Action Principles and Teleology

Michael Stöltzner

1. Introduction

In dealing with action principles, many textbooks on variational calculus or mechanics add a digression on history or even on philosophy. This is a remnant of a long tradition of discussion among physicists and philosophers, mainly on the so-called "Principle of Least Action". One must admit that this subject is not among the hottest topics in the contemporary debate on the foundations of physics. It is moreover a common opinion that the philosophical notions mentioned in connection with the principle, teleology and final causes, should be kept out of physics. A very decided statement of this spirit can be found in Ref. [1] (p. 155) as the conclusion of an entire book on variational principles: "Hence the teleological approach in exact science can no longer be a controversial issue; it is not only contrary to the whole orientation of theoretical physics, but presupposes that the variational principles themselves have mathematical characteristics which they *de facto* do not possess." For the authors variational principles are a mere reformulation of the equations of motion, which is physically equivalent to them.

A century ago, however, the Principle of Least Action was capturing the attention of many great physicists. Helmholtz in 1886 considered it as the leitmotif in striving for a unified theory of the physical world: "Already now it can be considered as highly probable that it is the universal law of all reversible processes in nature ..." ([2], p. 209ff.). After its successful application to electrodynamics and relativity, Max Planck in 1915 felt even more right to judge: "Amid the more or less general laws which mark the achievements of physical science during the course of the last centuries, the principle of least action is perhaps that which, as regards form and content, may claim to come nearest to that final aim of theoretical research" ([3], p. 68). The original proclamation by Maupertius in 1746 sounded even more exalted: "... a principle so wise, so worthy of the Supreme Being, a principle, to which nature seems to be constantly attached" ([4], p. 205). We shall see in what follows that this reference to God was not a mere formula, but shows one fatal element prevailing in the birth of the action principle.

In this contribution I do not intend to pursue the old line of reasoning, searching or discrediting teleology in physics by means of the action principle. The aim is rather to introduce action principles as a somewhat different way of looking at a physical problem than the equations of motion or the field equations imply. Whereas the latter focus on the local aspect of dynamics

Springer Series in Synergetics, Vol. 63 Inside Versus Outside
Editors: H. Atmanspacher, G.J. Dalenoort

an integral principle like the action principle searches for a global view on the problem and expresses a level of higher structure. Planck saw a smack of teleology in the fact that "the actual motion at a certain time is calculated by means of considering a later motion" ([3], p. 71). Often the variational method of determining the actual motion between two points is described anthropomorphically. The particle is said to already know its goal and to decide on the path of least expense. It "smells" or "looks at" other possible trajectories [5]. This view, however, presupposes a concept of teleology much poorer than that introduced by Kant, which directs science towards systematization and not a particle from the future. Instead of adhering to such a sort of pseudo-intentionality, the question of what directs the motion in the small might find an answer in the concept of chance that statistical theories had introduced during this century.

To study the prospects that a teleological look at action principles can still offer us today we are forced to inquire thoroughly what the misunderstandings between philosophy and physics were in the 18th century.[1] I shall try to divide with the help of Kant the things mixed up and ask how his concept of teleology as a regulative principle could help in interpreting some applications of variational principles. In order to show that the well-defined variational principle embraces more than the local view of the equations of motion, I will discuss cases that are usually neglected in textbooks on Langrangian and Hamiltonian mechanics. They will show the significance that action principles can still possess in statistical theories and the impact this could have on the notion of teleology.

2. The Failure of the Teleological World-View

2.1 The Best of All Possible Worlds – Classical Philosophical Teleology

The term *teleology* was introduced by Christian Wolff in his *Philosophia rationalis sive logica* in 1728 to define that part of philosophy not dealing with the efficient causes of natural things, but with the final causes considering them in the light of their purposes. Wolff here reflects a classical distinction of medieval scholastics that knew of four different *causae*: *causa efficiens*, *causa finalis*, *causa materialis* and *causa formalis*.[2] He extensively applies the new method, for instance when he speaks about the human eye. He first shows how its parts are interrelated and in what way they contribute to the function of the entire organ. Then he investigates the advantage of the spherical shape of the eye by assuming other possible forms and concludes that

[1] Here we more or less follow the excellent comprehensive study of Schramm [4].

[2] Roughly speaking, *causa efficiens* was that which brought things into existence; *causa finalis* defined the purpose for which something exists; *causa materialis* signified the material basis; *causa formalis* characterized the design of matter.

they would fail to fulfil the purpose of the eye so well. One can clearly see in Wolff's analysis two characteristic features of teleological arguments. Firstly, there exist relations of purpose among the parts in an organism and between the parts and the whole. Secondly, he uses hypothetical other worlds to show that the one realized is the best solution among the possible ones.

Material inner purposes – as one could call the first element – were widely used by Aristotle in his biology, for instance in *De partibus animalium* (On the parts of the animals). Such directly comparative arguments survived independently of philosophical teleology until modern biology was born with Darwin. But already Aristotle had found problems like the difference of sexes which could not be approached that way. At these places he applied *formal comparative teleology* showing that the existence of the difference is *better*. Here he refers to the notion *τὸ βέλτον* introduced by Plato in *Phaidon* (96a6-98b6), where Socrates asked Anaxagoras to show him whether the earth is spherical or plane by showing him which alternative is the better one.

Plato antithetically considers just given possibilities. Since Galileo, the newly emerging natural science used the notion of the natural law with its categorical and general validity. This idea, quite different from Greek thinking, opened up new possibilities in comparing phenomena. Newton's law of gravity allowed the construction of infinitely many solar systems. Why our solar system is the one actually realized now became a question of teleology, which turned into a variation of worlds.

Richard Bentley clearly states the method:

> ...we ought to consider every thing as not yet in Being; and then diligently examine, if it must needs have been at all, or what other ways it might have been as possibly as the present; and if we find a greater Good and Utility in the present constitution, than would have accrued either from the total Privation of it, or from other frames and structures that might have been as possibly as It: we may then reasonably conclude, that the present constitution proceeded neither from the necessity of material Causes nor the blind shuffles of imaginary Chance, but from an Intelligent and Good Being, that formed it in that particular way out of choice and design. ([6], p. 361)

The end of this quotation already contains the seed for the subsequent self-destruction of general teleology. From considering the disastrous implications on life that a highly eccentric orbit of the earth would have, Bentley inferred the necessity of a Supreme Being who had prevented that and put all so well. By their very structure there was no control of those arguments. Everything could be varied, judged best in some sense and used as a physicotheological proof of the existence of God. An enormously widespread trivialisation took place and the inflation of existence proofs devalued teleology.[3] So one can resume with Schramm ([4], p. 40): "It is not surprising that teleology as an embracing method of explanation, in which theology, philosophy and natural

[3] A deterring example is William Derham's *Physico-Theology, or a Demonstration of the Being and Attributes of God from his Works of Creation* (London 1713).

sciences felt united, came to an end, but that this lasted so long." It was the tradition of extremal principles in mathematics and optics that prolonged the life of teleology for another generation because it opened up the possibility of a formal teleology on a solid mathematical basis.

2.2 Extremal Principles in Optics

In the antique world it was a widespread belief that nature always acts with the least necessary effort. This so-called *lex parsimoniae* found a first useful application in natural science in Hero's derivation of the law of reflection (see Fig. 1). He supposed that the light takes the shortest way from the light source to the eye. By introducing a virtual image of the light source in P' he could use the argument that the straight line ORP' is the shortest path connecting O and P' and thus showed by the obvious equality of PR and $P'R$ that the actual ray goes PRO. Hence the angle of incidence equals that of reflection. Hero's law could be extended by introducing a tangential plane to the case of a spherical convex mirror. The spherical concave mirror was tacitly neglected already by the classical authors, because the light path can become maximal there.

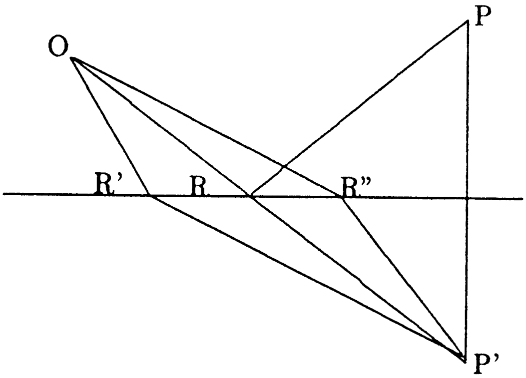

Fig. 1. Hero's derivation of the law of reflection.

