TELEOLOGY AND THE CONCEPTS OF CAUSATION

Ernst von Glasersfeld

Abstract

The paper suggests that the main disagreement about teleology is due to an ambiguity of the term and that one of the notions it covers is scientifically acceptable. The conceptual dichotomy goes back to Aristotle, whose categorization of 'causes' is discussed. Experiential finality is separated from metaphysical teleology, and shown to be derived from the notion of 'efficient cause'. Its use and interpretation in cybernetics is briefly sketched out.

The term 'teleology' was coined by the philosopher Christian Wolff in his Latin treatise of 1728 and he defined it as indicating the part of natural philosophy that explains the ends, or purposes, of things.[1] It has been used to describe a prominent feature of the theological views of Thomas Aquinas and the metaphysics of Aristotle. Some forty years ago, it was adopted by the founder of cybernetics to characterize a mechanism designed, among other things, to steer missiles to their target. Since the instrumental purpose of artificial devices would seem to belong to a domain that is somewhat different from the metaphysical purposes of Nature and its components, it is reasonable to suspect that there has been a shift of meaning or an ambiguity. Where ambiguities remain hidden, all kinds of nonsense can be generated.

In the present essay, I argue that the problem with 'teleology' goes back to one of the four 'explanatory principles' that resulted from Aristotle's examination of what we now call causation. After a brief survey of the contemporary uneasiness about teleology I present an exposition of Aristotle's classification of causes, suggesting that modern science has been somewhat cavalier in dismissing all but one of the four. I suggest that an adequate analysis of the concept of *efficient cause* opens a way to showing that the notion of *final cause*, usually condemned as

'teleological', is in fact twofold and that one of its components does not at all infringe the accepted rules of science. In the last section, I explain the connection between this scientifically viable concept and the cyberneticians' model of the 'negative feedback loop'. To summarize my position, I hold that it was unfortunate that the ascendency of a particularly narrow-minded school of psychology confounded the notion of goal-directed behavior with the independent metaphysical notion of teleology.

The Objection to 'Teleology'

Among those who consider themselves "hard-nosed" scientists, the term *teleology* is still quite often used to justify the unconditional rejection of an argument or explanation in question. The reason, as Ernest Nagel put it, is that "the notion is equated with the belief that future events are active agents in their own realization," a belief which he characterizes as "a species of superstition" (1965; p.24). Sometimes a historical note is added. It explains that the misguided notion goes back to Aristotle, remained current throughout the Middle Ages, and was discarded once and for all by the first modern scientists in the Renaissance.

For science, as it developed in the Western world, it is unacceptable that something that lies in the future - something that has not yet occurred - should have the power to determine what is happening now. It would imply belief in an agency *outside* the field of our actual experience, and although such alien agencies were commonplace in mythology and religion, the enterprise that came to be called science tries to manage without them.

This is not at all unreasonable. Contemporary scientists (at least the more thoughtful among them) no longer believe that their efforts will be able to clear up all mysteries presented by the world in which we find ourselves living. All the more, however, they believe that the tools they use, the concepts, hypotheses, and principles they formulate, must in some way be testable in the experiential world. These conceptual tools may originate as free inventions, but if they cannot be fitted into a more or less coherent network of demonstrable causes and effects they will not be considered scientific. "The scientist does not believe in effects without causes; not even when they happen in the brain" (Ashby, in Conant, 1981; p.425). Hence, an external agent with powers that override the constraints we run into in our experiential world, would have to be supernatural

and therefore out of bounds for science.

The Hidden Ambiguity

In spite of this unsympathetic attitude, the term *teleology* has, if anything, become more frequent in the discussions of scientists and even in the literature that is considered scientific. This, I would claim, is not due to an inconsistency in scientific thinking but rather to an inconsistency in the meaning attributed to the term. The inconsistency is not a random mutation but rather the surfacing of a distinction which, among others, was pushed out of sight by the particular view of causality inherent in Galileo's and Newton's mechanics.

The spectacular success of mechanistic thinking led to the conviction that the only type of causation relevant to the scientific enterprise was the type Aristotle had isolated and described as *efficient* cause. Richard Taylor summarized the historical development in Edwards' *Encyclopedia of Philosophy* (1967):

> The concepts of material and formal causes are archaic and now have little significance outside aesthetics. Final causes have likewise been expurgated from physics. (Vol.II, p.56)

To this influence of the history of physics on the Western way of thinking, one must add the thrust of the behavorist doctrine in this century. Skinner's campaign succeeded to an extraordinary extent in excluding from the scientific domain not only human mind and meaning, but also all manifestations of intelligence in animals. Living organisms were to be squeezed into the Newtonian clockwork image of the world (in spite of the fact that theoretical physics had already moved beyond that conception). Consequently, the powerful establishment of behaviorist psychology branded intention, purpose, and goal-directedness as dirty words, and did its best to expurgate them as relatives of the concept of final cause.[2]

But, as Taylor remarked, this was also due to: "... a widespread misconception of what a final cause or purpose is supposed to be" (*loc.cit.*). Goal-directed behavior could be shown empirically, even within the rather limited range of experiments customary among behaviorists. But the professional behaviorists' apparatus, such as the Skinner box, frequently relied on an automatic mechanisms to record the experimental animal's 'responses'. This fostered a kind of tunnel vision, and as a result, the experimenter paid no attention to what the animal actually

did. The apparatus keeps record of lever presses, but ignores *how* the animal moves to the lever in order to press it and that, at different times, it may be using a different paw. In a similar vein, Gerd Sommerhoff observed, once a rat has mastered a maze and found the place where the food is hidden, you can flood the maze, drop the rat at the entry, and it will *swim* to the foodbox. Thus there are things, beyond the 'chaining' of elementary muscular responses, to guide the rat's behavior. He concluded: "To list the 'responses' of animals in different experimental situations, therefore, is to list the goals rather than the movements" (Sommerhoff, 1974; p.7).

Such goal-directed behavior can not only be observed and documented in experiments with animals, but we encounter it innumerable times in our everyday experience. There are, for instance, the conscious or unconscious accommodations we have to make - and make quite successfully - in the thousands of trivial routines that are indispensable in our way of living, such as retrieving the toothpaste that has fallen behind the wash basin, looking up the telephone number of a person we want to meet, locating a book on a shelf, finding our misplaced car keys, negotiating the stairs to the garage during a power failure, etc., etc. In all these cases, we know what we want, we have a clear-cut goal, and we usually succeed in attaining it. It is also clear that it is not some future event that is guiding us, but rather experiences we have gathered in the past.

