Studies in History and Philosophy of Biological and Biomedical Sciences 43 (2012) 88–91

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

Introduction

On nature and normativity: Normativity, teleology, and mechanism in biological explanation

Lenny Moss [a], Daniel J. Nicholson [b]

[a] Department of Sociology and Philosophy, University of Exeter, Exeter EX44RJ, UK
[b] Konrad Lorenz Institute for Evolution and Cognition Research, Adolf Lorenz Gasse 2, Altenberg A-3422, Austria

1. Introduction

There is a spectre haunting contemporary Anglophone philosophy . . . and it is *not* teleology. The spectre is that of an intractable problem when it comes to reconciling two commitments that are each pervasively embraced by contemporary philosophers and yet ostensibly *ir*reconcilable, namely: (a) that there is no respectable alternative to some form of philosophical naturalism, and (b) that human life is saturated with norms in general and that philosophy itself is especially beholding to the norms of rationality in particular. One need look no further, for ample confirmation, than the recent collection edited by Mario De Caro and David Macarthur, entitled simply *Naturalism and Normativity* (2010). Featuring contributions by the likes of Hilary Putnam, Richard Rorty, T. M. Scanlon, Akeel Bilgrami, Huw Price, Paul Redding, Peter Godfrey-Smith and others, the common aspiration is clear: to avoid the Scylla of some form of reductive, normative eliminativism on the one side and the Charybdis of a dualistic *non*-naturalism on the other. Normativity, for these thinkers, is identified with the realm of the human, and the overarching strategy of the contributors is that of finding a 'naturalism' that is neither too reductively restrictive (to be able to countenance human normativity as natural) nor too permissively liberal (to be able to appear as *natural* in any presumably relevant sense). The good news seems to be that humans too are natural. The bad news seems to be that sailing this middle course is easier said than done. Any brand of naturalism that is sufficiently capacious as to do justice to the strong normativity of human practice appears to suffer such a radical disconnect from 'scientific naturalism' as to veer off into dualism, whereas any naturalism that stays faithful to the precepts of natural science appears inevitably to find its way to some form of normative eliminativism. All the many deflationary moves by contributors notwithstanding, the path not taken, nor scarcely even considered, is that of a more radical re-thinking of just what is the relationship of normativity to nature *all the way down*. Must our point of departure always be at the level of normativity as a distinctively *human* phenomenon? Might it not be the case that it is just such a presupposition that dooms the conciliatory enterprise to failure from the start?

Aristotle founded the systematic study of the living organism upon an understanding of intrinsic purposiveness (or finality) as a natural phenomenon. In Aristotle's hands, the fusion of form with finality as 'end-in-it-selfness' resulted in a highly fecund concept of the suitability (we would now say adaptability) of an organism's form to a stable way of life that was the lynchpin for elaborating a taxonomy, an anatomy and physiology, and a theory of *generation* (or, as we would now say, development). To meld biological (including behavioral) form with the logos of self-purpose is however to constitute a largely qualitative, i.e., not easily quantifiable, system of understanding. Aristotle's doctrine of the natural purposiveness of the living organism was of course entirely consonant with the legacy of Greek cosmology that cleaved to an organismic image of nature as a whole. It is largely uncontroversial that the radical Renaissance shift from the organismic cosmology of Greek *physis* to that of the mechanistic worldview that we call Modern Science brought with it important benefits. When formal features of nature were wholly, or even partially, stripped of imputations of immanent finality they became amenable to mathematic and related logics of understanding and analysis that conferred predictive, systematizing, and manipulative power. The cost of this benefit, however, was that of having to exile all that could not be denuded of immanent finality into a putatively non-natural realm of being—and this was not limited to the *human* domain. The phenomenon of life as a whole, having been constituted systematically by exactly that which cannot be wholly stripped of intrinsic purposiveness, was left in an equivocal position. Indeed, surely much of the history of biology since the seventeenth century could be reconstructed

E-mail addresses: Lenny.Moss@exeter.ac.uk (L. Moss), dan.j.nicholson@gmail.com (D.J. Nicholson)

1369-8486/$ - see front matter
doi:10.1016/j.shpsc.2011.05.007

as a kind of *pas de deux* between formal and finalistic (or functional) modes of explanation that have striven, by different measures and proportions, to somehow have it both ways: to retain the power of purely formalistic logics while still capturing the minimal finalistic features of a 'natural being' that are the *sine qua non* of its being alive. Though the logic of natural selection has done much to reconstitute an understanding of the intimate relationship between living form and the ends of life, yet all the while lacking an adequate theory of phenotypic (i.e., formal) variation (and thus innovation) (Kirschner and Gerhart, 2005), it has by no means eliminated the perception nor relevance of natural purposiveness from the domain of the living (nor would Darwin ever have imagined otherwise). While much of contemporary philosophy has long since been agonizing over the rigours of trying to reconcile the natural and the normative, the truly enabling work of elucidating the normative in nature has only just begun.

Drawing on a wide range of historical, philosophical, and empirical resources, the essays in this special issue consider the biological roots of normativity through an examination of life's ostensible purposiveness and accordingly reflect on the place of teleological considerations in biological explanation. The nine contributions can be clustered together under three relatively distinct thematic emphases—loosely described as theoretical, historical and critical (although each contribution expresses aspects of all three). The first cluster of articles, drawing on the idea of an autonomous system, locates the place and source of normativity in nature and the warrant for teleological concepts within a strictly physicalist vocabulary. The second cluster highlights the central role played by teleological reasoning in thinking about life as expressed in Kant's engagement with what became the origins of modern biology, in Hegel's attempt to go beyond a merely heuristic appropriation of Aristotle in grasping the impetus for an immanent movement within nature, and in the fascinating yet still poorly understood set of interactions and influence between Niels Bohr, Pascual Jordan, Max Delbrück and Erwin Schrödinger that in many ways gave birth to molecular biology. Finally, the third cluster critically examines the recent philosophical literature on the role played by mechanisms in contemporary biological and biomedical research and uses this discourse as a springboard to reconsider the relation between mechanism and teleology in biological explanation.

2. The place of normativity and teleology in nature

In the first article of the present volume, James Barham considers, in painstaking detail, the question that none of the contributors to the De Caro and Macarthur volume dared to ask: Is normativity really exclusive to the human domain? And if not, just what is the *scope* of normativity in nature? Following a detailed analysis of the concepts of 'normativity' and 'agency', and the ways in which they relate to one another, Barham's answer to what he calls the 'Scope Problem' is that normativity is co-extensive with life, in all its forms. In other words, the proper scope of application of normative agency in nature is living systems as such. To substantiate this thesis, Barham shows that the idea of normative agency is intimately connected to a family of interconnected concepts, such as 'purpose', 'value', 'need', and 'well-being', which find their most elementary applicability at the level of living systems. Drawing on a wealth of biological examples, Barham concludes that there is in fact no principled reason for maintaining that normativity and agency are properties of human beings alone, or even that they are attributable exclusively to the higher animals. Instead, Barham suggests, we should be prepared to accept that normativity and agency are real, objective properties of organisms as such.

Wayne Christensen advances a similar thesis in his article, in which he argues that normativity ought to be grounded in the distinctive organization of living systems, specifically in the form that generates their unity and hence explains their existence. Taking the concept of biological function as the point of departure for developing his naturalistic account of normativity, Christensen rejects the classic etiological theory that explains constituent 'proper functions' in terms of the action of natural selection, and advocates instead an account of 'functional normativity' based on the organizational autonomy of living systems. Christensen then proceeds to show through a series of thought experiments how his account can be extended to encompass key aspects of the normativity of practical reasons found at the human level. Implicit in this argumentative move is the recognition that normativity is not confined to the realm of rational agency and personhood, but is rather already fundamentally present in autonomous systems (i.e., organisms) as such. Overall, Christensen's discussion provides a compelling case for taking seriously the claim, defended likewise by Barham, that organisms, by virtue of their nature, constitute appropriate anchoring points for developing a general naturalistic understanding of normativity.

Georg Toepfer's article picks up where Barham's and Christensen's left off by reflecting on the implications of the normative nature of organisms for biology as a science. Like Barham and Christensen, Toepfer considers living systems to be inherently teleological. Indeed, the very concept of 'organism', Toepfer notes, is only truly comprehensible when it is understood in teleological terms. An organism is a cyclically organized system of interdependent causal processes that collectively constitute the whole and thereby contribute to its continued maintenance. The identity, unity, and functional operation of such a system is only understandable through a teleological perspective. Toepfer argues that teleology serves to identify and delimit organisms as the kind of natural systems that they are. For this reason, teleological reasoning plays an indispensable methodological role in biology, as it is what enables biologists to make conceptual sense of the objects they study. This leads Toepfer to declare, pace Dobzhanky, that 'nothing in biology makes sense except in the light of teleology', and much of his discussion is devoted to cashing out the implications of this view. Toepfer uses his conception of organisms to propose a new definition of function and he subsequently goes on to consider potential objections to his teleological understanding of the concepts of 'organism' and 'function'.

3. Historical perspectives on the role of teleology in biology

John Zammito's contribution is the first of three articles devoted to examining the role played by teleological reasoning in the history of biology. Zammito sets out to critically revisit the claims made by Timothy Lenoir thirty years ago in his seminal work on the deployment of teleological ideas in German biology in the late eighteenth and early nineteenth centuries (Lenoir, 1980, 1981, 1982). Specifically, Zammito takes issue with Lenoir's historical reconstruction of the reception of Kantian philosophy by German biologists through the figure of J. F. Blumenbach. Lenoir proposed the thesis that the so-called 'Göttingen School' around Blumenbach adopted Kant's methodological guidelines, and established a strictly heuristic (or in Kantian language, *regulative)* notion of 'teleo-mechanism' whereby the ascription of intrinsic purposiveness to organisms was not regarded as an objective scientific claim. Through a careful analysis of the writings of both Kant and Blumenbach, Zammito shows Lenoir's thesis to be fundamentally mistaken. Blumenbach and his followers actually took teleology to be an objectively ascertainable feature of organisms—a fact about the nature of living systems and not an epistemic presupposition

required for their conceptualization, as Kant would have it. Moreover, Zammito advances the claim that these biologists could not have possibly adopted Kant's prescriptions, as these are essentially incompatible with the empirical practice of a bona fide life science. This conclusion has important implications for recent reappraisals of Kant's philosophy of biology, as well as for the ongoing dispute over the status of natural teleology in contemporary biology.

Francesca Michelini turns her attention to Hegel's philosophical corpus in order to determine how he conceived of the nature of living systems. Up to now, Hegel has been almost completely ignored by modern philosophers of biology, so Michelini's lucid and detailed discussion of Hegel's biological ideas will undoubtedly be of wide interest. Indeed, far from the mystical, scientifically irrelevant figure that many cast him out to be, Hegel actually developed a surprisingly nuanced understanding of the ontological distinctiveness of living organisms. By skillfully conjugating Aristotelian and Kantian ideas of 'intrinsic purposiveness, Hegel formulated a new conception of the living state, which he characterized as the 'activity of deficiency'. As Michelini explains, Hegel used this expression to draw attention to the fact that the phenomenon of life itself is inextricably bound with what it lacks—its identity is at one with its own negation. This ontological situation, found in all living beings, is exemplified by the inevitable appeal to teleological concepts like 'need' and 'drive' when describing organisms. On the whole, Hegel's conception of life (which, we note in passing, anticipates the essence of our modern understanding of organisms as open systems far from thermodynamic equilibrium) provides a foundation for a naturalistic account of normativity and teleology in living beings not unlike the contemporary ones proposed in this special issue by Barham, Christensen, and Toepfer.

Philip Sloan reflects on the status of teleological reasoning in biology through the examination of a more recent episode in the history of biology, namely the reception of Niels Bohr's proposal to introduce a principle of 'complementarity' in the study of living systems by three theoretical physicists who played a pivotal role in the emergence of molecular biology in the second third of the twentieth century: Pascual Jordan, Erwin Schrödinger, and Max Delbrück. Influenced by Harald Høffding's realist interpretation of Kant's resolution of the *Antinomy of Teleological Judgment*, Bohr argued that a genuine understanding of organisms required the appeal to two mutually exclusive descriptive frameworks: the mechanistic (or reductionistic) and the teleological (or holistic). These two alternative characterizations of the living state are primary and complementary. Sloans hows how Jordan, Schrödinger, and Delbrück interpreted Bohr's philosophy of biology in strikingly different ways, and he illustrates how the debates sparked by these different interpretations were central in determining the fate of teleological reasoning in the biophysical research program that ultimately became molecular biology. Sloan also argues that Bohr's actual views on the complementarity of teleology and mechanism remain relevant and appealing, and that they should be seriously considered in current philosophical discussions.

4. Of mechanism and teleology

The last triad of articles engages critically with the *New Mechanism* research program that has emerged over the past decade with the objective of making sense of biologists' 'mechanism-talk' and the role it plays in biological explanation. It is rather surprising that despite the prominence of this new mechanismic discourse, there have been no attempts to investigate the connection between the contemporary appeal to mechanisms in biological explanation and the older *mechanistic* tradition in biology. Some important questions remain unanswered concerning this relation. For example, does the new mechanismic philosophy of biology represent some sort of continuation of the agenda of mechanistic philosophy as it applies to biology? Does the current appeal to mechanisms in biological research commit a biologist to a mechanistic view of life? Daniel Nicholson sets out to answer these questions by bringing a much-needed historical perspective to bear on the debate over the meaning of 'mechanism' in biology. Nicholson shows that 'mechanism' is actually an umbrella term for three distinct notions. It may refer to a philosophical thesis about the nature of life and biology (*mechanicism*), to the workings of a machine-like structure (*machine mechanism*), or to the causal explanation of a particular phenomenon (*causal mechanism*). Nicholson argues that the new mechanismic philosophers have inadvertently conflated these different meanings. Specifically, they have inappropriately endowed causal mechanisms (the sense of 'mechanism' most commonly invoked by biologists today) with the ontic status of machine mechanisms (the sense of 'mechanism' historically invoked by mechanistic biologists), and this inevitably results in confusion. Nicholson suggests that the ontic characterizations of causal mechanisms found in the current philosophical literature does not do justice to scientific practice, and that the notion of causal mechanism is better understood *epistemically* as a contingent explanatory model which heuristically abstracts away the complexity of a living system sufficiently to describe some localized causal process occurring within it that gives rise to a phenomenon of interest.

Lenny Moss brings teleology back into the picture through recourse to a phenomenological reconstruction of the background know-how and presuppositions of the competent biomedical research scientist for whom what even *counts* as a 'mechanism' (as opposed to an artifact) must always already be understood as an expression of the ostensible purposiveness of a living system (unlike, for example, the background know-how of a chemist or physicist for whom what counts as a 'mechanism' need not satisfy any such criterion). More damningly, Moss seeks to drive a wedge between the disciplinary motives of philosophical *hard naturalists* and those of the biological scientists for whom they claim to speak. Rather than grasping and elucidating the situated aims and practices of biologists themselves, Moss suggests that the philosophical investigation of the contemporary meaning of 'mechanism' in biology has been commandeered by the felt-disciplinary needs of such philosophers of science to replace the old deductive-nomological model of the so-called 'received view' with a new normative-explanatory gold-standard. He argues, however, that rather than an orientation toward an increasingly precise characterization of mechanisms as being an ultimate end in biological research, as the hard naturalists would have it, in actual biological practice 'mechanism' means different things in different contexts, pragmatically draws on our embodied know-how in the use of machines where and when it is useful, and is not, nor should be, any ultimate end of biological research. Moss argues that the kinds of entity and activity descriptions that the new mechanism philosophers have attempted to elevate as canons of normative practice in fact are taken up as merely plausibly-how, low-level building blocks, en route to an understanding of high-level purposive-regulatory functions that will inspire and require new, and very *unmachine-like*, forms of knowing.

In the final paper of the collection, Denis Walsh draws on currently accepted characterizations of mechanismic (or 'causal-mechanistic') explanation as a basis to develop a formal account of natural teleological explanation. Walsh argues that just as in causal-mechanistic explanations there is an invariance relation relation between the mechanism and the effect and an elucidative description that illustrates how the causal mechanism produces the effect, in teleological explanations there is likewise an invariance relation between the means (cause) and the goal (effect) as well as an elucidative description that illustrates the way the means conduces to the goal. Moreover, Walsh argues that both

causal-mechanistic and teleological explanations are complete in their own right, and are mutually autonomous in such a way that one cannot replace the other without explanatory loss. Walsh concludes from this that Jaegwon Kim's well-known explanatory exclusion principle (Kim, 1989) is incorrect, given that some natural phenomena (particularly those pertaining to the living realm) are in fact susceptible to more than one complete and autonomous form of explanation.

Acknowledgements

We would like to thank the Australian Research Council for providing financial support that enabled the preparation of this volume.

References

De Caro, M., & Macarthur, D. (2010). *Naturalism and normativity*. New York: Columbia University Press.

Kim, J. (1989). Mechanism, purpose and explanatory exclusion. In J. Tomberlin (Ed.), *Philosophical perspectives 3: Philosophy of mind and action theory* (pp. 77–108). Atascadero, Calif.: Ridgeview.

Kirschner, M. W., & Gerhart, J. C. (2005). *The plausibility of life: Resolving Darwin's dilemma*. New Haven: Yale University Press.

Lenoir, T. (1980). Kant, Blumenbach, and vital materialism in German biology. *Isis, 71*, 77–108.

Lenoir, T. (1981). Teleology without regrets. The transformation of physiology in Germany: 1790–1847. *Studies in History and Philosophy of Science, 12*, 293–354.

Lenoir, T. (1982). *The strategy of life: Teleology and mechanics in nineteenth-century German biology*. Dordecht: D. Riedel Publishing Company.

Studies in History and Philosophy of Biological and Biomedical Sciences 43 (2012) 92–103

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

Normativity, agency, and life

James Barham

Department of Philosophy, University of Notre Dame, 100 Molloy Hall, Notre Dame, IN 46556, USA

ARTICLE INFO

Article history:
Available online 21 June 2011

Keywords:
Normativity
Agency
Life
Teleology
Naturalism
Organism

ABSTRACT

There is an immense philosophical literature dealing with the notions of normativity and agency, as well as a sizeable and rapidly growing scientific literature on the topic of autonomous agents. However, there has been very little cross-fertilization between these two literatures. As a result, the philosophical literature tends to assume a somewhat outdated mechanistic image of living things, resulting in a quasi-dualistic picture in which only human beings, or the higher animals, can be normative agents properly speaking. From this perspective, the project of 'naturalizing normativity' becomes almost a contradiction in terms. At the same time, the scientific literature tends to misuse 'normativity,' 'agency,' and related terms, assuming that it is meaningful to ascribe these concepts to 'autonomous agents' conceived of as physical systems whose behavior is to be explained in terms of ordinary physical law. From this perspective, the true depth of the difficulty involved in understanding what makes living systems distinctive *qua* physical systems becomes occluded. In this essay, I begin the attempt to remedy this situation. After some preliminary discussion of terminology and situating of my project within the contemporary philosophical landscape, I make a distinction between two different aspects of the project of naturalizing normativity: (1) the 'Scope Problem,' which consists in saying how widely in nature our concept of normative agency may properly be applied; and (2) the 'Ground Problem,' which consists in rationalizing the phenomenon of normative agency in terms of the rest of our knowledge of nature. Then, in the remainder of this paper, I argue that the Scope Problem ought to be resolved in favor of attributing normative agency, in the proper sense of those words, to living things as such. The Ground Problem will be discussed in a companion paper at a later time.

1. Introduction

In this paper, I will explore the possibility of giving a realistic account of normative agency, properly so called, as an essential property of life. Needless to say, this is a highly ambitious and contentious thesis. I will not be able even to touch upon all of the many questions raised by my thesis here, much less provide anything like a proof. What I will do, however, is discuss two specific issues, which—together with a third issue I hope to discuss on a future occasion—I trust will constitute a *prima facie* case for at least according my thesis serious consideration.

First, in Section 2, below, I will deal with some key definitional issues. What exactly do we mean by the concepts of 'normativity' and 'agency'? How are the two concepts related? And what might it mean to 'naturalize' normativity and/or agency? In reply to this last question, I will distinguish eliminativist and epiphenomenalist versions of 'naturalized normativity' from the realistic project of giving an account of the place in nature of normativity, considered as an objectively existing phenomenon. Furthermore, I will argue that if we take the realistic project of naturalizing normativity seriously, then we must distinguish between what I will call the 'Scope Problem'—namely, the problem of determining the proper scope of application of our concept of normative agency—and the 'Ground Problem'—the problem of characterizing the physical ground of normativity in nature.

Then, in Section 3, I will investigate the Scope Problem, arguing that the proper scope of application of our concept of natural agency is to life—that is, to living systems, or organisms—as such. A similar investigation of the Ground Problem will be undertaken elsewhere.

E-mail address: jbarham@nd.edu

1369-8486/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsc.2011.05.008

2. What do we mean by 'normative agency' and what would it mean to 'naturalize' it?

The paradigm case of 'normativity' is undoubtedly moral prescription and proscription, expressed through the terms 'ought,' 'should,' 'must,' and related locutions. For example: 'Thou shalt not kill.' Nevertheless, it is not difficult to see that the moral 'ought' is only a species of a wider genus of normativity that applies to human actions generally. For example: 'You *ought* to use a hammer (to pound nails)'; 'You *should not* smoke (to avoid coming down with lung disease)'; 'You *must* practice, practice, practice (to get to Carnegie Hall)'; and so on. What all of these normative claims have in common is the prescription or proscription of an action, considered as the appropriate means to attaining an end. In this respect, we can see that norms are instrumental in character. They seem to be essentially involved with furthering the actualization of ends by specifying actions conducive to such actualization. That is, norms connect ends to the appropriate means, and wherever there is a means–end relationship, there is normativity in this sense. If norms are real, as opposed to merely notional, then the 'specifying'of appropriate actions that they do makes a real contribution to influencing or determining real events in the world. To this extent, then, norms are analogous to ordinary causes—physical forces—but, as I shall argue below, they cannot be construed as literally being ordinary causes or physical forces. In fact, the crux of the problem of normativity lies in understanding how something that is not an ordinary cause or physical force can nevertheless have a real influence or determinative power over events in the world.

The norms I have been discussing so far are clearly nonmoral, since actions attain a moral quality by virtue of their impact on the welfare of other human beings—an impact which actions like using a hammer, giving up smoking, and practicing one's musical instrument lack (at least directly). Moreover, moral and nonmoral norms are both 'instrumental oughts,' since they both connect ends to the appropriate means.[1] Following the customary terminology, we may distinguish 'moral actions' from merely 'prudential actions.' Let us call, then, nonmoral instances of prescription and proscription of actions instances of the 'prudential ought.' It follows that the genus 'instrumental ought' consists of two species, the 'moral ought' and the 'prudential ought.' And so the 'moral ought,' resident in our paradigm of normativity, is in fact only a fairly restricted special case of a much more general phenomenon. This is also evident from the fact that all 'moral oughts' prescribe or proscribe human actions, but not all prescriptions or proscriptions of human actions are moral in character. Many of them are prudential. In other words, outside of the sphere of moral action lies the vast sphere of prudential action where normativity is equally present under the guise of the 'instrumental ought.' This entitles me to ignore the 'moral ought' here, in spite of the fact that it is our paradigm of normativity. Everything I say hereafter about normativity should be understood as applying in the first instance to the 'prudential ought.'

Another issue that must be addressed is the nature of what I have been calling 'prescription' and 'proscription.' As we have seen, human beings often express normativity by means of such auxiliary verbs as 'must,' 'ought,' or 'should.' In addition, the imperative mood of the verb is often employed for this purpose. Moreover, norms may be codified in the form of written or unwritten laws, rules, maxims, and other types of commands, prohibitions, and recommendations. All of these types of normativity seem to involve language and human intentionality in a fundamental way. This is an issue that is orthogonal to the moral/prudential issue. That is, the seemingly linguistic character of normativity considered as prescription would seem to restrict the 'prudential ought' to human actions as surely as the 'moral ought' is so restricted. After all, how can there literally be prescriptions in the absence of a prescriber, commands in the absence of a commander, and so on?

And yet the notion of normativity does appear to be more widely applicable than just to the human case. For instance, it is natural to say things like: 'Dogs *ought* to get plenty of exercise'; 'Hearts *should* beat in sinus rhythm'; and 'Plants *must* have water.' This makes it seem as though there is a kind of *requirement* in some natural systems that has nothing to do essentially with either language or human intentionality. This notion of requirement is more generic than prescriptivity, or, in other words, human language-mediated prescriptivity stands in relation to this broader notion of normative requirement as species to genus. If that is so, then it is natural to ask: What is the nature of this more generic form of normative requirement? This is another way of posing the question that lies at core of this project, and will comprise the main topic of Section 3, below.

Yet another distinction I wish to make involves two different senses in which the terms 'normative' and 'normativity' are sometimes used. I will call them the 'narrow' and 'broad' senses. In the narrow sense, normativity is simply the 'instrumental ought' that we have been discussing up until now, namely, the idea of requirement—that is, the fact that there is something that an agent is required to do in a certain situation in order to attain a particular end. Though the notion of normative 'requirement' is already broad with respect to the narrower notion of 'prescriptivity,' it is nevertheless comparatively narrow in relation to another way that the term 'normativity' is sometimes used—namely, as an umbrella term to designate a family of closely related concepts for which we seem to have no collective name in colloquial English. We use the term 'normativity' in this broad sense *faute de mieux*, and the resulting ambiguity can give rise to confusion if we are not careful. The family of related concepts that are sometimes referred to as 'normative' in this broad sense is specified by the network of mutual implication existing among a number of concepts that are analytically contained in the concept of 'action' in the normative sense of 'acting for a reason' (as well as the concept of 'agency,' understood as the power to 'act for a reason'). 'Normativity' in this broad sense encompasses such concepts as purpose, value, well-being, need, and being a reason for action, in addition to the narrow 'instrumental ought.' In Section 3, below, I will attempt to justify the claim that there is in fact a natural kind corresponding to this umbrella concept of 'normativity.' For now, I would like to make a more limited point regarding the claim that normativity—in both the narrow and broad senses of the term—is intimately connected to agency.

First, take the narrow sense of normative requirement as the 'instrumental ought.' If normative requirement is the fact that an agent ought to (or should or must) do something in a given situation in order to attain a particular end, then normativity in the narrow sense clearly implies agency. But what about the converse case: Does agency imply normativity? If actions are held to be somehow controlled or guided by reasons, and if reasons are held to be metaphysically distinct from causes, then reasons may be said to indicate what should, or ought to, be done in a given situation. This does make it seem as though agency implies normativity. Unfortunately, there are two difficulties with this claim.

The first difficulty lies in determining the kinds of things to which the concept of normative agency may be properly applied.

[1] This is true even if one interprets 'moral oughts' as categorical imperatives, because the categoricity of a moral imperative lies in its supremacy over other imperatives (i.e., its unconditionality), not in its pointlessness. Categorical imperatives, too, prescribe or proscribe actions, and *ipso facto* connect ends to means (for example, where the end may be construed as 'doing one's duty').

Call this the 'Scope Problem.' The problem arises from the fact that many commentators feel that reasons may properly be said to exist only where the capacity for their conscious weighing, or rational deliberation, exists. Accepting this claim would of course mean that only human beings could qualify as agents in the normative sense. According to this way of thinking, one ought to take care to say that human beings 'act,' while other animals merely 'behave,' where actions are held to be guided by reasons, in contradistinction to behaviors, which are merely caused.[2]

Nevertheless, we find it natural to speak of the 'reasons' that (at least some) non-human animals have for doing the things that they do. For example, if I observe my cat jumping down from the windowsill and going into the kitchen, and I know that the kitchen is where her milk bowl is located, then I may infer the reason why she went into the kitchen: namely, to get a drink of milk. All of this seems closely analogous to my own behavior when I go into the kitchen from time to time to get a drink of water. If I say that getting a drink of water is the reason why I go into the kitchen, why should I not say that getting a drink of milk is the reason why my cat goes into the kitchen? It is true that my behavior may sometimes be complicated by the existence of countervailing reasons ('Shall I have a beer instead?') and the need to weigh them in a way that my cat's behavior is not. But I see little reason to doubt that our motivations in this case are basically similar—that when my cat is thirsty she experiences something similar to what I experience when I am thirsty; that the pleasure she takes in her milk is not so different from the pleasure I take in my glass of water; and so on. And, indeed, it may often happen that my behavior may be nearly as simple and unreflecting as hers (say, if I go into the kitchen on 'automatic pilot,' that is, with my mind on something else). If my unreflecting behavior nevertheless qualifies as acting for a reason—that is, qualifies as an action in the normative sense—then why should not her behavior so qualify? It may still be objected that I am trading on an ambiguity in the notion of a 'reason.' There is also a causal use of the concept, as in asking for the 'reason' for an airplane crash or a mining accident. Therefore, one might wonder why my cat's reason for going into the kitchen should not be construed as a purely causal reason of that sort. Of course, one would then have to explain why that construal of the concept should apply to my cat's behavior, but not to my own behavior. However, that would be a superficial reply. And, besides, there may be some readers who would be prepared to see my own reasons given this same sort of causal construal. Therefore, to address this worry adequately will mean digging deeper, and attempting to elucidate the fundamental difference between causes and normative reasons. Indeed, in a sense, that may be viewed as the central aim of this paper. But, in that case, I cannot accept the charge of equivocation, as that would amount to the claim that there is no important difference between causes and reasons, which would beg the main question at issue here.

If my cat's behavior really is so similar to mine as to justify counting it as a case of normative action, still it cannot be denied that it differs importantly from mine in that in my case the potential for rational deliberation is always there, while in her case it is not. This is certainly a significant difference, and it needs to be marked by a terminological distinction. Let us call the cat's form of acting 'subrational.' But then, the question arises: Is subrational action truly normative? To the extent that we are comfortable explaining the cat's behavior by reference to reasons, it would seem that it is. But if we accept this, then obviously we cannot associate the concept of acting for a reason with rational deliberation alone, nor can we sustain a distinction between action and behavior in the traditional way. There are several ways to go here. One would be to deny that subrational behavior is truly normative action. Another would be to say that not all action is truly normative, but that a sort of 'subnormative' action also exists. Yet another would be to bite the bullet and admit that our original distinction was misguided, and that the higher animals (at least) are fully capable of action in the normative sense. But since this last way involves rejecting the association of acting for a reason with rational deliberation, the question would then arise: How are we to understand the capacity of acting for a reason—that is, normative agency?