Snell's law of refraction was interpreted by René Descartes as a statement on light rays. Descartes even proposed a derivation assuming that light moves by a constant factor faster in the denser medium. The last point was doubted by Pierre de Fermat, who succeeded in reducing the law of refraction to a minimum principle. Supposing that the velocity of light is inversely proportional to the refraction index, one arrives at *Fermat's principle of least time*:

$$\int n ds = \int \frac{ds}{v} = \int dt \stackrel{!}{=} \min \tag{1}$$

2.3 Calculus of Variations

Under this notion, Leonhard Euler presented in 1756 a new method that J.L. Lagrange had invented shortly before.[4] Since the last century the term has been used in a more general sense indicated by the title of Euler's famous book *The art of finding curved lines which enjoy some property of minimum or maximum.*[5] There he studies the extremal properties of functionals, i.e., functions whose argument is again a function, by comparing them with test-functions that fulfil certain isoperimetric conditions. Such problems reach back even to the ancient Greeks. Zenodorus (around 200 B.C.) proved that the area of the circle is greater than that of any polygon of equal perimeter.

Although one can already find such isoperimetric methods in Newton's works and in Huygens' *Treatise on Light* the true beginning of variational calculus is marked by the brachistochrone[6] problem that John Bernoulli had posed in June 1696 and repeated in 1697. He encourages the "most ingenious mathematicians of the earth" to find the curve of fastest descent between two points in the gravitational field. His own solution, published in May 1697, uses the analogy with geometrical optics and considers successive refractions in thin layers of decreasing density. He relates the refraction index to the velocity the body would have at that height. He explicitly refers to Huygens and uses the techniques introduced by Fermat. His brother James Bernoulli was furthermore able to develop out of his own solution quite general methods for isoperimetric problems. So James Bernoulli could pose a problem to his brother which already reflected the general situation. Find the curve $q(t)$ of a given length between two points A and B minimizing or maximizing $\int f(t, q(t), \dot{q}(t))\, dt$, where f is an arbitrary continuous function. His solution supposes that the extremal property of the entire curve is also valid for an arbitrary small piece. So he can approximate the curve by polygons. By passing to the limit he obtains a differential equation from which he determines the nature of the curve. James Bernoulli's techniques became the nucleus of Euler's work that constituted the calculus of variations – in the broader sense – as an independent discipline of mathematics.

---

[4] "Calculus variationum est methodus inveniendi variationes quantitatum utcunque ex binis variabilis x et y conflatarum, quas patiuntur, si relatio inter x et y proposita infinite parum quomodocunque immutetur."(Lagrange quoted from [7], p. 93.)

[5] Euler: Methodus inveniendi lineas curvas maximi minive proprietate gaudentes sive solutio problemetis isoperimetrici latissimo sensu accepti (ed. by Constantin Carathéodory, Bern 1952).

[6] From Greek $\beta\rho\alpha\chi\iota\sigma\tau\acute{o}\varsigma$, short.

2.4 Minimum or Maximum? – The Lethal Question for Teleology

The mathematical methods had reached a considerable level of sophistication, so that one could also hope to successfully attack the problems of celestial mechanics. As the starting point for planetary motion Euler considered in 1744 the motion of projectiles in a medium without resistance. He was convinced "that all natural effects follow certain laws of the minimum or maximum... But what the property is, does not seem to be easily determinable a priori from metaphysical principles" ([4], p. 201).

Pierre Moreau de Maupertius, the first president of the Berlin Academy of Sciences took the plunge to derive on metaphysical grounds the quantity to be minimized. He calls $\int v ds$ the *action*[7] and states that it has to be minimized in all processes in nature.[8] In his first Berlin work *Les loix de mouvement et du repos déduites d'un principe métaphysique*[9] he focusses again exclusively on the minimum of action. He explicitly quotes Euler's work ignoring Euler's precaution in speaking about minimum or maximum. Maupertius extends his principle beyond physics even to the growth of plants.

The fatal point is that Maupertius does not content himself with finding it "worthy of the Supreme Being", but turns it into a proof of the existence of God: "One ought to realize that all laws of motion and of rest are based on the principle, appropriate to the highest degree, to see that they owe their establishing to an omnipotent and omniscient Being" ([4], p. 205). Although he had undertaken in his work a severe criticism of the outrageous physicotheology of Derham and others, he himself payed the price for the outgrowths.

It was very easy to attack Maupertius' concrete examples because they speak more against his principle than they support it. For instance in his discussion of scattering he does not really use the action principle but energy conservation. In 1751 Samuel Koenig en passant mentioned in a reply that Maupertius' quantity is the same as that already stated by Leibniz in a letter to the mathematician Jacob Hermann. This denial of priority and, moreover, Koenig's attempt to bring the Newtonian Maupertius on Leibnizian grounds provoked a fierce reaction by the president, which initiated a publicistic war between followers of Leibniz and of Newton, Voltaire and his former friend Frederick the Great. It culminated in Voltaire's *Diatribe du Docteur Akakia Medicin du Pape*, which by the order of the king was burned in public by the executioner on Cristmas Eve 1752. In the guise of the inquisitor, Voltaire accused Maupertius of having proven the existence of God from a formula –

---

[7] I will speak of *action principles* if physical problems are concerned, and of *variational principles* if the general method is meant.

[8] Sur l'accord de differentes lois de la nature qui avoient jusqu'ici paru incompatibles. Submitted to the Paris Academy of Sciences on April 15, 1744.

[9] Histoire de l'Académie Royale des Sciences et Belles Lettres, année 1746, Berlin 1748, pp. 267–294.

actually the lever law, one of Maupertius' examples. In this struggle teleological arguments finally lost all public credibility.

Let me try a first philosophical interpretation of why Maupertius' arguments had to fail. He saw that objective material purposes were not a good fundament for universally valid philosophical principles because obviously there were many purposes in natural objects and many things answering different purposes mutually excluding each other. So he tried to replace Bentley's general method of teleology by a formal principle. Insisting on the minimum he nevertheless took it as a substantial property of nature realized in all its actions and effects. Thus he could be easily attacked with counterexamples.

3. Philosophical Teleology: Leibniz Versus Kant

3.1 Leibniz – The Very Existence of Forms

So far I have neglected the universal genius of the epoque, who took part in all aspects of our subject: Gottfried Wilhelm Leibniz. He presented a solution to John Bernoulli's brachistochrone problem, which was based upon the same ideas as that of James Bernoulli. In his correspondence with John Bernoulli he even defined the right quantity of action $vds = v^2dt$. But he considered it only for the actual path of a motion and not as a quantity to be varied and minimized. In 1682 he tried to make Fermat's principle acceptable for the Newtonians who had a particle theory of light by introducing the quantity nds as the *difficultas* of an optical medium. Its minimization should be characteristic of the actual path.[10]

a) In the *Tentamen Anagogicum – Essay anagogique dans la recherche des causes*[11] Leibniz concedes to the Cartesians that

> all natural phenomena could be explained mechanically if we understood them well enough, but the principles of mechanics themselves cannot be explained geometrically, since they depend on more sublime principles which show the wisdom of the Author in the order and perfection of his work. ([8], p. 272; [9], p. 478[12])

The last notion, the perfection in nature, is the key to metaphysical explanations of why one of those situations, which are geometrically of equal necessity was realized. For Leibniz the world is only physically or hypothetically necessary, not absolutely or metaphysically. But there is nevertheless a guiding principle followed by the Author: *perfection*.

---

[10]Unicum Opticae, Catoptricae, et Dioptricae Principium, Autore G.G.L. In: Acta Eruditorum, publicata Lipsiae, cal. Junii, anno 1682, pp. 185–190.

[11]For Leibniz *anagogic* is the investigation of the supreme cause. Aristotle had used αναγωγή (bringing up) to describe the process of reducing incomplete syllogisms to those of the first figure.

[12]I do not always follow Loemker's translation.

b) In *De rerum originatione radicali* of 1697 Leibniz justifies ontologically that the largest effect is reached with the smallest outlay. The being in nature is not maximized simply in number, but in essence. For him the essence belongs to the region of ideas which exists in God himself. Like in a game, in which we have to fill a maximal number of fields with stones according to certain rules, possible things are brought into being, for being involves more perfection than non-being.

> Hence it can indeed be clearly understood that out of the infinitely many possible combinations and possible series that one exists by which the maximal essence or possibility is brought into existence. There is evidently in all things a principle of determination which is derived from a maximum or a minimum, such that without doubt the maximal effect is achieved at the least expense, so to speak. ([8], p. 303, [9], p. 487)

c) The end of the quote shows that Leibniz is aware of the problems of demanding a pure minimality. In the *Tentamen Anagogicum* he tries to avoid ambiguity by introducing the concepts of the *simplest* and the *most determined*. "... there are cases where one must have regard for the most simple or the most determined, without distinguishing whether it is the greatest or the smallest" ([8], p. 270; [9], p. 484). What is meant by that he shows by analyzing reflection and refraction on arbitrary mirrors. To any varied path there is a twin path ($PR'O$ and $PR''O$ in Fig. 1) of equal length. The actual solution is now characterized by the disappearance of their difference. So the twins reunite in the unique solution, which is therefore the most determined path, and also a simple one – if one reads simple as unique. Compared to Maupertius the notion of the most determined has the advantage of being applicable to a spherical concave mirror, too. But uniqueness of the solution is not given in all variational problems. The shortest path on the earth's surface from one point to its antipode is infinitely degenerate, since every meridian does the job.

d) Leibniz reduced the brachistochrone to a differential equation. This method assumes that the extremality property of the whole solution is present in an arbitrary small piece, too. "For the best of those forms and figures is not only found in the whole, but also within every part and it would not be sufficient in the whole without it." This belief is rooted in Leibniz's ideas about the structure of the universe. "... the principle of perfection is not limited to the general but descends also to the particulars of things and of phenomena..." (both citations [8], p. 272; [9], p. 478).

e) Forms are thus for Leibniz not only modes of perception, but they really exist in nature.