Hence, it was unfortunate, to say the least, that the term *teleological* was indiscriminately applied to the explanation of actions that are in no way determined by something that lies in the future, something that still awaits to be experienced.

The confusion concerning the conceptual structures that might be involved when we speak of 'purpose', 'goal', 'final causes', and 'teleology' has fueled more than one debate. A good example are the comments and rebuttals in the journal *Philosophy of Science* that followed upon the publication of the famous paper by Rosenblueth, Wiener, and Bigelow (1943) that is sometimes cited as the beginning of cybernetics. At the root of that debate was the contention that the explanation Wiener and his colleagues formulated in terms of purposive action was superfluous because the phenomena they cited could be adequately explained in terms of the traditional concept of 'efficient cause' customary in the mechanistic world view. (Taylor, 1950a, 1950b)

I shall return to this debate at a later point. However, in order to comment on it from the rather different perspective of constructivism, it is indispensable to provide at least a cursory exposition of how the various notions involved in the area of

causality can be seen as the results of conceptual construction.

A View of Causality

Twenty years after the first publication of his famous paper on the conservation of force[3] (1847), Hermann von Helmholtz wrote an addition to it and made the categoric statement: "the principle of causality is in fact nothing but the presupposition of lawfulness in all the appearances of nature." In 1878, he expanded the statement to a paragraph that shows the derivation of the idea from Hume and also gives some indication of how Helmholtz modified Kant's notion of "cause" and yet remained a Kantian in his fundamental epistemological orientation.

> Every inductive inference is based on trusting that an item of lawlike behavior, which has been observed up to now, will also prove true in all cases which have not yet come under observation. This is a trust in the lawlikeness of everything that happens. However, lawlikeness is the condition of comprehensibility. Trust in lawlikeness is thus at the same time trust in the comprehensibility of the appearances of nature. (1878/ 1977; p.142)

From my constructivist perspective, three points made in this statement are crucial: Lawlike behavior is abstracted from *experience* (observation); inductive abstractions are projected into future experience (as pertaining to what has *not yet* been observed); and what we deem to be comprehension, concerns the domain of the phenomenal (the *appearances* of nature) and not an ontological reality of the noumenal.

I shall return to the epistemological implications of this position. First, however, it is necessary to bring out the fact that the principle of causality is not as uniform as it might appear from the quoted statement by Helmholtz. Its *lawlikeness* springs from the trust that similar 'causes' will have similar 'effects'[4]; by itself it does not specify that all causes and the operative relation hat links them with effects must be of the same kind. Our topic, here, is *teleology*, a particular notion of causation that has been put into question since the rise of science. Considerations such as the following are characteristic of the modern attitude:

> When we ask "why?" concerning an event, we may mean either of two things. We may mean: "What purpose did this

> event serve?" or we may mean: "What earlier circumstances caused this event?" The answer to the former question is a teleological explanation, or an explanation by final causes; the answer to the latter question is a mechanistic explanation. I do not see how it could have been known in advance which of these two questions science ought to ask, or whether it ought to ask both. But experience has shown that the mechanistic question leads to scientific knowledge, while the teleological question does not. (Russell,1946, p. 86-87)

Knowing How to Explain

For the moment, I want to disregard the distinction Russell made by isolating "scientific knowledge". To get a handle on the structure of concepts we must, as a rule, try to unravel how they came to be used before they were specialized in a particular context. The notion of cause and effect is usually considered to go back to Aristotle, who provided the first detailed classification of different concepts in this area. According to most scholars, the term Aristotle used for the sort of concepts he was considering, would be more adequately translated as "explanatory factor" or "principle" rather than the standard word "cause", which was introduced only when he was translated into Latin. (e.g. Wheelright, 1935).

Aristotle's interest in these concepts sprang from his view concerning the nature of knowledge.

> Knowledge is the object of our inquiry, and men do not think they know a thing till they have grasped the 'why' of it (which is to grasp its primary cause). So clearly we too must do this as regards both coming to be and passing away and every kind of physical change, in order that, knowing their principles, we may try to refer to these principles each of our problems. (Aristotle, *Physics*, Book II, Ch.3, 194b).

From my point of view, it is important to note in this particular context that Aristotle explicitly states that *knowledge* is the object of his inquiry, not a specific aspect of the world to be known. Consequently, I feel justified in interpreting what he says about the different 'explanatory principles' he isolates as an analysis of the conceptual structures we use to "grasp" and to explain the sort of experience we call *change*.

Though Aristotle, as far as I know, did not dwell on this, *Change* itself is, of course, a compound notion because it involves a relation. One observation - a single frame of experience - provides no opportunity to conceive of change. We need at least two. The same applies to the notion of *state*. That is to say, neither of these two basic elements in the construction of our experiential world is conceivable unless we segment experience into separate observations - discrete frames - and then focus attention on similarities or differences between the segments. As Lawrence Frank (1948) put it, "We ... punctuate the order of events more or less arbitrarily, according to our preconceptions and methodological necessities, and then assume that these conceptual mechanisms are the basic processes of nature" (p.190).

Aristotle's Classification: Cause 1 and 2

The first two of Aristotle's four explanatory principles are now known as *material* and *formal causes*. Although they are generally considered "archaic" and inappropriate for our attempts at explanation, I am including them in this exposition of Aristoltle's classification, because there is an important area of modern science where the first of these *causes* plays a dominant role.

Aristotle lists the four categories into which he divides the concept of causation in the following order: Material, formal, efficient causes, and final cause, the one that is most important for his *teleological* world view, last. This is the very order I would have chosen for this exposition. Thus I shall follow Aristotle, presenting for each category the essentials of his characterization. To avoid space-consuming cross references, however, I shall discuss the first two together.

> 1. Material Cause: "That out of which a thing comes to be and which persists, ... e.g. the bronze of the statue, the silver of the bowl, and the genera of which the bronze and the silver are species."
>
> 2. Formal Cause: "The form or the archetype, i.e. the statement of the essence, and its genera, ... (e.g. the relation 2:1 of the octave, and generally number), and the parts in the definition." (Aristotle, *Physics*, Book II, ch.3, 194b)

On the face of it, it seems trivial to say that bronze is the cause of a bronze statue. Aristotle, however, had in mind the

relation between *matter* and *form*. Materials such as bronze and metals (the genus of which bronze is a species) have, in themselves, no form in which they could be said to *persist*. Conversely, the form of a statue has no way of *being* or *coming into being* except as the shape implemented in some kind of matter. This, as Bertrand Russell explained, can be further substantiated by invoking the concept of unity. Materials, such as bronze, water, or marble, have no unity (i.e. they are not unitary things), unless we take them in discrete amounts.