This brings us to the second difficulty involving the claim that agency implies normativity. This difficulty lies in understanding how something like normative agency can exist in nature at all, given the rest of the world picture painted for us by contemporary natural science. Call this the 'Ground Problem.' I note in passing that the Ground Problem is just as much a problem for those who hold that the concept of normative agency is essentially connected with rational deliberation as it is for those who would widen the concept's scope of applicability to include (at least) the higher animals. However we resolve the Scope Problem, the Ground Problem still remains—which is not to say, however, that some solutions to the Scope Problem may not lend themselves more readily than others to a solution to the Ground Problem.

In the next section, I will argue in favor of a radical solution to the Scope Problem that views normative agency as a property of living things as such. That is, I claim that all organisms are normative agents, and that only organisms are normative agents in a literal, original, and underived sense. This claim will be based on a conceptual argument that consists of several components, but which is ultimately based on the fact that what distinguishes organisms as a natural kind is that they must act in order to preserve themselves in existence.[3] Upon another occasion I hope to address the Ground Problem.

In this section, I have so far focused on explaining what I intend and do not intend by the terms 'normativity' and 'agency,' and I hope thereby to have clarified my aims in this paper. Before turning to the detailed investigation of the Scope Problem, however, there is one more preliminary matter that I would like to discuss in order to reduce still further the possibility of misunderstanding. Broadly speaking, this paper can be viewed as a contribution to the project of 'naturalizing normativity'—a project that is proceeding along a broad front of contemporary philosophy. And yet, for many philosophers the concepts 'normative' and 'natural' remain antithetical, and the idea of 'naturalized normativity' is an oxymoron. For this reason, a few words about what the project of naturalizing normativity does and does not entail are necessary.

The project of naturalizing normativity is a highly various and complex enterprise, but perhaps it would not be oversimplifying matters too much to distinguish three main approaches. The first approach is the effort to eliminate normativity from our ontology altogether. On this view, normativity is 'naturalized' by showing that it does not really exist, and that in reality the 'natural' (understood here as a contrast class to the 'normative') is all there is. This may be achieved, it is supposed, either

[2] Here, I shall say that an event is 'caused,' in the sense of 'efficient causation,' if it is determined solely by physical laws as currently understood by mainstream, contemporary natural science. This somewhat convoluted formula is intended to leave open the possibility that, while present-day natural science may lack a theoretical perspective apt for the proper understanding of acting for a reason, it is nevertheless conceivable that such a perspective may be developed in the future.

[3] I acknowledge many difficulties in specifying what is to count as an 'organism' (what do we say about viruses, colonial organisms, cancers, beehives, and other doubtful cases?), but cannot consider the problem in detail here. For present purposes, I assume that the individual prokaryotic cell is the paradigm organism.

by showing that the putative normative phenomena (such as actions) to which our normative concepts seem to refer can be ontologically 'reduced' to nonnormative phenomena, and so are redundant, or else by showing that the putative normative phenomena do not really exist in an objective sense, and are merely a subjective 'projection' of human concepts and behavioral response patterns onto the world—i.e., a sort of 'illusion.' The justification for the eliminative approach may be expressed by means of something like the following argument (*viz*, the 'Eliminative Argument'):

(1) The picture of the world painted for us by the present-day physical sciences (including chemistry and biology) is complete in all fundamentals. Call this the 'present physical picture.'
(2) Our ontology—that is, our list of the things that really exist in an objective sense—ought to correspond to the present physical picture.
(3) The present physical picture makes no mention of normative phenomena.
(4) Therefore, normative phenomena do not really exist in an objective sense, and ought to be eliminated from our ontology.

Now, this simple picture would have to be complicated in numerous ways if a faithful account of the state of play in the literature were our goal here. For one thing, it would have to be acknowledged that there are relatively few philosophers who explicitly embrace eliminativism (e.g., Churchland, 2007; P.S. Davies, 2009). This should not be surprising, since to deny flat-out that normativity exists is a very strong and highly counterintuitive claim. But it does mean that the many philosophers who subscribe to one form or another of 'reductionism' owe us a clear explanation of exactly what they take the ontological status of the 'reduced' higher-level entities to be. To see this, let us set aside the many complex epistemological and semantic issues, and look toward the metaphysical implications of the basic reductionist idea—that a higher-level 'reduced' entity is 'nothing but' or 'nothing over and above' the lower-level entities and relations of the reduction base.[4] It would seem that the reductionist is faced with a dilemma. After the 'reduction' has been carried out, the reductionist must say either that the higher-level 'reduced' entity still exists as a real entity with causal powers of its own, or that it does not. If the 'reduced' entity is held still to exist, then the position of the reductionist will be difficult to distinguish from that of the nonreductive physicalist (to be discussed below). If not, then the position of the reductionist will be difficult to distinguish from that of the eliminativist. Either way, the reductionist position will be revealed to be unstable.

In any case, my goal here is not to stake out a position on reductionism for its own sake, but rather to limn the conceptual alternatives available for 'naturalizing normativity.' For this purpose, it is enough to define 'reductionism' with respect to normativity as follows:

> *Normative Reduction:* To reduce a putative normative phenomenon is to give an account of the phenomenon that is both empirically and theoretically adequate and that neither employs nor presupposes any normative concepts.

If an empirically and theoretically adequate account of a putative normative phenomenon (such as action) could really be given in entirely nonnormative terms, then surely we would be entitled to deny the reality of the normativity of the putative normative phenomenon. Whether one takes an 'epiphenomenalist' or a frankly 'eliminativist' attitude toward the 'reduced' putatively normative phenomenon, then, would seem to be of comparatively small interest. What is of signal interest is that under the scenario we are considering we would appear to have little reason to allow the putative normative phenomenon onto our list of the real features of the world. For all intents and purposes, then, reductionism with respect to normativity is virtually indistinguishable from eliminativism, and so there is little reason for us to consider it here as an independent position within the conceptual landscape of 'naturalized normativity.'[5]

At the opposite extreme from eliminativism is so-called 'liberal' (McDowell, 1998) or 'naïve' (Hornsby, 1997) naturalism.[6] This second main approach to naturalizing normativity is a view that takes common sense rather than natural science as the arbiter of what is to count as 'natural,' i.e., as belonging to 'nature.' Liberal naturalism assumes that human beings are members in good standing of the natural world. This means that all the properties of human beings—indeed, all phenomena associated with, or pertaining to, human beings—are likewise natural. On this view, 'natural' contrasts with 'supernatural' (what 'transcends' nature), but not with 'normative.' The normative, as a feature of the human, is to be viewed as a subset of the natural. This of course raises the question of how the normative natural phenomena and the nonnormative natural phenomena (let us call them the 'physical phenomena') are related. However, liberal naturalism considers itself under no obligation to explain this relation. Rather, liberal naturalism is content to point out the limitations of natural science. Science is cognitively authoritative as far as it goes, but it only goes as far as the physical phenomena. Its writ simply does not extend to the entirety of nature. That is, liberal naturalism denies premise (2) of the Eliminative Argument outright. But while it is assuredly true that at present the normative phenomena lie beyond the ken of natural science, it is not clear why this limitation should be one of principle, true for all time. The problem with liberal naturalism is that by elevating the present limits of natural science to a matter of principle, it can seem to come perilously close to dualism. For if it is true that the normative is a part of nature, then there must be some connection between the normative and the physical, and what reason can there be in principle why natural science should be forever forbidden from coming to understand the nature of this connection?

In between the two extremes of eliminativism and liberal naturalism is nonreductive physicalism. This third main approach to naturalizing normativity exists in a great variety of different forms, but they all have in common the idea that premise (2) of the Eliminative Argument ought to be, not denied outright as in liberal naturalism, but relaxed in such a way as to make it possible for us to admit into our ontology the normative and other higher-level phenomena, which are conceived of as standing in a certain admissible relation to the present physical picture, even though they are not formally a part of that picture. The trick here is to specify the exact nature of the admissible relation between the normative phenomena and the present physical picture. The

[4] In a more adequate discussion, several different forms of reductionism would have to be distinguished: epistemological vs. ontological, and with respect to the latter, causal vs. compositional forms, to name only a few (see Gillett, 2007).

[5] For further discussion of these issues in terms of the realism/anti-realism debate, see Fine, 2002.

[6] One might suppose the opposite of eliminativism to be not liberal naturalism, but dualism—by which I mean the positing of a fundamental ontological discontinuity between normative and physical phenomena. For dualists, the natural is to be identified with the physical, understood as the 'nonnormative,' such that the 'normative' and the 'natural' become contraries. That being the case, it seems more appropriate to classify dualism, not as a pole within the naturalization project, but rather as the repudiation of that project altogether.

two main candidate relations are supervenience and emergence.[7] Unfortunately, there are good reasons to believe that the supervenience relation collapses back into epiphenomenalism—and hence, for all practical purposes, eliminativism—while the emergence relation has been criticized as being underspecified and mysterious (see Kim, 1998).

In this paper, I will pursue a strategy that has affinities with both liberal naturalism and nonreductive physicalism, but which accepts premise (2) of the Eliminative Argument according to the principle that it is desirable that our picture of the world be unified. Instead, I will deny premise (1). That is, I will claim that we have good reason to believe that the present physical picture is radically incomplete. Completing our physical picture will mean enlarging it to make room for the normative phenomena, considered as objectively real. Call this position 'normative realism.' No heavy-duty metaphysics is required to support normative realism; it merely requires being prepared to accord to normative phenomena the same ontological status that we ordinarily accord to nonnormative phenomena. In other words, ontological parity between normative and nonnormative phenomena will be realism enough for our purposes here.[8] In this way, we will be able to vindicate the liberal naturalist's insistence on according full ontological status and dignity to the normative phenomena, without walling them off from the physical phenomena on principle. At the same time, the nonreductive physicalist's postulate of a relation between the normative phenomena and the physical phenomena will be vindicated, and the relation itself clarified and shown to be admissible, by means of the notion of the nonreductive 'grounding'[9] of normative agency in physical phenomena of a certain sort that remains to be specified, but is capable in principle of being fully incorporated into our future scientific world-picture. With these various distinctions under our belt, let us now turn to the Scope Problem.

3. What is the proper scope of our concept of normative agency?

I begin with an informal argument for taking the proper scope of our concept of normative agency to be life itself, i.e., living systems, or organisms, as such. To simplify the presentation of the argument, however, I would like first to stipulate a definition of one of the concepts employed in it. The reason is that some of the concepts that might naturally be assumed to fall under the concept of normativity in the broad or umbrella sense (like the moral right; courage, justice, honesty, beneficence, and the other virtues; beauty; etc.) are of no relevance to our reflections here because their applicability is restricted to human beings. Therefore, it is convenient to define an intermediate class of normative concepts that lie in between normativity in the widest possible sense and normativity in the narrow sense of the 'instrumental ought.' I will call this intermediate group of normative concepts the class of 'elementary normative concepts,' which I define as follows:

> *Elementary normative concepts* are normative concepts connected to prudential instrumental action generally, exclusive of the normative concepts that imply human rational deliberation.

Examples of the elementary normative concepts are purpose, value, well-being, need, being a reason for action, and the 'instrumental ought' (normative requirement in the narrow sense). With this definition in hand, we are ready to proceed to the main argument of this section—let us call it the 'Scope Argument.'

First, although it is difficult to provide necessary and sufficient conditions for something's counting as 'normative,' nevertheless it is apparent that the elementary normative concepts are intimately related to one another conceptually. None of the concepts stands on its own two feet, as it were, but rather each leans heavily on its neighbors for support. Each of the elementary normative concepts is somehow incomplete on its own. For example, it is very hard to explain what we mean by 'purpose' without appealing to some notion of 'value' (Bedau, 1992). It seems, then, that the elementary normative concepts come as a package deal.

Second, though it is difficult to say precisely in what the 'family resemblance' among the elementary normative concepts consists, one feature that surely unites them as a group is that each of them is partly constitutive of agency, in the normative sense. That is to say, each elementary normative concept constitutes an aspect of our complex concept of normative agency. For example, 'having a purpose' is part of what we mean by 'acting' in the normative sense. A motion that had no purpose (in the sense of 'goal' or 'end') would not count as an 'action.' (Snowing is not an 'action.') Moreover, 'having a purpose,' or 'end,' implies a need to act—namely, to find and employ the 'means' appropriate to realizing the end. A state of affairs that no agent ever brought about by taking the appropriate instrumental actions would not count as an 'end.' (My snow-covered yard is merely the result, not the purpose or goal or end, of its having snowed.[10]) So, the logical entailment between purpose and action runs in both directions.[11]

Third, certain of the elementary normative concepts (e.g., purpose, need, well-being) are clearly properly ascribable to organisms as such.

From the foregoing considerations, we may conclude that all of the elementary normative concepts, as well as the concept of agency, are properly ascribable to organisms as such—i.e., organisms are properly regarded as agents in the full normative sense of the term. In other words, the proper scope of application of our concept of normative agency is living systems as such.

Let us now look more closely at each of these claims in turn.

[7] Supervenience is the relation between a higher-level ('supervenient') entity or property and a lower-level, acceptably physical ('subvenient') base such that there can be no change in the former without a corresponding change in the latter. It is important that the supervenience relation be conceived of as asymmetrical, in the sense that all causal influence flows from the base 'upwards' to the supervenient entity or property. (For discussion, see Savellos & Yalçin, 1995.) Emergence is conceived of in a variety of ways, but in its most important, synchronic sense, it is basically the denial of this last condition, such that at least some causal influence is conceived of as flowing 'downwards' from one or more higher-level entities or properties to the base. A further important component of the emergence relation is the idea that the higher-level entities and properties are not exhaustively determined by the causal properties of the base, which notion is often expressed by the slogan 'the whole is more than the sum of the parts.' (For discussion, see Bedau & Humphreys (2008), Clayton & Davies (2006), and Corradini & O'Connor (2010).)

[8] Thus, if someone were an anti-realist about scientific entities in general, but considered normative phenomena like normative action to be no less real (or more unreal) than nonnormative phenomena like matter, force, or energy, then that person would qualify as a 'normative realist' for present purposes.

[9] In the sense of Fine, 2002.

[10] To be sure, a snow-covered yard might be transformed into an end by human intentionality, as in a child's desire for a 'white Christmas,' and perhaps someday our improved control over the weather might even permit means to be taken to bring about such an end, but these examples only reinforce the tight conceptual link between purpose and action.

[11] It might be objected that I have simply stipulated that this be the case by excluding those concepts not constitutive of prudential instrumental action from my notion of an 'elementary normative concept' in the definition above. However, even in the wider case, moral concepts would seem to be just as closely linked to action as prudential instrumental concepts. After all, to be morally good is to act rightly (justly, beneficently, etc.) towards one's fellow human beings. While it is true that there may be a few normative concepts specific to the human domain for which the link to action seems looser (beauty comes to mind), nevertheless, the link seems very tight in the elementary cases, not just by definition, but rather due to inherent features of our concepts of normativity and agency. And, in any case, the elementary normative concepts are the ones that concern us here.

3.1. The elementary normative concepts are a package deal

This can be easily established by attending to the meaning of the concepts involved. I will begin with the concept of need. First, we must ask: Is it certain that need is in fact a normative concept at all?

It seems hard to deny that it is, at least in my own case. Satisfying my own vital needs appears to me as among the most peremptory of all the commands I am subject to. This fact becomes especially clear when one of them runs a risk of not being satisfied. For example, if I am lost in the desert, there is little that will appear to me under the aspect of a higher duty than that of securing some water to drink, in order to save my life.[12] It is true that one of the things that distinguish me from most if not all other life forms, is that there is indeed one thing that may appear to me as a higher duty than saving my own life, and that is saving another human being's life. So that if I happen to have a last swallow of water in my canteen, I may well give it to my wife or my child or my friend, or even a perfect stranger I happen to be thrown in with. But, notice that the point of my sacrifice is still to preserve life. I am unlikely simply to pour my last mouthful of water into the sand, at least so long as my reason and will do not fail me. Therefore, it seems that satisfying vital needs constitutes the highest of all normative imperatives, whether conceived of prudentially, in relation to the preservation of my own life, or morally, in relation to the preservation of the life of other human beings. Moreover, not only is need (at least in the vital sense we are investigating here) a normative concept itself, it can be shown to be very near to the *fons et origo* of all the other normative concepts. Let us see how.

From the concept of need immediately flows the concept of value: For a system to have needs is already for it to partition its environment into valenced categories. There are things to be pursued, and things to be avoided, that the needs may be satisfied. 'Good' and 'bad' are concepts of an immense semantic richness; nevertheless, there are really no more appropriate terms with which to describe these things that are to be pursued or avoided, based on our vital needs.[13] From the idea of pursuing the good proceeds directly that of end-directedness (or purposiveness), for what else does it mean to pursue the good than to have achieving a certain good (and thereby satisfying a certain need) as one's end or purpose? As Aquinas famously noted, the concept of value (good and bad) implies the concept of having a purpose or pursuing an end (*Summa Theologiae*, Ialiae.94.2): '*bonum est faciendum et prosequendum, et malum vitandum*' [the good is to be done and pursued, and the bad avoided]. From this, the 'instrumental ought' (normative requirement) follows immediately; indeed, the 'instrumental ought' is already tacitly relied upon in the grammatical form of Aquinas's formulation of this point: *faciendum* ... *prosequendum* ... *vitandum* [is to be done ... to be pursued ... to be avoided]. If one has the end or purpose of satisfying one's need for water (even short of saving one's life in the desert!), then one ought to seek water to drink. Which means, in turn, that the need for water provides an excellent reason for whatever steps must be taken to secure the water.

Conversely, as Burge has pointed out, good also implies should, or, as he puts it: 'goods generate shoulds' (Burge, 2003, p. 513), or, a little less apothegmatically, 'goods imply standards for achieving them' (*ibid*., p. 516). McLaughlin (2009) agrees, noting that (*ibid*., p. 98):

> When we view a causal chain as a series of means and ends, we presuppose something that stops the regress, something that has a good. And this applies whether it is an intentional agent, an organism, or simply anything that can be said to have interests—whether or not it consciously takes interest in them. We *presuppose* an entity somewhere down the line which has some kind of interests that (ceteris paribus) ought to be served. [original emphasis]

One way of summarizing much of the dense network of mutual implication formed by these concepts—a way that is pithy and highlights the central role of the notion of need—is the following (Lowe, 2008, p. 209):

> Just as a true belief is one which *corresponds to fact,* so a good action is one which *corresponds to need.* In another idiom, just as facts are the *truth-makers* of true beliefs, so needs are the *goodness-makers* of good actions. [original emphasis]

However, though the concept of vital need lies close to the center of normativity in the broad sense, it does not quite lie at the very center. Need is not quite basic. That is because most of the functions that we associate with vital needs are instrumental, that is, relative or conditional in character. For example, most living things need to consume water in some form or other. One might suppose that water is an intrinsic need of, say, human beings, if one judged solely from the pleasure that we derive from drinking water when we are thirsty. But of course we all know very well that it is not the quenching of thirst *per se* in which our vital need for water really consists. Rather, thirst is merely the sign by which our need for water is brought to our conscious awareness. A man lost in the desert might well be able to put up with mere thirst, no matter how terrible, if he did not know that the need represented by the thirst must be fulfilled if he is to go on living. The point is an obvious one that does not require belaboring. To put it in the most general way, I propose the following definition of (vital) need:

> *(Vital) Need:* A biological function is constituted as a (vital) need only in relation to a normative state of affairs such that the state of affairs can only be preserved by the proper exercise of the function.

This raises the fundamental question: What is the normative state of affairs that is logically prior to the concept of need? There are two obvious candidates. One is 'life' (or, perhaps, 'survival' or 'reproduction'). The other is 'well-being' (or 'welfare' or 'flourishing'). Detailed discussion of the definition of life must be postponed until a future occasion. For now, let us focus on the latter concept, of well-being or flourishing.

Kraut (2007) states the basic idea of well-being or flourishing very simply (*ibid*., p. 5): 'For most living things, to flourish is simply to be healthy: to be an organism that is unimpeded in its growth and normal functioning.' He goes on to show how the concept has nothing whatever to do with sapience or sentience, but is clearly properly ascribable even to plants (*ibid*., pp. 6–7):

[12] If anyone is tempted to say that the point of securing the water is primarily to satisfy my thirst, not to save my life, he is raising an interesting issue that opens out into a number of side-paths. For example, sometimes shipwrecked sailors may drink sea water, even in full knowledge that doing so spells death. There is no space to explore this complication adequately here, but let me make two quick points. First, the sailors will surely hold out against their thirst as long as possible, so long as their reason and will are intact. This proves that in their own minds the end of quenching their thirst is secondary and instrumental to the end of preserving their life. Second, at the end of the day we must explain the very existence of thirst in terms of the need of the organism for water, which again shows that the preservation of life is conceptually prior to the quenching of thirst.

[13] Stuart Kauffman offers the suggestion of 'yum' and 'yuck' (Kauffman, 2000; Kauffman & Clayton, 2006; Kauffman et al., 2008), which, in addition to wit, has the virtue of minimal ambiguity. His intended application of these terms to single cells may be controversial, but at least in human terms, who would deny that when I say 'yum,' I am saying of something that I find it 'good,' and likewise for 'yuck' and 'bad'?

> Such terms as 'welfare,' 'well-being,' and 'utility' are seldom, if ever, applied to plants. But it is just as obvious a point about plants as it is about animals that some things are good for them and other are not. If something can flourish or fall short of flourishing, that by itself shows that we can speak of what is good for it.

In another passage, he is even more explicit on the main point at issue (*ibid*., p. 9):

> Plants do not have minds. And yet some things are good for them: to grow, to thrive, to flourish, to live out the full term of their lives in good health. Whatever impedes this—diseases, droughts, excessive heat and cold—is bad for them.

In other words, logically speaking well-being is not essentially connected with sapience or sentience, but is rather connected with the fundamental vital functions as such.

Foot (2001) makes a very similar point, though she uses the slightly different terminology of 'natural goodness'; from the context, though, it is clear that she could just as well say 'well-being' or 'flourishing' (*ibid*., pp. 26–27):

> [. . .] 'natural' goodness, as I define it, which is attributable only to living things themselves and to their parts, characteristics, and operations, is intrinsic or 'autonomous' goodness in that it depends directly on the relation of an individual to the 'life form' of its species.

Here, we have finally reached rock bottom in our search for original or underived normativity. The notion of well-being or flourishing is as basic as it gets. The only way to go deeper is to pass from our everyday vocabulary altogether and venture onto the terrain of the natural sciences, in order to investigate in what the well-being and flourishing of living things consists, from a scientific point of view. That is, to go deeper we must pass from the Scope Problem to the Ground Problem, and inquire into the physical nature of life itself—a task that must be reserved for a future occasion.

3.2. The elementary normative concepts are constitutive of agency

The second consideration states that collectively the elementary normative concepts comprise or constitute our concept of (normative) agency. No detailed discussion is required here. This claim can be amply justified by simply observing that all of the elementary normative concepts discussed in the previous subsection are connected in one way or another with the concept of acting for a reason. Conversely, a direct analysis of the concept of acting for a reason reveals its fundamentally teleological (means-end) structure (behavior lacking a teleological structure does not count as action),[14] from which flow the concepts of value and the 'instrumental ought,' from which in turn flow the concepts of need and well-being. Agency—the capacity of acting for a reason—then, is implied by the elementary normative concepts, and the elementary normative concepts imply agency. Agency is not something over and above the elementary normative concepts. Rather, agency is a complex concept consisting of a number of different aspects, and some of these various aspects are captured by the individual elementary normative concepts.

3.3. Certain of the elementary normative concepts are properly ascribable to organisms as such

Everything that has been said so far tends to reinforce the intuition we began with—namely, that it is perfectly proper to ascribe normative concepts in a literal way to living systems as such. If only one or two of the concepts were clearly so ascribable—say, need or purpose—then one might perhaps dismiss that fact as a quirk of the language. But if all of the elementary normative concepts are so ascribable, and especially if all of them seem to stand in the same, densely interconnected, network-style relationship to one another when considered in their application to living systems generally as when considered in their application to human beings, then it becomes much more difficult to argue that the identity of the conditions of application of the concepts in the two cases is merely accidental, and of no importance for our understanding of the real nature of things. On the contrary, there seems to be a genuine mystery here that cries out for an explanation. Why do the world and our way of thinking and talking about it seem to conspire to give every appearance that normativity and agency are objectively real features of organisms, if in fact they are not?

We have already shown that some of the elementary normative concepts, such as purpose, need, and well-being, are clearly ascribable to some of the lower life forms, such as plants. Indeed, this is abundantly clear from ordinary language and our everyday experience of the world. Plants *need* water (need). Water is *good* for plants (value). It is *unhealthy* for a plant to go too long without water (well-being). Some plants turn their leaves toward the sun *in order to* capture more light (purpose). To capture more light is the *reason* why some plants turn their leaves toward the sun (having a reason for action). So much is, or ought to be, tolerably obvious.

Nevertheless, for many readers, I suppose that the conclusion of the Scope Argument—the proper scope of application of our concept of normative agency is living systems as such—will seem so difficult to believe as to constitute as *reductio* of the Scope Argument as a whole. If one looks for a claim to dispute as a result of taking the argument as a *reductio*, that claim will most likely be the one relating to the proper ascribability of any of the elementary normative concepts to organisms as such. For this reason, I will spend a little extra time attempting to provide independent motivation for the acceptance of this consideration.

The crucial point is to see that the ascription of normativity to living systems (organisms) as such is not only a matter of how we ordinarily speak. If that were the case, then indeed we could not accept the truth of this claim with such certainty. After all, ordinary language might be mistaken on this point, since it developed before so much was known about the material constitution of organisms. But it is not just ordinary language that sanctions the ascription of normativity to organisms, it is biological science itself. Let us see how.

3.3.1. The ascription of normativity: the case of bacterial chemotaxis

Take, for example, bacteria. Many bacteria, such as *E. coli,* swim about by means of a locomotory faculty called 'chemotaxis.'[15] Such bacteria are capable of engaging in two forms of locomotion, or 'motility.' In the first form (called 'running'), the bacteria swim in a straight line. In the second form (called 'tumbling'), they move about at random. At the molecular level, the bacteria contain a locomotory assemblage, which is basically a protein motor that causes external appendages called 'flagella' to rotate, either counterclockwise (for running) or clockwise (for tumbling). This motor is connected to a sensory assemblage, consisting of a complex, transmembrane, protein-receptor array that is sometimes referred to as a 'nanobrain' (e.g., Webre et al., 2003). The inner workings of this nanobrain, as well as its chemical linkages to the motor, are immensely complicated, but, in a nutshell, the organ enables the bacterium to sample

[14] See, e.g., Delancey (2006), Foot (2001), Okrent (2007), Schueler (2003), Sehon (2005), and Wilson (1989).

[15] For brief descriptions and interpretative discussion, see Shimizu & Bray (2003), Wadham & Armitage (2004), and Webre, Wolanin, & Stock (2003); for full technical details, see Stock & Surettte (1996).

its external environment for a large number of chemical compounds, to compare the concentrations of these compounds at different times, in this way to determine whether the concentration of a given compound is increasing or decreasing between samplings, and thus to determine whether it is traveling in a favorable or unfavorable direction (where 'favorable' means traveling toward an attractant or away from a repellent, and 'unfavorable' means the reverse). Finally, by means of its nanobrain the bacterium adjusts the setting of its motor so that if it finds itself swimming in a favorable direction it continues running (i.e., it continues traveling in the same direction) and if it finds itself swimming in an unfavorable direction it begins tumbling (i.e., it tries a different direction).

The elucidation of many of the molecular details of all of this, which are of staggering complexity, represents an outstanding achievement of contemporary science (even if many things remain to be worked out). How we should understand the relationship between those molecular details and the apparent normative agency of the bacterium in exercising its locomotory faculty is an important theme that I hope to address on another occasion. For now, I would like to point out just that the concepts of normativity and agency do indeed seem to apply in the case of bacterial motility, as just described.

Thus, we may begin with the observation that bacteria need various nutrients, such as lactose, sucrose, and other sugars. Without such nutrients, a bacterium will die. This of course presupposes that self-preservation in life is normative, and death something to be avoided. Indeed, 'health,' 'vigor,' 'vitality,' 'viability'—all of these are descriptors that scientists commonly use to refer to the well-being of living things, including individual cells. For example, Campbell (2008, p. 2386) claims that '[m]echanical forces, generated while cells migrate, are important for maintaining a healthy cell,' while Lloyd and Hayes (1995) expressly ascribe the notions of 'vigor,' 'vitality,' and 'viability' to microorganisms.[16] Given this norm of well-being and the needs generated by it, nutrients then may be said to be good for a bacterium—that is, they are 'to be pursued.' Thus, the bacterium's motility is end-directed, or purposive. Moreover, a bacterium 'should' swim toward its nutrients (if it does not, there is something wrong with it). If it senses that it is swimming in the right direction (toward its nutrients), then it has reason to continue swimming in the same direction, that is, to run (by rotating its flagella counterclockwise). All of this makes it seem natural to say that swimming toward its food is something that the bacterium does, not something that happens to it. In short, bacteria act.

All of this may be said quite naturally, without in any way forcing the language. There is no sense that in describing a bacterium's swimming toward its food as the bacterium's acting, we have slipped somewhere from speaking the literal truth to speaking in poetic fancies or metaphors. That is not to say, of course, that how such descriptions sound to the untutored ear settles the matter. There are certainly objections that can be raised at this point, and I will address some of them presently. Nevertheless, in the ensuing discussion, it is important for us to keep in mind that this way of describing the faculty of motility in even the lowly bacterium is perfectly natural, and that this fact is a significant one.