> [If] nature were obliged in general to construct a triangle and that for this purpose *(effect)* only the perimeter or the sum of the sides were given, and nothing else; then nature would construct an equilateral triangle. This example shows the difference between architectonic and geometric determinations. ([8], p. 278; [9], p. 484)

Architectonical determinations belong to the realm of wisdom where everything is explained by final causes, whereas in the realm of power constituted by geometrical determinations everything is explained mechanically by efficient causes.

We find two levels in Leibniz's attitude towards final causes. Firstly, with regard to particular problems final causes work as heuristics. They can guide research successfully since they are founded on the metaphysical level of non-mechanical principles. How this method could be applied to gain new insight Leibniz tried to show with his notion of the *difficultas* in optics. He writes:[13]

> Therefore those are wrong who with Descartes reject final causes in physics, although except for the admiration of God they present the prettiest principle for finding (*principium inveniendi*) also those properties of the matters whose inner nature is not yet known to us so clearly that we could use the closest efficient causes and explain the mechanisms (*machinas*) which the Creator had applied to produce those effects and to obtain his purposes.

Secondly, Leibniz was convinced that the concept of the most determined was a principle actually governing nature, i.e., not only methodological but metaphysical. That it was only formal did not limit its validity, it was even therefore closer to the metaphysical principle of perfection. In *De origine* he writes that "everything in the world takes place in accordance with the laws of the eternal truths and not merely geometric but also metaphysical laws; that is, not merely according to material necessities but also according to formal reasons" ([8], p. 305; [8], p. 488). Those laws constitute "divine mathematics".

Leibniz needs God to realize the possibilities, which are present as essences on the level of ideas in God. These can be realized because in the essence of God his existence is already contained; His existence is metaphysically necessary. Here Leibniz stands on the same grounds of medieval scholastics like Descartes who needed God to ensure the existence of a world of things outside our *res cogitans*.

3.2 Kant – Teleology as a Regulative Principle

For Leibniz the Principle of Least Time was a formal principle firmly grounded in ontology, since forms were ideas and existed in God himself. Particularly this ontological status of ideas was subjected to the strong criticism of Kant's *Critique of Pure Reason*. It inquires into the conditions of the possibility of knowledge (German: Bedingungen der Möglichkeit von Erkenntnis).

a) *Understanding* (German: Verstand), our first faculty of knowledge (German: Erkenntnisvermögen), produces our experience (German: Erfahrung). Here intuitions (German: Anschauungen) and concepts (German: Be-

[13]From Nova Acta Eruditorum of June 1682, p. 185–190 (see footnote 10); quoted according to [4], p. 81 and 203.

griffe) meet and constitute empirical laws about nature. But there are conditions a priori for this, which are the necessary forms of all experience: the categories. They are found in the transcendental deduction, which shows the way how concepts are a priori related to objects. How things are in themselves is thus a question which can never find an answer in science. The second faculty of knowledge is *reason*:

> Reason is never in immediate relation to an object, but only to the understanding... It does not, therefore, *create* concepts (of objects) but only *orders* them, and gives them that unity which they can have only if they be employed in their widest possible application, that is, with a view to obtaining totality in the various series. ([10], A 643; B 671)
> Just as the understanding unifies the manifold in the object by means of concepts, so reason unifies the manifold of concepts by means of ideas. ([10], A 644; B 672)

Ideas never constitute concepts of objects, their employment is just *regulative*. They direct the understanding towards a certain goal, which is just a *focus imaginarius* beyond possible experience, but useful to unify our knowledge.

b) The link between understanding and reason is *judgement* (German: Urteilskraft):

> Judgement in general is the faculty of thinking the particular as contained under the universal. If the universal (the rule, principle, or law) is given, then the judgement which subsumes the particular under it is *determinant* (German: bestimmend). If, however, only the particular is given and the universal has to be found for it, then the judgement is simply *reflective* (German: reflektierend). The determinant judgement determines under universal transcendental laws furnished by understanding and is subsumptive only; the law is marked out for it *a priori*... ([11], A XXIII,XXIV; B XXV,XXVI; M 18)[14]

But reflective judgement can only give a principle from *and* to itself. For such a law cannot be taken from experience because the latter provides only particular laws without giving a rule of their generalization. Since reflective judgement has no legislative authority for nature, the central element of this principle is the *as-if-structure*.

> ... as universal laws of nature have their ground in our understanding, which prescribes them to nature ..., particular empirical laws must be regarded, in respect of that which is left undetermined in them by these universal laws, according to unity such as they would have if an understanding (though it be not ours) had supplied them for the benefit of our cognitive faculties, so as to render possible a system of experience according to particular natural laws. ([11], A XXV; B XXVII; M 19)

This structure leads to the concept of *finality* (German: Zweckmäßigkeit).

[14]I cite Kant in the usual way using the paging of the first and second edition. Unfortunately the translation of Meredith [11] does not do so, hence I also indicate his page numbers using the letter M.

> Now the concept of an Object, so far as it contains at the same time the ground of the actuality of this Object, is called its *end* (German: Zweck), and the agreement of a thing with that constitution of things which is only possible according to ends, is called the finality of its form. Accordingly the principle of judgement, in respect of the form of the things of nature under empirical laws generally is the *finality of nature* in its multiplicity. ([11], A XXVI; B XXVIII; M 19)

The principle of formal finality of nature is a transcendental, but only subjective principle a priori, a *maxim* of the reflective judgement. Among Kant's examples are the lex parsimoniae and the law that principles should not be multiplied without necessity.

c) Hence the formal finality in nature, that it makes up a system, is always given *for* our judgement and not in itself as a finality actually realized in nature. According to Kant the comprehensibility of nature is a precondition that we impose a priori to fulfil the interest of reason to get systematic knowledge. Its abstract elements were already discussed by Kant in the *Critique of Pure Reason*:

> The logical principle of *genera*, which postulates identity, is balanced by another principle, namely, that of *species*, which calls for manifoldness and diversity in things, notwithstanding their agreement as coming under the same genus, and which prescribes to the understanding that it attends to the diversity no less than to the identity. ([10], A 654; B 682)

So reason shows two competing interests, one of *extent*, where understanding thinks *under* its concepts, and one of *content*, where it thinks *in* them. So the transcendental principle of homogeneity, which tries to unify in a few simple laws gets its counterpart in the principle of *specification*.

d) The subjective finality of nature presents us this system of experience *as if* all particular experiences were shaped for our judgement. In the *Analytic of Teleological Judgement* this as-if structure is interpreted as a formal mean-purpose (or mean-end) relation. This suggests that one might also submit the things in nature to such a relation. In a problematic sense we are thus allowed to think nature teleologically and suppose its *objective finality*. But it is essential that this does not introduce a new type of causality into natural science because finality is not a constitutive principle of the determinant judgement, but a regulative principle of reflective judgement.

e) Kant divides objective finality into a formal and a material one. Formal objective finality shows up in geometry:

> All geometrical figures drawn on a principle display an objective finality which takes many directions and has often been admired. This finality is one of convenience on the part of the figure for solving a number of problems by a single principle, and even for solving each one of the problems in an infinite number of ways. ([11], A 267; B 271; M II,7)

Even though it is objective this finality is not based on an end (German: Zweck). It is moreover a finality without any end, hence purely formal. This structure it shares with the beautiful in aesthetics. Whereas Leibniz had distinguished the equilateral triangle as the simplest and most determined, Kant focusses on the principal unity of all triangles. Are there no further determinations given – in Leibniz's terms: there is no further necessity – there is no need that teleology chooses among the equal alternatives. On the contrary, teleology directs us to see the unity of all the triangles which are described by Thales' circle. So its interest is the general and global structure of the problem. Since we know many geometries today, we would certainly deny the objective character of geometry.

f) Objective material finality is given if "we are only able to see uniformity in [the cause-effect] relation on introducing into the causal principle the idea of the effect and making it the source of the causality and the basal condition on which the effect is possible" ([11], A 275; B 279; M II,12sq.). This can be done either by regarding the effect as *relative* to some other object employing it for its purpose, or by considering it in its *inner* finality. Taking the former as an argument for the existence of God had finally discredited teleology in the 1750s. Inner finality takes a thing as if it were an art-product.