> The part of a block of marble which afterwards becomes a statue is, as yet, not separated from the rest of the marble; it is not yet a 'thing', and has not yet any unity. After the sculptor has made the statue, it has unity, which it derives from its shape. (Russell, 1940; p.193)

The way the sculptor imposes form, is by chipping away the bits of marble that do not fit into his vision. The way form is imposed on the bronze, is by pouring it into a constraining mold. Potentially, the material could end up in innumerable other forms, but the procedure of statue-making, be it chipping or casting, eliminates all but one.

This Aristotelian insight is not a bit archaic, it merely tends to be overlooked nowadays. Most of the questions of how the theory of evolution relates to other scientific theories could be answered by the simple observation that it implicitly involves the principle Aristotle characterized as the *material cause*. Natural selection bears on populations of organisms whose individual properties and capabilities are provided by accidental variation. The selective procedure operates through the environment whose constrain determine which among the individually different organisms survive and which do not. The environment cannot generate the *material*, i.e., the population that comprises viable and non-viable variations, it only 'chisels away' those that do not fit. The analogy, of course, does not stretch to include the sculptor. The natural environment that carries out the selective process has no more a vision of the forms that are left than the sculptor's chisel has a vision of the statue it helps to peel out of the marble. Such a vision may be attributed to the sculptor. It would constitute a *telos* or goal, which will be discussed when we come to discuss *final* causes.

Two further points have to be made. The first is that Aristotle saw the premises of a syllogism as the *material cause* of the conclusion. This was his way of saying what logicians have reiterated ever since, namely that the deductive procedure does

not generate *new* knowledge, but merely picks out or recombines elements that are already contained in the premises. (This, of course, in no way precludes the fact that the deductive procedure often brings out features of whose presence in the premises one was unaware.)

The second is that Aristotle, as we saw above, held that definitions are to be considered the *formal* cause of the entity that is defined. The example of the bronze statue again offers a useful image. The shape of the statue is quite literally 'defined' by the mold, in the sense that the mold constrains and delimits where the liquid metal can flow. Analogously, the 'parts of a definition' constrain and delimit both conceptual construction and the application of concepts. Definitions set conditions which the abstract *form* we call concept must satisfy, and it is the same conditions that some experiential material must fit in order to be accepted as a proper instantiation of the concept.

The Cause in Mechanics

3. Efficient Cause: "The primary source of the change or coming to rest; ... and generally what makes of what is made and what causes change of what is changed." (Aristotle, *Physics*, Book II, ch.3, 194b)

The notion of *change* seems to go back to the very beginnings of human conceptualization. It must have been a commonplace notion already in the days of Heraclitus, who placed it at the foundation of the experiential world. A quotation from the famous fragments of his philosophy is usually given as "one cannot step twice into the same river". In fact it continues, "for fresh waters are ever flowing in upon you".[5] Taken in its completeness, it is a perfect metaphor of 'change'. It shows that the concept requires at least *two* experiences, an item that is identical in both, (the *same* river), and a difference (the water that is *not* the same in the second experience). If no difference is noticed, and we find whatever we related to the item in the first experience related to it also in the second, we would conceive the combination as a state. (Stretching the metaphor: One *can* step twice into the same puddle.)

According to this break-down, change is a relational concept and, as such, requires more than one segment of experience, as well as a comparison. It also manifests what I consider a basic presupposition of all conceptual analysis: Segments of experience, insofar as they reach the level of conceptualization and

rational description, always appear sequential. This sequence is usually interpreted as a temporal one.[6] We can therefore assign individual moments to the segments from which we derive 'change', and we can map them as t_1 and t_n. It was Silvio Ceccato, the founder of the Italian Operationalist School, who in the 1950s devised a simple way of mapping such sequential conceptual structures.[7] In his notation, *change* would be mapped as follows:

t_1		t_2
X	≡	X
-a	≠	+a

Where "X" is the item considered the identical individual and "a" the attribute whose absence in t_1 and presence in t_2 constitutes the difference. The peculiarity of this structure is that it comprises both the relation of identity and the relation of difference. The apparent conflict between the two relations generates, if I may borrow a Piagetian term, a 'perturbation'.[8] The conceptual equilibrium is disturbed by the fact that, at t_2, the item X should be *the same* as well as *different*. To re-establish balance, the difference must be justified in some way, that is to say, something must be found at t_1 that can be held responsible for the change (Ceccato, 1966, called this 'looking for *un sanatore*', i.e., for a curing agent.)

When this search for what might have brought about the observed change becomes systematic, in that it lists elements that can be considered copresent with X at t_1, it develops into the 'scientific' method that aims at isolating one variable (or several) as the active *cause*. In experimental science this is done by successively 'controlling' for all variables that could possibly be candidates. If one is found, it can be added to the notation for *change*, and we thus obtain a graphic representation of the cause/effect relation.

t_1		t_2
X	≡	X
-a	≠	+a
b,c,d,(e) ↗		

Where "b,c,d,e" are other elements found present at t_1 and "e"

has been isolated as the element that is to be considered the *primary source* or the *efficient* cause responsible for the change in "X".

To this must be added that science always strives to go beyond the mere correlation of cause and effect and tries to conceive of a 'model' that shows *how* the change could be brought about.

I have presented this approach to a conceptual analysis of causation in some detail, because it opens a way of sorting out the confusion which, in my view, has bedeviled the discussion about *teleology*. Though the temporal frames that compose the concept of efficient cause are obviously abstracted from prior experience and therefore lie in the past, they can, and often are, projected as *predictions* into areas that have not yet been actually experienced. As Burgers observed in his essay on *Causality* and *anticipation*:

> "... there is no reference to the future as a determining agent. ... All predictions that can be made are still statements which relate the present to the past". ... The notion of a coming future involves an extrapolation from past experience to possible future events. (Burgers, 1975; p.194-195)

The procedure of making predictions on the strength of past experience is not only the mainstay of science but also underlies almost every action we carry out in our everyday lives. However, whereas scientists as a rule explicitly formulate their predictions, the ordinary human subject leaves them implicit in the successful ways of acting that have been turned into habits. When we open the door to the dark staircase that leads to the basement, we do not 'predict' that our feet will find steps to walk down, we simply *trust* that the stairs are still there.