There is one objection that may be advanced against the preceding argument, which can, I believe, be dispensed with fairly quickly. One might say that the biologists themselves do not use this sort of normative language to describe bacterial motility. Or, to be more precise, they attempt to avoid using such language wherever possible, though they are seldom successful in suppressing normative vocabulary entirely for any length of time.[17] Still why not take our cue from the biologists' own practice? Rather than speak of the bacterium's 'pursuing the good,' or even 'swimming toward its food,' why not just speak of its 'following a positive attractant gradient'? But notice that this locution is itself a metaphor. After all, bacteria are not 'attracted' up a chemical gradient in the same way that iron filings are 'attracted' to a magnet.[18] Bacterial motility is not a matter of a direct reaction to impressed forces or of a tight coupling to an external field. Chemical gradients do not 'pull' bacteria along; rather, bacteria carry their own principle of motion within them. They move, as we might say, 'of their own accord.' That is, they control what they do in such a way that they swim up only those gradients that are good for them. Therefore, motility is not something that merely happens to bacteria, but rather something that bacteria achieve or accomplish. And that is just another way of saying that bacteria 'act.' Therefore, in point of fact, it is the commonsense normative, agential descriptors of bacterial motility that are literal, and the descriptors that employ physico-chemical terminology known not to be strictly applicable that are metaphorical. Such metaphors amount to a kind of euphemism—an effort to avoid the natural way of describing phenomena such as bacterial motility in terms of normativity and agency.

However, there is a more penetrating form of the foregoing objection that cannot be dismissed so easily. Some might claim that, rather than focusing on whether bacterial motility at the whole-system level is more properly described as 'pursuing the good' *versus* 'following an attractant gradient,' we ought rather to consider the fact that both sorts of descriptions have (supposedly) been rendered redundant by our knowledge of the molecular details of the chemotaxis subsystem. The idea would be that both sorts of whole-system–level descriptions are little more than convenient verbal summaries that stand in for the myriad physical and chemical details of what is transpiring at the molecular level. In principle, then, if not in practice, one should be able to explain bacterial motility by referring to events exclusively at the molecular level. And indeed if it were true that all the causal work was being done at that level, then, by the parsimony principle, we really should avoid ascribing any ontological significance to whatever purely verbal formulations we may use to summarize those events for our own convenience at the whole-system level.

This sort of objection might seem to be open to the same reply as before—namely, that living systems are not passively swept along by external causes, but rather are active in the pursuit of their own interests. However, this time, when the objection is expressed in its more radical form, a ready rejoinder becomes apparent. That is the following claim. Science has now (for all practical purposes) fully explained in molecular detail how organismic subsystems like the bacterial chemotaxis locomotory system work. That is, we are now in possession of a (for all practical purposes) complete understanding of the internal 'mechanisms' that give rise to the behavior of bacterial motility. While it is true that that type of behavior is very different in detail from the movement of iron filings in a magnetic field, nevertheless—so the argument goes—we are now in a position to see that there is no deep difference in principle. Everything is still happening according to the laws of physics and chemistry; it is just that those laws work themselves out in a special way in certain kinds of systems, which we call 'organisms.' But that is no problem, because we can fully

[16] Of course, such usage of normative concepts by scientists does not in itself show that the concepts cannot be given a reductive analysis. While there is an extensive philosophical literature on the concept of 'health' (Ereshefsky, 2009), most of it focuses solely on human beings, and simply presupposes the natural/normative dichotomy at issue here. Wachbroit (1994) importantly shows that the notion of biological 'normality' is irreducible to a nonnormative, statistical concept.

[17] Cf. almost any page of any molecular or cell biology textbook, to say nothing of works on physiology or animal behavior.

[18] Historically, the metaphor must have run the other way—from personal or sexual attraction to magnetic 'attraction.' But if biologists today speak of a bacterium's food as an 'attractant,' it is surely in order to assimilate its behavior more closely to that of iron filings, and not that of young lovers.

explain that special way the laws of physics and chemistry have of working themselves out in the case of organisms, by supplementing those laws with a few metaphysically unproblematic auxiliary concepts, such as 'negative feedback control,' 'fitness,' 'natural selection,' and a few others. The capstone of this line of thinking is the observation that we ascribe normative, agential descriptors to manmade machines, as well as to organisms. For example, I might well say that my car 'needs' gasoline; that the 'purpose' of the gasoline is to make the car go; that if the fuel tank is nearly empty, then gasoline 'should' be added; that a nearly empty fuel tank is a 'reason' for gasoline to be added; etc. And an automobile, too, is not ordinarily moved about willy-nilly by external forces, but rather contains its own principle of motion within it. In this sense, it too moves 'of its own accord.'

Since the 'machinery' of bacteria is now known to be no different, in principle, from the machinery of automobiles—or so it is claimed—and since we ascribe the same sort of normative, agential descriptors to both kinds of systems, should we not then view organisms and machines as belonging to the same natural kind? Not to put too fine a point on it: Shouldn't we simply say that organisms are machines? And if that is so, then we need not worry about which vocabulary we use. Just as I feel free to say that my car 'needs' gasoline, all the while realizing that this is just an elliptical way of describing how the car operates internally, so too (on this view) I should feel free to say that *E. coli* 'need' sucrose, all the while realizing that this is just an elliptical way of describing how bacteria operate internally.

There are two kinds of responses that one might make to this suggestion. One would be to retreat to the position that there is no fundamental difference between organisms and machines, after all, and give up the aim of naturalizing normativity altogether, except by elimination. This is the way urged upon us by Lenman (2005). In a penetrating discussion of McDowell, Foot, Hursthouse, and other 'liberal' naturalist authors, he refuses to accept their finding of normativity in the natural inclinations of living things. For example, he writes (*ibid*., p. 46):

> A nurturing polar bear father ... is certainly behaving in a way that may surprise ethologists and we may classify it accordingly as defective in a *very* deflated sense of that word. But surely that's just classification. How does something that deserves to be called *authority* get into this picture? That's the mystery. A greenhouse full of plants is a space full of healthy and less healthy specimens, specimens that promise to reproduce and live a long time, and specimens that do not. Sure it does. But, except when you are inside it, there are no *reasons* in your greenhose. No *normativity*, certainly no *authority*, merely a space in which certain natural dispositional properties are distributed in certain ways. [original emphasis]

On the next page, Lenman goes on to invoke Williams's (1995, p. 110) dictum that the complete absence of teleology from nature is the 'first and hardest lesson of Darwinism,' one which we have yet to take sufficiently to heart.

Lenman's paper is of the first importance because it poses in stark and vivid terms the precise challenge to which any realistic effort to naturalize normativity must respond. But it is not as though there were an actual argument in the quoted passage; rather, Lenman simply assumes that organisms are mechanistic systems to which normative concepts may not properly be ascribed. But of course that is the very point at issue. The reason he is able to get away with such flagrant question-begging is that he is working against the background of near-universal agreement with his presupposition that organisms are machines.[19] Therefore, in the final analysis there is no way to respond to Lenman's challenge effectively other than by providing an alternative account of what organisms could be, such that normative agency might be properly ascribable to them.

The other type of response would be to take the bull by the horns and explain why organisms are not machines—that is, why organisms constitute a natural kind, but manmade machines do not. It is easy enough to say (what is obviously true) that organisms have 'original' or 'intrinsic' normativity, while machines have 'derived' or 'extrinsic' normativity. But what does that mean? What is original or intrinsic normativity? After all, organisms are physical systems, are they not? How, then, exactly, do they differ from machines?

This is the master question. To pose this question is to ask about the ultimate ground of normativity in nature. Unfortunately, the detailed investigation of this question must await a future occasion. For now, let us turn to the conclusion of the Scope Argument before concluding.

3.4. The proper scope of our concept of normative agency is living systems as such

In addition to defending the above considerations, I shall support the Scope Argument by attempting to defend the conclusion directly. If the conclusion can be rendered more plausible on independent grounds, then this suspicion that the overall argument amounts to a *reductio* will pose less of a problem.

To this end, I would like to review some considerations that have been introduced recently into the literature on the philosophy of action by some of our most distinguished philosophers working in this field. For the most part, they take their arguments to apply to the higher animals alone, but after reviewing some of them, I will show that they are equally applicable to organisms as such.

Let us start with a distinction of Railton's (2009). He notes that much of our action is the result, not of rational deliberation, but rather of more or less automatic practical skills or competences—what he calls 'fluent agency.' Then, he notes that rational deliberation presupposes fluent agency (*ibid*., p. 103):

> I have no quarrel with treating deliberate choice as one paradigm in the theory of rational or autonomous action—it is certainly an important phenomenon for any such theory to explain. My argument instead is that it cannot be the fundamental phenomenon, for it is built up from, and at every step involves, the operation of countless non-deliberative processes that are—and must be—quite unlike choice. These processes are not self-aware or reflective, yet they are intelligent and responsive to reasons *qua* reasons. They make us the agents we are, and give our agency its capacity for rational, autonomous self-expression.

Railton does not discuss the other animals, but his notion of fluent agency would seem to apply to them as well. Certainly, such notions as automatic skills or competences and fluidity of motion would seem to apply to the pouncings of cats and the acrobatics of squirrels in a perfectly literal way. There remains the issue of whether such behaviors are responsive to reasons *qua* reasons. This is, of course, the crucial point. As it happens, a number of philosophers have recently begun to argue that the behaviors of at least the higher animals are responsive to reasons in the right way, and thus do qualify as 'actions' in the normative sense.

[19] Davidson's seminal contributions (e.g., 2001a, 2001b) played an important historical role in framing the action debate in this way. For argument that Davidson's position is indeed question-begging in essential respects, see Finkelstein (2007, especially, p. 267).

For instance, Steward (2009a) believes that it is not necessary to ascribe intentions to the higher animals in order to accept that they are in an important respect the authors of their own actions. Thus, she writes the following, appealing essentially to our commonsense way of speaking and thinking about animals (Steward, 2009b, pp. 303–304):

> And I should like to insist that the idea that an *animal* might be able to produce a bodily movement, so far from being a strange piece of metaphysical lunacy seems to be part and parcel of an everyday picture of the world with which we are very comfortable. It is not at all obvious that there must be something deeply wrong with it. Animals have many powers—what is so strange about the idea that one of the types of powers of which they are possessed is the power to control in certain respects movements (and other changes) in their own bodies? [original emphasis]

Korsgaard's (2009) view of the matter is similar. Though she is more willing than Steward is to ascribe intentions to the higher animals, her reasoning here, like Steward's, remains anchored in our commonsense way of understanding animal behavior (Korsgaard, *ibid*., p. 90):

> Human beings are, after all, not the only creatures who act. The distinction between actions and events also applies to the other animals. A non-human action, no less than a human one, is in some way ascribed to the acting animal herself. The movements are her own. When a cat chases a mouse, that is not something that happens to the cat, but something that she does. To this extent, we regard the other animals as being the authors of their own actions, and as having something like volition.

Glock (2009) is still more explicit about the propriety of ascribing intentional states to the higher animals (*ibid*., p. 242):

> Both in everyday life and in science we explain the behaviour of higher animals by reference to their beliefs, desires, intentions, goals, purposes. These psychological explanations are not causal, at least not in the sense of efficient or mechanical causation. Instead they are intentional in the sense explained above, just as our explanations of human behaviour. In both cases we employ intentional verbs, and we explain the behaviour by reference to the fact that *A* believes that *p*, desires *X*, wants to Φ, etc.

Boyle and Lavin (2010, p. 178) agree, observing that the general form of explanation of which intentional explanation is an instance 'can apply to nonrational animals and indeed to plants. Its application marks the feature of living things we are tracking when we say that what goes on with them is subject to teleological explanation.'

Finally, Hurley (2003) has addressed the issue of rational deliberation in this way (*ibid*., p. 231):

> [. . .] acting for reasons does not require conceptual abilities—not, at least, the full-fledged context-free conceptual abilities associated with theoretical rationality and inferential promiscuity. I appeal to practical reasons in particular to argue that the space of reasons is the space of actions, not the space of conceptualized inference or theorizing.

Hurley goes on to raise the issue of whether we can properly speak of a non-human animal's reasons for action as being the animal's own reasons, as opposed to its behavior's being merely conformable to reasons supplied by a human observer, as suggested by Dennett's (1987) notion of the 'intentional stance.' Here is how she puts this point (*ibid*.; 233):

> It may still be objected that while there may *be* reasons to act that an agent has not conceptualized, these cannot be the agent's *own* reasons, reasons for the agent, at the personal or animal level (see and cf. Dennett, 1996, chap. 5, 6). [original emphasis]

And here is what she says immediately in reply:

> I disagree. I understand reasons for action at the personal or animal level in terms of the requirements of holism and normativity. Perceptual information leads to no invariant response, but explains actions only in the context set by intentions and the constraints of at least primitive forms of practical rationality.

In these passages, Hurley corroborates my conclusion that subrational animals may properly be said to act intentionally, and to be agents.

So far, I have only discussed reasonsfor ascribing literal normative agency to the higher animals. Apart from Steward, the reason cited was basically that the higher animals appear to have intentional states like ours. This material was rehearsed in order to respond to the traditional concerns of many if not most philosophers of action who have usually assumed that literal normative agency ought to be ascribed only to rational beings like us. But even if the position of Steward and the others were to be accepted, that would still leave me only half-way to my stated goal. For, I wish to claim, not just that normativity and agency exist objectively in relation to the higher animals, but that they exist objectively in relation to organisms as such. That is a bridge too far for Steward and the others, and is denied with a greater or lesser degree of explicitness by all of them.

What are some of their reasons for resisting the more radical move I am urging? Interestingly, it does not seem to be the issue of intentionality that is of primary concern to them (that is to say, none of them argues that action is conceptually linked to conscious intentions). Rather, they make two basic points.

The first point is that they are reluctant to ascribe normative agency to living systems that do not meet some threshold of flexibility of behavior, or 'intelligence.' The idea seems to be that if the system's behavior is sufficiently stereotyped, then it is simply 'automatic' or 'mechanical,' and no longer meets the criterion of normative agency. Thus, Hurley (2003) contrasts animals with intentions to those supposedly operating according to 'invariant' stimulus-response relations (*ibid*., pp. 235–236).

There are two different kinds of responses that one might give to this worry. First, as the details of the chemotaxis system outlined above suggest, the behavior of lower organisms is not really as stereotyped as one might think. In fact, it has been observed that no two bacteria can be counted on to respond in precisely the same way to identical environmental circumstances, not even if they are genetically identical (Zimmer, 2008, pp. 44–49).[20] In general, one may say that the idea of a rigid stimulus-response relation in the lower organisms is something of a myth. Most of the behavior even of the lower organisms is in fact endogenously generated (Brembs, 2010; Heisenberg, 2009; Maye, Hsieh, Sugihara, & Brembs, 2007; Prete, 2004; Simons, 1992; Trewavas, 2009). Moreover, it is now beginning to be acknowledged that the capacity for flexible, purposive behavior is the key to the 'robustness,' or stability, of the cell, and ultimately of all living things. For example, Kirschner and Gerhart (2005, pp. 107–108) have put this point as follows:

> The organism is not robust because it has been built in such a rigid manner that it does not buckle under stress. Its robustness stems from a physiology that is adaptive. It stays the same, not

[20] See, also, Trewavas, 1999.

> because it cannot change but because it compensates for change around it. The secret of the phenotype is dynamic restoration.

Indeed, Kirschner (2010, p. 3803) goes so far as to claim that 'all of biology is built on the dynamic and adaptive capacity of the cell.'[21] On this view, 'adaptive capacity' is tantamount to an elementary form of 'cognition' or 'intelligence' that is an inherent property of living things as such.[22]

Nevertheless, it would of course be foolish to deny that the behavior of bacteria is relatively speaking far more stereotyped than that of higher organisms like cats and dogs. It is important, therefore, to add—and this is the second response to the first worry—that intelligence is not really a relevant criterion for assessing whether agency is properly ascribed to a system. Rather, responsiveness to reasons is the relevant criterion. And as we have seen above, however limited a bacterium's behavioral repertoire may be compared to a higher animal's, it clearly passes that test with flying colors.

The second worry raised by several of our authors relates to the fact that we commonly ascribe agency only to whole animals, and not to their component parts. Thus, Hurley (2003) clearly states that '[...] I understand the subpersonal level as the level of causal/functional description at which talk of normative constraints and reasons no longer applies' (*ibid.*, p. 234), and the other authors make similar remarks.

This makes intuitive sense, and does reflect common sense, which has been one of our chief guides so far. However, we must be attentive here to a distinction that is too easily blurred. It is one thing to say that agency is properly ascribable to whole organisms, and not to their parts. It is something else to say that whole organisms are endowed with a power of agency only over the movements of their bodies as a whole, or over the movements of the external parts of their bodies, and not over the processes internal to their bodies. I am going to argue that there is no good reason in principle to withhold ascription of objectively normative agency to an organism's control of its own internal processes.

I agree, of course, that agency is conceptually linked to the capacities of a system as a whole (and I will examine in detail what this condition amounts to, in a future publication). But it does not follow that internal processes cannot be actions of a system, for there remains the possibility that the system as a whole may actively control its own component parts.[23]

Burge (2009) gives us a clear account of what this holistic requirement involves (*ibid.*, p. 260):

> I think that the relevant notion of action is grounded in functioning, coordinated behavior by the whole organism, issuing from the individual's central behavioral capacities, not purely from sub-systems.

This criterion can clearly be met with respect to the active control of a whole system's component parts, just so long as the parts are controlled by the whole system, and not the other way around. For example, consider the difference between voluntary and involuntary actions within your own body.

We have voluntary control over several of the component parts of our body. Examples include the thoracic diaphragm (breathing), the esophagus (swallowing, belching), the bladder, and the rectal sphincter, not to mention the skeletal muscles.[24] Let us consider breathing. No one, I take it, will deny that by holding my breath for a minute while I am under water, I am acting. And yet, the same internal part (namely, the thoracic diaphragm) is being controlled just as surely when that control is involuntary (i.e., unconscious) as when it is voluntary (conscious). In both cases, the control has exactly the same function—that is, it occurs for basically the same reasons. In both cases, the reason for the occurrence of the internal processes is the introduction of air (containing oxygen) into the respiratory and eventually the circulatory systems. The only difference is that voluntary breathing permits an additional layer of control, permitting greater responsiveness to environmental contingencies, like the need to exclude water or other foreign substances. In short, from the point of view of why the body does what it does, voluntary control of breathing is just more of the same of what is already provided by involuntary control of breathing. Therefore, it is hard to see what principled reason one could give for saying that the voluntary control of breathing qualifies as a normative action while the involuntary control of breathing does not.

I conclude from this example that there is no good reason to deny that, in principle, the whole organism can be in control of its component parts.

4. Conclusion

On the basis of commonsense linguistic usage and conceptual analysis, as well as the Scope Argument, I conclude that there is no principled reason for maintaining that normativity and agency are properties of human beings alone, or even that they are properties of the higher animals only. If that is the case, then we are faced with a decision (assuming we do not wish to be outright dualists) between accepting eliminativism and seeing ourselves as mere machines devoid of any genuine normativity, on the one hand, and seeing all living systems (organisms) without exception as normative agents, on the other. Nothing I have said here excludes our taking the eliminativist path. However, assuming that we opt to follow common sense in viewing ourselves as genuine normative agents, then the arguments I have deployed in this essay lead to the conclusion that we must also accept the objective existence of normativity and agency in organisms as such.

Acknowledgements

I would like to express my deepest gratitude to Lenny Moss and Phillip R. Sloan for their unwavering loyalty and support over the years.

References

Albrecht-Buehler, G. (2009). Cell intelligence website. <http://www.basic.northwestern.edu/g-buehler/FRAME.HTM> Accessed 11.04.10 (last updated 24.07.09).

Bedau, M. A. (1992). Goal-directed systems and the good. *Monist, 75*, 34–51.

Bedau, M. A., & Humphreys, P. (Eds.). (2008). *Emergence: Contemporary readings in philosophy and science*. Cambridge, MA: Bradford Books/MIT Press.

[21] See, also, Harold (2001). Piersma & van Gils (2011), and Turner (2007), take a similar view of the adaptive capacity of higher animals.

[22] There is no space here to analyze this controversial claim, but for the idea that 'intelligence' may be properly ascribed to living things as such, see Albrecht-Buehler (2009), Ben-Jacob (2009a, 2009b), Ben-Jacob & Levine (2006), Ford (2009), Shapiro (2007), and Trewavas (2003, 2005, 2010). For the closely related view that living processes are inherently 'cognitive,' see Calvo & Keijzer (2009), Heschl (1990), Lyon (2006), Stewart (1996), and van Duijn, Keijzer, & Franken (2006).

[23] Frankfurt (1997) raises an objection to this line of reasoning when he asserts that the concept of control or guidance is intuitively linked to the conscious actions of whole persons. As he remarks of pupil dilation (*ibid.*, p. 46): 'The guidance in this case is attributable only to the operation of some mechanism with which [the person] cannot be identified.' But this objection fails to take into account the fact that it is the whole organism, not the person *qua* rational agent, with which such subpersonal instances of control are to be identified, as well as the fact that such control (or 'regulation') is routinely attributed by scientists to biological systems.

[24] The case of the skeletal muscles contains the complication that the voluntary control of the internal part (the muscle) is simultaneously manifested externally (by the movement of the corresponding limb), and some might wish to ascribe the agent's control in such cases solely to the external manifestation. For simplicity's sake, I set this case aside.

Ben-Jacob, E. (2009a). Learning from bacteria about natural information processing. In G. Witzany (Ed.). *Natural genetic engineering and natural genome editing (=Annals of the New York Academy of Sciences)* (Vol. 1178, pp. 78–90). Boston, MA: Blackwell Publishers on behalf of the New York Academy of Sciences.

Ben-Jacob, E. (2009b). Bacterial complexity: More is different on all levels. In S. Nakanishi, R. Kageyama, & D. Watanabe (Eds.), *Systems biology: The challenge of complexity* (pp. 25–35). Tokyo: Springer.

Ben-Jacob, E., & Levine, H. (2006). Self-engineering capabilities of bacteria. *Journal of the Royal Society Interface, 3*, 197–214.

Boyle, M., & Lavin, D. (2010). Goodness and desire. In S. Tenenbaum (Ed.), *Desire, practical reason, and the good* (pp. 161–201). Oxford: Oxford University Press.

Brembs, B. (2010). Towards a scientific concept of free will as a biological trait: Spontaneous actions and decision-making in invertebrates. *Proceedings of the Royal Society of London B, 278*, 930–939.

Burge, T. (2003). Perceptual entitlement. *Philosophy and Phenomenological Research, 67*, 503–548.

Burge, T. (2009). Primitive agency and natural norms. *Philosophy and Phenomenological Research, 79*, 251–278.

Calvo, P., & Keijzer, F. (2009). Cognition in plants. In F. Baluška (Ed.), *Plant–environment interactions: From sensory plant biology to active plant behavior* (pp. 247–266). Berlin: Springer.

Campbell, I. D. (2008). Croonian lecture 2006: Structure of the living cell. *Philosophical Transactions of the Royal Society of London B, 363*, 2379–2391.

Clayton, P., & Davies, P. (Eds.). (2006). *The re-emergence of emergence: The emergentist hypothesis from science to religion*. Oxford: Oxford University Press.

Churchland, P. M. (2007). *Neurophilosophy at work*. Cambridge: Cambridge University Press.

Corradini, A., & O'Connor, T. (2010). *Emergence in science and philosophy*. New York: Routledge.

Davidson, D. (2001a). Mental events. In *idem, Essays on actions and events* (2nd ed., pp. 207–224). Oxford: Clarendon Press (Originally published in Foster, L. & Swanson, J. W. (Eds.). *Experience and theory* (pp. 79-101). Amherst: University of Massachusetts Press, 1970).

Davidson, D. (2001b). Rational animals. In *idem, Subjective, intersubjective, objective* (pp. 95–105). Oxford: Clarendon Press (Originally published in *Dialectica*, 1982, 36, 317–327).

Davies, P. S. (2009). *Subjects of the world: Darwin's rhetoric and the study of agency in nature*. Chicago: University of Chicago Press.

Delancey, C. (2006). Action, the scientific worldview, and being-in-the-world. In H. L. Dreyfus & M. A. Wrathall (Eds.), *A companion to phenomenology and existentialism* (pp. 356–376). Malden, MA: Blackwell.

Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: Bradford Books/MIT Press.

Dennett, D. C. (1996). *Kinds of minds: Toward an understanding of consciousness*. New York: Basic Books.

Ereshefsky, M. (2009). Defining 'health' and 'disease'. *Studies in History and Philosophy of the Biological and Biomedical Sciences, 40*, 221–227.

Fine, K. (2002). The question of realism. In A. Bottani, M. Carrara, & P. Giaretta (Eds.), *Individuals, essence and identity: Themes of analytic metaphysics* (pp. 3–48). Dordrecht, Netherlands: Kluwer Academic.

Finkelstein, D. H. (2007). Holism and animal minds. In A. Crary (Ed.), *Wittgenstein and the moral life: Essays in honor of Cora Diamond* (pp. 251–278). Cambridge, MA: Bradford Books/MIT Press.

Foot, P. (2001). *Natural goodness*. Oxford: Clarendon Press.

Ford, B. J. (2009). On intelligence in cells: The case for whole cell biology. *Interdisciplinary Science Reviews, 34*, 350–365.

Frankfurt, H. G. (1997). The problem of action. In A. R. Mele (Ed.), *The philosophy of action* (pp. 42–52). Oxford: Oxford University Press (Originally published in *American Philosophical Quarterly*, 1978, 15, 157–162).

Gillett, C. (2007). Understanding the new reductionism: The metaphysics of science and compositional reduction. *Journal of Philosophy, 104*, 193–216.

Glock, H.-J. (2009). Can animals act for reasons? *Inquiry, 52*, 232–254.

Harold, F. M. (2001). *The way of the cell: Molecules, organisms, and the order of life*. Oxford: University of Oxford Press.

Heisenberg, M. (2009). Is free will an illusion? *Nature, 459*, 164–165.

Heschl, A. (1990). L=C: A simple equation with astonishing consequences. *Journal of Theoretical Biology, 145*, 13–40.

Hornsby, J. (1997). *Simple mindedness: In defense of naïve naturalism in the philosophy of mind*. Cambridge, MA: Harvard University Press.

Hurley, S. (2003). Animal action in the space of reasons. *Mind and Language, 18*, 231–256 (Reprinted with revisions as 'Making Sense of Animals,' in Hurley S. & Nudds, M. (Eds.). *Rational animals*? (pp. 139–171). Oxford: Oxford University Press, 200.).

Kauffman, S. A. (2000). *Investigations*. Oxford: Oxford University Press.

Kauffman, S., & Clayton, P. (2006). On emergence, agency, and organization. *Biology and Philosophy, 21*, 501–521.

Kauffman, S., Logan, R. K., Este, R., Goebel, R., Hobill, D., & Shmulevich, I. (2008). Propagating organization: An inquiry. *Biology and Philosophy, 23*, 27–45.

Kim, J. (1998). *Mind in a physical world: An essay on the mind-body problem and mental causation*. Cambridge, MA: Bradford Books/MIT Press.

Kirschner, M. (2010). Cell biology as a world view. *Molecular Biology of the Cell, 21*, 3803.

Kirschner, M. W., & Gerhart, J. C. (2005). *The plausibility of life: Resolving Darwin's dilemma*. New Haven, CT: Yale University Press.

Korsgaard, C. M. (2009). *Self-constitution: Agency, identity, and integrity*. Oxford: Oxford University Press.

Kraut, R. (2007). *What is good and why: The ethics of well-being*. Cambridge, MA: Harvard University Press.

Lenman, J. (2005). The Saucer of Mud, the Kudzu Vine and the Uxorious Cheetah: Against Neo-Aristotelian Naturalism in Metaethics. *European Journal of Analytical Philosophy, 1*(2), 37–50.

Lloyd, D., & Hayes, A. J. (1995). Vigour, vitality and viability of microorganisms. *FEMS Microbiology Letters, 133*, 1–7.

Lowe, E. J. (2008). *Personal agency: The metaphysics of mind and action*. Oxford: Oxford University Press.

Lyon, P. (2006). The Biogenic Approach to Cognition. *Cognitive Processing, 7*, 11–29.

Maye, A., Hsieh, C., Sugihara, G., & Brembs, B. (2007). Order in spontaneous behavior. *PLoS ONE, 2*(5), e443.

McDowell, J. (1998). Two Sorts of Naturalism. *In idem, Mind, value, and reality* (pp. 167–197). Cambridge, MA: Harvard University Press (Originally published in Hursthouse, R., Lawrence, G., & Quinn, W., (Eds.). *Virtues and reasons: Philippa Foot and moral theory* (pp. 149–179). Oxford: Clarendon Press, 1996).

McLaughlin, P. (2009). Functions and norms. In U. Krohs & P. Kroes (Eds.), *Functions in biological and artificial worlds: Comparative philosophical perspectives* (pp. 93–102). Cambridge, MA: MIT Press.

Okrent, M. (2007). *Rational animals: The teleological roots of intentionality*. Athens, OH: University of Ohio Press.

Piersma, T., & van Gils, J. A. (2011). *The flexible phenotype: A body-centered integration of ecology, physiology, and behaviour*. Oxford: Oxford University Press.

Prete, F. R. (Ed.). (2004). *Complex worlds from simpler nervous systems*. Cambridge, MA: Bradford Books/MIT Press.

Railton, P. (2009). Practical competence and fluent agency. In D. Sobel & S. Wall (Eds.), *Reasons for action* (pp. 81–115). Cambridge: Cambridge University Press.

Savellos, E. E., & Yalçin, U. D. (Eds.). (1995). *Supervenience: New essays*. Cambridge: Cambridge University Press.

Schueler, G. F. (2003). *Reasons and purposes: Human rationality and the teleological explanation of action*. Oxford: Clarendon Press.

Sehon, S. (2005). *Teleological realism: Mind, agency, and explanation*. Cambridge, MA: Bradford Books/MIT Press.

Shapiro, J. A. (2007). Bacteria are small but not stupid: Cognition, natural genetic engineering, and socio-bacteriology. *Studies in History and Philosophy of the Biological and Biomedical Sciences, 38*, 807–819.

Shimizu, T. S., & Bray, D. (2003). Modelling the bacterial chemotaxis receptor complex. In G. Bock & J. A. Goode (Eds.), *In Silico Simulation of Biological Processes* (pp. 162–177). Chichester, UK: John Wiley & Sons (Novartis Foundation Symposium 247).

Simons, P. (1992). *The action plant: Movement and nervous behaviour in plants*. Oxford: Blackwell.