To perceive that a thing is only possible as an end, it is required that its form is not possible on purely natural laws, which are given by understanding, "but that, on the contrary, even to know it empirically in respect of its cause and effect presupposes conceptions of reason" ([11], A 281; B 284; M II,16 sq.). The only thing that can possibly fulfil this condition is an organism, "an organized natural product ... in which every part is reciprocally both end and means" ([11], A 292; B 296; M II,24). The parts of an organism are possible only in their relation to the whole. This inner finality in our perception of an organism, however, does not suffice to consider a thing a physical end (German: Naturzweck) in itself without outer causality. To achieve this, a

> second requisite is involved, namely, that the parts of the thing combine of themselves into the unity of a whole by being reciprocally cause and effect of their form. For this is the only way in which it is possible that the idea of the whole may conversely, or reciprocally, determine in its turn the form and combination of all parts, not as a cause – for that would make it an art-product – but as the epistemological basis upon which the systematic unity of the form and combination of all the manifold contained in the given matter becomes cognizable for the person estimating it. ([11], A 287; B 291; M II,21)

So we consider nature as technical, but only in an analogy *for* our judgement, because the parts have to produce the whole out of themselves.[15] An organism as a physical end is both an *organized* and a *self-organizing* being. The latter distinguishes the organism from a clock. It is interesting that even in Kant's time Blumenbach [12] had already found reorganization in animals. For Kant this "organization of nature has nothing analogous to any causality known to

[15] Aristotle had also emphasized this aspect in his Physic (B8, 199 b 26–30). He explains a physical end as if the art of naval architecture is already present in the wood forming itself into the ship.

us" ([11], A 290; B 294; M II,23). But the concept of a thing as intrinsically a physical end is of regulative validity only and does not introduce a vitalistic force.

g) From the organism we are led to the idea of a system of nature, of its subjective formal finality. This regulative idea is useful to extend the physical science, "yet without interfering with the principle of the mechanism of physical causes"([11], A 298; B 301; M II,28) or introducing a new causality. So we suppose that the idea of a system of nature is already present in its construction, such that we can investigate nature *architectonically*. We consider nature as a technical product, but in mere analogy with the organism, hence without making nature an intelligent being or setting another intelligent being as its architect. While Leibniz had used the idea of a creator to ground the architecture of nature, in Kant's view we think it as an intrinsic one and try to reconstruct the system like in a Gedankenexperiment for our knowledge. So the analogy with the technical use of reason gives us teleology as a rule "upon which certain natural products are to be investigated ([11], A 305; B 309; MII,34).

Restricting ourselves to this regulative use, the antinomy of reflective judgement dissolves. We cannot decide whether nature as a whole is reducible to mechanics or not. But for our knowledge, which takes both thesis and antithesis only as maxims, there is in fact no contradiction. We cannot explain organisms mechanically, but on the other hand we have no chance ever to understand nature's architectonic structure.

h) The *Critique of Judgement* does not content itself with dissolving the antinomy, but aims at the unification of mechanism and teleology. So philosophy has to search for an *end of nature*. It is found in man, not in his happiness (German: Glückseligkeit), since this would be conditioned by nature fulfilling or preventing it, but in his "aptitude for setting ends before himself at all... The production in a rational being of an aptitude for any ends whatever of his own choosing, consequently of the aptitude of a being in his freedom, is *culture*" ([11], A 386 sq.; B 391; M II,94).

Through his free will man is the creator of culture and a subject of morality. Judgement has thus achieved to bring the theoretical back to the practical philosophy and closes the building of the three Kantian critics.

To sum up the points of Kant's teleology[16] essential for this paper: Kant gave teleology a more modest, but stable foundation as a regulative principle. It is no more a causality determining special facts but a principle of systematization, either in an organism as an independent entity or towards a final structure in our empirical laws. While Leibniz saw the most simple in the unique (most determined), for Kant the general description, like the method of constructing all triangles, is the simpler form. This is consistent with the

[16]In what follows I will embrace all types of finality mentioned by Kant in the term teleology. As Engfer [13] remarks, Kant's terminology in the Critique of Judgement is not completely strict.

significance of the simple unifying law that the physics of our century always had in mind. But the structure is not finished with the greatest homogeneity, as the above citation (in (c)) from the *Critique of Pure Reason* had shown. We need specification and continuity as further principles. So if teleology is applied to a problem in natural science, it does not seek a single cause, like an efficient cause, but looks more globally at the problem. It is interested in its global structure more than just in determining a solution.

4. Some Mathematical and Physical Examples

Kant's *Critique of Judgement* was first published in 1790 two years after Lagrange's *Mechanique analytique* of 1788. At that time the philosophical and mathematical discussions were already separated, and they remained so in the course of the entire 19th century, when great progress was achieved in mathematics and theoretical mechanics on the fundament laid by Euler and Lagrange.

I do not aim to give a history of mathematics in this paper and refer to the literature. In this chapter I want to show that variational principles express a more global view than the mere differential equations. This means on the other hand that a well-defined action principle is stronger than the Euler-Lagrange equations derived from it.[17] The sufficient conditions for a solution of a variational problem are mainly ignored by physics textbooks. A second point to be mentioned is the structure of the varied curves. In the third place I shall briefly discuss what conclusions can already be drawn on the level of an action given. The fourth and the fifth example will show that defining an action in some situations selects a particular theory that is not expressed at the equation level. With the Feynman path integral I will not present an exceptional case but touch the very problem of what the notion of competing trajectories or variations could mean in a theory where the concept of a path familiar from mechanics becomes at least problematic. In the last case, Eigen's hypercyclic theory of evolution, there is no variational principle in the strict sense at hand, but the problem nevertheless shows features typical of formal teleology. I am hereby aware of the danger of tackling subtle points without having laid out a broad basis for the subject.

4.1 Necessary and Sufficient Conditions

Let us consider a twice continuously differentiable function $L(t, q(t), \dot{q}(t))$ on a compact interval $I = [a, b]$. By defining the norm of q as $\| q \| = \sup_{t \in I}\{|q(t)|, |\dot{q}(t)|\}$, we obtain the Banach space of continuously differentiable functions $q : I \longrightarrow \mathbb{R}$. As the varied curves (the potential trajectories) all

[17] There are certainly cases, for instance in two-dimensional field theory, where it has not yet been possible to define an action.

$q \in E_1(I) = \{q = C^1(I) \mid q(a) = q_S, q(b) = q_E\}$ are allowed, which coincide at the ends. The curves to be compared with each other are constructed by variations $h \in E_1^0(I) = \{h \in C^1(I) \mid h(a) = h(b) = 0\}$. Obviously $q + h \in E_1$, too. By this construction it is tacitly assumed that time is not varied. One can extend differential calculus with some modifications to Banach spaces. To search for extrema in normal calculus one considers the first derivative. One defines the first variation of W as the Fréchet derivative DW:[18]

$$DW[q](h) = \int_a^b [L_q(t)h(t) + L_{\dot{q}}(t)\dot{h}(t)]\, dt. \tag{2}$$

Integrating by parts and recognizing that $\int [.]h(t)dt = 0 \Longrightarrow [.] = 0$, since h is arbitrary (lemma of DuBois-Reymond), we get the *Euler-Lagrange equations* as a necessary condition for the action to be minimal:

$$L_q(t, q(t), \dot{q}(t)) = \frac{d}{dt} L_{\dot{q}}(t, q(t), \dot{q}(t)) \tag{3}$$

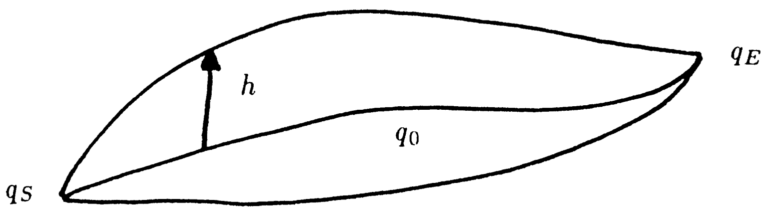

Fig. 2. Variations of the curve q.

The generation of Lagrange had believed that the Euler-Lagrange equations always provide the solution of the variational problem. A famous counterexample of Weierstraß showed that this was in fact not the case.[19] For the existence of a local minimum of the variational problem it suffices that D^2W is strictly positive. Turning this condition into one for L allows us to formulate the

Principle of Least Action: If a trajectory $q_0 \in E_1(I)$ satisfies the Euler-Lagrange equations (3) and the following conditions:

1) $A^{(q_0)} \in C(I, \mathbb{R}), B^{(q_0)} \in C^1(I, \mathbb{R}), \quad A = L_{qq} - \frac{d}{dt}L_{q\dot{q}}, \quad B = L_{\dot{q}\dot{q}}$,
2) $B^{(q_0)} > 0$ for all $t \in I = [a, b]$,
3) $A^{(q_0)}$ and $B^{(q_0)}$ satisfy the *Jacobi condition* on $[a, b]$, i.e., q_0 solves the Jacobi equation $\frac{d}{dt}[B\dot{q}] + Aq = 0$ and vanishes nowhere in $[a, b]$,

[18] With the abbreviations $L_q = \frac{\partial L}{\partial q}(t, q(t), \dot{q}(t))$ etc.