This trust must be no less, but need be no more, than Hume held to be necessary:

> If there be any Suspicion, that the Course of Nature may change, and that the past may be no Rule for the future, all Experience becomes useless, and can give rise to no Inferences or Conclusions. (Hume, 1750, Essay IV, Part II)

Needless to say, suspicion about the continuing stability of the world we live in would not only vitiate the formulation of inferences and conclusions but would also turn every step we take and every movement we make into an exploration of *terra in-*

cognita. Prediction, in one form or another, permeates our living, and the expectation that the *efficient causes* we have isolated in the past will have their *effects* also in the future, is the key to whatever success we have in managing our experience. In the next section, I shall try to show that the projection of efficient causes provides a key also to the problem of goal-directed actions.

The Concept of 'Goal'

4. Final Cause: "That for the sake of which a thing is done, e.g. health is the cause of walking about. ('Why is he walking about?' we say. 'To be healthy', and having said that, we think we have assigned the cause.) The same is true also of all the intermediate steps which are brought about through the action of something else as means towards the end, e.g. ... drugs, or surgical instruments are means towards health. All these things are 'for the sake of' the end, though they differ from one another in that some are activities, others instruments." (Aristotle, *Physics*, Book II, ch.3, 194b-195a)

In the next chapter of his *Physics*, where Aristotle adds some further remarks on his classification of causes, he says:

> Some things cause each other reciprocally, e.g. hard work causes fitness and vice versa, but not in the same way, but the one as end, the other as the origin of change. (Physics, Book II, ch.3, 195a)

This, I suggest, is an important clue that Aristotle was using the *end (telos)* of his *final causation* in two somewhat different senses. In fact, one paragraph later on the same page, he says: "..; for 'that for the sake of which' means what is best and the end of the things that lead up to it."

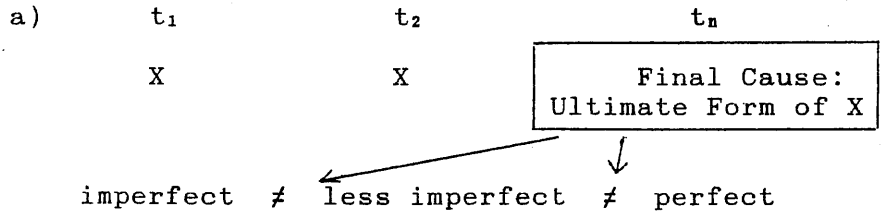
For Aristotle, 'what is best' is the *highest good*, i.e. the ultimate *form* towards which, according to the *teleology* in his metaphysical belief, the world is drawn.

> Matter is conceived as a potentiality of form; all change is what we should call 'evolution', in the sense that after the change the thing in question has more form than before ... the universe and everything in it is developing towards something continually better than what went before. (Russell, 1940; p.189)

For any experiencing subject, *the things that lead up to this end* lie in the future and are therefore unknown, except for their general direction. In contrast, those who jog *for the sake of their health* do so because they believe that jogging is, indeed, an *efficient cause* that leads up to that end by generating it as an *effect*. They wish to improve their health, consequently they enact the cause which, in their past experience, has reliably brought about the desired end. What prompts them to act in this way, is not an end that lies in the future but the connection of an *efficient cause* and its *effect* which they have in one way or another come to believe in and which they are able to remember and *re-present* to themselves as a means to change the state of their health.

How such a belief comes about, has been suggested by many authors. Ernst Mach, towards the end of the last century, wrote: "In speaking of cause and effect we arbitrarily give relief to those elements to whose connection we have to attend in the reproduction of a fact in the respect in which it is important to us" (Newman, 1988; Vol.3, 1760). Bertrand Russell expressed the same idea early in this century: "The mere fact that something has happened a certain number of times causes animals and men to expect that it will happen again" (1912; p.63).

Using the notation of temporal frames, we can graphically represent the difference between Aristotle's two senses of 'final' by mapping both.



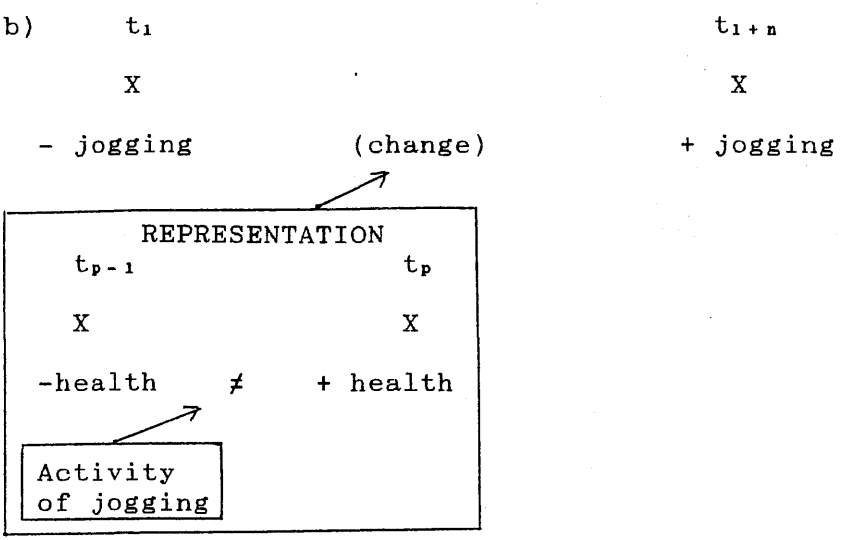
The causal agent held responsible for the change in "X" is located at a distant point in the future, tn. Consequently, something that has not yet entered the experiential world is assumed to have the power to determine an experience that is present. One might suggest that this would create no difficulty, if the notion of time that we abstract from the succession of our experiences had nothing to do with the sequence of events in a universe whose goings-on are independent of any experiencer. This assumption is not at all alien to 'instrumentalists' and 'constructivists' who hold that whatever we want to call knowledge inevitably applies *only* to the experiential world and war-

rants no inferences about an independent ontological reality. But even for this school of thought the oddity of this type of causation remains.

To connect two experiential items by a concept of cause and effect, one must at least be able to re-present both to oneself. The prisoner who plots to escape presumably has some vision of an outside, a vision of what it would be like to be rid of the constraints that obstruct his freedom at present. Any such vision is a re-presentation[9] of things he knows or dreams of, and it is this re-presentation that constitutes *final cause*, the end 'for the sake of which' he is now engaging in actions which, on the basis of past experience, he believes will bring that end about.