Steward, H. (2009a). Animal agency. *Inquiry, 52*, 217–231.

Steward, H. (2009b). Sub-intentional actions and the over-mentalization of agency. In C. Sandis (Ed.), *New essays on the explanation of action* (pp. 295–312). London: Palgrave/Macmillan.

Stewart, J. (1996). Cognition = life: Implications for higher-level cognition. *Behavioural Processes, 35*, 311–326.

Stock, J. B., & Surette, M. G. (1996). Chemotaxis. In F. C. Neidhardt (Ed.). *Escherichia coli and salmonella: Cellular and molecular biology* (2nd ed., Vol. I, pp. 1103–1129). Washington, DC: ASM Press.

Trewavas, A. (1999). The importance of individuality. In H. R. Lerner (Ed.), *Plant responses to environmental stresses: From phytohormones to genome reorganization* (pp. 27–42). New York: Marcel Dekker.

Trewavas, A. (2003). Aspects of plant intelligence. *Annals of Botany, 92*, 1–20.

Trewavas, A. (2005). Plant intelligence. *Naturwissenschaften, 92*, 401–413.

Trewavas, A. (2009). What is plant behaviour? *Plant, Cell and Environment, 32*, 606–616.

Trewavas, A. (2010). The green plant as an intelligent organism. In F. Baluška, S. Mancuso, & D. Volkmann (Eds.), *Communication in plants: Neuronal aspects of plant life* (pp. 1–18). Berlin: Springer.

Turner, J. S. (2007). *The tinkerer's accomplice: How design emerges from life itself*. Cambridge, MA: Harvard University Press.

van Duijn, M., Keijzer, F., & Franken, D. (2006). Principles of minimal cognition: Casting cognition as sensorimotor coordination. *Adaptive Behavior, 14*, 157–170.

Wachbroit, R. (1994). Normality as a biological concept. *Philosophy of Science, 61*, 579–591.

Wadham, G. H., & Armitage, J. P. (2004). Making sense of it all: Bacterial chemotaxis. *Nature Reviews: Molecular & Cell Biology, 5*, 1024–1037.

Webre, D. J., Wolanin, P. M., & Stock, J. B. (2003). Bacterial chemotaxis. *Current Biology, 13*(2), R47–R49.

Williams, B. (1995). Evolution, ethics, and the representation problem. In *idem, Making sense of humanity, other philosophical papers, 1982–1993* (pp. 100–110). Cambridge: Cambridge University Press (Originally published in Bendall, D. S. (Ed.). *Evolution from molecules to men* (pp. 554–566). Cambridge: Cambridge University Press, 1983).

Wilson, G. M. (1989). *The intentionality of human action, revised and enlarged ed*. Stanford, CA: Stanford University Press (Originally published in 1980).

Zimmer, C. (2008). *Microcosm: E. coli and the new science of life*. New York: Pantheon.

Studies in History and Philosophy of Biological and Biomedical Sciences 43 (2012) 104–112

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

Natural sources of normativity

Wayne Christensen

Konrad Lorenz Institute for Evolution and Cognition Research, Adolf Lorenz Gasse, 2, Altenberg A-3422, Austria

ARTICLE INFO

Article history:
Available online 22 June 2011

Keywords:
Normativity
Naturalism
Autonomous systems
Functions
Reasons
Persons

ABSTRACT

Normativity is widely regarded as being naturalistically problematic. Teleosemantic theories aimed to provide a naturalistic grounding for the normativity of mental representation in biological proper function, but have been subject to a variety of criticisms and would in any case provide only a thin naturalist platform for grounding normativity more generally. Here I present an account that identifies a basic form of valuational normativity in autonomous systems, and show how the account can be extended to encompass key aspects of the normativity of functions and practical reasons.

1. Introduction: normativity and naturalism

Normativity is paradigmatically a matter of right and wrong, good and bad. Philosophical work on normativity seeks to understand the nature of normative claims, the nature of justification for such claims, and the fundamental sources of normativity. One common view is that there is nothing in the natural world, accessible by scientific means, which grounds normative claims. The most influential arguments to this effect are due to Hume and G. E. Moore: Hume argued that no normative conclusion can be validly derived from descriptive premises (Hume, 1978), whilst Moore's 'open question' argument asserts that any attempt to identify a normative property (e.g., goodness) with a natural property (e.g., pleasure) is always open to doubt, thus showing that conceptually the two cannot be identical (Moore, 1971). The popularity of this view is probably due to a more complex set of influences than just the force of these arguments, however. Lurking in the background are a pair of ideas that tend to work hand-in-hand: on the one hand, the idea that modern science replaced Aristotelian teleology with mechanistic explanation, and on the other, the idea that normativity is a very special feature of human agency, linked to consciousness and perhaps the capacity for reflection.

Whatever the exact reasons, it is often thought that naturalistic theory should not stray over the putative fact/value boundary. Yet naturalist theory in this mode must overcome a major obstacle, which is that normativity seems to be an endemic and very important feature of human agency. Not only moral agency, but cognitive agency more broadly. Representations can misrepresent, words can be used wrongly, people can leap to irrational conclusions, and they can act unwisely. If adopting a scientifically based perspective means giving up normativity, this is giving up a lot. Naturalists practicing an austere norm eliminativism aim to show that these phenomena can be understood without appeal to normative concepts, despite appearances to the contrary, but it is not unreasonable to doubt that the project can succeed. Normative eliminativism may be an unnecessary straightjacket, however. Here I will sketch a naturalist approach that follows Aristotle in recognizing relatively rich forms of normativity in living systems. Specifically, it sees normativity as inherent in the organization or form of living systems, specifically in the form that generates their unity and hence explains their existence.

The most immediate point of comparison for this account is the etiological theory of normative function. The approach to functional normativity advocated here differs in fundamental ways with the etiological theory, and indeed with most other contemporary accounts of normative function, inasmuch as it begins with a different explanatory agenda. On the usual conception the task of function theories is to explain how functions are assigned to parts,

E-mail address: wayne.christensen@gmail.com

1369-8486/$ - see front matter © 2011 Published by Elsevier Ltd.
doi:10.1016/j.shpsc.2011.05.009

whereas the approach taken here instead focuses on explaining value in relation to systems, and much of the emphasis is on identifying the relevant class of systems. This is done by means of a theory of the fundamental organization of living systems. The basic idea is not especially novel: as noted, it treads in the footsteps of Aristotle, and there are a variety of contemporary theories that attempt to give an account of the organization of living systems, which often assume that functional normativity pertains to these systems in virtue of their organizational structure.[1] Here I attempt to flesh out the intuition in a way that relates it to a broader understanding of normativity.

2. Normativity: some basic distinctions

Before proceeding further it will help to sketch out the nature of normativity in a little more detail. This cannot be done in an uncontroversial way, but the following distinctions capture at least approximately some of the major forms of normativity that have been discussed (see e.g. Darwall, 2001; Glüer & Wikforss, 2009; Schroeder, 2008). The initial description given above associates normativity with evaluation and prescription, but some have identified a kind of normativity referred to as 'descriptive.' Descriptive or 'non-evaluative' norms are such that it is possible to specify conformance or departure from the norm, but there is no reason from this alone to think that there ought to be conformance to the norm, or that nonconformance is bad. Etiological proper functions (discussed in the next section) are thought by most proponents to have descriptive normativity (Neander, 2009). If we include such non-evaluative norms within the realm of the normative then the minimal kind of normativity may simply involve some kind of non-arbitrary framework allowing comparison between actual and alternative states. There is room to doubt that this is sufficient for normativity, but it may at least be necessary.

In the case of 'evaluative' normativity the comparison between actual and alternative states takes the form of a valenced assessment. 'Valuation' (traditionally addressed by axiology) involves assessments such as 'good,' 'better than,' and 'worse than.' 'Prescription' (traditionally addressed by deontic theory) specifies what ought or ought not to happen, with the biblical commandment 'thou shalt not kill' being a paradigm example of a (candidate) prescriptive norm. 'Constitutive norms' specify rules which must hold if something is to exist, such as the rules of a game like chess. They are per se non-evaluative, though they can inform evaluations in conjunction with other information, such as an agreement (perhaps tacit) to play by the rules. With regard to games and other activities we can further distinguish 'performance norms,' concerned with how well the game or activity is conducted, with winning, losing and 'playing well' counting as paradigmatic performance norms.

3. Etiological theories of normative function

Since the mind is often thought to be entirely or at least substantially functional in nature, theories of normative function are an obvious starting point for developing naturalist accounts of the normativity of cognitive phenomena. The teleosemantic program takes this route, attempting to ground the normativity of mental representation in biological function (Millikan, 1984; Papineau, 1984). Causal theories of representation, such as that of Dretske (1981), attempt to explain to explain mental representation in terms of causally based correlations. Thus, activity in a toad's retina is correlated with events in the world, and thereby represents those events. The familiar problem is that understanding representation in terms of causal correlation leaves no room for misrepresentation, because correlations either exist or they do not, they cannot be 'false.' But there do seem to be false representations. Teleosemantics offered a solution by appealing to an etiological theory of normative function. It specifies what a representation is supposed to represent in terms of the 'proper function' of the mechanism doing the representing; thus, toads will respond to a long dark horizontally moving stimulus as if it is a worm, and it seems reasonable to think that this is what the detection system in their brain is supposed to indicate. In the lab they respond to artificial stimuli created by the scientist, but in these cases they are misrepresenting. The etiological theory of normative function explains proper function in terms of natural selection: the proper function of an item is the function it is adapted to perform. This has been an appealing pathway for a naturalist account of normativity because normativity is explained by appeal to a natural phenomenon (evolutionary adaptation) that is relatively well understood, clearly of great importance, and is intuitively normative (as the putative basis of 'biological design').

Nevertheless, for some progress with this approach has not met expectations (e.g. Godfrey-Smith, 2006). Specific difficulties in the analysis of representational content need not concern us here, however some of the deeper and thornier issues stem from the basic source of normativity. As noted above, the normativity of etiological functions is supposed to be descriptive rather than evaluative. Thus, on Millikan's account the proper function of a heart is to do what ancestor hearts did that made them the target of selection. But identifying this putative proper function will not allow us to conclude that this heart now *ought* to do what it's ancestor hearts did, or that it is bad if it does not. By avoiding evaluation the theory evades Hume and Moore, however the result is a very thin and somewhat peculiar kind of normativity. Deviance from an ancestral state subject to positive selection is called 'malfunction,' but malfunction defined this way is not really 'mal': there is nothing inherently bad about it (cf. Ferguson, 2007). Indeed, an etiologically defined malfunction may be functionally advantageous in the current context. It would be clearer and more accurate to replace the terms 'proper function' and 'malfunction' defined according to etiological theory with technical labels that have no evaluative associations. For instance, we could replace 'proper function' with 'AS-function' (for 'ancestrally-selected function'), and replace 'malfunction' with 'C-function' (for 'changed function'). With these substitutions the etiological theory no longer appears normative, which suggests that it is getting illegitimate normative 'oomph' by means of evocative labels. Without this oomph the grip provided by the theory is unconvincing: as we saw, the etiological theory is the grounding point for the teleosemantic account of misrepresentation, but misrepresentation defined this way is not really malfunctioning, or 'incorrect' functioning, it is just different functioning. If we think there is something genuinely incorrect about misrepresentation then we need more resources than the etiological account is providing.

The etiological approach is pseudo-prescriptivist in the sense that it gives something of the flavor of prescriptivity without the actual prescriptivity. It does not aim to explain valuation, nor does it support valuational assessment for the reasons just given. Yet, at least on first appearances, biological functioning seems to involve valuational normativity: an organ can function well or poorly, and an organism can be healthy or sick. Since etiological theory has nothing to say about these kinds of phenomena (again, talk of 'malfunction' is deceptive), it would seem to be at best

[1] These theories usually focusing on self-maintenance and/or self-production; see e.g. Schrödinger (1944), Maturana & Varela (1980), Bickhard (1993), Christensen & Hooker (2000a, 2000b), Christensen & Bickhard (2002), Kauffman (2003), Moreno, Etxeberria, & Umerez (2008), Barandiaran, Di Paolo, & Rohde (2009) and Toepfer (this volume).

incomplete insofar as it is supposed to account for functional normativity in biological systems. Certain theoretical alternatives have been rejected on grounds that will be considered next, but for reasons that are unclear the literature on normative function has been fixated on a prescriptivist model of functional normativity, according to which functional normativity consists in the possession by an item of a proper function, which the item is in some sense 'supposed' to perform (Wright, 1973).[2] The issues of normative perspective and functional value have been largely neglected.

4. Autonomous systems and normative function

Proponents of the etiological account of normative function have been drawn to it in part because they are skeptical of system theories of function. Cummins (1975) is thought to have provided the canonical account of system-based analysis of function, and on his account functional analysis is interest relative, in the sense that more or less anything can be given a systems functional analysis and ascribed function on that basis. Just as we can analyze the propensity of hearts to pump blood, we can analyze biologically irrelevant relations such as the contribution of the heart to body mass (Sober, 1993), or the propensity of mice to explode in space (Millikan, 1989). A particularly crucial claim is that the boundaries of a biological system cannot be identified on the basis of causal relations; only by identifying proper functions can a biological system be individuated in a principled way (Millikan, 1999). The importance of this claim is that it identifies a putative in-principle limitation. Cummins did not attempt to give a principled account of system individuation or normative function analysis, but this is not a reason to think it cannot be done. However if biological systems cannot be causally individuated that is a reason to think that any purely systemic analysis must be arbitrary in focus.

There are compelling reasons to reject the claim, however. As a matter of epistemology it cannot be the case that the principled identification of biological systems depends on the prior identification of etiological proper functions, because a theory of the adaptive origins of a trait is generated by analyzing the effect of the trait on the organism's ability to survive and reproduce in the ancestral environment in which it appeared (Griffiths, 1993; Stotz & Griffiths, 2001). As a matter of ontology we should expect that living systems are physically individuated, because if they were not then there would be no physically distinct entities for natural selection to select over. If there are physical principles of individuation they can be used to supplement the systems framework for functional analysis that Cummins describes, providing a non-arbitrary basis for functional analysis.

This is the approach taken by Christensen and Bickhard (2002), who argue that an account of autonomy can serve as the basis for a suitable theory of individuation and system organization.[3] This account has an Aristotelean flavor because it relates the organization systems to their unity and existence conditions. In a broad sense the organization of a system is just how it is arranged, and a basic question to ask for natural systems is how their organization is related to their existence. Often, many aspects of a system's organization make little difference to its ongoing existence. Consider a collection of rocks scattered across the floor of a dry cave: the rocks have many different shapes and sizes, but for the most part the differences in shape, size and location will not affect the ongoing existence of the rocks in this very stable environment. On the other hand, on an open plain differences in shape, size and location will tend to have a stronger effect on rock existence because of their effects on weathering processes.

For some systems there is a very special relationship between their organization and their existence because (unlike rocks) they actively construct the conditions which give them unity and ongoing persistence. The concept of autonomy is intended to capture this idea, and according to the analysis of autonomy given by Christensen and Bickhard a system is autonomous if it tends to generate the conditions for its persistence, and if it has infrastructure that contributes to this self-maintenance. Infrastructure here refers to persistent, relatively stable structure that shapes more dynamic system-maintaining processes, with the cell membrane of living cells being a paradigm example. The infrastructure requirement rules out simple positive feedback systems, and indeed more complex feedback systems like tornados, which are self-maintaining but lack infrastructure that supports self-maintenance.

Establishing a principled basis for system individuation is a crucial anchor point for an account of normative function, because normative evaluations of function can be made relative to system identities. That is, functional relations can be assessed in terms of their effect on the system. The contribution of the heart to body mass clearly has little significance for the organism, whereas the contribution of the heart to fluid transport has profound importance. Christensen and Bickhard (2002) develop an account of normative function of this kind. It is not intended to explain how parts 'have' functions that they 'should' perform, it is instead intended to provide a valuational account of functional relations relative to the system as a whole.

Normative properties such as benefit and dysfunction are characterized. An item is beneficial for an autonomous system if it contributes positively to the autonomy of the system, and this can be so regardless of the whether the item 'has' the function of making this contribution. Similarly, if a system is autonomous it will be composed of a network of interdependent processes, and we can understand dysfunction in terms of these interdependencies. If the heart stops beating then there will be a cascade of failures as physiological processes that depend on fluid transport cease to function, leading to the death of the organism. The dysfunction here is systemic—a property of the pattern of network dependencies—and as such *not* attributable to the heart in isolation. If an alternative mechanism for fluid transport appears, such as an artificial heart, the dysfunction goes away. Again, these network dependencies can be analyzed quite independently of what functions 'belong' to the various parts.

One line of response the etiologist might take is that these kinds of network analyses are merely descriptive. Mirroring the critique of the etiological account given above, it could be said that terms like 'benefit' and 'dysfunction' are a colorful and misleading way of describing purely physical relations. We can say that a given item is or is not affecting the system in some way, but without the etiological account there is no basis for comparing actual performance against a normative benchmark. The etiological account

[2] The problem of pseudo-normativity is not specific to the etiological theory; it arguably afflicts any account that tries to assign proper functions without a well-founded normative perspective. Thus, the system-based account of proper functions proposed by Schlosser (1998), and the development-based account of design (and proper function) by Krohs (2009), are both vulnerable to the type of criticism here leveled at the etiological account: the putative proper functions are not actually normative, and the use of the evaluative terms 'proper function' and 'malfunction' is misleading. Schlosser (1998) offers a system-based theory of normative function, but his account is structurally similar to the etiological account in that it aims to explain proper functions, and it doesn't give an account of normative perspective or functional value. Krohs (2009) treats development rather than evolutionary history as the source of design for biological systems, and he considers this design to be non-intentional. One worry is whether the concept of non-intentional design really makes sense, but even if we accept the idea it will not support any substantive normativity.

[3] The particular account of autonomy they use is previously developed in Christensen & Hooker (2000a, 2000b). Bickhard's account of self-maintenant systems is described in Bickhard (1993, 2000). More generally, this theory of autonomy is one of a family of theories that aim to give a systemic account of living systems—see footnote 1 above. Autonomy as it is used in this context must of course be distinguished from personal autonomy (Buss, 2008).

provides at least this basic form of normativity by means of the distinction between what an item does and what it is for, and it is unique amongst naturalistic forms of function ascription in doing so.

This line of criticism is not strong, however. The autonomous systems account does provide a basis for normative comparison: using the autonomy of the system as a whole as the grounding point, we can compare actual state with alternative states that would be better or worse. Moreover, in this kind of analysis words like 'benefit' and 'dysfunction' are not misleading. One way to gauge this is by the fact that it is not possible to replace them with technical terms that have no evaluative content. We can certainly replace the words, but if we leave out the evaluative content we lose information. Thus, the systems account has a firmer normative basis than the etiological theory: it uses evaluative concepts in an informative way.

Another possible line of criticism is that autonomous systems do not have a legitimate normative perspective. We can talk about certain things being 'good' or 'bad' for these kinds of autonomous systems, but this is ultimately just as empty as talking about certain things as being 'good' or 'bad' for a rock. We can imaginatively think of the breaking of a rock as bad for the rock, but rocks do not really have the kind of normative perspective that warrants such evaluation. An initial response to this criticism is to point out that according to the theory autonomous systems are causally special in a way that makes them unlike rocks. Autonomous systems are organized such that they tend to be self-perpetuating, and they have infrastructure which supports this self-perpetuation. We cannot properly understand the causal structure of these systems if we do not recognize that they are organized in a way that achieves self-perpetuation. And to understand this self-perpetuation at a more fine-grained level we must characterize the relations between the persistence of the system as a whole, and the constituent structures and processes.[4] Thus, the use of evaluative concepts is not simply an imaginative projection, it is required to properly characterize the causal structure of these kinds of systems. Moreover, in many cases the infrastructure possessed by these systems is regulative: it repairs, avoids, seeks, etc. In a limited but significant way living systems are doing their own evaluation, which is a persuasive reason for treating them as having a genuinely normative perspective. The etiologist may point out that these systems have infrastructure for self-perpetuation largely as a result of an evolutionary history. The autonomous systems account does not deny this, but nevertheless insists that the key perspective for normative evaluation of function is the current system rather than past selection. Regulation does not succeed by making parts function as they did in the past, it succeeds by making the system work well in present conditions.

This is only a provisional response; a detailed theory of what it is to have normative perspective is needed. The autonomy-based account at least takes some steps in this direction, and in this respect does more than the etiological theory. The etiological theory offers no explicit account of normative perspective, and the normative perspective that appears in the account is rather dubious. It is perhaps also worth briefly noting that the kind of normative perspective proposed by the autonomous systems account differs from a traditional consequentialist view in ethics in that it does not depend on the experience of pleasure and pain. It also does not appeal to a capacity for reflection, as with personal autonomy, though it does emphasize self-governance, albeit of a much simpler kind. The autonomous systems account does not aim to explain moral responsibility, or why certain entities should be the object of moral regard, so the type of normative perspective it proposes should be distinguished from these kinds.

5. Design and purpose

One reason why the autonomous systems account may *seem* less normative than the etiological account is because it does not appeal to design. Intuitions about design are a major buttress for the etiological account, and in particular the intuition that intricate functional structure of biological systems is a lot like the functional structure of artifacts. One of the key sources of normativity for artifacts (and their parts) is the intentions of the designer. An artifact is constructed so that the parts will interrelate in a way that conforms to the plan of the designer, and conforming to this plan is supposed to allow the artifact to perform its intended functions. By analogy we can regard biological organization as 'virtual design' (cf. Dawkins, 1986; Kitcher, 1993). In contrast, the autonomous systems account makes no appeal to design, and so is approaching the issue of normative function in biological organization from a very different direction.

Although the appeal to design helps give intuitive force to the etiological approach, it is not clear that this support is legitimate for the kind of reasons described in the previous section. That is, the analogy between biological systems and artifacts is questionable in exactly the ways that bear on normative function: 'mother nature' is not a real agent with real design intentions. In any case, though, even for artifacts, where there are real agents, we should not treat design intentions as the sole source of normativity, and perhaps not even the most fundamental source. An artifact can fail functionally even though it conforms to its design plan, and it may be functionally successful despite not conforming to the designer's intentions. Valuational normativity must be more fundamental than design normativity because we need it to understand design itself: things do not work well just because they have been designed—designers try to design things that will work well.

Rather than base the normativity of artifacts on design we could base it on purposes (Franssen, 2006), treating design as just one source of purposes, with the users of artifacts being another. This would help explain how an artifact might be functionally successful despite not performing according to its design, because normativity is also being conferred by the purposes of users. However a very broad understanding of purposes will be needed. Our interaction with artifacts is complex, especially in the case of artifacts that are themselves complex systems, such as buildings, aircraft, power stations, rail systems, computers, computer networks, etc. It can be difficult and sometimes impossible for the designers of such artifacts to understand in detail how the users will respond to the artifact, and the users themselves may not understand their interactions with the artifact particularly well, not least because tacit learning plays a major role. Moreover, we often adjust our purposes through experience with an artifact, both abandoning goals and discovering new ones. So the normativity of artifacts will not be well captured if it is thought of as entirely dependent on explicit psychological purposes. Although explicit purposes are undoubtedly an important source of normativity for artifacts, there also appears to be a form of normativity involving broader relations to human activity.

If purposes themselves are normatively constrained, such that we can have the wrong purposes, and through learning acquire better ones, then we need a deeper form of normativity. The autonomous systems account supplies one proposal for what this normativity might be. Artifacts are not themselves autonomous systems,[5] and only acquire normativity through their relations to users, who are autonomous systems. Admittedly human autonomy has rich psychological structure, but this psychological structure,

[4] For a more detailed discussion see Christensen & Bickhard (2002).

[5] In the future we may develop the ability to construct autonomous systems.

rather being the fundamental source of normativity, is itself normatively constrained in virtue of being embedded in an autonomous system, which it helps to steer. The sources of normativity for artifacts and living systems are thus quite different: artifacts derive their normativity from their relations to the living systems that use them, whereas living systems have their normativity indigenously, in virtue of being autonomous systems.

6. Reasons and persons

At this point it will help to revisit the question of what normativity is. On one view normativity is connected to reasons. Thus, '[a]spects of the world are normative in as much as they or their existence constitute reasons for persons, i.e. grounds which make certain beliefs, moods, emotions, intentions or actions appropriate or inappropriate' (Raz, 1999, p. 354). This way of framing normativity is helpful because it highlights three key issues: (i) the perspectives for which things matter (persons, according to Raz), (ii) the nature of mattering (that which makes facts about the world relevant to entities with a normative perspective, these relevance relations constituting normative facts), and (iii) the mechanisms by which the entities with normative perspective respond to normative facts (which Raz associates with rationality).

Because it emphasizes personhood as the basis for normative perspective, and rationality as the mechanism by which persons respond to normative facts, this way of conceptualizing normativity may seem to favor non-naturalism.[6] It at least confines normativity to the realm of rational agency, and thereby separates it from the broader natural world, even if we seek to treat persons and rationality naturalistically. This would be bad news for the autonomous systems approach because, although artifacts might derive normativity from their relationship to humans, much of the functional structure found in the living world is independent of humans and would not be normative.

However, the schematic structure of the conception of normativity given by Raz is congenial to the autonomous systems account, which proposes a similar structure involving more basic entities and mechanisms (Table 1). The autonomous systems theorist must reject the idea that personhood and rationality are normatively fundamental, and propose instead that these are grounded in the more basic kind of normativity identified by the autonomous systems account. Indeed, two kinds of grounding are on offer: origins and constitution (Fig. 1). With regard to origins, the basic idea is to treat personhood as just a particular kind of agency, and more specifically as a cognitively sophisticated form of agency that has evolved from more basic non-cognitive forms of agency. The normativity of personhood is an elaboration of the normativity of these simpler forms of agency, with special features arising from the psychological attributes of personhood, but also with a great deal of continuity. With regard to constitution, persons are not just descended from autonomous agents, they *are* autonomous agents: a person is constituted as a certain kind of autonomous agent in the base sense of autonomy, and this makes an important contribution to the normativity of personhood. The normativity of reasons is in part the normativity of the functional organization that constitutes the autonomous system that is the person. Psychological mechanisms that respond to this normativity are complemented by biological regulatory mechanisms.

Non-naturalists are likely to agree that there is some important story to be told about the relations described in 1a and 1b. However, they will restrict normativity to the far side of the arrows in each case. The idea will be that, although in the larger picture we are interested in knowing how cognitive agents have causally appeared in the world, and how they are causally instantiated, this is not the subject matter of normative theory as such, which is concerned only with the boxes the arrows point to. There are at least two kinds of response to a restrictionist view of this kind: (1) argue that structural parallels between the two boxes warrant extending normativity to encompass them both, and (2) argue that the restrictionist view will render its own subject matter incomplete and mysterious. The parallels described in Table 1 are a starting point for an argument of the first kind. Arguments for (2) can focus on the following two claims: (a) biological and psychological normativity is integrated in cognitive agents, and (b) psychological and biological mechanisms form a single normative response system (albeit imperfectly integrated) with numerous interdependencies, just as one would expect, given (a).

The following arguments provide some support for (a) and (b): (i) Psychological and biological mechanisms respond to the same normative facts. Thus, persons have reason to avoid consuming things that will make them ill. Decisions to avoid particular foods based on acquired knowledge and experience are a cognitive means for responding to this normative fact, whereas vomiting after ingestion is a biological regulatory mechanism for responding to the same fact. (ii) Biological mechanisms can respond to normative facts without the aid of psychological mechanisms. For instance, you eat food you believe is OK, but your body detects toxins and reacts with vomiting. (iii) Biological mechanisms can train cognitive mechanisms on which normative facts to recognize. You will for example learn to avoid foods that make you nauseous.

7. Applying the autonomous systems theory

Applying the autonomous systems theory of normative function to several examples will help make its structure clearer, and hopefully show its usefulness.

7.1. The remarkable re-wired ferrets

Experiments in ferrets have shown surprising functional plasticity in sensory processing areas of the brain. By deafferenting the auditory thalamus in ferrets at birth, Sharma, Angelucci, and Sur (2000) were able to induce retinal axons to innervate the medial geniculate nucleus (MGN), which is a relay to the ferret primary auditory cortex (A1). In other words, visual input in these 'rewired' ferrets was directed to the first cortical area involved in auditory processing. Histological examination of the affected cortical area showed that it had taken on structural characteristics (orientation modules) similar to primary visual cortex (V1) and unlike the typical organization of auditory cortex. Follow-up experiments reported in von Melchner, Pallas, and Sur (2000) addressed the question of how this highly abnormal area of primary sensory cortex was treated by downstream neural processing. A particularly intriguing question was whether, from the point of view of the ferret, input to the rewired sensory cortex would be treated as if it were visual or auditory. Given that no cortical areas downstream of A1 were directly affected by the intervention the most parsimonious prediction is that the ferrets will respond to stimuli processed through the re-wired A1 as if it is auditory.

Because only a known section of the retina was rerouted, with the rest of the retina connecting by the normal pathways to visual cortex, the issue could be tested by selectively presenting stimuli to the affected retinal areas. Restraining the ferrets in an apparatus ensured that a visual stimulus presented in external space would

[6] Olson (2009) argues that the recent emphasis on reasons has bolstered non-naturalism in meta-ethics, though not for the reasons suggested above. He argues that a reasons-based conception of normativity appears (wrongly) to make non-naturalism less metaphysically problematic than the Moorean version based on goodness.

Table 1
Structural parallels between reasons and autonomous systems conceptions of normativity.