[19] See for example [2], p. 79.

then the action functional $W[q] = \int_I L(t, q, \dot{q})\, dt$ has a strict local minimum for the trajectory q_0.

A zero of q_0 in $(a, b]$ would be called a *conjugate point* to a with respect to the Jacobi equation. Since $q = 0$ also solves the Jacobi equation, there would be at least two curves passing through a and its conjugate point. Exactly this happens in the geodesic problem on the sphere. The antipode is a conjugate point to the starting point. Therefore the variational problem loses its uniqueness. This underlines the fact that we must make global assumptions about the structure of the field of curves to ensure the validity of the Principle of Least Action.

4.2 The Rolling Ball

I have already mentioned above that it is important to construct the varied curves properly. In this section I would like to shed more light on this issue from a physical point of view and show two different ways of relating the actual and the varied curves. In the preceding section I had already tacitly introduced Hamilton's principle with $h = \delta q$

$$\delta \int L dt = \delta \int (T - V)\, dt = 0 \tag{4}$$

where L is called the Lagrangian, $T = p^2/2$ is the kinetic energy, and V the potential of the forces acting on the system.[20] The essential feature of variations according to Hamilton's principle is that time is not varied, i.e., $\delta t = 0$. Two related points are reached by a moving particle at the same moment. Yet there is another possible identification. Demanding that both points should have equal total energy the varied path is allowed to take arbitrary time. Considering the general variations equivalent to D'Alembert's differential principle, it can be shown that by this identification we arrive at the principle of minimal action as formulated by Maupertius and Euler:

$$\int \delta(T dt) = \delta \int T dt = 0 \tag{5}$$

This formulation allows only time-independent potentials. It does not lead to the Euler-Lagrange equations. Instead, one uses the energy condition to derive the equations of motion.

Another aspect is the physical characterization of the varied curve. Here constrained systems can show a surprising property. Constraints are physical conditions that a system or a single point particle has to obey. We have met

[20]Usually V depends only on the coordinates, but it is also possible to define L for potentials depending on time or velocity like in electrodynamics. The integral is then correctly called *Hamilton's principal function*, but the word *action* has become common in this case, too.

them already in the isoperimetric problem (Sec. 2.3). As an example one could confine a particle to stay on the surface of a balloon smoothly distended. In this case the constraints can be put in the form $\Omega_i(\mathbf{x_1}\dots\mathbf{x_n};t) = 0 \quad (i = 1\dots r)$. Such constraints are called *holonomic.* The constraints describing a ball of radius a rolling on a plane are not integrable to the form given above. With x, y as the coordinates on the surface and θ, ψ, ϕ as the Euler angles[21] they read:

$$dx - a\sin\psi d\theta + a\cos\psi\sin\theta d\phi = 0 \tag{6}$$

$$dy + a\cos\psi d\theta + a\sin\psi\sin\theta d\phi = 0 \tag{7}$$

There are two ways to handle these constraints, of which only the first yields a correct result. If one takes the variations as virtual displacements, i.e., one only considers pure rollings, one obtains a three-dimensional manifold of possible motions. The constraint is only applied *after* the variation in order to get the equations of motion and reduces the dimension of possible motions to one.[22] Allowing *from the very beginning* only those variations that lead to curves obeying the constraints yields a different result. Curves solving this minimal problem form a higher-dimensional manifold.

Since we had to plug in the constraints after the variation to get the right equations of motion, the varied curves are no longer of the same type as the actual motion. So one obtains a pure rolling after comparison with other motions including slipping. Hölder ([15], p. 126) states the problem of interpretation:

> By other reasons we all along have got used to conceive the Principle of Least Action and Hamilton's principle only such that the variation of an integral [...] is set to zero. By that the name of the Principle of Least Action, however, does not go well any more together with the content.

Obviously an interpretation that conceives the actual curve as the best in some sense or as the most determinate, because it has the property fulfilling the constraints, would be circular. As the right method has shown, the constraint is put in after the variation and thus it distinguishes the actual path from the others. It is not specified naturally by the variational calculus itself. For the rolling ball the embedding into a systematic approach as Kant's subjective formal finality intends is achieved by modern geometry which provides the concept of a principal bundle. The space of possible motions of the ball is no longer a simple product of its centre of mass motion and its rotation.

[21] I define them as done by A. Budó: *Theoretische Mechanik*, Berlin 1980.

[22] A generalized version of the principle of least action or Hamilton's principle could be used equivalently.

4.3 Symmetry Considerations

A century ago, any study of a physical problem was centred on the equations of motion. At the time of Planck the study of the action functional itself became useful to gain global insight into the structure of the problem without solving the equations of motion. An important step forward was made by Emmy Noether, who proved that if the Lagrangian of the system was invariant under a one-parameter group of transformations, one automatically obtains conservation laws. I will sketch Noether's theorem in field theory. There are basically two types of symmetries. Firstly, if L is invariant under space-time translations or global rotations Noether's theorem yields energy-momentum conservation or the conservation of the total angular momentum. On the other hand the Lagrangian may exhibit certain inner symmetries, invariances under the action of gauge groups. For instance a complex scalar field with

$$L = (\partial_\mu \phi)^* (\partial^\mu \phi) - m^2 \phi^* \phi \tag{8}$$

is invariant under transformations of the form $\phi(x) \to \mathrm{e}^{-ie\alpha}\phi(x)$. This invariance leads to a conserved current.

Modern elementary particle physics is to a great extent the theory of gauge fields. They mediate the fundamental forces in nature. The standard model is specified by a certain gauge group and a Lagrangian invariant under it. Using the machinery of group theory it is possible to classify all experimentally known particles. Various interactions can be introduced already on the Lagrangian level. A large variety of generalizations of the standard model can be formulated that way.

4.4 Nonuniqueness of the Lagrangian

It is a widespread belief that in mechanics the equations of motion define the Lagrangian, from which they are derived, up to a total time derivative. That this is not true is shown by the following example, about which I learned from G. Marmo [16]:

$$L_1 = (\dot{q}_1 \dot{q}_2 - q_1 q_2) + \frac{1}{2}(\dot{q}_3^2 - q_3^2) \tag{9}$$

$$L_2 = \frac{1}{2}(\dot{q}_1^2 + \dot{q}_2^2 + \dot{q}_3^2 - q_1^2 - q_2^2 - q_3^2) \tag{10}$$

The Euler-Lagrange equations give the same equations of motion $\ddot{q}_k + q_k = 0$ ($k = 1, 2, 3$), but the difference $L_1 - L_2 \neq \frac{d}{dt}G$ is not a total time derivative. Both Lagrangians show angular momentum conservation, but the groups yielding this by Noether's theorem are not identical. The angular momentum in the second case is the generator of the ordinary rotation group $SO(3)$, but in the first case it generates the three-dimensional Lorentz group $SO(2,1)$, which is non-compact. The two Lagrangians give different quantum theories.

While for L_2 the spectrum of the angular momenta is purely discrete due to the compactness of $SO(3)$, there is always one direction with a purely continuous spectrum for the quantum theory obtained from L_1. Hence different Lagrangians, and in consequence different actions, really have a different physical meaning, even though their classical limit leads to the same equations of motion. This difference can be characterized by a global property, the symmetry group.

4.5 Surface Terms in General Relativity

To give a simple example in which the effects of the boundary cannot be neglected, consider an action:

$$S[\phi] = \int \phi \,\Delta\, \phi \qquad (\delta\phi)_{\partial\Omega} = 0. \tag{11}$$

Its variation and integration by parts yield

$$\begin{aligned}\delta S &= \int \delta\phi \,\Delta\, \phi + \int \phi \,\Delta\, \delta\phi = \int \delta\phi \,\Delta\, \phi + \int \nabla(\phi\nabla\delta\phi) - \int \nabla\phi\nabla\delta\phi \\ &= 2\int \delta\phi \,\Delta\, \phi + \oint_{\partial\Omega} \phi\nabla(\delta\phi) - \oint_{\partial\Omega} \delta\phi\nabla\phi \end{aligned} \tag{12}$$

where the last surface term again vanishes because of $\delta\phi\lceil_{\partial\Omega} = 0$. However, $\nabla\delta\phi$ does not vanish at the boundary. In order to obtain the Laplace equation $\Delta\phi = 0$ one has to subtract $\oint_{\partial\Omega} \phi\nabla\phi$ already from the action to be varied in order to cancel the first surface term. If one is just interested in the equations of motion this might seem harmless. However, in many situations one would like to ascribe a physical meaning to the action itself. For instance, in static solutions of Einstein's equations the action simply corresponds to the energy.[23] The Hamiltonian formulation of general relativity shows that the redefinition is not only a simple subtraction.[24] Speaking about a single object like a star or a black hole, one supposes the metric to be asymptotically flat. The same is done in the case of an open universe. The boundary conditions are replaced by the demand that the fields decrease fast enough at spatial infinity. In a closed universe, however, there are no cosmological surface terms, simply because it has no boundary. If one considers the modified Hamiltonian

$$H = H_0 + E[g_{ij}], \tag{13}$$

the usual Hamiltonian H_0 is zero if the field equations are satisfied. Hence the total energy of the gravitational field is the value of the surface term $E[g_{ij}] = \oint d^2s_k(g_{ik,i} - g_{ii,k})$. With that in mind, the redefinition of the action is not simply a technicality.