In contrast, the ultimate good that Aristotle posits as the final cause in his metaphysics is not, and cannot be, the re-presentation of something already experienced. It will manifest itself only when it is reached. Consequently there is no past experience of steps that led towards it or actions that brought it about. There are no abstracted cause-effect relations that one could try to implement to reach an unfathomable end. In short, even for those who believe that knowledge does not pertain to anything beyond the realm of experience, a cause that lies outside has no explanatory power.

The conceptual structure is a different one in situations where there is experiential knowledge of how to bring the end about. This knowledge can be mapped as the re-presentation of an established cause-effect connection, and it is this re-presentation, *projected* into the future, that now becomes the cause of the activity that is believed to bring about the end.



Here "X" is the person who engages in the activity of jogging *because* (for the sake of) his or her health. That is to say, the acting person has the *goal* of improved health. From past experience (t_{p-1} ... t_p), the person has the belief that jogging is an efficient cause that will bring about the goal of better health as its *effect*. It is the re-presentation of this cause-effect connection that now *causes* the actor's engagement in the activity.

The Source of Induction

One of the most striking slogans of Skinner's campaign was: "Behavior is shaped and maintained by its consequences" (1972, p.16). Since a consequence refers to something that happens after what it is thought to be a consequence of, the slogan might sound teleological. This is, of course not intended. Rather, one has to interpret it as: Behavior is shaped and maintained by the consequences it had when it occurred in the past. Understood in this way, it is hardly more original than the old proverb, "Once burned, twice shy", which one can see confirmed daily by children, dogs, and other living organisms. A more explicit rephrasing of the idea may bring it closer to our topic:

> The living system, due to its circular organization, is an inductive system and functions always in a predicitive manner: what occurred once will occur again. Its organization (both genetic and otherwise) is conservative and re-

peats only that which works. (Maturana, 1970, p. 15-16)

An organism's actions, thus, are selected and shaped according to what worked in the past. This is claimed to be the case, both in phylogenetic and ontogenetic development. There is, however, one big difference between the two domains. In the evolutionary domain, 'that which works' is simply that which survives. Everything that does not work, dies out. There is no 'induction' on tha part of the organisms; only outside observers may describe the phylogenetic procedure as inductive, and if they do, it is somewhat metaphorical. The same goes for certain mechanisms in biological organisms, that act to maintain certain internal conditions in equilibrium. Walter Cannon (1932/1963) provided examples of such 'homeostatic' phenomena (e.g. control of the heartbeat, sugar level, etc.). These mechanisms are sometimes described as though they were ingenious inventions of the evolutionary process, but this is a metaphorical inversion. The Darwinian theory explains them as the result of fortuitous mutations that bestowed increased fitness on the organisms that happened to have them.

In the individual history of an organism, however, the situation is not quite so stark. With regard to most situations an organism gets itself into, there is a 'twice', and the organism can do something to avoid the burns of the first occasion. If the avoidance manouvre is successful, it will be repeated whenever a similar situation is met. Thus it is, indeed, an inductive procedure, because 'what works' is seen from the organism's point of view and selected within the organism's own experience. But this means that there must be something in the organism that decides whether an action does or does not 'work'. This assessment may be as automatic and unconscious as it is with the rat that avoids electric shock or the toddler who stays away from the hot stove; but, conscious or not, there is something that decides that actions which lead to shocks or burning sensations do not 'work' and, consequently, should not be repeated.

To my mind, this way of analysing behavior is illuminating because it brings out the governing principle of induction. But it is incomplete because it persistently avoids mentioning that in order to answer the question *what works*, one needs a notion of value, no matter how elementary or unconscious that notion might be. If there is no preference for *not* having pain and getting blisters on one's fingers, there is no reason why the toddler should not touch the hot stove every time it happens to be near enough.

This point is crucial to any discussion of our concept of

causality. If we claim that the causal connections we inductively make between segments of experience are 'shaped and maintained' because they work, we must go on to explain that 'work', in this context means that these connections enable us, at least to some extent, to avoid certain situations and to engender others that we prefer. In other words, the cuasal connections, fragile though they may be, provide some control over our experience. From this, one is led to conclude that the scientific search of *efficient causes* is fueled largely, if not entirely, by our intention to use them for the attainment of goals. Of course, one must add that the goals need not be 'utilitarian' in the common, materialistic sense, but can include *understanding or explanation* at any level of abstraction.

This way of thinking leads to the realization that the ambiguity of the term *teleology* is at least threefold:

First, there is the scientifically useless metaphysical teleology that involves a non-experiential agent that pulls all changing and developing towards a future end.

Second, there is the goal-directedness of action that living organisms embark on because these action have in their past experience led to satisfactory ends.

Third, but closely related to the second, is the endeavor to establish further causal connections, an endeavor that is unquestionably goal-directed although the goals are as yet unknown.

The most radical expression of this third sense of teleology is Spencer Brown's statement:

> "There can be no distinction without a motive, and there can be no motive unless contents are seen to differ in value." (1969, p. 1)

Teleology Revived

In our century, debates about teleology arose in two distinct areas of science. On the one side, proposals to introduce *èlan vital* or some other external guiding force as supplement to the Darwinian theory of evolution were, and are as a rule, vigorously opposed as unscientific teleology by biologists. Since this opposition is felt very strongly, it has frequently led them to dismiss also the notion of goal-directedness, because this, too, was covered by the ambiguous term 'teleology'. On the other side, the founders of cybernetics, intent upon mechanizing goal-directed behavior in artificial devices, helped to perpetuate the ambiguity

by calling their constructs 'teleological'. By choosing that term, they provoked the immediate antagonism of people who might otherwise have been quite favorably disposed towards the novel enterprise.

As a non-biologist, I can deal only briefly with the teleology debate in that area. For the reasons I have mentioned in connection with Aristotle's metaphysical beliefs, the assumption of a directive force that cannot be integrated with other networks of causal explanation is held to be unacceptable by most scientists. If no model has been conceived to illustrate *how* such a force might work and if there is not even a mathematical method of approaching it quantitatively, it is treated as an *ad hoc* hypothesis that explains nothing because it could be suggested as an answer to almost any question.[10]

Biologists, however, continually find themselves describing processes and developments whose end-states are so well known that they can be taken for granted. Yet, for fear that they *might* be led into the heresy of 'teleological explanation', they tend to avoid reference to these end-states in their descriptions.