	Normative perspective	Basis of mattering	Mechanism for responding to normative facts
Reasons conception	Persons	Relevance to the person	Rationality
Autonomous systems conception	Autonomous systems	Relevance to the autonomous system	Regulation

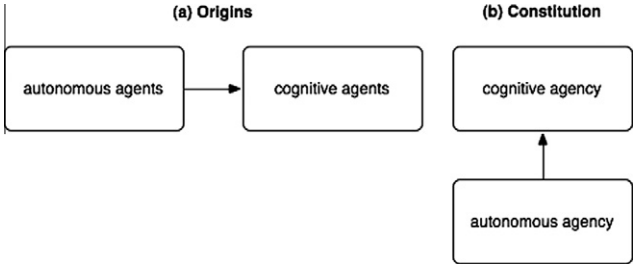

Fig. 1. Two kinds of naturalist grounding for the normativity of personhood.

be processed by a restricted area of the retina. Using signals routed through normal sensory cortex the ferrets were trained to visit a reward spout on their left for an auditory stimulus and a reward spout on their right for a visual stimulus. The key question was, when presented with stimulus that was selectively processed by the modified retinal area, whether the ferrets would respond as if they had received an auditory stimulus or a visual stimulus. The results indicated that they treated it as visual: they went to the right reward spout. This result is striking because not only did the affected area of primary auditory cortex remodel itself for visual processing, somehow downstream cortical areas were able to detect that the information stream was visual rather than auditory, and exploit the information functionally in the control of behavior. Nevertheless, though dramatic, these findings are consistent with a wide range of evidence indicating high levels of plasticity in neural processing (see e.g. Elbert, Heim, & Rockstroh, 2001).

This example illustrates some of the limitations of assigning functions to parts without regard to the whole system. Humans and animals show a robust ability to recover from serious brain injury, often by constructing highly unusual functional circuitry. What will work well in the here-and-now may be substantially different to what has worked in the past, and adaptive plasticity is a crucial mechanism that allows organisms to construct workable solutions in the here-and-now. It highlights the fact that what really matters, functionally, is that the current system work well.

Indeed, adaptive plasticity is one of the more important empirical phenomena that the autonomous systems account can help illuminate. There has been interest in the immune system for its role in distinguishing 'self' and 'non-self' (Tauber, 2010), but less appreciation that functional regulation in general poses questions about system identity and 'better' versus 'worse' normativity. Regulation alters system state in ways that, to be adaptive, must count as improvement for the system. It also acts to restrict what is incorporated into the system, and eject things that will have a negative effect. Cell membranes and the skin of multicellular organisms serve as regulated boundaries, whilst the regulation of material ingress via the mouth is especially complex and sophisticated, for obvious reasons. Sensory and motor systems provide more distal regulation of intake by means of approach and avoid behavior. Plasticity allows adjustment to local circumstances, but to be adaptive plasticity must be regulated so that it is shaped into functional forms that are beneficial for the system. The kind of dramatic neural plasticity seen in ferrets illustrates just how subtle and powerful the regulation can be. But behavioral learning is also a form of adaptive plasticity with profound effects. Almost all animals are capable of at least simple forms of learning, and many are capable of very complex forms (see e.g. Moore, 2004).

To develop a fundamental theoretical understanding of adaptive plasticity we need an account of the ontology of biological systems, and what counts as better or worse for them. Regulatory mechanisms depend on proxy information, e.g. a looming visual stimulus as a proxy for danger. The relation between the proxy signals and the underlying system conditions they regulate can be indirect and imperfect, but to be adaptive it must be the case that the regulatory mechanisms tend to have a beneficial effect. Thus, to understand the evolution of such mechanisms we need to understand: (a) the proximal discriminations made by regulatory mechanism and the alterations they induce in system state, and (b) what the system *actually* is, and what *actually* counts as better or worse for it.

7.2. Reconstructing Joe

In the not-too-distant future these theoretical questions concerning normative function are likely to gain increasing practical relevance. Consider Joe, a fighter pilot of the late twenty-first century who has been involved in a collision during combat training in low earth orbit. He is extracted from his damaged craft and rushed by ambulance shuttle to an earth hospital. Joe's main injury is that his right front cortex is smashed in from the anterior prefrontal cortex to the motor cortex. Fortunately, brain reconstruction can restore a high level of function, though Joe will never be quite the same as he was before the crash. His long-term declarative memories are largely intact, but a great deal of his motor skill is lost, as is much of the substrate for his higher order emotional and cognitive control. Because the reconstruction cannot exactly duplicate the organization of the lost neural tissue, and because Joe's intact brain will reorganize during the reconstruction process, there will inevitably be significant functional discontinuity between Joe's brain before the accident and after the reconstruction. Joe's doctors cannot simply recreate the functional system as it was before the accident, they have to create new functional structure that aims as best possible for the continuity of Joe-the-person in his new circumstances. To know how to intervene in Joe's brain his doctors will need to develop a rich understanding of Joe-the-person.

Joe's brain is unique; it has been molded by his genetics, development and idiosyncratic learning experience. Using scaffolded neurogenesis the doctors will begin to reconstruct the basic structures of the right frontal cortex, and induce the major projections to and from other cortical areas. But the generic neuroanatomical templates must be adapted to the specific structure of Joe's brain, and the fine-grained structure of the rebuilt neural systems must mesh well with the rest of his brain. As a fighter pilot Joe had extremely well developed higher cognitive control: excellent control of attention, high ability to maintain spatial and task awareness, excellent task management ability, and so on. A great deal of this cognitive control was provided by the brain areas now destroyed, and brain areas that were previously under complex patterns of regulation from the right frontal cortex are no longer experiencing this regulation, with the result that they will tend to disorganize and organize for other functions. Unless Joe's doctors take steps

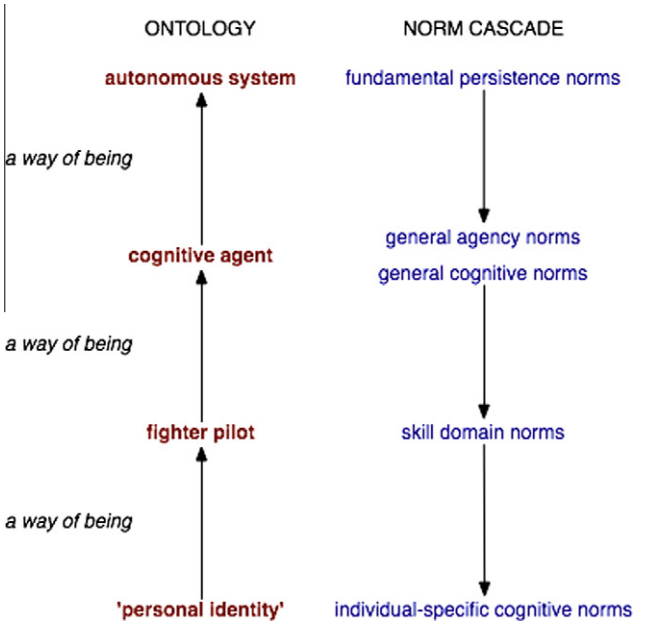

Fig. 2. System-based specification of Joe's ontology, together with associated norms.

to prevent it Joe's brain may start to form bizarre and dysfunctional patterns during the course of the reconstruction.

Thus, to achieve functionally successful reconstruction Joe's doctors will need to do more than regenerate the basic neuroanatomy of the right frontal cortex, they will need to exert precisely structured influence on the re-growing neural systems to shape them in relation to higher cognitive functions and the overall organization of Joe's brain, and Joe-the-person. Fig. 2 sketches a systems-based ontology for Joe, together with some of the kinds of norms that come into play with various aspects of the ontology. According to this ontology Joe-the-person is an individual with a particular kind of lifestyle, namely being a fighter pilot. In turn, as a fighter pilot Joe is at a more basic level a cognitive agent, and at an even more basic level he is an autonomous system. These are hierarchically structured forms of organization, each of which impose normative constraints. As an autonomous system Joe is subject to the very basic norms of existence for an autonomous system. As a cognitive agent Joe is subject to general norms for agency and cognition. Many of the core functional norms that apply in the reconstruction of Joe's brain come from here: to be a competent cognitive agent Joe will need amongst other things functional working memory, reasoning and higher order emotional regulation. As a fighter pilot norms for this skill domain apply, with excellent visuo-spatial working memory and fine motor control being some of the more important. If he is to return to his previous life the doctors will need to rebuild these capacities. There are also norms specific to Joe: Joe-the-person has a particular history and personality, specific friends and family, and particular cognitive skills acquired through idiosyncratic learning shaped by his particular cognitive strengths and weaknesses. To return to his previous life he needs to maintain and continue to develop his individual cognitive and social style, and his personal relations. He must not only be good at being a fighter pilot, but also good at being Joe.

It may be that it is not feasible to re-make Joe in a way that allows him to return to his old life, or at least not obvious that this is the right thing to do. The reconstruction of the basic neuroanatomy is likely to take many months, followed by a much longer period of therapy and training designed to induce the formation of the appropriate fine-grained functional organization, much as with the post-acute phase of stroke rehabilitation now. In the latter part of this century brain regenerative techniques are much more precise and effective than they are currently, involving carefully targeted cognitive and behavioral therapies complemented with direct neural therapies which include an array of implants that provide electrical stimulation, targeted delivery of growth factors, and structural scaffolding. Even so, after such a massive injury it will take years for Joe years to regain his former elite abilities, by which time the nature of his job will have changed and his former companions moved on. Joe's left, language-oriented hemisphere is intact, and he has always been witty and verbally gifted. He read extensively, and had talked with his wife and friends about becoming a writer after retiring from the military.

This presents Joe and his doctors with alternatives that become increasingly distinct as the rehabilitation process progresses. If the aim is to make him a fighter pilot then neurocognitive rehabilitation should focus on developing a quite different set of cognitive abilities than those required to be a writer. The training needed to develop either suite of abilities to an advanced level is intensive and protracted. Training that aims to help him become a writer will amongst other things emphasize higher order emotional and social cognition, rather than visuo-spatial cognition, motor control, and task management under time pressure. Joe is allowed to regain consciousness less than a week after the accident, but with heavy stabilization to damp disorganized brain activity in response to the injury, and traumatic psychological response to his situation. With his right frontal cortex destroyed his cognitive and emotional regulation is substantially impaired, and he will not be competent to make complex choices about his future until much later in the reconstruction. His wife and family are consulted extensively, and the doctors delay putting a particular functional orientation in the reconstruction until Joe's basic decision making ability improves.

The holistic structure of the normative constraints in this situation can be characterized in terms of a notional decision cycle (Fig. 3). First, there is an assessment to determine whether the system is or can be an autonomous system. In the case of Joe, the initial question is whether he will survive. The next step in the cycle is to determine, given the current system capacities, and intervention capacities, the best available state for the system. Here goals for the reconstruction are determined: re-establishing Joe as a competent cognitive agent, remaking Joe-the-person, giving Joe the abilities he will need for a suitable career. These goals guide the detailed neural interventions, and as the goals become increasingly specific Joe's doctors can work out in increasingly precise ways how particular parts of Joe's brain should be functioning during and after the reconstruction. To adapt to changing low and high order information Joe's doctors will go through many decision cycles with a similar structure to Fig. 3, though the emphasis will shift from whether he can survive to how he should be.

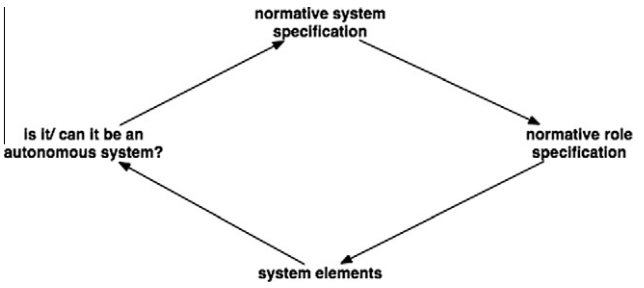

Fig. 3. Anotional decision cycle for determining how to intervene to improve function.

The dominant role of the second step (determining the best available system state) in shaping the third step (normative role assignment for system elements) reflects the holistic structure of the normativity of functional value, which has its source in the fact that the system as a whole constitutes the key normative perspective. Parts have no particular value or normative role independently, and history exerts no direct normative influence. History certainly exerts a powerful indirect influence: Joe has an extensive legacy from his life before the accident, and this legacy—including his intact memories and knowledge, existing personal relations, home and possessions, for example—tends to be more supportive of some ways of being compared with others. It in particular tends to be more supportive of his former way of life relative to alternatives. But Joe's past exerts normative influence only by shaping the ways of being available for Joe now, and the legacy from his past may not be decisive in favoring Joe's previous way of life, especially in the drastically altered circumstances he is currently faced with. Thus, to reiterate, the crucial normative perspective is Joe-now, and what possibilities there are for Joe-now.

Fig. 3 goes beyond the account given in Section 4 by incorporating a form of normative role assignment. The account of Section 4 only addresses value. In the human context, where we are considering direct intervention to modify function, the cognitive apparatus at work generates goals and plans, and role assignment. It may be that more limited forms of role assignment can be characterized in purely biological cases, but I will set aside this issue. The basic account of normative function described in Section 4 is agnostic with regard to role assignment.

7.3. Matilda's slippery relation to the truth

A very different kind of example may help to further clarify the system-relative nature of functional norms. Hattiangadi (2006) argues in favor of naturalism for meaning. She accepts the argument that if meaning has prescriptive normativity then it will violate Humean limitations on naturalist explanation, and that we will therefore need to take a non-naturalist approach to meaning. Her aim is to show that the normativity of meaning is descriptive rather than prescriptive,[7] and that naturalism is consequently safe. The gist of her argument that meaning normativity is descriptive is that, although meaning has correctness conditions—it is correct to apply the term 'horse' to X if and only if X is a horse—meaning is not prescriptive inasmuch as speakers are not obliged to use terms correctly.

Hattiangadi distinguishes hypothetical means/ends prescriptions, which specify conditional relations of the kind 'If you want to get to the airport on time, take a taxi' from categorical prescriptions, which are not conditional on ends. She claims that hypothetical means/ends prescriptions do not pose a difficulty for naturalism because they are not really prescriptive. Even the combination of a hypothetical means/ends prescription and the appropriate end is not prescriptive, because it might be the case that the end should be abandoned. Thus, even if it is true that, if you want to get to the airport on time, you should take a taxi, and it is also true that you want to get to the airport on time, it still does not follow you should take a taxi, because there might be reasons that make it better for you to abandon the goal of getting to the airport on time. Perhaps, rather than going on holiday, you should stay to look after your sick parent. Hattiangadi claims that the normativity of meaning is hypothetical and consequently merely descriptive. Accordingly, Matilda, who tells terrible lies, should use her words correctly if she wants to tell the truth, but she is not obliged to tell the truth.

However, even if Hattiangadi successfully shows that meaning is not prescriptive her argument leaves naturalism in an uncomfortable situation. After all, surely there will come a point where Matilda ought to do *something*. Either there is not, in which case it looks like there is nothing anyone ought to do, or there is, and it looks like naturalism will be in trouble at that point. Not perhaps in virtue of meaning normativity, but possibly in virtue of some norm of practical reason. Of course, the standard Humean answer is that what Matilda ought to do is determined by applying the maximizing principle to her total set of desires. This is naturalistically safe, supposedly, because no objective norms are appealed to, only Matilda's desires. However the Humean answer has some unattractive features. The maximizing principle itself appears to be a normative principle, so there is the threat of inconsistency (Wallace, 2008), but the approach also rules out the possibility that Matilda could have the wrong desires. It will fault her if her desires are inconsistent, but it does not allow that Matilda might have a consistent set of desires that are misguided.

It seems on the face of it quite possible that Matilda might have a misguided set of desires, and the autonomous systems view provides an account of what this might amount to. As a person she is a particular kind of autonomous system with a complex normative structure something like that depicted in Fig. 2. Her desires are proxies for what is actually good for her, and they can be misaligned with what is good for her, as well as with each other. This idea runs contrary to the Humean separation of fact and value, but the benefits are substantial. By not making purposes normatively primitive we get to understand the functional relations between psychological structures and the systems they steer. This helps us understand how these psychological structures have evolved, and how they can be improved in the here-and-now. Furthermore, the account of valuational normativity draws on naturalistically unproblematic resources.

The autonomous systems account challenges Hattiangadi's framing assumptions, but on the other hand it supports her idea that prescriptive normativity should not be directly associated with meaning. According to the autonomous systems account such normativity can only be assigned after taking into account the larger system. Within this framework meaning norms might be interpreted as constitutive norms pertaining to a certain kind of 'game,' specifically the 'game' of rational cognition and communication. Matilda-the-person is not obliged to play this game on all occasions, but as a cognitive agent she is structurally committed to being rational at least some of the time. That is, if she is not rational at least some of the time she will cease to be a cognitive agent. The constitutive norms of meaning gain valuational and prescriptive normative force for her because of this structural commitment: in thought and language it is sometimes good for her to use concepts and words correctly. In effect, the constitutive norms of meaning are also partly constitutive for her as a cognitive agent, and because she has normative perspective her constitutive norms have evaluative normativity (for her). But her agency does not depend on perfect conformance to the constitutive norms of meaning, and whether she should adhere to such norms on any given occasion is a function of a complex array of personal and situational factors. Taken individually and without regard to context, meaning norms are not prescriptive.

[7] She does not distinguish valuational from prescriptive normativity.

8. Conclusion: natural sources of normativity

Naturalist approaches to function, cognition and agency may have hobbled themselves unnecessarily by restricting themselves to 'descriptive' normativity. The putative normativity is thin at best, and without an account of valuational normativity we are left with an incomplete understanding of key phenomena like regulation and adaptive plasticity. Conversely, approaches that ground normativity in high-level features of human agency, such as personhood or purposes, also leave us with an incomplete and somewhat mysterious picture. The structures and capacities that support high level agency are themselves, arguably, constrained by broader forms of functional normativity. A naturalist approach that tackles evaluative normativity head-on, rather than skirting it, can provide a more coherent and informative picture.

The discussion here has aimed to make these claims plausible, but a detailed theory will need to address a number of difficult issues. The ontological analysis of autonomous systems must be defended, and a detailed argument linking the features of autonomy to normative perspective is required, together with a more general account of normative perspective. It will also be important to specify more closely how different global system states are to be evaluated as better or worse; in other words an account of flourishing is needed, and it must avoid succumbing to circularity or stipulation. The extension of the ontology of autonomous systems to encompass personhood, along the lines suggested in Fig. 1, will confront many highly contentious issues. It is not unreasonable to worry that the approach may not be able to satisfactorily carry through on these tasks, but equally, it seems worth trying. There is a prima facie basis for thinking that the sources of normativity, at least of the kind considered here, are natural systems amenable to broadly scientific understanding.

Acknowledgements

Preparation of this manuscript was supported by the Australian Research Council grant 'From Neuron to Self: Human Nature and the New Cognitive Sciences'. The ideas presented here build on earlier collaborative work conducted with Cliff Hooker and Mark Bickhard.

References

Barandiaran, X., Di Paolo, E., & Rohde, M. (2009). Defining agency: Individuality, normativity, asymmetry, and spatio-temporality in action. *Adaptive Behavior, 17*, 367–386.

Bickhard, M. (1993). Representational content in humans and machines. *Journal of Experimental and Theoretical Artificial Intelligence, 5*, 285–333.

Bickhard, M. (2000). Autonomy, function, and representation. *Communication and Cognition: Artificial Intelligence, 17*(3–4), 111–131.

Buss, S. (2008). Personal autonomy. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2008 Edition). <http://www.plato.stanford.edu/archives/fall2008/entries/personal-autonomy/>.

Christensen, W., & Bickhard, M. (2002). The process dynamics of normative function. *Monist, 85*, 3–28.

Christensen, W., & Hooker, C. (2000a). An interactivist–constructivist approach to intelligence: Self-directed anticipative learning. *Philosophical Psychology, 13*, 5–45.

Christensen, W. D., & Hooker, C. (2000b). Autonomy and the emergence of intelligence: Organised interactive construction. *Communication and Cognition: Artificial Intelligence, 17*(3–4), 133–157.

Cummins, R. (1975). Functional analysis. *The Journal of Philosophy, 72*(20), 741–765.

Darwall, S. (2001). Normativity. In E. Craig (Ed.), *Routledge encyclopedia of philosophy*. London: Routledge. <http://www.rep.routledge.com/article/L135> (retrieved 24.09.2003).

Dawkins, R. (1986). *The blind watchmaker*. Norton & Co.

Dretske, F. (1981). *Knowledge and the flow of information*. Oxford: Blackwell.

Elbert, T., Heim, S., & Rockstroh, B. (2001). Neural plasticity and development. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience*. MIT Press.

Ferguson, K. (2007). Biological function and normativity. *Philo: A Journal of Philosophy, 10*, 17–26.

Franssen, M. (2006). The normativity of artefacts. *Studies in History and Philosophy of Science, 37*, 42–57.

Godfrey-Smith, P. (2006). Mental representation, naturalism, and teleosemantics. In G. Macdonald & D. Papineau (Eds.), *Teleosemantics: New philosophical essays* (pp. 42–68). Oxford: Oxford University Press.

Glüer, K., & Wikforss, A. (2009). The normativity of meaning and content. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2009 Edition). <http://www.plato.stanford.edu/archives/sum2009/entries/meaning-normativity/>.

Griffiths, P. (1993). Functional analysis and proper functions. *British Journal for the Philosophy of Science, 44*, 409–422.

Hattiangadi, A. (2006). Is meaning normative? *Mind and Language, 21*, 220–249.

Hume, D. (1978). *A treatise of human nature*. Oxford: Clarendon Press (First published 1739).

Kauffman, S. (2003). Molecular autonomous agents. *Philosophical Transactions of the Royal Society of London A, 361*, 1089–1099.

Kitcher, P. (1993). Function and design. *Midwest Studies in Philosophy, 18*, 379–397.

Krohs, U. (2009). Functions as based on a concept of general design. *Synthese, 166*, 69–89.

Maturana, H., & Varela, F. (1980). *Autopoiesis and cognition: The realization of the living*. D. Reidel Publishing Company.

Millikan, R. (1984). *Language, thought, and other biological categories: New foundations for realism*. Cambridge, MA: MIT Press.

Millikan, R. (1989). In defense of proper functions. *Philosophy of Science, 56*, 288–302.

Millikan, R. (1999). Wings, spoons, pills and quills. *Journal of Philosophy, 96*, 191–218.

Moore, B. (2004). The evolution of learning. *Biological Reviews of the Cambridge Philosophical Society, 79*, 301–335.

Moore, G. (1971). *Principia ethica*. Cambridge: Cambridge University Press (First published 1903).

Moreno, A., Etxeberria, A., & Umerez, J. (2008). The autonomy of biological individuals and artificial models. *BioSystems, 91*, 309–319.

Neander, K. (2009). Teleological theories of mental content. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2009 Edition). <http://www.plato.stanford.edu/archives/win2009/entries/content-teleological/>.

Olson, J. (2009). Reasons and the new non-naturalism. In S. Robertson, J. Skorupski, & J. Timmermann (Eds.), *Spheres of reason*. Oxford: OUP.

Papineau, D. (1984). Representation and explanation. *Philosophy of Science, 51*, 550–572.

Raz, J. (1999). Explaining normativity: On rationality and the justification of reason. *Ratio, 12*, 354–379.

Schlosser, G. (1998). Self-re-production and functionality. *Synthese, 116*, 303–354.

Schrödinger, E. (1944). *What is life?* Cambridge: Cambridge University Press.

Schroeder, M. (2008). Value theory. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2008 Edition). <http://www.plato.stanford.edu/archives/fall2008/entries/value-theory/>.

Sharma, J., Angelucci, A., & Sur, M. (2000). Induction of visual orientation modules in auditory cortex. *Nature, 404*(6780), 841–847.

Sober, E. (1993). *Philosophy of biology*. Boulder, CO: Westview Press.

Stotz, K., & Griffiths, P. (2001). Dancing in the dark: Evolutionary psychology and the argument from design. In S. Scher & M. Rauscher (Eds.), *Evolutionary psychology: Alternative approaches*. Dordrecht: Kluwer.

Tauber, A. (2010). The biological notion of self and non-self. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2010 Edition). <http://www.plato.stanford.edu/archives/sum2010/entries/biology-self/>.

Toepfer, G. (this volume). *Teleology and its constitutive role for biology as the science of organized systems in nature*.

von Melchner, L., Pallas, S., & Sur, M. (2000). Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature, 404*(6780), 871–876.

Wallace, R. Jay, (2009). Practical Reason. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2009 Edition). <http://plato.stanford.edu/archives/sum2009/entries/practical-reason/>.

Wright, L. (1973). Functions. *Philosophical Review, 82*, 139–168.

Studies in History and Philosophy of Biological and Biomedical Sciences 43 (2012) 113–119

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

Teleology and its constitutive role for biology as the science of organized systems in nature

Georg Toepfer

Collaborative Research Centre 644 "Transformations of Antiquity", Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Germany

ARTICLE INFO

Article history:
Available online 29 June 2011

Keywords:
Teleology
Organism
Biology
Function
Cyclical systems

ABSTRACT

'Nothing in biology makes sense, except in the light of teleology'. This could be the first sentence in a textbook about the methodology of biology. The fundamental concepts in biology, e.g. 'organism' and 'ecosystem', are only intelligible given a teleological framework. Since early modern times, teleology has often been considered methodologically unscientific. With the acceptance of evolutionary theory, one popular strategy for accommodating teleological reasoning was to explain it by reference to selection in the past: functions were reconstructed as 'selected effects'. But the theory of evolution obviously presupposes the existence of organisms as organized and regulated, i.e. functional systems. Therefore, evolutionary theory cannot provide the foundation for teleology. The underlying reason for the central methodological role of teleology in biology is not its potential to offer particular forms of (evolutionary) explanations for the presence of parts, but rather an ontological one: organisms and other basic biological entities do not exist as physical bodies do, as amounts of matter with a definite form. Rather, they are dynamic systems in stable equilibrium; despite changes of their matter and form (in metabolism and metamorphosis) they maintain their identity. What remains constant in these kinds of systems is their 'organization', i.e. the causal pattern of interdependence of parts with certain effects of each part being relevant for the working of the system. Teleological analysis consists in the identification of these system-relevant effects and at the same time of the system as a whole. Therefore, the identity of biological systems cannot be specified without teleological reasoning.

1. Introduction

Teleology plays many roles in biology. Teleological notions are important as heuristic devices in guiding research, they serve to analyse or decompose organisms and their behaviour into functional types, and they are integral parts in the explanation of the presence of parts in living beings as selected systems. In this paper, I defend the view that teleology is closely connected to the concept of the organism and therefore has its most fundamental role in the very definition of biology as a particular science of a special class of natural objects. This view has, of course, a long tradition going back at least as far as Immanuel Kant's *Critique of Teleological Judgement* (1790; Kant, 1913), in which he analyzes organisms as 'natural purposes' and assigns a methodological role to teleology in his complex epistemology of organized beings. Following Kant, teleology as a regulative idea played some role in the conceptualization of organisms by leading physiologists of the 19th century, e.g. Johannes Müller (Müller, 1833–1840, I, p. 18 f.) or Claude Bernard (Bernard, 1878–1879, I, p. 340). In the Neo-Kantian movement at the turn of the 20th century, there were also several attempts to explicate and refine the Kantian insight, e.g. by stating that 'the concept of the organism is essentially teleological, built on the concept of purpose and purposiveness, inconceivable and incomprehensible without the idea of purpose' (Liebmann, 1899, p. 236) or: 'We even have to define this science [i.e. biology] as the science of bodies whose parts combine to a teleological 'unity'. This concept of unity is inseparable from the concept of the organism, such that only because of the teleological coherence we call living beings 'organisms'. Biology would, therefore, if it avoided all teleology, cease to be the science of organisms as organisms' (Rickert, 1929, p. 412).

E-mail address: ToepferG@rz.hu-berlin.de

1369-8486/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsc.2011.05.010

Yet in the extensive discussions on teleology in biology of the second half of the 20th century, this methodological understanding of teleology played virtually no role. Teleology was almost exclusively discussed as a heuristic principle for initiating or guiding research or as a means for providing explanations. Consequently, Dennis Walsh wrote at the very beginning of his recent review of teleology: 'Teleology is a mode of explanation in which the presence, occurrence, or nature of some phenomenon is explained by appeal to the goal or end to which it contributes' (Walsh, 2008, p. 113). What has virtually disappeared from the recent discussions is the most fundamental role of teleology in biology: its constitutive function for the concept of the organism. In this role, the main aim of teleological reflection is not to *explain* something but to *identify* or *delimit* a particular kind of system. Teleological reasoning provides the foundational framework for a particular kind of system. These systems do not exist outside teleological reflections. In the light of this understanding, teleology is not just one mode of reasoning or an explanatory strategy for biologists to analyse their objects. It is what enables biologists to come to terms with their objects. As a biologist, one conceptualises objects as functionally integrated systems and asks functional questions. Teleology is what makes biology a special science; it is central to its methodology. Therefore, I will call this understanding of teleology *methodological*.

In this paper, I will investigate this methodological role of teleology by focusing on the concept of the organism (Section 2), relating it to systems-theoretical accounts of the concept of function in biology (Section 3), analysing the relation of teleology to organismic activities that do not contribute to the integrity of the organism, especially *reproduction* (Section 4), and, finally, stressing the essential role of cycles for the constitution of organized systems and by discussing the problem of inorganic cycles (Section 5).

2. The teleological conception of the organism

In a general sense, an organism may be defined as a causal system of interdependent parts that is able to perform certain complex activities, including nutrition, growth, reproduction, and, at least to a certain extent, locomotion and perception. There are two central features of this definition: the *interdependence* of parts or processes, and the performance of *complex activities*.

Since Antiquity, this set of complex activities has formed the central part in every definition of living beings (following Aristotle, *De anima*412a). They were connected to the central principle of life, the soul and its parts. After the mid-17th century the soul was gradually replaced by concepts surrounding 'organization' and by the idea that the complex activities of living beings emerged from nothing but the interaction of their parts. In the context of mechanistic physiology in the early 18th century there appeared definitions of the concept of the organism (or 'organic body') that were rooted in the idea of interdependence of parts in a system. One early example is Herman Boerhaave's definition at the beginning of his *Historia plantarum*: 'An organic body was composed of clearly different parts [...] whose actions mutually depend on each other' (Boerhaave, 1727, p. 3 [Prooemium]).[1]

Building on these mechanistic conceptions, the causal structure of organisms was described as a cycle of processes. One such description appeared in 1754 in the fourth volume of the French *Encyclopédie*: 'Animated bodies' were analysed as 'a kind of circle [*cercle*] in which every part could be regarded as the beginning or the end, these parts respond to each other and they all aim at each other [*elles tiennent toutes les unes aux autres*]' (Tarin, 1754, p. 1046).