---

[23] Recall that the velocity of light is set dimensionless to 1.

[24] The following argument is due to Regge and Teitelboim [17]. Hamilton's principle is used here in the form $\delta\int dt(p_i\dot{q}^i - H) = 0$ over phase space.

Two conclusions can be drawn from the preceding example. Firstly, if the action itself is given a physical meaning, surface terms may be of importance and the right variational principle is no longer a mere tool to derive the field equation. Secondly, in cosmology the assumption as to whether the universe is open or closed is directly reflected in the definition of the action. Or vice versa, a well-formulated cosmological action principle already expresses the hypothesis chosen.

4.6 The Feynman Path Integral

As Noether's theorem opened up the perspective of a general analysis of physical models, group theory furthermore equipped us with a more abstract language for a modern formulation of quantum theory. The viewpoint of an initial value problem (differential equations and prescribed boundary values for their solutions) was replaced by the concept of the state of a system represented by a vector $\psi(q,t) = \mid q,t >$ and its time evolution $U_t = e^{-iHt}$ generated by a Hamiltonian $H(t,q,p)$. The experimental features described by quantum theory, such as atomic excitations, the decay of nuclei, or particle scattering are understood as transitions between different states. Instead of considering the motion of a particle from one point A in phase space to another point B, one is now interested in transition amplitudes (or probabilities) $< q',t' \mid q,t >$.

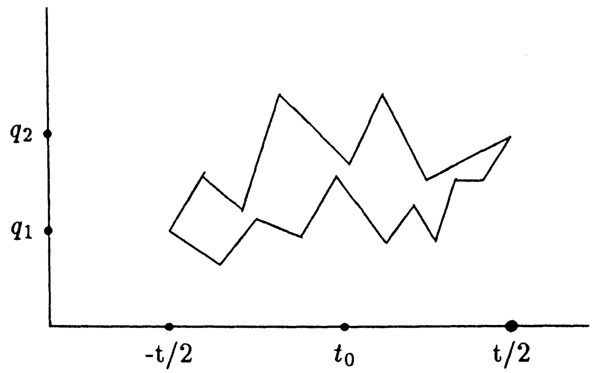

Fig. 3. Two paths of the path integral.

The double slit experiment already shows the germ of quantum physics. We are left with probabilities for possible paths and the question about the definite path, about a single event remains an unspeakable problem in the realm of conventional quantum physics. If we fix an initial and a final state we thus have to consider all possible paths connecting them and count them with a certain weight. It is, however, not only necessary to collect paths connecting discrete lattice sites in a sum as Fig. 3 suggests, but to integrate over a continuum of them. The usual argument of physics textbooks first considers transition amplitudes $< q_i, t_i \mid q_{i-1}, t_{i-1} >$ between N small time intervals and then passes to the limit obtaining the Feynman path integral:

$$< q', t \mid q, t_0 > = \lim_{N \to \infty} \int \prod_{i=1}^{N} dq_i \; e^{\frac{i}{\hbar} \int_{t_0}^{t} L(t', q(t'), \dot{q}(t'))\, dt} \qquad (14)$$

However, the Lebesgue integral does not exist in the infinite-dimensional case. In order to define the integration properly one has to find an appropriate measure. If one already includes the kinetic term, i.e., the free motion $H_0 = p^2/2$, in the measure, one obtains the so-called Wiener measure $dW^t_{q,q'}$, which is essentially constructed out of Gaussian distributions. For the definition of the Wiener measure it is not necessary to fix a second point such that all paths are touched. So for $H_0, V, H = H_0 + V$ essentially self-adjoint and bounded from below, one arrives via the Feynman-Kac formula at the Euclidian formulation:

$$(e^{-tH}\psi)(q) = \int \psi(q')\, dq' \int dW^t_{q,q'} \; e^{-\int_{-\frac{t}{2}}^{\frac{t}{2}} V(q(s))\, ds} \qquad (15)$$

In the mathematically well-based formulation we do not simply find the Lagrangian in the action principle, but the kinetic term and the potential play different roles. The classical path does not remain the only actual one any more. But it is still distinguished among all possible paths because its action is extremal. Since the action appears as a phase in the Feynman formulation, large differences of the order $\Delta S \approx \pi$ cancel out. Hence it is mainly the paths near the classical path that contribute to the path integral.

Looking at the path integral, the interpretation that a particle moves goal-directed, in the sense that it behaves as if it acted intentionally, breaks down. The problem turns out to be of fundamentally stochastic nature. Not intention but chance determines the path locally. It is, however, not the single path that counts, but the whole ensemble, for which a stochastic description is valid. One can still perform the classical limit of the path integral and regains the classical path. The ensemble of paths thus remains structured. In the case of the harmonic oscillator one can even split the integral into the classical path and oscillations around it. An appropriate view of the path integral formulation of quantum mechanics therefore contains a certain globality, not in the sense that we were forced to consider boundary terms, but that we ought to take it as representing an ensemble in the statistical sense.

I had interpreted teleology following Kant as the search for systematization. This concept stands in opposition to a world void of any structure. The dichotomy *structured* – *structureless* replaces the old one, as expressed by Bentley, between planning and chance. If we recognize the achievements of modern science that essential results are formulated statistically, we learn that chance does not lead to the negation of structure at all. In the next example I will try to show that under some circumstances it is in fact the only way to create new structure in self-organization processes.

4.7 The Hypercycle and Molecular Darwinism

In this section I want to discuss the model of hypercycles of Manfred Eigen and Peter Schuster [18]. The philosophical implications of their theory of *molecular Darwinism* have mainly been studied by Bernd-Olaf Küppers [19].

4.7.1 Teleo...? – Some Remarks Concerning Philosophical Misunderstandings

In philosophical statements of biologists *teleology* is largely avoided and has a negative connotation. As Robert Spaemann [20] points out this is due to the fact that it is still associated with Leibniz's understanding, or even the physicotheology of Derham and others. Hence biologists like Pittendrigh [21] tried to replace teleology by the concept of *teleonomy*, which defines finality as goal-directed by a program in the sense of information theory. Of course the discovery of the genetic code as a blue print for organisms is behind this concept. Also Monod [22] speaks about teleonomical processes in a way which does not contradict Kant's concept of teleology.

Küppers ([19], p. 33ff.) emphasizes that teleonomy is a *descriptive* concept for the finality found in organisms. It does not explain the cause of existence like the final end determines the shape of an artificial object. Explicitly quoting Kant in a reference ([19], p. 263ff.) even his examples for teleonomical structures, the eye, the cell, and biological macromolecules like hemoglobin, show that he is on the ground of Kant's concept of objective material finality. The only difference, however, consists in the plan coded in the DNA that guides the formation of these structures. The hypercycle will lead us back in evolution to the first formation of long DNA-chains themselves, i.e., to the creation of the plan. So I will step back to the two conditions Kant had given for objective material finality: the reciprocal part-whole relation and the self-organization of the organism.

Küppers calls theories proposing an irreducible principle that guides the formation of life teleological. One of his examples is Walter Elsasser's *creative selection*. This principle is neither expressible in mathematical terms nor decucible from chemistry or physics. I have shown above that the Kantian concept of teleology is not necessarily linked to anti-reductionism.

4.7.2 Darwinian Systems

There are three necessary properties for matter to yield Darwinian behaviour:

1. *Metabolism*: The system has to escape from entropy death by using free energy from its environment. Thus it must be sufficiently far from thermal

equilibrium and it must show both formation and degradation. Since the system is open, microreversibility between these processes is broken. Thus formation and degradation are independent of each other.

2. *Self-reproduction*: The structures of the system have an inherent ability of instructing their own synthesis. They show *autocatalytic* properties. The system constitutes a certain amount of information. Their self-copying is indispensible to prevent its loss in the steady degradation.
3. *Mutability*: Errors in copying provide the only source of new information. They are inevitable because of thermal or quantum fluctuations.