> Biologists for a while were prepared to say a turtle came ashore *and* laid its eggs. These verbal scruples were intended as a rejection of teleology but were based on the mistaken view that the efficiency of final causes is necessarily implied by the simple *description* of an end-directed mechanism. ... The biologists long-standing confusion would be removed if all end-directed systems were described by some other term, like 'teleonomic', in order to emphasize that recognition and description of end-directedness does not carry a commitment to Aristotelian teleology as an efficient causal principle. (Pittendrigh, 1958, p. 393-394; italics in the original).

Ernst Mayr (1965) cited Pittendrigh and criticized him for not making a "clear distinction between the two teleologies of Aristotle." The analysis of Aristotle's ambiguity clearly justifies this criticism. In spite of it, Mayr was willing to adopt Pittendrigh's new term, but supplied his own definition: "It would seem useful to rigidly restrict the term teleonomic to systems operating on the basis of a program of coded information" (p. 42).

It would go beyond the scope of this paper to investigate the causative properties of codes with regard to their interpretation. I would merely observe that if one sees the genetic code as the determinant of biological development and behavior, it is a cause

that lies not in the future but inside the organism and is also prior to any of the effects one attributes to it. Mayr's definition, thus, successfully avoids the accusation of metaphysical teleology. The way this is achieved, however, would strictly limit the application of the new term to phenomena that can reasonably be explained as governed by the genetic code - and behaviors or actions that are genetically determined could not at the same time be considered the result of induction.

As far as I know, the introduction of 'teleonomy' has not had much resonance in the biological literature. Jacques Monod (1970) adopted it, but made no mention of either Pittendrigh's or Mayr's definition. He suggested using the term for "objects endowed with a purpose or project." Instead of clarifying the issue, this merely compounds the ambiguity, because 'purpose' was traditionally included in the proscription against teleology. In the twenty essays by prominent biologists collected in *Studies in the philosophy of biology* (Ayala & Dobzhansky, 1974), the term 'teleonomy' occurs only twice. It is mentioned by Monod himself, and also by Montalenti, who understands it as a substitute for the "finalism" which, he claims, "cannot be denied in biological affairs" (p. 10). In contrast, disapproval of 'teleology' is emphasized by nearly all the contributors to the book.

From this outsider's point of view, the failure to resolve discussions about causality in biology springs from two roots: unawareness of the fact that the theory of evolution involves *material* rather than *efficient causation*; and insufficient awareness of the ambiguity in the use of the term 'teleological', mentioned by Mayr, but still not ad-equately clarified on the conceptual level.

To end this cursory review of the teleology debate in biology, I want to return to a passage in Aristotle which would deserve to be mentioned in the context of evolution. In chapter 8 of Book II on Physics, Aristotle sets up a straw man in order to argue for his belief that the world as a whole has a purpose and is teleologically guided. The straw man is a remarkable anticipation of the theory of evolution in terms of natural selection. From our present point of view, the argument he brings up *against* this theory is unconvincing because it is metaphysical, not scientific. This is not surprising. What *is* surprising, is that in this passage Darwin and Wallace could have found all the essential components of their theory: spontaneous, accidental variation, fit with the environment, and the elimination of the unfit by natural selection - Aristotle's closing argument, though conceived more than two thousand years ago, crystallized a theme that still reverberates in natural philosophy today:

> It is absurd to suppose that purpose is not present because we do not observe the agent deliberating. Art does not deliberate. If the ship-building art were in the wood, it would produce the same results *by nature*. If, therefore, purpose is present in art. it is present also in nature. The best illustration is a doctor doctoring himself: nature is like that.
> It is plain that nature is a cause, a cause that operates for a purpose. (*Physics*, Book II, ch. 8, 199b)

The Cybernetic Approach

This all-embracing purposiveness that Aristotle ascribed to the universe as a whole sprang from a metaphysical world view incompatible with the engineering interests of the cyberneticians who generated their own definition of 'purpose' when they launched the new discipline some decades ago.

> Any theory starts off with an observer or experimenter. He has in mind a collection of abstract models with predictive capabilities. Using various criteria of relevance, he selects one of them. In order to actually make predictions, this model must be interpreted and identified with a real assembly to form a theory. The interpretation may be prescriptive or predictive, as when the model is used like a blueprint for *designing* a machine and predicting its states. On the other hand, it may be descriptive and predictive as it is when the model is used to explain and predict the behaviour of a *given* organism. (Pask, 1969; p. 23; italics in the original).

Gordon Pask links the prescriptive purpose with the preposition 'for', the descriptive with 'of', and he completes this by adding that "any system with a purpose for it ... also has a purpose in it" (*loc.cit.*, p. 24). A simple example may illustrate this. Having got tired of buying matches, someone may decide to design a cigarette lighter. Lighting cigarettes will be the purpose prescribed *for* the gadget. "People do not build purposeless machines" (*loc.cit.*). Prescriptive purposes, therefore, are there prior to their embodiment, which then has the particular purpose *in* it. In contrast, if archaeologists, in digging up remnants of a bygone civilization, find an unknown item and discover that it generates a flame when it is handled in a particular way, they may conclude that this was indeed its purpose. This would be

conceived as the purpose *of* the item, in their description.

The descriptive purpose *of* is what Pittendrigh attempted to capture when he introduced the term 'teleonomy'. In biology this would be a concept that pertains to an observed function of an organ or behavior and it in no way entails that the organ or behavior was *designed for* the particular function. Where evolution is concerned, then, there is no harm in using 'purpose of' as a descriptive tool, provided one does not mistake it for the purpose *for*, which would imply a guiding outside force that intentionally designed the thing one is describing.

This distinction clears some of the ground, but it does not illuminate *how* the behavior of an organism might, in fact, be directed towards a goals. This is the very question that led to the conception of 'negative feedback' and the foundation of cybernetics.

As I mentioned earlier, it was unfortunate that Rosenblueth, Wiener, and Bigelow (1943) mentioned *teleology* in the title of their seminal paper. (In the pages that follow, I shall refer to the authors of this paper as RWB.) In their attempt to define 'purposeful' behavior, they gave a variety of examples which, rather than clarify, confounded the concepts of prescriptive and descriptive purpose. Thus they distinguish 'voluntary' from 'involuntary reactions' and discuss the notion "that all machines are purposeful" (p. 19). I would argue that the structures we call 'machines' are, as a rule, designed to carry out quite specific functions which, applying the above dichotomy, must be categorized as *prescriptive* purposes. The only mechanical structures I know that are deliberately devoid of function are those designed by contemporary artists such as Pol Bury and Jean Tinguely, created for their effect as symbolic constructs and called 'machines' in a metaphorical vein.