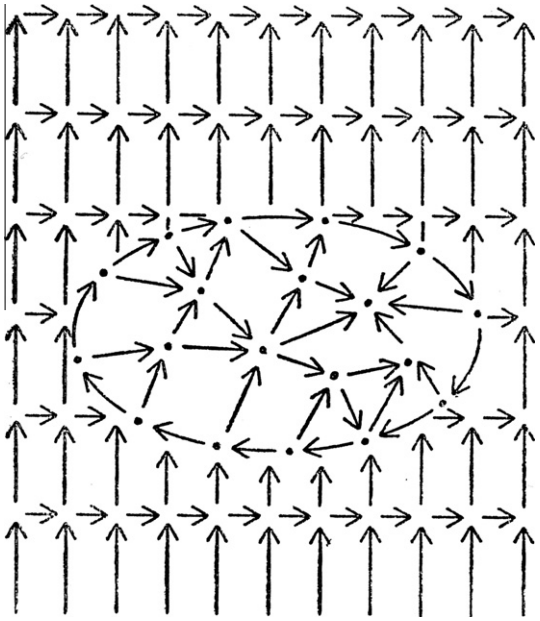

Fig. 1. The organism represented as a closed causal system functionally separated from the stream of causal events in its environment (Rothschuh, 1963, p. 34).

Based on this causal reciprocity of the parts, an organism can be represented as a closed causal system that is functionally separated from the stream of causal events in its environment. Surprisingly, not many attempts have been made to visualize the essential causal structure of organisms. Fig. 1 shows one proposal from mid-20th century.

The causal links constituting the organism—symbolized by the central oval—form a pattern of interaction: each element is simultaneously cause and effect of other elements. The essential causal structure of an organism is a cycle of processes.

This straightforward diagram is to some extent plausible, but at the same time, it is not clear what it really represents. To make it more comprehensible, the systems must be made more concrete and specific. One well-known example of a concrete cyclical system is Stuart Kaufman's model for a network of autocatalytic reactions (Fig. 2).

Here, the dots represent the reactants, and the lines the steps of synthesis and the influence of the catalysts. The network is more or less closed in the sense that every reactant is simultaneously catalyst and product, i.e. at the same time the means and end of a reaction. For this reason, the network forms a cycle, a cycle of production: every component of the system participates in the production of other parts, and therefore, via the other parts, finally is engaged in the re-production of itself—or at least the reproduction of other parts of its own kind.[2]

In real biological systems there are many biochemical cycles that comply approximately with this model, e.g. the *Krebs*- or *Calvin*-cycle. They differ from Kauffman's model in that they do not produce the catalysts that are essential for their reactions, but nevertheless they are *production cycles* in the sense that, on the one hand, the fabrication of every chemical compound in the cycle

[1] 'Erat corpus Organicum ex diversis planè partibus compositum [...] & sic harum partium actiones ab invicem dependent'. Boerhaave used the past tense to indicate that his definition was already well established but he did not quote any sources (cf. Müller-Wille, 1999, p. 122). Kant certainly knew definitions of this kind because via Boerhaave's students in Königsberg, M. E. Boretius and J. C. Bohlius, he was well acquainted with Boerhaave's work on physiology and botany (cf. Löw, 1980, p. 87).

[2] Kauffman explicitly refers to his model as an explication of Kant's idea of causal reciprocity: 'an autocatalytic set of molecules is perhaps the simplest image one can have of Kant's holism' (Kauffman, 1995, p. 69).

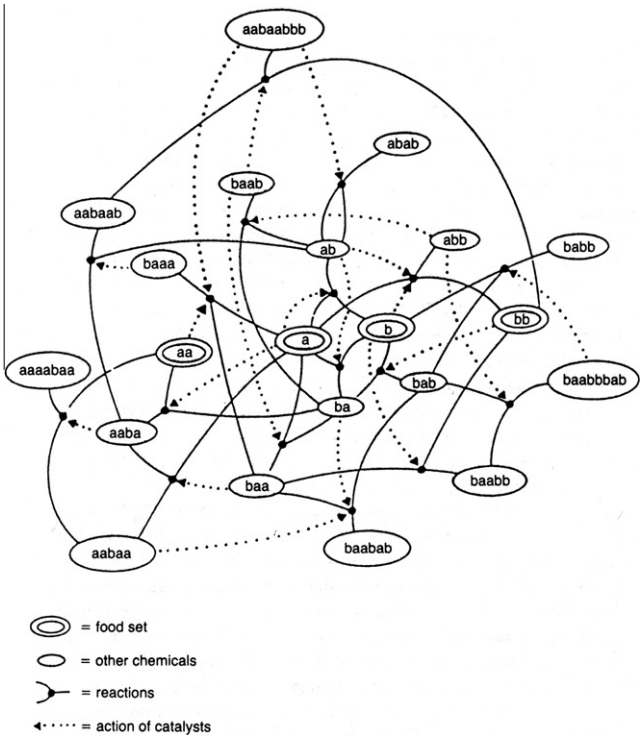

Fig. 2. A network of catalytic reactions (Kauffman, 1995, p. 65).

depends on the other compounds and, on the other, every compound makes the production of the other compounds possible.

This means that every part of these systems of interdependence only exists within the network of the other parts. Every part only exists as a component of the system. Therefore, every component of the system needs to be specified not only by its particular internal structure but also by its particular role it plays in the perpetuation of the system as a whole.

3. Functions as feedback effects of parts in cyclically organized systems

It is a central claim of this paper that roles of parts in a system can be called *functions* precisely because they are integrated into a cyclically organized system, i.e. a system of interdependent parts. In this respect, a function is the result of a process (or the activity of a part) that has an impact on the future performance of the same process (or the same part) or processes (parts) of its kind.

Accordingly, my general definition of 'function' would be:

> *Functions are system-relevant effects of parts (or sub-processes) in systems of mutually dependent parts, i.e. those effects of any part that contribute to the maintenance of the other parts, and via them, feed back onto their own maintenance or perpetuation.*

For this definition of function, the concept of *organism*, or *organized system*, is fundamental because the attribution of functions to objects consists in the integration of these objects into a system of interdependent parts. Any talk of function, therefore, is linked to the assumption of an organized system of mutually dependent elements, parts or processes. And it also works the other way round: an organism only exists as these interdependent functional relations.

The definition of the concept of organism is based on teleological reasoning because identifying organisms as a special kind of natural systems includes the attribution of functions or purposes to their parts.[3] This attribution is only possible by constituting a whole cycle (or network) of interactions and interdependencies because functions always come in sets as components of interdependent systems. Functions never occur in isolation; they are always integrated into a system together with other functions. As Marcel Weber notes in his *coherence account of function*, what is essential is the horizontal interaction of functions with other functions (Weber, 2005).[4] This can be best visualized by cyclical causal models, in which functions occur as elements of a cycle together with other functions. Thus every function has to have co-functions at the same level, but there does not necessarily have to be a vertical order, i.e. a hierarchy of functions. Functional hierarchies can come into play when complex biological functions like nutrition, protection, or parental care are related to the so-called basic functions of self-preservation and reproduction or when functions on lower levels are investigated.

The coherence of functions in an organism can be expressed by saying that 'organism' is a functional concept: organisms do not exist as definite amounts of matter—the exchange of matter in metabolism is essential for a system to be an organism. Additionally, organisms do not exist as definite forms given that the change of form, metamorphosis, is a common phenomenon. However, despite the changes in its matter and form, an organism may remain the same individual. What defines the starting and end point of its existence is not the coherence of a certain amount of matter or the maintenance of a certain form but the cyclical causal structure of the system, i.e. the duration of the interaction of its components and activities. This means that, beyond the functional perspective, which consists in specifying the system by fixing the roles of its parts, the organism does not even *exist* as a definite entity.

As 'organism' is one of the basic concepts in biology, teleology plays a constitutive methodological role in this science. This is because talking of organisms presupposes a certain notion of teleology. 'Function' and 'purpose' are not just descriptive or explanatory concepts, they play a methodological role in specifying biology as a particular science. With this understanding of 'teleology' in mind, it becomes possible to modify Theodosius Dobzhansky's famous dictum 'nothing makes sense in biology except in the light of evolution' (Dobzhansky, 1964, p. 449, 1973) to: 'nothing in biology makes sense except in the light of teleology'. Evolution surely is necessary for the explanation of biological phenomena, but it is teleology that provides the very subject to biology. 'Organism' as a concept is essentially teleological.

But what exactly is teleology? Basically, the teleological perspective is a way of defining and analyzing causal processes by looking at the final state of these processes. The final state, outcome, or effect of a process represents the focus of teleological thinking. By fixing the outcomes (i.e., keeping them constant) teleological analysis establishes equivalence classes for causal processes (Luhmann, 1962, p. 623). The similarity of outcome is the criterion for the equivalence of processes in a functional class. For example, there may be many different ways of ingesting food; what makes them all elements of

[3] Recently, Marcel Quarfood has defended a similar view, the 'identificatory account of teleology', in his interpretation of Kant's teleology: '[A]ccording to the identificatory account of teleology, the notion of organism itself is dependent on teleology, so that a non-teleological consideration of such objects could only identify them as complexly built aggregates of matter. [...] teleology provides the objects of biological science' (Quarfood, 2006, p. 743). According to this interpretation of Kant, teleological reflection provides biology with its objects and in this respect plays a constitutive role. Still, all explanations have to be mechanical. So, in the epistemology of organisms there are two distinct methodological levels involved: teleology for their identification and mechanism for their explanation.

[4] '[I]t is the place of certain capacities in a coherent system of capacities that underwrites their status as functions. On this view, nothing counts as a function unless there are lots of other things that are also functions and this system of functions provides the best explanation for the organism's capacity to self-reproduce' (Weber, 2005, p. 196f.). In relying on the capacity of self-reproduction (in the sense of reproducing the parts in an individual organism) this view is similar to Gerhard Schlosser's systems-theoretical account of biological teleology (Schlosser, 1998).

one functional class is their overall effect: the intake of food material from the environment. By this type of analysis, causal processes are made comparable with respect to the requirements of a system, e.g. the requirement of ingesting food.

There is one simple and obvious reason why teleological thinking is of special relevance for the concept of the organism. The focus on effects or outcomes is important in systems of interdependent parts because in these systems the end state of one process is necessary for the performance of the other processes. This special causal relevance of outcomes for the system is the reason why we emphasize them in biological thinking and why biological descriptions include so many functional classes (like types of organs or units of behaviour). And because the persistence of the organism depends on the persistence of the functional roles of its components, the ensemble of functions forms the basis of the identity of the organism.

Teleology should not be seen primarily as an explanatory strategy for giving reasons for the presence or occurrence of a part in a system, but primarily as a descriptive tool for analyzing systems in terms of their functional components and thereby enabling a proper conceptualization of them. In this respect, teleology plays an important methodological role: teleological analysis provides the identity criteria for organized systems. In modifying a well-known sentence by Paul Griffiths, one could say: 'Wherever there is organization, i.e. cyclicity or closure of causal relationships in a system of interdependent parts, there is teleology'.[5]

This understanding of biological function roughly corresponds to a systems-theoretical notion of function. It is, first of all, related to Robert Cummins' account of functions as causal roles in complex systems. According to Cummins, '[t]o ascribe a function to something is to ascribe a capacity to it which is singled out by its role in an analysis of some capacity of a containing system' (Cummins, 1975, p. 765). This means that to ascribe a function to an entity one has to focus on the causal structure of a system, not on its individual or evolutionary history. Yet as the many discussions of Cummins' analysis have shown, his criterion of function is too liberal. There are many capacities of systems that are not functions. For example, in most contexts it is not a function of the heart to emit sounds, although this is one of the capacities that contribute to the overall effects an organism produces.

Another systems-theoretical account of functions, which is more similar to that which is proposed here, can be found in Peter McLaughlin's book, *What Functions Explain* (2001). For McLaughlin, a function is a part in a system of regeneration or self-reproduction, as he calls it. According to him, for a part to be a function it must engage in its own future fabrication: 'The particular item *x* ascribed the function of doing [...] *Y* actually is a reproduction of *itself* and actually did [...] something like *Y* in the past and by doing this actually contributed to—or was part of the causal explanation of—its own reproduction' (McLaughlin, 2001, p. 167). In this account of function there is an intra-generational feedback operating for each function.[6] The important point consists in the integration of functional parts in the individual organism as a self-reproducing system. The organism is viewed as a system in flux, a system that constantly changes and literally rebuilds and regenerates its parts. As a matter of fact, all the cells that together form our bodies are persistently renewed. This kind of individual regeneration forms the physiological basis upon which Peter McLaughlin builds his account.

There is, nevertheless, an important objection that could be raised against McLaughlin's account pertaining to the notions of 'production' and 'reproduction'. Production and reproduction can be one form of causal interaction that is relevant for assigning functions to parts in organized systems. However, not all functional relations in systems of mutually dependent parts are relations of production and reproduction. It is not necessary for a system with functional parts that these parts constantly build and rebuild each other. The capacity of regeneration or self-reproduction is a physiological detail in living organisms—it is not the essential character that makes their parts functional.

Indeed, it is possible to imagine an entirely or partly artificial organism consisting of parts that are created separately and are only later assembled together to form one coordinated and integrated whole, e.g. a human individual with an artificial heart. Although there is no mutual production of the parts in such a system it is nevertheless an organism because its working order consists of mutually dependent processes: the parts of this organism—e.g. the heart—are not produced by the other parts and are not constantly rebuilt—but nevertheless they are functional insofar as they are engaged in mutually dependent processes.

As functions are, according to my understanding, elements of systems of interdependent parts, each function is always on a level with other functions that together constitute the organized system. Every function, therefore, is a *co-function*, though for a system to be functional there is no necessity for the presence of overarching functions. Ecosystems, for example, can be seen as functional systems that consist of interdependent components with specific causal roles (and which exist as a unity only within this functional perspective)—but for the attribution of functions to its components it is not necessary to assume an overarching function for the whole system.

In an organized system that is conceptualized as a unity, each function represents a capacity of the system that is coordinated to other capacities, which together make up the system. Hence, the organic activities that do not fit into this network of interdependent processes cannot count as functions.

4. Function and reproduction

The most prominent organismic activity that has no co-function but is commonly considered as an overarching function is *reproduction*. Reproduction is the generation of offspring, i.e. the coming into being of new organisms due to the activity of existing organisms. This process is one of the ultimate ends for all organisms and since antiquity it has been conceived as one of the two fundamental aims of life, along with self-preservation,[7] or even the defining characteristic of all living beings.[8] Biologists' emphasis on reproduction in their analysis of organisms is empirically and theoretically well justified, mainly for two reasons: firstly, because reproduction is as a matter of fact one of the most consistent of all organismic activities[9];

[5] Griffiths wrote in 1993, defending an evolutionary account of the biological function concept, that 'wherever there is selection, there is teleology' (Griffiths, 1993, p. 422).

[6] Claus Emmeche has proposed a similar form of feedback or 'operational closure' as the underlying structure of systems with functional parts: 'functionality is only possible under a closure of operations [...] Only when the causal chain from one part to the next closes or feeds back in a closed loop—at once a feedback on the level of parts and an emergent function defined [...] as a part-whole relation—can we talk about a genuine function' (Emmeche, 2000, p. 195). Note that Emmeche's formulation is more abstract as he does not talk of 'production' and 'reproduction'. He therefore aims at a broader conception of feedback and interdependence than McLaughlin does.

[7] Aristotle states that '[W]hile one part of living consists [...] in the activities to do with the production of young, a further and different part consists in those to do with food; for these two objects in fact engage the efforts and lives of all animals' (Aristotle, Historia animalium, 589a2-5).

[8] For example, 'The most reliable test of whether a thing is alive is whether it can reproduce its like indefinitely if given the proper food' (Haldane, 1940, p. 20), or '[A] living system is any self-reproducing and mutating system which reproduces its mutations, and which exercises some degree of environmental control' (Shklovskii & Sagan, 1966, p. 197).

[9] 'Everything in a living being is centred on reproduction. A bacterium, an amoeba, a fern—what destiny can they dream of other than forming two bacteria, two amoeba or several more ferns?' (Jacob, 1976, p. 4).

and secondly, because reproduction is the starting point of population-level processes that in the long run have the most fundamental effect on the structure and functioning of organic beings: evolution.

However, for any systems-theoretical account of function, reproduction constitutes a problem. This is because reproduction is not a process whose effect is primarily directed towards the (individual) system that initiated it. Rather, reproduction forms the starting point for a new entity, a linear causal chain that does not lead back to its initiator and that has no definite future in the course of evolution. Reproduction breaks open the cyclical causal organization of processes that constitutes the organism, and consequently it does not, conceptually speaking, form part of the understanding of the organism as a functionally closed system. As a result, it cannot be taken to constitute a proper function according to the account proposed here.

This exceptional status and rank reproduction has in relation to the other organismic activities has long been noted: Already Immanuel Kant pointed out that reproduction 'does not necessarily belong to the concept of the organism but is an empirical addition' to it (Kant, 1938, p. 547). 'Birth and death [...] are only synthetically attached to life', as Edgar Singer put it at the beginning of the 20th century (Singer, 1914, p. 655). Thus, we may conclude that reproduction is not constitutive for the organization of the living. In other words, it is 'not intrinsic to the minimal logic of the living' (Varela, 1991, p. 81). The particular status of reproduction is exemplified by the fact that organisms that do not reproduce are still able to exist as entities with a functional organization.

Although reproduction may not belong to the defining characteristics of an organism it has, since Aristotle, been standard practice in biology and philosophy of biology to place reproduction on the same functional level as self-preservation. These two are considered to be the ultimate goals of organisms and, consequently, as the final ends of functional reasoning in biology.[10] By uniting these two ultimate functions it would be possible to perceive reproduction as a kind of self-preservation[11]—not, of course, of individual organisms but of their special organization. In this sense, the prevalence of reproduction may be explained as a result of natural selection because it is the most effective way of preserving a structure ('preservation by multiplication'). Indeed, the pervasiveness of reproduction in organic beings is itself the outcome of evolutionary processes given that it is a basic analytic truth of evolutionary theory that those types of organisms that reproduce the most will spread in populations. However, strictly speaking, this mode of preservation does not actually contribute to the individual organism's own integrity and maintenance. Reproduction does contribute to the maintenance of species, but, ultimately, also leads to their transformation. It is therefore ironic that this most effective means of preservation has actually resulted in the immense transformations of organisms that have taken place on earth since life began.

The question is: How is it possible to make functional sense of an organismic activity that does not contribute to an organism's integrity and maintenance? One possible answer is that although reproduction is not an element of the cycle of interdependent processes that constitute an individual organism, and consequentially not a function of this system, it may still be the function of another system, for example of the so-called 'life-cycle'. As a component of a life-cycle, reproduction contributes to the maintenance of this cycle. The co-function of reproduction in the life-cycle could be what we commonly refer to as 'development'. In this way, a cycle of reproduction and development is formed in which every instance of reproduction of an organism results in a new young organism developing to a reproducing adult organism. This cycle clearly exists since the origin of life, but the problem is that it is not really distinct. There is no single system that embodies this cycle comparable to an organism's cycle of physiological processes or functional activities.

Consequently there is an important disanalogy here. Whereas the activity of pumping blood by the heart feeds back onto this same heart and makes its future pumping possible, the activity of reproduction of one organism feeds back only onto an organism of the same type, making the general activity-type of reproduction possible by its performance. Therefore, this second type of feedback, an inter-generational feedback loop, already presupposes an ontology of types. The reproduction of one organism does not make this same organism and its further reproduction possible; it only makes future organisms of the same type and their reproduction possible. In a sense, the very concept of 'life-cycle' clearly presupposes an ontology of types. In a life-cycle there is no single entity that is involved in cyclical transformations. In contrast to this, the feedback loop involved in physiological processes can be made explicit within an ontology of individuals. Indeed, it is one and the same heart that benefits (in the sense of being maintained) from its own activity.

Having said this, an ontology of types does not seem to constitute a real problem in biology. For biologists it is perfectly fine to consider reproduction as a function within a life-cycle. And in any case, in the context of my analysis, the attribution of function is possible so far as there is a cycle involved, be it a cycle with a feedback loop leading back to the same individual, or an individual of the same type from which it started. Nevertheless, this tolerance brings about a number of problems, which I will address in the final section of this paper.

5. Functions and cycles

Cyclical processes are found in many places in nature, not only in living beings. One example of an inorganic cycle is the water cycle. It is a cycle of three interdependent processes, *evaporation*, *condensation* and *precipitation*. Because of their interdependence it seems in principle possible to assign functions to these different processes: the function of precipitation would be, for example, the refilling of the water reservoirs on the surface of the earth; it sets the conditions and starting point for evaporation. Because of its cyclical connection with the other processes, each instance of precipitation feeds back onto future instances. Following the analysis of functions as system-relevant effects in systems of mutually dependent processes proposed here, it is legitimate to assign functions to the sub-processes of the water cycle. For instance, it is a function of rain to sustain the water cycle.

Nevertheless, there has to be an explanation for the fact that normally rain and clouds and other meteorological entities are not considered as functional or as organs.[12] Two possible explanations can be given. The first is simply a lack of understanding of the true nature of the phenomenon. 'Rain' is often identified by its intrinsic features as an isolated physical process, similar to other

[10] To cite only two authors: 'There may be no serious objection to saying that the two basic 'purposes' of living organisms—to maintain themselves and to perpetuate their kind—underlie the whole panorama of evolution' (Goudge, 1961, p. 196f.); 'we may speak of survival and reproductive success as the ultimate purpose served by individual biological adaptations, i.e., the reason why they have come about' (Ayala, 1998, p. 46).

[11] This way of relating reproduction and self-preservation was proposed by W. Ostwald in 1902. Ostwald thought it would be methodologically more appropriate to take reproduction as *part* of self-preservation than to put it on the same level (1902, p. 316). For Ostwald, self-preservation is the characteristic feature of the organic world, and reproduction forms only one aspect of it.

[12] There are some exceptions, most famously in the context of the 'Gaia-hypothesis'. Already in 1790, Georg Christoph Lichtenberg wrote the air could be judged to be 'an organ for the production of rain' (1994, p. 253). In discussing Larry Wright's 'etiological account' of functions, Michael Ruse explicitly rejects such an understanding (Ruse, 1978, p. 201).

events that are not part of cycles, like the sunrays, simply something that comes down from the skies. In contrast to organic parts that are always integrated into some well-arranged body, all the elements of the water-cycle appear at first sight to be isolated phenomena, as rain, surface water or clouds. They are not described as relational processes that are integrated into a large, and at first sight invisible cycle. Consequently, because of its usual description as an isolated phenomenon we hesitate to assign a function to it.

The second explanation for the common practice of avoiding function ascriptions in inorganic cycles is epistemological and is related to the first: because of their identification by intrinsic features, the sub-processes of the watercycle (and other geochemical cycles) are conceptually distinct from organic activities. Organic activities are conceptualized as relational processes already at a descriptive level: the intake of food is related to digestion, digestion to circulation, and so forth. In the case of organic activities, the causal relatedness of processes finds its expression in the relational conceptualisation of these activities. Causal interdependence is marked by conceptual interdetermination: each process is identified and determined in relation to the other processes in the system. In contrast to this, the interdependence of processes in the case of abiotic cycles, like the water cycle, is not expressed in the way we conceptualize these processes. And this is the reason why we do not assign functions to them. If the rain were to be described and identified relationally as an essential element in the watercycle, whose activity to on its own future performance, it would be clearly analogous to organic processes and it would therefore appear acceptable to assign a function to it.

There are, of course, other disanalogies between organisms and the watercycle as well. Organisms are concrete, individual, localized entities, whereas watercycles are systems of scattered, dispersed processes. But this is not really a relevant difference in this context. Ecosystems could be seen as something in between organisms and geochemical cycles: they are large and scattered, with unclear boundaries, but they clearly have causal roles and it is common practice in biology to assign functions to the bearers of ecological roles, for example to the 'producers' or 'reducers' of organic matter. This may be thought is because living beings are involved here. However, on the account proposed here, the real reason is because ecosystems are essentially cycles of processes and causal dependencies.

The basic parallel between organisms and geochemical cycles is their inexistence (as a particular class of entities) in purely physical descriptions. To be sure, these systems are 'nothing else but' physical entities, but the cognition of their wholeness and unity presupposes concepts that are alien to physics. With the conceptual tools of physics, organic systems and inorganic causal cycles can be analyzed as aggregates of transforming matter, but not as integrated systems with functional roles. The ascription of roles depends on the construction of unitary systems consisting of physically heterogeneous (but interdependent) processes and the functional decomposition of these systems.

To admit that there are functionally organized systems in the inorganic world also implies that the concept of function is not an exclusive property of biology. The concept is necessary for biology and is part of biological methodology because it allows organisms to be identified as persisting systems despite changes of matter and form, but it is also necessary for specifying the identity of cyclical systems that are not living such as the water-cycle or ecosystems. The case of geochemical cycles shows that there are functional systems in nature that are neither exclusively biological because they do not necessarily constitute living beings nor exclusively physical because they only exist as patterns of interdependent processes.

Overall, it is cyclicity and closure of operations that drives and justifies functional talk. Functional language is thus elicited in the identification of parts as elements in a system that are mutually cause and effect for each other and that depend in their very existence on their integration into the system. In systems of this kind, living or not, the conceptualization of processes by focusing on their end-states (i.e. the emphasis on effects) is essential because in them the outcome of each component feeds back onto its future performance and consequently on the perpetuation of the system.

6. Conclusions

The first aim of this paper has been to show the constitutive role of teleology for biology. I have argued that there is a methodological link between teleology and biology. In discussions over the last decades the special status of biology among the sciences was attributed to the central place of evolutionary theory in biology rather than to teleology. But, there can be, and for centuries there has been, if not a 'biology' then at least biological thinking without evolutionary theory. Therefore, it does not hold true that nothing in biology makes sense except in the light of evolution. In contrast to this, there never has been and never will be biology without teleological dimensions. This is because teleology, in a certain sense, is deeply rooted in the descriptive language of biology. Most biological objects do not even exist as definite entities apart from the teleological perspective. This is because biological systems are not given as definite amounts of matter or structures with a certain form. They instead persist as functionally integrated entities while their matter and form changes. The period of existence of an organism is not determined by the conservation of its matter or form, but by the preservation of the cycle of its activities. As the unity of this cycle is given by relating functional processes to each other, teleology plays a *synthetic* role for biology and has *ontological* consequences. The identity conditions of biological systems are given by functional analysis, not by chemical or physical descriptions. The same holds true for their parts and sub-processes: they are not individuated by physical or chemical methods as transformation processes but are specified by decomposing the system into functional roles. Biologists can identify in every organism devices for protection, feeding, reproduction or parental care irrespective of their material realization. These functional categories play the most crucial role in biological analyses. Consequently, functions can be seen, in the first instance, as descriptive tools for the constitution and decomposition of dynamic systems. The basic aim of function talk is therefore not to *explain* the occurrence of a part in a system, but to *identify* the system as a whole and analyse it into functional components.

The basic causal pattern of a system that only exists because of the interdependence of its parts is a causal cycle consisting of interdependent sub-processes. In such a system it makes sense to conceptualize each process by its outcome because the outcome is the relevant point for the other processes and consequentially for the maintenance of the system as a whole. As the activity of each component of the cycle enables the activity of the other components and is enabled in turn by them, there is a feedback loop between an activity and its own perpetuation. In biological systems, this feedback can be incorporated in a single body (physiological processes of an individual organism mutually maintaining each other) or it can be distributed among several distinct entities (organisms with different functional roles united in an ecosystem or organisms of the same species united in a 'life-cycle' in which each instance of reproduction enables future instances). From the perspective of an individual organism, reproduction makes no functional sense because it does not feed back onto the system from which it started. But, within the framework of a different per-

spective in which distinct (but structurally similar) entities are united in a 'life-cycle', it is possible to make functional sense of reproduction because as an activity of a certain type it feeds back onto itself.

It is important to note that, as there are also abiotic systems with a cyclical causal structure, such as the geochemical cycles, 'function' is not an exclusive concept of biology. To the extent that cycles of causal processes can be identified in other areas, the concept applies to systems outside of biology as well. In their ontological status these systems are similar to organized systems in biology because their identity is not given by their matter or form but by the cyclical recurrence of processes mutually depending on each other.

The model of cyclicity offers a simple causal theory for explicating teleology and function in holistic systems. Systems of mutually dependent components and with holistic properties can be identified and decomposed into recurrent causal links as their elements. The model aims at a simple theory of function that is causal and holistic at the same time. According to this theory, it is not by virtue of organisms as selected entities but by virtue of their causal structure as cycles of mutually dependent parts that the function concept is justified for their analysis. By means of selection, the closure of operations in an organism is enforced and as a result its organized structure is enormously enriched. However, the ascription of functions depends on the identification of organisms as cyclical causal structures, not as the products of selection.

Acknowledgements

I thank Lenny Moss for organizing the panel on 'The Place of Normativity and Teleology in Nature' at the ISHPSSB 2009 meeting, and two reviewers for their very helpful comments on an earlier version of this paper. Special thanks to Dan Nicholson for improving my English.