These three conditions are only necessary, not sufficient, for a definition of life. Even a crystal growing in a solution exchanges energy with it and is never free of defects. Transfering the conditions into a mathematical model one finds that it shows selective behaviour under limited resources:

$$\dot{x}_i = (A_i Q_i - D_i) x_i + \sum_{k \neq i} w_{ik} x_k + \phi_i \tag{16}$$

x_i is the respective concentration of a self-reproductive unit, i.e., a particular DNA sequence written as a vector. We will call each x_i a species. A_i is the velocity of its spontaneous formation, D_i of its degradation. The quality of copying $0 < Q_i < 1$ represents its mutability. The constraint of constant population $\sum_k x_k = c = const$ is regulated with the outside flux. It expresses that ressources are limited. Diagonalization leads to

$$\dot{v}_i = (\lambda_i - \overline{\lambda(t)}) v_i \qquad \text{with} \qquad \overline{\lambda(t)} = c^{-1} \sum_{j=1}^{N} \lambda_j v_j(t) \tag{17}$$

where $\lambda_i = \lambda_i(A_i, D_i, Q_i, w_{ik})$ is the selective value of the i-th quasi-species[25]. The growth rate $\dot{v}_i$ is positive for all v_i with selection parameter $\lambda_i > \overline{\lambda}$, the others die out. Thus $\overline{\lambda(t)}$ is adjusted to higher values and converges to the maximum $\lambda_{max} = \lim_{t \to \infty} \overline{\lambda(t)}$. The *fittest* finally gets selected. The ascent of $\overline{\lambda(t)}$ can also be expressed by an extremal principle of the form:[26]

$$\int_{t_0}^{t_1} \dot{\overline{\lambda}}(t) dt = \int d\overline{\lambda} \stackrel{!}{=} \max. \tag{18}$$

All equilibria reached by this maximum principle are metastable with respect to new mutations leading out of the sample of N species. If one supposes that selection is much faster than evolution, any new mutant with $\lambda_{N+1} > \lambda_{max}$ created at t_1 leads to new selection. So the extremal principle is in fact local,

[25] A quasi-species corresponds to a phenotype, a biological species, and contains a sample of genotypes, called species here.

[26] The mathematical argument given in [23], p. 69, shows that this is understood as a differential maximum principle and not in the sense variational principles are defined here.

it only gives the gradient from one local maximum to another one, but does not globally determine a path of evolution.

One can now ask how the quality of reproduction and the velocity parameters are related. Defining the selective advantage $\sigma_m = A_m/(D_m + \overline{E}_{k\neq m})$ (where $\overline{E}_{k\neq m}$ stands for the average excess of mutants) and fixing the quality of copying q_m, one finds the following qualitative behaviour for the system (16). Too high a percentage of errors destroys any information. If there are only few errors then evolution proceeds very slowly. The fastest evolution is reached just below the threshold of information stability, which is given by the maximum information content of a species:

$$\nu_{max} = \frac{\ln \sigma_m}{1 - q_m}. \tag{19}$$

4.7.3 The Hypercycle

Sequences of the order of 10 nucleotides can spontaneously form in a primordial soup. But for longer sequences the information threshold (19) puts a limit because the hydrogen bonds between nucleotides allow only an exactness $\overline{q} < 0.99$ which is not sufficient to build up sequences of length 100. This length corresponds to the simplest part of the cellular replication machinery, t-RNA. We have to construct a model that avoids selection pressure between short chains and allows cooperative behaviour. The simplest way to introduce a coupling between two self-replicative units is to add a quadratic term cx_ix_{i-1} as in spin models. Every unit couples to its nearest neighbours via some catalytic process, where the coupling to the (i–1)-th unit favours the i-th unit.

$$\dot{x_i} = a_ix_i + b_ix_ix_j - c^{-1}x_i \sum_{l\neq i}(a_lx_l + b_lx_lx_k) \tag{20}$$

If one builds up only a linear chain, one still has a selection of one of the ends. The idea of the hypercycle is now to close the chain by the conditions: $j = i - 1 + N\delta_{i1}, k = l - 1 + N\delta_{i1}$.[27] The quadratic terms induce hyperbolic growth, which is faster than the exponential growth of (16). This leads to a *once-and-forever selection* of one particular species. Whereas in the Darwinian system (16) every fitter mutant, which is not accidentally exterminated by a second mutation before it starts reproduction, can become dominant, the hyperbolic growth no longer allows this after a certain time.[28] The singularity, the infinite growth at a finite time, poses no problem because

[27] $\delta_{ab} = 1$ if $a = b$, $\delta_{ab} = 0$ otherwise.

[28] This could only happen if a sudden selective advantage of some orders of magnitude is supposed. Such a change, however, would contradict the continuity of parameters for the genotype within a phenotype (quasi-species). The latter in fact connects the species to the biological boundary conditions.

at some point – at latest at the cellular level – enzyme-induced compartmentalization takes place and stops the growth. It is followed by individualization leading to the protocell. The hypercycle is in one aspect non-Darwinian. It does not allow a variety of species, which represents an important element in Darwin's theory because it suppresses the selection of mutants in one of its components during the hyperbolic growth.

Why did I mention Eigen's evolution theory in our context? In the use of a variational principle the model (16) is not very spectacular. It describes a motion in a landscape of selective values, which are given like a potential in physics. From the biological point of view these values are boundary conditions for life. But as Küppers insists ([19], V.3.) life is not reducible to them because they are in some sense created together with the system. To this end they allow mutations on a larger timescale, which so far were not present in the N parameters λ_i of the system (16). Those are scrutinized in the sense of finality, every approach to the goal gives selective advantage. Eigen ([24], p. 1071) and Küppers ([19], p. 261) insist that only the differential advantage counts regardless of whether there is a goal or not. But in our conceptual framework we clearly observe that finality is used in a regulative sense to study how a given species could have come to exist. Due to the feedback between a biological system and its environment, it makes no sense to determine selective values for all possible DNA-sequences. We thus take the path from one λ_{max} to the higher one after a certain equilibrium. Locally the path is taken by chance, determined by the occurence of new mutations. Also the hypercycle is similar to variational constructions; the end, however, is not a point in an abstract space but a bit-limit to be reached. The origin of life, the evolution of the cell is a historical event, i.e. a singular path taken. If we believe in the hypercyclic model, we cannot leave the global point of view that shows the possibility of an origin of life from self-organization. We could redetermine the actual path taken only if we found special fossile remnants, thus changing the structure of the problem. So like in the case of the functional integral we are forced to consider an ensemble of paths, but the special structure and the unique classical path is lost.

5. Conclusions

How can we profit from the long march undertaken through philosophy, mathematics and physics? I have shown by examples the sense in which the variational principle is more general and more global than the field equations or the equations of motion. If it is well-defined it even contains more information. To a large extent this is due to the fact that it always comprises a certain structuring of the ensemble of possible paths. But the global interpretation may still be applied to problems, in which the statistical characterization of the ensemble, still present in the path integral, breaks down.

I have proposed to replace an anthropomorphic interpretation, which attributes goal-directed behaviour and intentionality to the particle, by the constructive role of chance. In Eigen's hypercycle chance is not only a statistical element that produces fluctuations, but is constitutive for the origin of life. There is an actual history of nature, but the lack of an archaeology of prebiotic time forces us to speak about all possible paths without distinguishing any particular path among them.

I have considered various variational principles in this article, and different actions have been constructed out of Lagrangians. In this wide applicability of the action principle, Helmholtz [2] and Planck [3] saw its qualification to be a universal law. Similarly, Kneser [25] considered its main value in its simplicity and indeterminacy. In our examples we have met the need always to specify carefully the quantity of the action by a Laplacian or otherwise, the boundary conditions assumed, and the variations allowed. Thus the action principle in general can be considered as a law only in a regulative sense because in order to become a definite law it needs further specification. So any program to express all physical laws in the form of a variational principle could at best be considered as a regulative idea. Success in explanation of physical phenomena decides on its usefulness.

However, there is nothing wrong with the fact that a candidate for a universal law is not a world formula from which everything follows by pure deduction. Even if we considered physics as the quest for such a unifying simple law, a theory of everything, we could justify this point of convergence only regulatively in the sense of Kant's subjective formal finality. As Stoeltzner and Thirring [26] point out, the pyramid of physical laws cannot be simply determined from the top, but needs careful investigation of all its interrelations. If we consider the contemporary candidates for a theory of everything and try to step down from the top we find a lack of initial conditions which are fundamental for the physics on the lower level. For instance, higher-dimensional cosmologies do not tell us which dimensions collapsed into internal degrees of freedom.

6. Epilogue: Endo/Exo-Physics and Teleology

In this last section I want to relate the concept of teleology outlined here to Primas' use of the word in the realm of the problem of endophysics and exophysics. The result will be twofold: Whereas the distinction of exophysics and endophysics has many features in common with the Kantian perspective I have taken, Primas tightly connects teleology to the temporal cause-effect structure and thus loses some powerful aspects of Kant's concept for his argument.

6.1 Regulative Principles

Primas ventures the working hypothesis "that quantum mechanics is an intrinsically holistic theory"([27], p. 7) that applies to all physical, chemical and biological processes. Since a complete description of the world is logically excluded by Gödel's theorem, this "leads to the necessity to distinguish between internal and external viewpoints" ([27], p. 14). According to Primas we need, in addition to the holistic endophysical theory, *normative regulative principles.* The step towards exophysics is performed by two types of symmetry breakings. Firstly, an observer is introduced as an abstract concept distinct from the endoworld. Secondly, the time-reversal symmetry of quantum mechanics is broken. I will postpone the second aspect to the next section.