RWB state that "mechanical devices such as roulette (are) designed precisely for purposelessness" (*loc.cit.*), forgetting that a description of a roulette table would be somewhat incomplete if one did not mention that it was designed to produce random numbers and would be considered a costly failure if it did not do this.

The authors make the further distinction of "active" versus "passive" behavior, based on the internal or external origin of the energy that serves as motor; and they cite a "final condition" (or event) towards which the behavior of purposeful devices strives (*loc.cit.*). From an engineering point of view, these are no doubt useful distinctions, but they do not help the clarification of the notion of purpose.

The important feature of the RWB paper is the introduction

of the concept of *negative feedback*. Having explained "positive" feedback by means of the example of an electrical amplifier, they describe its positive counterpart as follows:

> The term feed-back is also employed in a more restricted sense to signify that the behavior of an object is controlled by the margin of error at which the objects stands at a given time with reference to a relatively specific goal. The feed-back is then negative, that is, the signals from the goal are used to restrict outputs which would otherwise go beyond the goal.

In his frequently cited critique of RWB, Richard Taylor (1950a) discusses at great length the dubious examples given in RWB. In retrospect, many of his objections seem justified. However, what he says about objects controlled by negative feedback, reveals an apparent misunderstanding: "... such a mechanism is so designed that the *effects* of its behavior themselves enter as causal factors *on* its behavior..." (p. 315, italics in the original).

This is even more simplistic than the behaviorists adage stating that behavior is shaped by its consequence. They at least acknowledged that certain consequences were 'reinforcing' whereas others were not. As they quickly discovered, one and the same thing might be reinforcing under certain circumstances (e.g. meat pellets, when the rat was hungry) and not reinforcing under others (e.g. when the rat was well fed). Thus there was no proper causal connection between reinforcement and subsequent behavior, because it was to some extent the rat who decided what it considered reinforcing and what not.

Taylor wants to turn the *effects* of a feedback mechanism's behavior into a 'causal factor', but he overlooks that one and the same effect does not always generate the same subsequent behavior. This has to be so, because, as RWB say, the behavior of these mechanisms is controlled not by an external energy or signal but by the *relationship* between the received signal and an internal reference value or goal.

This is the reason why the needle of a magnetic compass, cited by Taylor (1950a, p. 316), is a particularly misleading example. The equilibrium to which it returns after any disturbance is as wholly determined by the external magnetic field as the direction of a vane by the wind or the direction of a cart by the horse that pulls it. Rosenblueth and Wiener (1950), in their reply to Taylor's critique, categorize the compass needle's behavior as 'passive', but concede that "it appears to be desirable

to regard the motions of a magnetic compass that has been deviated from its resting position as purposeful, with the final resting orientation as the goals."

The example shows the usefulness of Pask's conceptual distinctions. The compass was designed *for* the purpose of indicating where north is. Consequently that purpose is *in* the compass, irrespective of how it manages to serve it.

With living organisms, the question of *how* they manage to attain goals is crucial, if one wants to maintain that they have goals. In this regard, William Powers' work on negative feedback provided a clarification. The title of his book put it succinctly: *Behavior, the control of perception* (1973). That is to say, a feedback mechanism acts or behaves only to maintain or re-establish a *fit* between what it senses and the 'reference' constituted by the representation of a goal-state that was set for it. This goal is neither outside the mechanism, nor does it lie in the future. Consequently, the conception of control through negative feedback is perfectly compatible with the analysis of goal-directed behavior I suggested as type (b) in the discussion of *final cause*. The two new features it adds are the capabilities of sensing the present state of affairs and of comparing it to a representation of the desired state.

As far as mechanical or electronic devices are concerned, from the humble thermostat in our refrigerators to the automatic pilot and target-finding projectiles, the goal or reference state is, of course, always set by the user of the device. Hence it is quite legitimate to call the attainment and maintainment of the projected state the *purpose in* the mechanism.

In regard to living organisms, the negative feedback principle offers a powerful model for the description of many of their behaviors. For instance, the various actions the prisoner engages in to implement his escape can all be seen as actions directed towards intermediary goals which, in his view, are steps towards regaining the final goal, i.e., to be rid of the constraints he suffers at present. All the reference states to which, at every moment, he compares his present state, are re-presentations constructed from past experiences to which he himself assigns the role of goal.

There is nothing mysterious about this procedure, nothing that requires belief in a metaphysical teleology. It is a chain of actions, each of which has in the actor's past experience led reliably to a particular result. He now wants to generate these results, and therefore uses the same actions as *efficient causes* to generate them.

In this sense we can now understand that "purposive ac-

tivity does not involve a special kind of causality but only a special organization of ordinary causal processes" (Edwards, 1967, Vol. 8, p. 90).

Conclusion

The claim that goal-directed activities can be conceptualized in a way that neatly separates them from the metaphysical notion of teleology is derived from an operational analysis of the constructs of 'cause'. This analysis shows that if cognitive agents have the ability to re-present to themselves the outcome of actions they carried out in the past, they are also able to carry out actions which they believe, on the basis of past experience, will cause desirable results. In other words, it makes sense for them to say, for instance, that they go for walks *for the sake of their health.*

The approach I have taken in this analysis is explicitly based on the elementary form of induction as it was understood by Hume. Thus it is deliberately simplified - perhaps oversimplified in its application to living organisms. It is intended as a working hypothesis which, if found useful, will undoubtedly have to be extended and adapted. Above all, it leaves out of account the creativity living organisms manifest in their efforts to solve problems and to attain goals for which their past experience has not yet provided a reliable recipe. Peirce's notion of *abduction*, i.e., the spontaneous invention of possible regularities and their subsequent testing and conversion to inductive rules would have to be incorporated into any more ambitious model of human behavior. On the other hand, I believe that before such a model can be successfully formulated, it is essential to get rid of the notion of metaphysical teleology without eliminating the notion of goal-directed behavior.

Scientific Reasoning Research Institute
University of Massachusetts

NOTES

1. The title of Wolff's treatise was *Philosophia rationalis sive logica* and his definition of 'teleology', according to my dictionary of philosophy, is in §85 of his Preliminary Discourse.
2. Powerful though the behaviorist establishment was, in forming the general image of psychology, it is only fair to

report that there were quite a few eminent psychologists and biologists who, each in his own way, swam against the current; Wolfgang Koehler, Karl Lashley, Donald Hebb, George Kelly, Jean Piaget, to mention just a few. But, at least in the United States, it was only in the 1970s that the balance began to shift.

