is sufficient to make the important point concerning the causal relevance of the future.

[15] At least since Plato's Sophist this causal inertia of the abstracts has been the chief ontological motive behind scepticism about their existence. Another was the epistemological one: the only way of knowing the abstracts is through mental operations like proof, and provability falls short of justifying the unrestricted use of the bivalence principle with respect to a domain in question, which on these accounts (like Leibniz's or Ingarden's or Dummett's) is the mark of reality as opposed e.g. to fiction.

[16] The idea of concepts as a kind of mental pictures of things viewed under the aspect of their general features can be considered an extremely raw, primordial version of this subjective construal. However, even in this primitive view we can discern elements pointing in the right direction: concepts (subjectively understood) are certain characters of a class of creatures, allowing them e.g. to cognize reality or to act in an organized, purposive way.

[17] L. Wittgenstein, Bemerkungen über die Grundlagen der Mathematik, Suhrkamp, Frankfurt a. M.

1994, p. 408.

[18] One is tempted to interpret Wittgenstein's rule-following topos as an exposition of general scepticism about meaning and concepts. I propose instead to read it as directed against a certain conception of concepts, namely as dispositional properties that can be fully instantiated at points in space-time.

[19] See for example T. Filk, "Temporal Non-locality", Foundations of Physics, 43, pp. 533-547.

[20] This is just a more "philosophical" way of putting essentially the same as what is captured by the so-called weak-causality principle: "A macroscopic cause must always precede its macroscopic effects in any reference frame.", as opposed to the strong-causality principle: "A cause must always precede all of its effects in any reference frame" (J.G. Cramer, "Generalized absorber theory and the Einstein-Podolsky-Rosen paradox", op. cit., p. 367). However, I think that the very distinction of macroscopic and microscopic causes and effects is a merely conceptual one as all causal work goes on at the micro-level. The fact that our conceptualization of the world in terms of macroscopic things and events favours forward causality can be seen as an effect of two factors: the real asymmetry of time and our psychological tendency to think of the past as fixed and of the future as an open space of possibilities.

References

- Black, Max, "Why Cannot an Effect Precede its Cause?", *Analysis*, 16, pp. 49-58
- Cramer, John G., "Generalized absorber theory and the Einstein-Podolsky-Rosen paradox", *Physical Reviev D*, 22, pp. 362-376.
- Cramer, John G., "The transactional interpretation of quantum mechanics", *Review of Modern Physics*, 58, pp. 647-688.
- Dummett, Michael, "Can an Effect Precede its Cause?", *Proceedings of the Aristotelian Society*, 28 (Supplement), pp. 27-44....
- Dummett, Michael, "Bringing about the Past", *Philosophical Review*, 73, pp. 338-359.
- Filk, Thomas, "Temporal Non-locality", *Foundations of Physics*, 43, pp. 533-547.
- Forrest, Peter, "Backward Causation in Defence of Free Will", *Mind*, 374, pp. 210-217
- Kant, Immanuel, *Critique of Pure Reason*, trans. P. Guyer, A.W. Wood, Cambridge University Press 1998.

- Schmidt, Jan Hendrik, "Newcomb's Paradox Realized with Backward Causation", *British Journal for the Philosophy of Science*, 49, pp. 67-87.
- Swinburne, Richard, "Time and Causation", *American Philosophical Quarterly*, 51, pp. 233-245.
- Wittgenstein, Ludwig. *Bemerkungen über die Grundlagen der Mathematik*, Suhrkamp, Frankfurt a. M.1994.

Declarations

Funding: Grant Nr. 2019/33/B/HS1/03003 of the National Science Centre, Poland.

Potential competing interests: No potential competing interests to declare.