How are Primas' normative regulative principles related to those of Kant? At least the starting point of Kant's critical philosophy is a totally different one. Its basis lies in the study of our faculties of knowledge. Inquiring into the process of gaining experience, it turns out that Kantian judgement necessarily structures the empirical laws according to regulative principles provided by reason. The two approaches seem to be maximally opposed. But one should avoid falling into an oversimplified dualism. The endoworld of Primas is "a theoretical construct"([27], p. 19) that has nothing to do with Kant's *Ding an sich.* It is axiomatically defined by a holistic quantum theory whereas the latter is absolutely undetermined. Once we adopt Bacon's *dissecare naturam,* i.e., the need for us to divide reality that is taken for granted by both Kant and Primas, regulative principles are necessary preconditions for the observer. According to Primas "an interpretation of a physical theory is characterized by a set of normative regulative principles"([27], p. 15). For instance an ontic interpretation suggests maximal symmetry as a principle for quantum endophysics. This example is very much in the spirit of Kant's subjective formal finality because symmetries give structure to our experience. For Primas there are further regulative principles in the exoworld that lead to epistemic interpretations. Here the analogy to Kant breaks down because even at the level of the observer, to whom we would like to ascribe Kantian understanding, we cannot escape the regulative principles and we have no schematism (like in determinant judgement) for understanding.

6.2 Time and Teleology

Primas intends to show that the Baconian rejection of teleology and final causes does not follow from the first principles (the endophysics) of quantum theory. Moreover, since they are time-reversal invariant, symmetry breaking leads to a necessary split because "in order to set apart cause and effect, temporally one-sided phenomena like irreversible processes are inevitable" ([27], p. 10). Hence a choice is made of whether the theory is backward deterministic and forward purely non-deterministic (Baconian) or forward deterministic and backward purely non-deterministic (teleological in Primas' sense). Primas thus ties teleology exclusively to the direction of time. It is therefore not surprising that he finds no good in a teleological interpretation of variational or minimal principles: "Curiously enough, some scientists thought they could find instances of finalistic causes already in classical Hamiltonian mechanics..." ([27], p. 9). Primas here opposes an interpretation in the pre-Kantian style, which lets the particle consider different paths, and regards the motion as predetermined by the future. I have tried to show in this article, why this picture is wrong and that teleology implies a structuring view on the variational principle, which is not in opposition to causality but supplements it. Hamilton's principle is the object of symmetry studies, which were examples for Primas' concept of regulative principles. Lagrangian formulations of field theories consider time just as a coordinate. In deriving Einstein's equation from the Hilbert action $\int R$ the entire metric is varied. Variational principles in general are not limited to closed systems. Only the principle of least action in the form $\int T dt$ assumes energy conservation. In quantum field theory only the combined reversal of charge, parity, and time is a symmetry (CPT-invariance).

Primas' project to go *beyond Baconian quantum endophysics* could in my view profit from the more general concept of teleology introduced here. Not only can no decision in favour of Baconian or teleological (forward deterministic) processes be made on the level of endophysics, but even in generic open systems both types are present. They might contain entropy-increasing (Baconian) or entropy-decreasing (teleological in Primas' sense) subsystems. Typical entropy-decreasing processes show the phenomenon of self-organization. Kant's concept of objective material finality emphasizes that self-organization and a reciprocal part-whole relation are intimately linked concepts. Pointing this out might be of help for a better understanding of larger systems and their open subsystems, since: "Nowadays we have the tools to classify open systems and to develop teleological descriptions in a conceptually sound and mathematically rigorous way" ([27], p. 29).

Acknowledgements. My interest in this field was initiated by Matthias Schramm during my studies in Tübingen. In spite of the distance he remained a most valuable partner in discussions. I thank Walter Thirring for his critical comments and the patience he had with my other work being delayed by this paper. To Piotr Bizon I owe thanks for pointing out to me the problem of surface terms in relativity. Karin Berndl, Michael Müger and Christoph Schweigert were critical readers in various stages of the paper's progress. During my studies I was supported by the Studienstiftung des Deutschen Volkes. Last but not least, Harald Atmanspacher's invitation to the conference on endo/exo-problems is responsible that the paper was written at that time.

References

1. W. Yourgrau and S. Mandelstam: *Variational Principles in Dynamics and Quantum Theory.* Pitman, London 1960 (second edition)
2. H. von Helmholtz: Ueber die physikalische Bedeutung des Princips der kleinsten Wirkung. In *v. Helmholtz' wissenschaftliche Abhandlungen*, Leipzig 1895, Vol. III
3. M. Planck: Das Prinzip der kleinsten Wirkung. In *Wege zur physikalischen Erkenntnis.* München 1944
4. M. Schramm: *Natur ohne Sinn ? – Das Ende des teleologischen Weltbildes.* Styria, Graz 1985
5. R.P. Feynman, R.B. Leighton, and M. Sands: *The Feynman Lectures on Physics, Vols. I, II.* Addison-Wesley, Reading (Mass.) 1964
6. R.A. Bentley: *A Confutation of Atheism (II)*, London 1693. Reprinted in *Isaac Newton's Papers and Letters on Natural Philosophy.* Edited by I. Bernard Cohen, University Press, Cambridge 1958
7. C. Carathéodory: The beginning of research in the calculus of variations. In *Gesammelte mathematische Schriften.* München 1954–1957, Vol. II, pp. 93–107
8. G.W. Leibniz: *Die philosophischen Schriften.* Ed. by G.J. Gerhardt, Berlin 1890 (reprint: Olms, Hildesheim 1961), Vol. VII
9. G.W. Leibniz: *Philosophical Papers and Letters.* A selection translated and edited with an introduction by L.E. Loemker, Reidel, Dordrecht 1969 (second edition)
10. I. Kant: *The Critique of Pure Reason.* Translated by N.K. Smith, Macmillan, London 1990
11. I. Kant: *The Critique of Judgement.* Translated by J.C. Meredith, Clarendon Press, Oxford 1991
12. F. Blumenbach: *Über den Bildungstrieb und das Zeugungsgeschäfte*, Göttingen 1781. Reprinted by Gustav Fischer Verlag, Stuttgart 1971
13. H.-J. Engfer: Über die Unabdingbarkeit teleologischen Denkens. In *Formen teleologischen Denkens. Philosophische und wissenschaftstheoretische Analysen*, Ed. by H. Poser, Berlin 1981
14. P. Blanchard and E. Brüning: *Variational Methods in Mathematical Physics.* Springer, Berlin 1992
15. O. Hölder: Über die Prinzipien von Hamilton und Maupertius. In *Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen. mathematisch-physikalische Klasse.* Göttingen 1896, Vol. 2, pp. 122–136
16. G. Marmo and E.J. Saletan: Ambiguities in the Lagrangian and Hamiltonian Formalism: Transformation Properties. *Il Nuovo Cimento* **40B**, 67–89 (1977)
17. T. Regge and C. Teitelboim: Role of Surface Integrals in the Hamiltonian Formulation of General Relativity. *Annals of Physics* **88**, 286–318 (1974)
18. M. Eigen and P. Schuster: *The Hypercycle – A Principle of Natural Self-Organization.* Springer, Berlin 1979
19. B.-O. Küppers: *Der Ursprung biologischer Information. Zur Naturphilosophie der Lebensentstehung.* Piper, München 1990
20. R. Spaemann: Teleologie und Teleonomie. In *Die Frage Wozu?*, ed. by R. Spaemann and R. Löw, Piper, München 1991, pp. 300–310
21. C.S. Pittendrigh: Adaption, Natural Selection, and Behavior. In *Behavior and Evolution.* ed. by A. Roe and G.G. Simpson, Yale University Press, New Haven 1958, pp. 390–416
22. J. Monod: *Le hasard et la nécessité.* Éditions du Seuil, Paris 1970

23. B.-O. Küppers: *Molecular Theory of Evolution*. Springer, Berlin 1985 (second edition)
24. M. Eigen: Wie entsteht Information ? – Prinzipien der Selbstorganisation in der Biologie. In *Berichte der Bunsen-Gesellschaft für physikalische Chemie* **80**, 1059–1081 (1976)
25. A. Kneser: Das Prinzip der kleinsten Wirkung von Leibniz bis zur Gegenwart. In *Wissenschaftliche Grundfragen IX*, ed. by Hönigswald, Teubner, Leipzig 1928
26. M. Stöltzner and W. Thirring: Entstehen neuer Gesetze in der Evolution der Welt. Submitted to *Naturwissenschaften*.
27. H. Primas: Time-Asymmetric Phenomena in Biology – Complementary Exophysical Descriptions Arising from Deterministic Quantum Endophysics, *Open Systems & Information Dynamics* **1**, 3–34 (1992)