3. Like Julius Robert Mayer (1842), a pioneer in the formulation of the principle of energy conservation, Helmholtz used the word *Kraft* for what later was called "energy".
4. Analytic philosophers have raised the difficulty of establishing what is 'similar' and what not. Rather than pursue this line of argumentation here, I would repeat that we are concerned with appearances, which is to say with conceptions and perceptions, and that the problem of similarity can be resolved by the Piagetian notion of 'assimilation'.
5. cf. Russell, 1946, p. 63; and Diels, fragments 12 & 49a.
6. This inevitable sequentiality of experience was what led Kant to the assumption that time had to be an *a priori* conception (*Wahrnehmungsform*). From the constructivist point of view, sequentiality is indeed a basic condition of conscious experience, but it is only one element in the construction of 'time'.
7. Ceccato's method was used by all of us who worked on semantic analyses at the Center for Cybernetics in those years (cf. Glasersfeld, 1961/63, 1972, and 1974), but Ceccato, to my knowledge, published a reference to it only in 1980.
8. The need to establish an individual identity in spite of an observed difference was mentioned by Russell (1917): "We have now to consider what leads us to speak of another set of appearances as belonging to the same 'thing' at a different time" (p.169).
9. I spell 're-presentation' with a hyphen to indicate that it should be interpreted as the replay of a past experience, as opposed to an icon or picture of an original. The hyphenated word would be translated into German as *Vorstellung*, the other as *Darstellung*. The conflation of the two concepts in English has created endless confusion.
10. It is interesting to note that hypnosis has by and large been accepted as a viable notion in spite of the fact there is neither a model of how it works nor a quantitative approach. The reason presumably is that one can actual witness both the causative agent and the effect.

REFERENCES

Aristotle, Physics (Book II). In R. McKeon (Ed.), *Introduction to Aristotle* (translation by W.D.Ross). New York: Random House Modern Library, 1947.

Ashby, W.R. (1981) see Conant.

Burgers, J.M. (1975) *Causality and anticipation*, Science, 189, 194-198.

Cannon, W. (1963) *The wisdom of the body*. New York: Norton (Enlarged edition of original of 1932).

Ceccato, S.(1966) *Un tecnico fra i filosofi*, Vol.2. Padova: Marsilio.

Ceccato, S. & Zonta, B. (1980) *Linguaggio, consapevolezza, pensiero*. Milan: Feltrinelli.

Conant, R. (1981, Editor) *Mechanisms of intelligence: Ross Ashby's writings on cybernetics*. Seaside, CA: Intersystems Publications.

Edwards, P. (1967, Editor) *The encyclopedia of philosophy*. New York: Macmillan.

Frank, L.K. (1948) Foreword. In *Teleological mechanisms*. Annals of the New York Academy of Sciences, Vol.50, Article 4, 189-196.

Glasersfeld, E. von (1961) *Operational semantics: Analysis of meaning in terms of operations*. Report EUR 296e. Brussels: EURATOM Publications (reprinted, 1963).

Glasersfeld, E. von (1972), Semantic analysis in terms of conceptual situations, *Linguistics*, 94, 90-107.

Glasersfeld, E. von (1974). 'Because' and the concepts of causation, *Semiotica*, 12(2), 129-144.

Helmholtz, H. von (1881) *Ueber die Erhaltung der Kraft: eine physikalische Abhandlung* (first published 1847); reprinted with additions, Leipzig: Ostwald, 1899.

Helmholtz, H. von (1878) Die Tatsachen in der Wahrnehmung. In P.Hertz & M. Schlick (Eds.) *Hermann von Helmholtz, Epistemological Writings* (transl. M.F.Lowe). Dordrecht, Holland: Reidel, 1977.

Hume, D. (1750) *An enquiry concerning human understanding*. New York: Washington Square Press, 1963.

Mach, E. (1917) *Erkenntnis und Irrtum* (3rd edition). Leipzig: J.A.Barth.

Maturana, H. (1970) Neurophysiology of cognition. In P.Garvin (Ed.), *Cognition: A multiple view* (pp. 3-23). New York: Spartan Books.

Mayer, J.L. (1842) Bemerkungen über Kräfte. In J.Liebig (Ed.) *Annalen der Chemie und Pharmacie*, Munich.

Mayr, E. (1965) Cause and effect in biology. In D. Lerner (Ed.),

Cause and effect (pp.33-50). New York: Free Press.
Monod, J. (1972) *Chance and necessity*. New York: Vintage.
Nagel, E. (1965) Types of causal explanation in science. In D.Lerner, *Cause and effect* (pp.11-32). New York: Free Press.
Newman, J.R. (Editor, 1988) *The world of mathematics*, Vol.3. Redmont, Washington: Tempus/Microsoft Press.
Pask, G. (1969) The meaning of cybernetics in the behavioural sciences (The cybernetics of behaviour and cognition; extending the meaning of "goal"). In J.Rose (Ed.) *Progress of cybernetics* (pp. 15-44), London/New York: Gordon & Breach.
Pittendrigh, C.S. (1958) Adaptation, natural selection, and behavior. In A.Roe & G.G.Simpson (Eds.), *Behavior and evolution* (pp.390-416). New Haven: Yale University Press.
Powers, W. (1973) *Behavior: The control of perception*. Chicago, Aldine.
Rosenblueth, A. & Wiener, N. (1950), Purposeful and non-purposeful behavior, *Philosophy of Science*, 17, 318-326.
Rosenblueth, A., Wiener, N., & Bigelow J. (1943), Behavior, purpose and teleology, *Philosophy of Science*, 10, 18-24.
Russell, B. (1946) *A history of Western philosophy*. London: Allen & Unwin.
Russell, B. (1917) *Mysticism and Logic*. London: Allen & Unwin.
Russell, B. (1912) *The problems of philosophy*. New York: Oxford University Press (Galaxy Edition, 1959).
Skinner, B.F. (1972) *Beyond freedom and dignity*. New York: Bantam/Vintage.
Sommerhoff, G. (1974) *Logic of the living brain*. London/New York: Wiley & Sons.
Spencer Brown, G. (1969) *Laws of form*. London: Allen & Unwin.
Taylor, R. (1950a) Comments on a mechanistic conception of purposefulness, *Philosophy of Science*, 17, 310-317.
Taylor, R. (1950b) Purposeful and non-purposeful behavior: A rejoinder, *Philosophy of Science*, 17, 326-332.
Wheelright, P. (1935) Introduction. In *Aristotle* (Translation by Wheelright), Indianapolis: Bobbs-Merrill.