References

Aristotle (1991). *Historia animalium (books VII–X, transl. D.M. Balme)*. Cambridge, MA: Harvard University Press.
Ayala, F. J. (1998). Teleological explanations *versus* teleology. *History and Philosophy of the Life Sciences, 20*, 41–50.
Bernard, C. (1878–1879). *Leçons sur les phénomènes de la vie communs aux animaux et aux végétaux* (Vol. 2 Vols.). Paris: Baillière.
Boerhaave, H. (1727). *Historia plantarum*. Romae: Gonzaga.
Cummins, R. (1975). Functional analysis. *Journal of Philosophy, 72*, 741–765.
Dobzhansky, T. (1964). Biology, molecular and organismic. *American Zoologist, 4*, 443–452.
Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *American Biology Teacher, 35*, 125–129.
Emmeche, C. (2000). Closure, function, emergence, semiosis and life: The same idea? In J. L. R. Chandler & G. van de Vijver (Eds.), *Closure. Emergent organizations and their dynamics* (pp. 187–197). New York: The New York Academy of Sciences.
Goudge, T. A. (1961). *The ascent of life. A philosophical study of the theory of evolution*. Toronto: University of Toronto Press.
Griffiths, P. E. (1993). Functional analysis and proper functions. *British Journal for the Philosophy of Science, 44*, 409–422.
Haldane, J. B. S. (1940). Can we make life? *Keeping cool and other essays*. London: British Publishers Guide 1944 (pp. 19–23).
Jacob, F. (1976). *The logic of life: A history of heredity*. New York: Vintage Books.
Kant, I. (1913). Kritik der Urtheilskraft (1st ed., 1790). In Königlich Preußische Akademie der Wissenschaften (Ed.). *Kant's gesammelte Schriften* (Vol. V, pp. 165–485). Berlin: Reimer.
Kant, I. (1938). Opus postumum, Vol. 2. In A. Buchenau (Ed.), *Kant's Opus postumum, 2 Vols. (=Akademie Ausgabe, Vols. XXI–XXII)*. Berlin: de Gruyter.
Kauffman, S. A. (1995). *At home in the universe: Search for laws of self-organization and complexity*. New York: Oxford University Press.
Lichtenberg, G. C. (1994). Sudelbücher, Heft J. In W. Promies (Ed.). *Georg Christoph Lichtenberg, Schriften und Briefe* (Vol. 2, pp. 227–397). Frankfurt/M.: Zweitausendeins.
Liebmann, O. (1899). Organische natur und teleologie. *Gedanken und Thatsachen. Philosophische Abhandlungen, Aphorismen und Studien, Zweites Heft*. Straßburg: Trübner (pp. 230–275).
Löw, R. (1980). *Philosophie des Lebendigen. Der Begriff des Organischen bei Kant, sein Grund und seine Aktualität*. Frankfurt/M.: Suhrkamp.
Luhmann, N. (1962). Funktion und Kausalität. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 14*, 617–644.
McLaughlin, P. (2001). *What functions explain: Functional explanation and self-reproducing systems*. Cambridge: Cambridge University Press.
Müller, J. (1833–1840). *Handbuch der Physiologie des Menschen für Vorlesungen* (Vol. 2 Vols.). Coblenz: Hölscher.
Müller-Wille, S. (1999). *Botanik und weltweiter Handel. Zur Begründung eines Natürlichen Systems der Pflanzen durch Carl von Linné (1707–1778)*. Berlin: Verlag für Wissenschaft und Bildung.
Ostwald, W. (1902). *Vorlesungen über Naturphilosophie*. Leipzig: Veit & Comp..
Quarfood, M. (2006). Kant on biological teleology: Towards a two-level interpretation. *Studies in History and Philosophy of Biological and Biomedical Sciences, 37*, 735–747.
Rickert, H. (1929). *Die Grenzen der naturwissenschaftlichen Begriffsbildung. Eine logische Einleitung in die historischen Wissenschaften* (1st ed., 1896–1902). Tübingen: Mohr (Siebeck).
Rothschuh, K. E. (1963). *Theorie des Organismus. Bios. Psyche. Pathos* (1st ed., 1959) (2nd ed.). München: Urban & Schwarzenberg.
Ruse, M. (1978). Critical notice. Andrew Woodfield, Teleology. *Canadian Journal of Philosophy, 8*, 191–203.
Schlosser, G. (1998). Self-re-production and functionality. A systems-theoretical approach to teleological explanation. *Synthese, 116*, 303–354.
Shklovskii, I. S., & Sagan, C. (1966). *Intelligent life in the universe*. San Francisco: Holden-Day.
Singer, E. A. (1914). The pulse of life. *The Journal of Philosophy, Psychology and Scientific Methods, 11*, 645–655.
Tarin, P. (1754). Dissection. In D. Diderot & J. d'Alembert (Eds.). *Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers* (Vol. 4, pp. 1046–1047). Paris: Briasson.
Varela, F. J. (1991). Organism: A meshwork of selfless selves. In A. I. Tauber (Ed.), *Organism and the origins of self* (pp. 79–107). Dordrecht: Kluwer.
Walsh, D. (2008). Teleology. In M. Ruse (Ed.), *The Oxford handbook of philosophy of biology* (pp. 113–137). Oxford: Oxford University Press.
Weber, M. (2005). Holism, coherence and the dispositional concept of functions. *Annals of the History and Philosophy of Biology, 10*, 189–201.

Studies in History and Philosophy of Biological and Biomedical Sciences 43 (2012) 120–132

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

The Lenoir thesis revisited: Blumenbach and Kant

John H. Zammito

Department of History, MS-42, Rice University, P.O. Box 1892, Houston, TX 77251-1892, USA

ARTICLE INFO

Article history:
Available online 25 June 2011

Keywords:
Timothy Lenoir
Johann Friedrich Blumenbach
Immanuel Kant
Teleomechanism
Reflective judgment
Intrinsic purposiveness

ABSTRACT

Timothy Lenoir launched the historical study of German life science at the end of the 18th century with the claim that J. F. Blumenbach's approach was shaped by his reception of the philosophy of Immanuel Kant: a 'teleomechanism' that adopted a strictly 'regulative' approach to the character of organisms. It now appears that Lenoir was wrong about Blumenbach's understanding of Kant, for Blumenbach's *Bildungstrieb* entailed an actual empirical claim. Moreover, he had worked out the decisive contours of his theory and he had exerted his maximal influence on the so-called 'Göttingen School' before 1795, when Lenoir posits the main influence of Kant's thought took hold. This has crucial significance for the historical reconstruction of the German life sciences in the period. The Lenoir thesis can no longer serve as the point of departure for that reconstruction.

1. Introduction

Some thirty years ago, now, in pioneering work on the emergence of biology in Germany at the end of the 18th century, Timothy Lenoir formulated the thesis that the so-called 'Göttingen School' around Johann Friedrich Blumenbach took up methodological guidelines developed by Immanuel Kant and established a *strictly heuristic* (or in Kantian language, *regulative*) notion of 'teleo-mechanism,' whereby the imputation of natural teleology (immanent purposiveness) to organisms was never an objective scientific knowledge claim.[1] Lenoir organized his reconstruction of German life science from 1790 to 1860 into three periods: those of 'vital materialism,' of 'developmental morphology,' and of 'functional morphology.' (Lenoir, 1981b, p. 298, 1989) My critique will concern his claims specifically concerning the 'vital materialism' of Blumenbach and the 'Göttingen School' in the 1790s.[2] While Lenoir has many interesting claims concerning 'teleo-mechanism' in the nineteenth century, these will not enter into consideration here.[3]

E-mail address: zammito@rice.edu

[1] 'My principal thesis is that the development of biology in Germany during the first half of the nineteenth century was guided by a core of ideas and a program for research set forth initially during the 1790s. The clearest formulation of those ideas is to be found in the writings of the philosopher Immanuel Kant. I do not claim that German biologists discovered a program of research in Kant's writings which they set out to realize in practice rather that in the latter part of the eighteenth century a number of biologists were seeking to establish a foundation for constructing a consistent body of unified theory for the life sciences which could adapt the methods and conceptual framework of Newtonian science to the special requirements of investigating biological organisms. Kant stepped into this ongoing dialogue and set forth a clear synthesis. It was through Blumenbach and his students that Kant's special brand of teleology entered biology.' (Lenoir, 1989, pp. 2–3) See also Lenoir (1978, 1980, 1981a, 1981b, 1988). For thoughtful critiques, see Caneva (1990), Richards (2002).

[2] I thank the peer reviewer on my piece for urging me to specify exactly what in Lenoir's thesis I am disputing and to cite sources. In that light, since my emphasis is on the Kant-Blumenbach relationship, I would point out that it is Lenoir's articles, not his book of 1982/1989, that spell out the details of his view, as he admits in the book: 'Elsewhere I have attempted to document in detail the relationship between these two men and the extent to which Blumenbach incorporated Kant's work into the mature formulation of his ideas.' (Lenoir, 1989, p. 22) He refers the reader specifically to Lenoir (1980). It is with that text (and the other articles) that I will be primarily engaging.

[3] Of course, Kant was for nineteenth-century Germans an eminence fervently to be invoked as a model or warrant. (See Friedman & Nordmann, 2006). Yet whether the Kant they invoked was the historical Kant, and whether, more significantly, they really needed (or even used, rather than mentioned) Kant for their undertakings: these are matters that allow for more than one reading. That is, first: whatever they *thought* Kant meant (assuming we can establish *that*), there is a whole guild devoted to jousting over what Kant *really meant*, and jousters sometimes decide the best way to win the day is to proclaim what he *should have meant*. Usually that involves dismissing what others have thought Kant meant/should have meant. The nineteenth-century Kant reception falls fully within this conspectus, and Lenoir's book documents some striking instances of this. Second: the arguments of nineteenth-century biologists typically involved both theoretical and empirical elements that were simply not part of Kant's possible intellectual horizon, and these, rather than any direct appeal to Kant, appear far more plausibly central to their actual arguments, though Kant was always a rhetorical trump card, if he could be plausibly invoked. In any event, these are not questions here to be pursued.

1369-8486/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsc.2011.05.011

In Lenoir's view, Kant's philosophy of science played a major role 'in helping to shape the theoretical foundations of the life sciences' led by Blumenbach after 1790. (Lenoir, 1980, p. 77) 'Initiated by Kant's probing insights, the goal of uniting the teleological and mechanical frameworks of explanation was a topic of central importance in discussion on the philosophy of nature in the 1790s.' (Lenoir, 1980, p. 83n) Concretely, Lenoir claimed: 'from the late 1780s to the late 1790s Blumenbach's ideas on natural history underwent a thorough revision in light of Kant's analysis of the conceptual foundations for the construction of a scientific theory of organic form.' (Lenoir, 1980, p. 77) Lenoir found evidence of 'a revolution in [Blumenbach's] whole manner of thinking about the phenomena of natural history' in the years 1795–1797. (Lenoir, 1980, p. 77) Thus, 'Blumenbach's most significant achievement, from our point of view, was to synthesize some of the best elements of Enlightenment thought on biology [...] in terms of a view of biological organization that he found in the writings of Kant.' (Lenoir, 1981a, p. 115) That thesis has remained a powerful influence on the field to this day, but it has serious problems both as a historical claim about the 'Göttingen School' and its founder, Blumenbach, as well as for the larger question of the place of natural teleology in the history of modern biology and even for its status as a special science today.

Lenoir notes: 'It cannot be argued that Blumenbach fashioned himself a follower of Kant.' Instead, 'Kant's main contribution to Blumenbach's work was in making explicit the quite extraordinary assumptions behind the model of the *Bildungstrieb*.' (Lenoir, 1989, pp. 22, 24) My claim is that these were *not* Blumenbach's assumptions, and that he could never assimilate them *as* assumptions, even after he became aware of Kant's 'contribution.'[4] I dispute that any 'revolution in [Blumenbach's] whole manner of thinking' took place, or that the essential features of Blumenbach's life science derived from Kant. On the contrary, I propose to demonstrate here that Blumenbach and his school actually took natural teleology to be an objectively ascertainable feature of biological organisms. Lenoir himself equivocates: Kant's ideas 'only came to be embraced fully by Blumenbach in the period between 1795 and 1797.' (Lenoir, 1980, p. 90) Two points are clear: first, by then, Blumenbach had worked out almost all his important ideas, hence, Kant could not have been 'embraced fully' in their constitution; second, Blumenbach's influence upon the Göttingen School came primarily in the years before 1795. So what is left of the decisive continuity that Lenoir claims, and what of Kant's preeminent place? This systematically undercuts Lenoir's central contention that Kant's philosophy of biology formed the 'hard core' of the 'research programme' (in the Lakatosian sense) of the 'Göttingen School.' (Lenoir, 1981b, 1989, pp. 12–13)

The period between 1786 and 1797 brought the Göttingen physiologist and the Königsberg philosopher into direct communication, and there is clear evidence that Blumenbach assimilated many aspects of Kantianism into his scientific writings. The fullest incorporation of Kant's ideas, entailing abandonment of ideas Blumenbach had long held, came in his *theory of race* after 1797. Lenoir pointed to Blumenbach's completely reorganized third edition of the dissertation on human variety (1795) and the 1797 and 1799 editions of the *Handbuch der Naturgeschichte*. Robert Bernasconi similarly identifies dramatic revisions in Blumenbach's theory of race after 1795 which he associates with Kant. (Bernasconi, 2001a, 2001b, 2001c; compare Lagier, 2004) Phillip Sloan sees a substantial influence of Kant on Blumenbach's ideas about species and organic form in these years, especially via the work of his student and associate, Christoph Girtanner, (Sloan, 1979)[5] There is also evidence in the converse direction, i.e., Kant's assimilation of Blumenbach's scientific work into his own exposition of philosophy of science. Was this a real convergence or was it a mutual misunderstanding? (Richards, 2000; compare Jardine, 2000, pp. 11–55)

There is no question that Blumenbach increasingly *inflected* his theory of the *Bildungstrieb* in language taken from Kant. There is similarly no question that he *incorporated* a great deal of Kant's theory of race into his later writing. (See esp, Blumenbach, 1795) My question is whether Blumenbach actually *understood* and *accepted* the epistemological prescriptions of Kant for biological science. Robert Richards has suggested that Blumenbach's practice was in fact inconsistent with Kant's prescriptions, and that Kant improperly assimilated Blumenbach's practices to his prescriptions. (Richards, 2000, 2002, pp. 221–237) I agree with Richards. (Zammito, 2003) I believe that Lenoir misunderstands both Kant and Blumenbach at crucial points, enabling a false assimilation of their positions. Lenoir does detect a crucial metaphysical and methodological agreement between Kant and Blumenbach: 'it is not possible to reduce life to physics or explain biological organization in terms of physical principles. Rather, organization must be accepted as the primary given [...] At the limits of mechanical explanation in biology we must assume the presence of other types of forces following different types of laws than those of physics. These forces can never be constructed *a priori* from other natural forces, but they can be the object of research. Within the organic realm the various empirical regularities associated with functional organisms can be investigated.' (Lenoir, 1981b, p. 305) Lenoir goes further, however: 'the origin of these original forms themselves can never be the subject of theoretical treatment.' (Ibid., p. 306) But if, as Lenoir elsewhere argues, 'the task of biology is to uncover the laws in terms of which those forces in the organic realm operate' (Lenoir, 1989, p. 33), then, as Robert Richards rightly insists, 'Blumenbach wanted to explain the origin of organization in the first place.' (Richards, 2002, p. 235)[6]

I question Lenoir's conception of empirical science and especially of life science. Lenoir gets off on the wrong foot by suggesting that 'the solution to this problem lies in determining whether the notion of *Naturzweck* is capable of generating *a priori* deductive statements constitutive of experience.' Of course 'it is not possible to offer a deductive, *a priori* scientific treatment of organic forms.' (Lenoir, 1989, p. 28) The fallacy, here, is to believe that *any* substantial part of empirical science—*including physics*—can be deduced *a priori*. Lenoir writes: 'biology as a science must have a completely different character from physics. Biology must always be an empirical science. Its first principles must ultimately be found in experience [...] This contrasts sharply with physics.' (Lenoir, 1980, p. 306, 1989, p. 29) Kant certainly insisted that (some) physics could be deduced *a priori*, but instead of taking Kant's postures about *a priori* science in the *Metaphysical Foundations of Natural Science* (1786) as having any long-term staying power, we must recognize that the turn to

[4] I concur entirely with Robert Richards on this score: 'Blumenbach's *Bldungstrieb* [...] directed the formation of anatomical structures and the operations of physiological processes of the organism [...] Kant would have rejected any such force pretending to be constitutive of nature [...] For Kant, [...] the *Bildungstrieb* could only be a regulative concept [...] But for Blumenbach, [...] [it] was a teleological cause fully resident in nature.' (Richards, 2002, pp. 220–221) While Richards agrees that Blumenbach 'continued gradually to alter and refine the core of the concept,' he denies that this 'turn[ed] the *Bildungstrieb* into what Lenoir has called a teleomechanistic principle.' (Ibid., pp. 226–227) Even in the later editions of his work, 'the *Bildungstrieb* was thus not a Kantian 'as if' cause but a real teleological cause [...] known only through the ends it achieved.' (Ibid., p. 229)

[5] On the other hand, Sloan has been taken to have affirmed a substantial disparity on the question of species between Kant and Blumenbach. (Richards, 2002, p. 235n)

[6] And so does modern biology; see the enormous literature on the problem of the origin of life.

empirical laws as *discovered*, not *deduced*, as contingent, not *a priori*, was the essential advance of the sciences in the modern era. The contrast of biology with physics in *a priori* terms is a function of Kant's *metaphysical* agenda, not a legacy we should embrace.[7] Nor was it one that Blumenbach or his school could embrace. To be sure, they distinguished their science from physics, but not because they believed physics was a deductive *a priori* science and not because they believed that biology was an inherently defective empirical science. They, as Lenoir himself noted, wanted biology to be a legitimate, if special empirical science in a broadly Newtonian unity of science. (Lenoir, 1989, pp. 2–3) Kant preached that biology could *never* be a science at all.[8]

Lenoir's claim boils down to this: 'Kant's formulation of the notion of generic preformationism was an exact, if unhappy, expression of the fundamental idea behind Blumenbach's *Bildungstrieb*,' and Blumenbach acknowledged this. (Lenoir, 1980, p. 91) No, it wasn't, and no, he didn't. To dispute Lenoir's thesis, what is required is a finer-grained consideration of how Blumenbach changed his positions and of the degree to which these can be seen as *accurate* and *informed* adoptions of Kant's views. My strategy will be, first, to lay out Kant's position in a brief sketch (Section 2), and then explore the changes in Blumenbach's thought in the era Lenoir proposes as crucial to his assimilation of Kant's principles (Section 3). Following this, I will turn my attention to the 'Göttingen School' of Blumenbach's students and associates, focusing primarily on Christoph Girtanner, and to the actual course of biological research in the era after 1790 (Section 4).

2. Kant's stipulative methodology for life science[9]

Kant first mentioned Blumenbach in a footnote to his 1788 rejoinder to Georg Forster, *On the Use of Teleological Principles in Philosophy*. (Kant, 1788, p. 180n) He invoked Blumenbach's authority to dismiss the transformation of the great chain of being from a taxonomy to a phylogeny—that is, what later, in the *Critique of Judgment*, he would call a 'daring adventure of reason.' (Kant, 1790, 419n)[10] Forster had questioned this 'widely cherished notion preeminently advanced by Bonnet' and Kant was happy to report that, under the critical scrutiny of Blumenbach's *Handbuch der Naturgeschichte*, all the weaknesses of that position had been exposed. (Kant, 1788, p. 180n)[11] Then he added the observation: 'this insightful man also ascribes the *Bildungstrieb*, through which he has shed so much light on the doctrine of generation, not to inorganic matter but solely to the members of organic being.' (Kant, 1788, p. 180n)[12] In 1790, in the *Critique of Judgment*, Kant elaborated:

> He makes organic substance the starting point for physical explanation of these formations. For to suppose that crude matter, obeying mechanical laws, was originally its own architect, that life could have sprung up from the nature of what is void of life, and matter have spontaneously adopted the form of a self-maintaining finality, he justly declares to be contrary to reason. (Kant, 1790, pp. 378–379)

There were few ideas Kant struggled to keep divided more than life and matter. It is the idea of *hylozoism*—of any radical spontaneity in matter itself—that Kant could not abide.[13] Kant denied that we could even think of nature as alive: 'the possibility of living matter cannot even be thought; its concept involves a contradiction, because lifelessness, *inertia*, constitutes the essential character of matter.' (Kant, 1790, p. 394) He elaborated: 'life means the capacity of a substance to determine itself to act from an internal principle, of a finite substance to determine itself to change, and of a material substance to determine itself to motion or rest as change of its state.' (Kant, 1786, p. 544)[14] Consequently, he wished to secure the distinction of organic life from the inorganic, affirming the uniqueness and mystery of organisms as phenomena of empirical nature, and upholding the utter inexplicability of the origins of life.[15]

Marcel Quarfood sets the discussion of Kant's conceptualizations of organism as *Naturzweck* in the proper frame by asserting: 'The distinctive feature of Kant's view is [...] an *epistemic presupposition* constitutive for the study of life, rather than a definite *ontological* commitment.' (Quarfood, 2004, p. 145) Joan Steigerwald agrees Kant was concerned with the 'epistemic conditions of our estimation of living beings, the conditions of the possibility of our cognition of them, not with the nature of living beings.' (Steigerwald, 2006, pp. 2–3; now, more extensively: Zuckert, 2007) That might be a possible posture for a *philosopher* of science, but it is *not* a stance that can have any appeal to practicing life-scientists, for their inquiry *must* be into the 'nature of living beings' and to be denied cognitive access to it is to be stipulatively stripped of a scientific domain.[16] Quarfood has gone so far as to suggest that what Kant really meant was that *transcendental philosophers* should consider the conceptualization of organisms as merely 'regulative' but that he recognized that for practicing biologists it had to be 'constitutive.' (Quarfood, 2006) Unfortunately, that is not true, but it would certainly have made Kant more amenable to practitioners of life science.

The 'marvelous properties of organized creatures,' which Kant adumbrates with confidence in the 'Analytic' of his 'Critique of Teleological Judgment,' are part of the empirical-experiential data available to human investigators trying to comprehend the order

[7] On Kant's philosophy of science and its 'looseness of fit' with the critical philosophy, see especially Buchdahl (1965, 1967, 1969a, 1971, 1981, 1986, 1991); see also Friedman (1986, 1990, 1991, 1992a, 1992b); Allison (1991, 1994), Guyer (2001, 2003, 2005), Kitcher (1983, 1986, 1994), Morrison (1989), Okruhlik (1983), Butts (1990).

[8] Thus, Richards writes: 'Most biologists of the period [...] thought their disciplines could be developed into sciences and could, in that respect, come to stand as certainly on that pinnacle of human accomplishment as Newton's physics. They believed [...] that teleological processes could be found governing natural phenomena and that valid laws could be formulated to capture such relationships.' (Richards, 2002, p. 231) He adds, in a note: 'That Kant excluded biology from the realm of real science (*Wissenschaft*) is, I think, indisputable.' (Ibid., p. 231n) That point needs to be hammered heavily: it pierces not only Lenoir's thesis but the whole effort to retrieve Kant as the basis for a more sophisticated philosophy of biology today. See Zammito (2006c).

[9] For an overview of the field today, see Heidemann, ed. (2009), Huneman, ed. (2007), Steigerwald, ed. (2006).

[10] On this historicization of the 'great chain of being,' see the classic Lovejoy (1936)

[11] He cited the first edition of the *Handbuch der Naturgeschichte* (1779), which he owned.

[12] It is not entirely clear when or how Kant came to know about Blumenbach's *Bildungstrieb*. It was not mentioned in the 1779 edition, but there were numerous other formulations—the original article version in the *Göttingisches Magazin* (1780), the first book version on the *Bildungstrieb* of 1781, the second edition of the *Handbuch der Naturgeschichte*—any of which Kant might well have perused, for he read voraciously and from all quarters—or the Latin versions (Blumenbach, 1785, 1787) explaining his discovery, which appeared before Kant's 1788 essay.

[13] This constituted a decisive influence on Kant's receptivity towards the theory of epigenesis. See Zammito (2006a, 2007).

[14] Hence Kant situated himself squarely in the tradition of the new scientific rationalism. For an old but still trenchant assessment of this view, see Burtt (1954). For a more recent, penetrating analysis, see Buchdahl (1969b).

[15] For a recent study of Kant's theory of organic form, see Löw (1980), esp. 138ff. For the older literature, see Menzer (1911), Roretz (1922), pp. 112–150; Ungerer (1922), 64–132; Bommersheim (1919, 1927), Lieber (1950), Baumanns (1965).

[16] Biology is a special science concerned with actual entities in the physical world; it is not reasonable to pursue such an enterprise if it is *in principle* not possible to explain those entities. It may well be that such explanations are contingent and fallible, but biologists *must* resist any imposition by philosophy that would stipulate the impossibility of the venture. See Zammito (2003, 2006b); for an alternative view, see Breitenbach (2009).

of nature. (Kant, 1790, p. 371)[17] That is, Kant appears to consider their *phenomenal description* unproblematic. But how these 'marvelous properties' can be explained as actual entities—and how they can be *integrated* into a unified system of empirical laws as the 'order of nature'—remains, for Kant, a philosophical conundrum. (Zammito, 2003) As Quarfood explains, 'organisms like all objects of experience are subject to the causal principle,' but 'there are features of organisms that appear to be intractable for the kind of explanations in terms of causal laws appropriate for ordinary physical objects' and thus 'there is no explanation (or 'law') for how matter comes together in the ways characteristic for organisms.' (Quarfood, 2004, p. 146) Kant characterizes what presents itself as an organism 'provisionally, [as] a thing [that] is *both cause and effect of itself*.' (Kant, 1790, p. 370) While we can 'think this causality without contradiction, we cannot grasp [*begreifen*] it.' (Kant, 1790, p. 371) That is, we cannot bring it under concepts of the understanding.

That an entity can be cause and effect of itself, Kant argued, is beyond discursive rationality. To take teleology as explanatory would 'introduce a new causality into natural science, even though in fact we only borrow this causality from ourselves' (Kant, 1790, p. 361) This would be a quite 'special kind of causality, or at least a quite distinct lawfulness [*Gesetzmäßigkeit*] of nature' and 'even experience cannot prove that there actually are such purposes [*die Wirklichkeit derselben* [...] *beweisen*].' (Kant, 1790, p. 359) Kant insists 'natural purpose' is *our construction*, not an empirical given. What is empirically given is a problem, an anomaly, not a fact. Steigerwald stresses that Kant claimed we could only grasp 'living beings by reference to our own purposive activity,' i.e. he maintained only the *analogy to human purpose* gave us conceptual access to organic form. (Steigerwald, 2006, pp. 1–3) Technically, Kant had to deny that teleology can explain anything in phenomenal nature. (cf. Flasch, 1997; Fricke, 1990; Ginsborg, 1987; Warnke, 1992) What teleology is alone permitted to do is offer an *analogy* of some *heuristic* utility. It is even less than an empirical *conjecture*.

> We perhaps approach nearer to this inscrutable property if we describe it as an *analogon of life*, but then we must either endow matter, as mere matter, with a property which contradicts its very being (hylozoism) or associate therewith an alien principle *standing in communion* with it (a soul). But in the latter case we must, if such a product is to be a natural product, either presuppose organized matter as the instrument of that soul, which does not make the soul a whit more comprehensible, or regard the soul as artificer of this structure, and so remove the product from (corporeal) nature. (Kant, 1790, pp. 374–375)

In short, 'strictly speaking, [...] the organization of nature has nothing analogous to any causality known to us,' that is, '*intrinsic natural perfection*, as possessed by those things that are possible only as *natural purposes* and that are hence called organized beings, is not conceivable or explicable on any analogy to any known physical ability, i.e., ability of nature, not even—since we belong to nature in the broadest sense—on a precisely fitting analogy to human art.' (Kant, 1790, p. 375)

Consequently, Kant's notion of *organism* is broader than that of *life*, and the failure of these two terms to have the same extension expresses the insufficiency Kant acknowledged in his 'analogy of life' for natural purpose. (Dörflinger, 2000; Ingensiep, 2004) How do we construe the residual *dis*analogy for biology? Plants epitomize Kant's conceptual discrimination of life from organism. They are very hard to reconcile with Kant's stipulative formulation of 'life' and yet they are unquestionably 'natural purposes' in Kant's technical sense. In the opening exposition of the 'Critique of Teleological Judgment' in the third *Critique*, he illustrated the features of organism precisely by a plant—a tree. Even plants have a *Bildungstrieb*, not just *Bildungskraft*, in the discrimination Kant adopted from Blumenbach. (Kant, 1790, p. 424) That is, they have some 'internal, quasi-spontaneous principle of motion.' (Ingensiep, 2004, p. 128) The question of *Trieb* in Kant's notion of organism denotes the element unaccounted for even by Kant's analogy of life. Organisms were clearly identified with *Trieb*. But what was a *Trieb* for Kant, and how did he distinguish it from a *Kraft*? How could an 'inner' force be actual for scientific inquiry? Lenoir is comfortable with a regulative, heuristic formulation of the matter, but that oversimplified Kant's actual endeavors, especially if we consider his entire career as a *Naturforscher*.[18]

Kant did explicitly develop a scientific theory about such inner organismic forces in his essays on race. Lenoir is misleading (or misled) in suggesting that for Kant this was all simply 'subjective' or 'heuristic' in a manner that disowned empirical assertion. He writes: 'Kant's *Stamm* [...] is an Ideal Type, a transcendental idea whose only significance is regulative.' (Lenoir, 1978, p. 68) The *Stamm* 'is not to be conceived as an ancestral form.' (Lenoir, 1988, p. 107) 'The *Stamm* was a hypothetical construct of reason [...] it contained schematically all possible morphological structures within a given order' (Lenoir, 1978, p. 69) 'Rather than seeing these organic unities reconstructed by comparative anatomy as potential historical ancestors, it is more appropriate to view them as *plans of organization*, as the particular ways in which the forces constituting the organic world can be assembled into functional organs and systems capable of surviving.' (Lenoir, 1981b, pp. 308–309) In a late article, published after the original edition of his book, Lenoir writes: 'Kant advocated the construction of morphotypes or organizational plans to be arrived at through comparative anatomy and physiology.' (Lenoir, 1988, p. 107) That, I submit, is 'rational reconstruction' (or in blunter German, *hineinlesen*) with a vengeance. Lenoir may wish to *interpret* Kant as holding this, but there is no such explicit advocacy in Kant's texts. Kant, in contrast, developed an explicit *empirical* hypothesis, alleging actual causal relations in the physical world.[19] Lenoir in one article does recognize that 'Kant had gone on to provide a mechanical model [...] in a set of *Keime* and *Anlagen* present in the generative fluid.' (Lenoir 1981b, p. 307), but for the most part he wants to claim that Kant restricted himself to a heuristic, a regulative 'as if.' On the contrary, Kant's theory of *Keime* and *Anlagen* was, like *all* empirical hypotheses, a matter of construction ('a model'), involving theoretical terms to account for observable macrophenomena, and hence dependent upon empirical confirmation, if only holistically. Bluntly, *Stammgattung* is not *simply* an 'ideal' in Kant's technical sense; it is a theoretical concept to which is imputed a determinate historical actuality.[20]

[17] But to claim, as Lenoir does, 'that such 'natural purposes' exist is an objective fact of experience according to Kant.' (Lenoir, 1989, p. 25) is in flat contradiction to what Kant wrote in the *Critique of Judgment*: 'even experience cannot prove that there actually are such purposes [*die Wirklichkeit derselben ... beweisen*].' (1790, p. 359) For Kant, the whole problem is about the *recognition* of an anomaly in the empirical order of nature and its *conception* as a 'natural purpose.' The anomaly violates the categorial framework of the understanding, and the conception is a *subjective recourse* to deal with it.

[18] Long ago, Erich Adickes wrote extensively about Kant's sense of himself as a 'Naturforscher.' Adickes judged Kant's expertise in natural history and physical anthropology far more harshly than more recent commentators, and it turns out he was more apt than they. (Adickes, 1924, pp. 406–459)

[19] Over the 1780s, as Kant worked up the critical philosophy, Rafael Lagier has argued quite persuasively that this *empirical* component waned in the fact of increasing epistemological scruples. See Lagier (2004, p. 140).

[20] A better way to make sense of this whole problem of idealizations in scientific model building and its relation to scientific 'objectivity' is developed especially in the treatment of eighteenth-century science in Daston & Galison (2007).

Kant conceived of *Keime* [germs] and *natürliche Anlagen* [natural potentialities] as real forces in human variation. (Kant, 1775–77)[21] Clark Zumbach observes, for example: '*Keime*, as part of the generative force [*Zeugungskraft*], are postulated [...] as the inner mechanisms for development in future circumstances [T]hey control the permanence of phenotypic traits and are 'kept back or unfolded' depending on the situation at hand.' (Zumbach, 1984, p. 102) Through them Kant sought to characterize the mysterious 'inner possibility' of organic form in its objective reality or real possibility. What kind of 'theoretical terms' did they constitute, and what sorts of observational evidence could substantiate them? The cognitive status of these concepts is all the more pressing since Kant postulated an *original* or *ancestral* form [*Stammgattung*] which, at least in the case of humans and in all likelihood for any other life forms, no longer persisted in the present.[22] Without some *empirically determinate* principle of the derivation of current species from these ancestors, the whole approach would be less than an art, it would be arrant speculation.[23] In Kantian terms, what made these *'real* possibilities' and not just wild hypotheses irreconcilable with 'proper Newtonian science?'[24]

To grasp that, we must consider Kant's advocacy of a newly emergent empirical science in the late eighteenth century, for which he proposed to appropriate the going concept, *Naturgeschichte*. The term 'natural history' in German science before Kant had really only signified natural *description*. It was heuristic and classificatory, as exemplified above all by Linnaeus. Kant, taking up impulses from Buffon, suggested in 1777 this could be displaced by a real and genetic conception of the order of living forms (*Naturgattungen* in place of *Schulgattungen*), making *history* central to the project of the life sciences. (Kant, 1775–77; Zammito, 2010) But Kant came increasingly to doubt the efficacy of this new empirical science. Above all, 'how this stock [of *Keime*] arose, is an assignment which lies entirely beyond the borders of humanly possible *natural philosophy*, within which I believe I must contain myself,' Kant proclaimed. (Kant, 1788, p. 179)[25] 'Chance or general mechanical laws can never bring about such adaptation. Therefore we must see such developments which appear accidental according to them, as *predetermined* [*vorgebildet*].' External factors could be occasions, but not direct causes of changes that could be inherited through generation. 'As little as chance or physical-mechanical causes can generate [*hervorbringen*] an organic body, so little will they be able to effect in them a modification of their reproductive powers which can be inherited.' (Kant, 1790, p. 435) This was the essential postulate to which Kant had committed himself in his second essay on race (1785), and the stakes were not small: without some fixity in the power of generation [*Zeugungskraft*], the prospect of the scientific reconstruction of the connection between current and originating species—*Naturgeschichte,* as Kant formulated it in his first essay on race (1775–77), or the 'archaeology of nature' as he would call it in the third *Critique* (Kant, 1790, p. 424)—would be altogether dim.

Yet it was not simply a *methodological* issue, however dire. There was also an essential *metaphysical* component. Kant was adamant that the *ultimate* origin of 'organization' or of the formative drive [*Bildungstrieb*] required a *metaphysical*, not a physical, account. (see Zammito 2003, 2006c, 2007, 2009, forthcoming; cf. Rang, 1998; Quarfood, 2004; Steigerwald, 2006; Zuckert, 2007; Breitenbach, 2009; Beihart, 2009) All organic form had to be fundamentally distinguished from mere matter. 'Organization' demanded separate creation. 'This inscrutable *principle* of an original *organization*' lay beyond natural science. (Kant, 1790, p. 424) That put *life* science beyond the pale of empirical science. Organisms, as empirically given—indeed, *pervasive*—occurrences in nature, became literally indecipherable once the concept *life* was removed, leaving us to grope after them by analogies. In the third *Critique* Kant would twice insist that there would never be a Newton of even a 'blade of grass.' (Kant, 1790, pp. 400, 429) Robert Richards says what needs to be said: 'the *Kritik der Urteilskraft* delivered up a profound indictment of any biological discipline attempting to become a science.' (Richards, 2002, p. 229) Eternal inscrutability was preferable to any 'monstrous' conjectures of hylozoism and transformationism that made reason flinch. (Kant, 1785a, p. 54)[26] Kant invoked Blumenbach for support in these *metaphysical* reservations. (Kant, 1788, p. 180n; 1790, p. 424) The leading life scientist of the day seemed to be affirming just the same metaphysical and methodological discriminations that Kant himself demanded. But could Blumenbach, whose whole career exemplified a 'biological discipline attempting to become a science,' really have embraced such a philosophy of science? That disconnect puts Kant's appropriation of Blumenbach—and *a fortiori* Lenoir's assimilation of the two of them—starkly in question.

3. Blumenbach's life science and Kant's influence

Blumenbach began serious consideration of the philosophy of Kant in 1786 as a direct consequence of the dispute surrounding Kant's reviews of Herder's *Ideen zur Philosophie der Geschichte der Menschheit*, especially Kant's controversy with Georg Forster. (Forster, 1786; see Riedel, 1980; Lüsebrink, 1994; Schmied-Kowarzik, 1994; Strack, 2001; Weingarten, 1982; and above all, van Hoorn, 2004) But already five years before, in 1781, Blumenbach proposed the most important revision in the 18th-century fields of embryology and physiology with his idea of the *Bildungstrieb* and his implied endorsement of epigenesis. (Blumenbach, 1781) How did Blumenbach respond to Kant's appropriation of his ideas? Blumenbach's first major publication after Kant's essay appeared, the third edition of the *Handbuch der Naturgeschichte*, was dated March 1788, and it unsurprisingly gives no evidence of Blumenbach's attention to Kantian ideas. (Blumenbach, 1788) But less than a year later, in January 1789, he published his revised version of *Über den Bildungstrieb* and sent Kant a copy of this work in acknowledgment

[21] This theory was reasserted without revision in Kant's reviews of Herder in 1784/5 and in the 1785 reprise of Kant's treatment of race, then defended against Georg Forster in 1788. It remains (vestigially) in the *Critique of Judgment*. See Zammito (2006b).

[22] 'Indeed, if we depart from this principle, we cannot know with certainty whether several parts of the form which is now apparent in a species have not a contingent and unpurposive origin; and the principle of teleology: to judge nothing in an organized being as unpurposive which maintains it in its propagation, would be very unreliable in its application and would be reliable solely for the original stock (of which we have no further knowledge).' (Kant, 1790, p. 420)

[23] Here I am invoking the language from the Preface to Kant (1786, pp. 467–469).

[24] Here I am invoking the framework proposed by Buchdahl (1965) etc. See Zammito (2003, 2006c).

[25] '[I]f some magical power of imagination [...] were capable of modifying ... the reproductive faculty itself, of transforming Nature's original model or of making additions to it, [...] we should no longer know from what original Nature had begun, nor how far the alteration of that original may proceed, nor [...] into what grotesqueries of form species might eventually be transmogrified [...] I for my part adopt it as a fundamental principle to recognize no power ... to meddle with the reproductive work of Nature [...] [to] effect changes in the ancient original of a species in any such way as to implant those changes in the reproductive process and make them hereditary.' (Kant, 1785, p. 97; tr. in Lovejoy 1959, p. 184)

[26] The connection between this reaction to Herder, Kant's equivocations in the debate with Forster, and his eventual discussion of the 'daring adventure of reason' is crucial to an understanding of this whole configuration. In my view, Lenoir (1981a, pp. 150–514), gets Kant's argument in §§ 80–81 of the *Critique of Judgment* altogether wrong. He thinks Kant was *affirming* what he was in fact problematizing. Ironically, Lenoir's misreading tallies with that of most creative life scientists in the 1790s. But that means it was not Kant but their misinterpretation of Kant that was the driving force here. See Sloan (2006).

of Kant's references to him in the 1788 essay. (Blumenbach, 1789)[27] The Preface to this second edition of his essay on the *Bildungstrieb* advised readers that his earlier version was 'immature.' (Blumenbach, 1789, p. A4)

What did Blumenbach intend by his Preface of January 1789, and by routine appeals in later versions of his *Handbuch* and of his dissertation on human variety, 'not to confuse [this second edition] with the immature treatise that appeared under a similar title in 1781'? (Blumenbach, 1789, p. A4)[28] Can we take it for granted that this was 'immaturity' by Kantian standards? Lenoir explicitly claims Blumenbach's 'mature formulation resulted from his encounter with Kant's work.' (Lenoir, 1980, p. 84n)[29] That is not historically defensible for the Preface of January 1789, and it is quite problematic for later incorporations of Kantian language. I suggest that we must regard Blumenbach's judgment of his earlier work in a more complex light. He was already making changes in his 1788 *Handbuch*, before we have any reason to suspect Kantian influence. He had encountered significant resistance to his ideas—and from two fronts: the die-hard preformationists (Bonnet, Spallanzani, Caldani), but also the more aggressively naturalistic epigenesists—Thomas Sömmerring and Georg Forster and, of far greater importance, Caspar Friedrich Wolff.[30] If we consider the texts of 1781 and 1789 in juxtaposition, what is foremost is the clarity with which Blumenbach characterizes his central innovation. The structure of the argument is considerably clearer: after the historical background leading up to his own discovery, Blumenbach presents a thorough drubbing of the arguments for preformation, followed by a clear account of the advantages of his *Bildungstrieb* theory. He is far more comfortable that he has made a major breakthrough and that he has defeated his rivals on that front. That is, Blumenbach believed he had dramatically improved the exposition of his *scientific* position by 1789, not—or not *just*—his sophistication about philosophy of science.

One of the most important aspects of his argument in 1781 was that the *Bildungstrieb* encompassed and explained three vital functions—generation, nutrition, and regeneration. In the 1789 version, nutrition gets scant attention. It is generation and regeneration that Blumenbach believes offer the best support for his theory in comparison with others. But it may also be that he had addressed the nutrition issue separately, in a prize-winning essay submitted to a competition sponsored by the Academy of Sciences in St. Petersburg, and presided over by his rival epigenesist, Caspar Friedrich Wolff. (Blumenbach, 1789b) While Wolff awarded Blumenbach the prize, he published a far lengthier work of his own on the topic, taking a sharply critical posture towards Blumenbach's views. (Wolff, 1789) There are thus grounds to think that there is another presence besides Kant whose appraisal of his work loomed large for Blumenbach in 1789, namely Wolff.

And this might well explain the most important methodological clarification in the 1789 version, which did bring Blumenbach happily into alignment with Kant: the radical separation of organic from inorganic form and the repudiation of any hylozoism. Blumenbach embraced a fundamental ontological distinction between the general order of nature and the specific order of the organic. In the 1789 version of his *Bildungstrieb* book, Blumenbach made this very clear: 'No one could be more totally convinced by something than I am of the mighty abyss which nature has entrenched [*befestigt*] between the living and the lifeless creation, between the organized and the unorganized creatures.' (Blumenbach, 1789, p. 79) Indeed, Blumenbach shared Kant's skepticism about a bridge from the inorganic to the organic and about the phylogenetic continuity of life forms. What bound them most together was their commitment to the fixity of species and their rejection of the reality of the *scala naturae*. Yet Blumenbach drew neither of these commitments from Kant. They were already expressed with clarity in his dissertation of 1775 and especially the first edition of his *Handbuch* of 1779. These were basic issues for anyone taking up natural history or life science in the 18th century. It is far more likely that Blumenbach adopted them from Albrecht von Haller than from Kant.[31] What remains is to consider whether the *reasons* for Blumenbach's commitments were the same as the reasons for Kant's commitments to these same positions.

When he first presented his notion of the *Bildungstrieb*, Blumenbach concentrated on how it answered certain physiological problems in organisms better than the alternative theories of preformation and of epigenesis. He did not dwell yet on the methodological or epistemological status of his concept. In the 1782 edition of his *Handbuch der Naturgeschichte*, his treatment of the idea of the *Bildungstrieb* once again gave no attention to this epistemological issue. He simply carried forward with his empirical exposition. Perhaps he came to regard this as one of the 'immature' features of his work. He changed already in the 1788 edition of the *Handbuch*—presumably before he could have absorbed very much of Kant's methodological thinking. He introduced a new section, immediately after defining his *Bildungstrieb*, with the following language:

> The *cause* of this formative drive can admittedly be so little adduced as that of attraction or gravity and other such generally recognized natural forces. It is enough that it is a distinctive force whose undeniable existence and broad influence throughout all of nature is revealed by experience, and whose constant phenomena offer a far more ready and clear insight into generation and many other of the most important topics of natural history than other theories offered for their explanation. (Blumenbach, 1788, p. 14)

There is, here, a tacit Newtonian analogy, without the mention of Newton by name. Moreover, the argument is presented in terms of the *general* order of nature: no strong distinction is made between the organic and the inorganic realms in terms of the nature of such forces, though, clearly, this particular force operates in generation and organic forms.

In the second edition of his *Bildungstrieb* book, Blumenbach became far more explicit about the Newtonian connection: 'The *term* formative drive, just like the *terms* attraction and gravity, etc. serve no more and no less than to denote a force whose constant effect is recognized but whose *cause* just as little as the causes of the other, nonetheless so generally recognized natural forces, remains for us a *qualitas occulta*. That does not hinder us in any way whatsoever, however, from attempting to investigate the effects of this force through empirical observations and to bring them under general laws.' (Blumenbach, 1789, pp. 32–33) In the attached footnotes, Blumenbach referred directly to Newton, and then, in the context of the phrase *qualitas occulta*, to Voltaire's exposition of Newton,

[27] Blumenbach's transmission to Kant in 1789 is acknowledged by Kant in his letter to Blumenbach, August 5, 1790, in Kant, B, AA:11, pp. 176–177.

[28] For later avowals along the same lines see, for instance, Blumenbach (1791, p. 13; 1797, p. 17n).

[29] Lenoir argues that Blumenbach's 'mature theory' was composed only 'after he had begun to wrestle with Kant's philosophy of organic form,' and ostensibly upon that basis. (Lenoir, 1980, p. 83)

[30] For one seminal discussion of the epigenesis controversy in Germany, see Shirley Roe (1981). See also: Duchesneau (1979, 1985), Roger (1963, 1968, 1980), Gaisinnovich (1968), Müller-Sievers (1993, 1997); and with specific reference to Kant, Ginsborg (1987, 2001, 2004), Genova (1974), McLaughlin (1990), Hunemann (2002), Huneman, ed. (2007), Wubnig (1968/69).

[31] Blumenbach commented: 'I think it says a lot—but, as I see it, not too much—when I maintain that Haller was the greatest among all recently deceased scholars who have been working in Europe since Leibniz's death. He was the greatest scholar as concerns variety as well as quantity and depth of his knowledge.' (Blumenbach, *Medicinische Bibliothek* 2 (Göttingen, 1785, p. 177)

in particular to the passage where Voltaire argued that from a mere 'blade of grass' to the order of the stars, *all* causes (physical as well as biological) were simply occult qualities. (Blumenbach, 1789, pp. 32n, 33n)[32] This was standard epistemology of science in the wake of John Locke's discrimination of 'nominal' from 'real' essences, of empirical (external) observation from 'inner' or ultimate reality of nature. (Locke, 1698) It is important to stress that Kant was hardly a necessary influence for Blumenbach in making this Newtonian appeal. It was common practice among all innovative life scientists. Haller and Buffon had done it, and so had Caspar Friedrich Wolff. (Wolff, 1764; see, e.g., Gaissinovich, 1968; Roe, 1979) As Peter McLaughlin has argued, making the Newtonian appeal was constitutive for the emergent life sciences in the late 18th century. (McLaughlin, 1982) While it was epistemologically expedient, this may well have been disingenuous in many cases, for the forces were taken quite straightforwardly as real, even if the *ultimate* causes remained mysterious. That anything like Kant's critical epistemology was in play must be open to considerable doubt.

Ultimately, then, what major break was there between Blumenbach's 1781 formulation and the new 'mature' formulation of 1789? McLaughlin has set this inquiry on the proper path by a very close reading of Blumenbach's various formulations of the notion of the *Bildungstrieb* in successive publications. As McLaughlin is quite right to maintain, Blumenbach did not do a very good job in explicating his *Bildungstrieb*: 'what that is supposed to mean exactly is nowhere systematically elaborated.' (McLaughlin, 1982, p. 364)[33] But McLaughlin offers three avenues to clarify the concept: first, how Blumenbach contrasted it with other theories and other forces; second, how he specified its typical laws of operation; and, finally, how he used it to explain other phenomena in natural history. (McLaughlin, 1982, p. 364) For McLaughlin, the contrast with C. F. Wolff is most illuminating, and the issue of the relation of the formative drive to organic matter is central. I think that is exactly the right line of attack, though I deviate somewhat from McLaughlin in the interpretation of these matters.

In his *Handbuch* of 1782, Blumenbach wrote that 'a particular, innate *drive*, active throughout its life, lies in every organized body.' (Blumenbach, 1782, p. 15) In the *Handbuch* of 1788, he wrote that one could find 'throughout all nature the most unmistakable traces of a virtually general drive to give matter a determinate form, which already in the inorganic realm is of striking effectiveness.' (Blumenbach, 1788, pp. 12–13) As McLaughlin properly observes: 'In fact, the *only clear substantive difference* in the key formulations of the theory of the *Bildungstrieb* between the 'more mature' and the 'immature' phase is the replacement of an 'innate' drive by a 'general' drive.' (McLaughlin, 1982, p. 371) As the editor of the reprint of Blumenbach's classic comments, though Blumenbach called the earlier version immature, 'nevertheless even stylistically the essential statements are hardly changed' in the later ones. (Karolyi, 1971, p. vi) What Karolyi discerns is a clearer self-assertion versus Haller and Wolff, but 'argumentation, examples, and the core of the statement remain unchanged.' (Karolyi, 1971, p. xii) There were, to be sure, 'in part more refined, more differentiated formulations and some additions to the exposition of the first edition,' but for Karolyi these hardly amounted to the 'completely new construction of the theme' alleged by Robert Herrlinger in his preface to the reprint of the work of C. F. Wolff. (Karolyi, 1971, p. xi) Herrlinger had implied Blumenbach *needed* such a new formulation in light of Wolff's criticisms. (Herrlinger, 1966, p. 19n) In short, there is more to the tension between Blumenbach and C. F. Wolff than to the affinity of Blumenbach to Kant that needs to be considered in Blumenbach's discomfort with the 'immaturity' of his work of 1781.

This would make no sense if Blumenbach really believed that C. F. Wolff was a 'mystical' vitalist, as Lenoir strangely conceives him.[34] Rather, it is the notion of a continuity from the inorganic (in Wolff, the chemical) to the organic—i.e., a materialist naturalism or 'hylozoism'—in Wolff that Blumenbach wishes to distance himself from. Blumenbach found Wolff's notion of epigenesis problematic as much—or more—for the metaphysical quandaries as for the methodological ones. There is a high level of ambivalence and ambiguity in his critique of Wolff and in his assimilation of Kantian principles over the 1780s and early 1790s, such that his own position has occasioned widely divergent reconstructions.[35] There is good reason to question whether his ultimate version of epigenesis diverged that substantially from Wolff's, despite all his efforts to uphold a difference.[36] That professed difference, nonetheless, proved central for his affiliation with Kant.

McLaughlin identifies crucial changes that Blumenbach introduced in 1791, after he had absorbed Kant's ideas not only from the 1788 essay but from the *Critique of Judgment* which Kant had sent him. As we have noted, in 1788 Blumenbach found 'throughout all nature the most unmistakable traces of a virtually general drive to give matter a determinate form, which already in the inorganic realm is of striking effectiveness.' (Blumenbach, 1788, pp. 12–13) In 1791, Blumenbach pruned the line as follows: one finds 'in the entirety of organic nature the most unmistakable traces of a generally distributed drive to give matter a determinate form.' (Blumenbach, 1791, p. 14) The appended clause from 1788 was eliminated altogether. In 1789, as we have noted, Blumenbach compared the *Bildungstrieb* to 'the *terms* attraction and gravity [...] generally recognized natural forces' But in 1797, he changed this to: 'The term *Bildungstrieb* just like all other life forces' (Blumenbach, 1797, p. 18) The point, here, is that Blumenbach wished his formative drive to be considered only in comparison with other *life-forces*. The thrust, as McLaughlin notes, was to make a radical distinction between the organic and the inorganic realms and to assign the drive exclusively to the former.

The point that McLaughlin wishes to derive from this shift in position in Blumenbach by 1791 is that the *Bildungstrieb* is not the *cause* of life but rather its *consequence*. (McLaughlin, 1982, p. 359) That is, what all the earlier (materialist/naturalist) proponents of epigenesis sought to explain (life as an emergent property arising out of matter itself) in Blumenbach becomes an inexplicable presupposition. For La Mettrie, Buffon and Holbach,

[32] One wonders whether it was not this passage from Voltaire that provoked in Kant the famous passage that there would never be a Newton of the blade of grass.

[33] Jardine moves too quickly from the correct observation that Blumenbach 'offers no positive account of the nature of the formative drive' to the inference that 'it is proposed as a heuristic in the search for empirical laws ...' (Jardine, 2000, p. 26) The Newtonian analogy did not minimize at all the actuality of the formative drive, but only denied access to its *ultimate* cause. This is a vital discrimination if we are to understand the relation between Blumenbach and Kant.

[34] I think in several of his publications Lenoir misunderstands Wolff in a manner than sets up his misconstrual of the whole epoch of life science from the late 18th to the mid-19th centuries, because he identifies vitalism with 'idealism'—i.e., animism. We must rescue 'vital materialism' from Lenoir's residual positivism. See Reill (2005). Further, Lenoir imputes to the imaginative construction of hypotheses in life science a 'mystical' propensity—or a (privatively) 'aesthetic' one—that deeply misprises (as irrational) the *interpretive* idea of science that was being developed by its most brilliant eighteenth-century expositors—Buffon, Daubenton, Diderot, Camper, Goethe, and Herder. See Daston & Galison (2007).

[35] Thus different interpreters see Blumenbach moving towards vitalism or away from it, as achieving the clear distinctions between constitutive and regulative that Kant required and as dissolving these, e.g., McLaughlin vs Lenoir on the first, Larson vs. Lenoir on the second. See Larson (1979).

[36] Most commentators are hard-pressed to uphold, though they clearly try to articulate, what Blumenbach thought distinguished himself from Wolff. For a good discussion, see McLaughlin (1982, pp. 365–367).

according to McLaughlin, 'life was a mechanical result of organization'—that is, of the general order of nature grounded in physics and chemistry. Blumenbach, by contrast, aimed 'to explain organic *form* through organic *matter*.' That is, an *organic* force is 'a force that only *has effect within* organic matter, not a force that somehow *causes* the transition from inorganic to organic matter.' The *Bildungstrieb* did not explain life but rather presumed it. (McLaughlin, 1982, p. 357)[37] While there was organization already in inorganic matter, there was something extra about organic matter, which John Hunter called a 'supplementary force,' something 'applied in addition.' (McLaughlin, 1982, p. 359)[38]

For McLaughlin, 'Wolff's essential force was a *chemical* attraction-repulsion force.' (McLaughlin, 1982, p. 365) Thus, for Wolff, matter was heterogeneous, i.e., it achieved various *levels* of organization, and once it passed a certain threshold, there ensued something of a chemical chain-reaction that initiated life. The important inquiry was into the component constraints that directed the chain-reaction. (Wolff, 1764; see the divergent views on Wolff: Aulie, 1961; Gaissinovich, 1968, 1990; Herrlinger, 1959; Lukina, 1975; Mocek, 1995; Roe, 1979, 1981; Schuster, 1941; Uschmann, 1955) For Blumenbach, by contrast, McLaughlin believes the important question was the *inherent* relation between a distinctively organic matter and the forces unique to it. That did not mean one could not draw *analogies* from the inorganic to the organic, for, Blumenbach wrote,

> even in the inorganic realm the traces of formative forces are so unmistakable and so general. Of formative *forces*—but not by far of the formative *drive* (*nisus formativus*) in the sense this term assumes in the current study, for it is a *life*-force [*Lebenskraft*] and accordingly as such inconceivable in inorganic creation—rather of other formative forces, which provide the clearest proof in this inorganic realm of nature of determinate and everywhere regular formations [*Gestaltungen*] shaped out of a previously formless matter. (Blumenbach, 1789, p. 80; my emphases)

This distinction between the formative *forces* [*Kräfte*] that structure the inorganic realm and the formative *drive* (*Trieb*; note that it is always singular in Blumenbach's usage) which is unique to organic life, and indeed a *Lebenskraft* among others, proved crucial for Kant.

This was what Kant found most gratifying in the new book, as he reported in his letter of acknowledgment to Blumenbach, August 5, 1790. (Kant, B, AA: 11, pp. 176–177) In the *Critique of Judgment* he elaborated:

> Blumenbach [...] rightly declares it to be contrary to reason that raw matter should originally have formed itself in accordance with mechanical laws, that life should have arisen from the nature of the lifeless, and that matter should have been able to assemble itself into the form of a self-preserving purposiveness by itself; at the same time, however, he leaves natural mechanism an indeterminable but at the same time also unmistakable role under this inscrutable *principle* of an original *organization*, on account of which he calls the faculty in the matter in an organized body (in distinction from the merely mechanical *formative power* [*Bildungskraft*] that is present in all matter) a *formative drive* [*Bildungstrieb*] (standing, as it were, under the guidance and direction of that former principle). (Kant, 1790, p. 424)

This passage in the *Critique of Judgment* makes the distinction between formative force and formative drive prominent.[39] Yet it remains problematic within Kant's own philosophical system on two counts. First, how Kant relates the orders of the two suggests that the formative *forces* (of general, physical nature) constrain the formative *drive*. This is a plausible scientific claim, but it goes against the metaphysical thrust of his whole argument, which is to suggest that organisms as natural purposes urge us toward the notion that there is a larger purpose in nature as a whole which constrains the physical order (a 'supersensible substrate'). (Kant, 1790, pp. 377–378, 398–399) Some translators of this key passage have been so motivated by this larger concern that they have mistranslated Kant's text. Second, it is not clear how Kant conceives of the notion of drive (*Trieb*) in his philosophy: in what measure is it really different conceptually from force (*Kraft*)? Are they not all equally 'inscrutable,' or is there a supplementary inscrutability about *life*-forces? More importantly for my argument, Kant is simply appropriating Blumenbach for philosophical purposes alien to Blumenbach's own scientific practice. Blumenbach n*ever* considered his formative drive anything but an actual force in nature. To be sure, he found Kant's suggestion that he brought teleological and mechanical explanations together in his scientific practice quite pleasing, but it is not clear that he understood Kant's painstaking argument for their radically different roles in scientific explanation. In short, notwithstanding Lenoir (and Jardine), Blumenbach's affiliation with Kant is best understood as a *misunderstanding*. But it was a *creative* misunderstanding, because it enabled Blumenbach and his followers to continue with even greater energy the development of that new science of *Naturgeschichte*, that 'daring adventure of reason,' that Kant by 1790 found deeply problematic. To illustrate this, we must turn briefly to the wider 'Göttingen School.'

4. The 'Gottingen School' and Kant's 'Daring Adventure of Reason'

Christoph Girtanner's *Über das Kantische Prinzip für die Naturgeschichte* (1796) offers insight into how Kant was being understood by Blumenbach and the Göttingen school at the decisive moment. He began learning about Kant around the same time Blumenbach did, and, like Blumenbach himself, he was stimulated by Kant's controversy with Herder and Forster, which drew the attention of most of the leading life scientists in Germany. (See documentation in Fambach, 1959, III, pp. 357–397) In 1787 Girtanner corresponded regarding Kant's philosophy of science with Karl Reinhold, who in 1786 had converted from a defender of Herder into the decisive popularizer of Kant. (Sloan, 1979, p. 138; Lenoir, 1980, p. 99) In 1788, Girtanner formed a personal acquaintance in Edinburgh with one of Kant's disciples, Johann Jachmann, who would serve as an intermediary between Blumenbach and Kant in the 1790s. (Sloan, 1979, p. 138; see Jachmann to Kant, October 14, 1790, in Kant, B, AA 11, pp. 201–213) Once back in Göttingen from 1790 onward, Girtanner participated in the Blumenbach circle during the years of the composition and reception of his work, which he dedicated to Blumenbach as a contribution to the assimilation of Kantianism by Blumenbach and his school.

Girtanner presented Kant's thought as the paradigm for a new research program in the life sciences under the rubric of *Naturgeschichte*. Girtanner's extension of Kant's work followed just the

[37] 'The formative drive is not the *cause* of this leap [from inorganic to organic], but rather its *expression*.' (McLaughlin, 1982, p. 364) I share the view of Robert Richards (2002, p. 221ff) that the relation in Blumenbach in fact tended to flow in the other direction, even if Blumenbach's metaphysical preferences inclined him to want to see it as McLaughlin reconstructs.

[38] See Duchesneau (1985). This, of course, becomes the Achilles heel of 'vitalism' in historical retrospect.

[39] Christoph Girtanner would pick this up explicitly. See Girtanner (1796). On Girtanner, see Wegelin (1957), Querner (1990).

vein that Kant himself had indicated his theory of race would require were it to become a serious scientific research program.[40] This new research program would ask, in Girtanner's words, 'what the primal form of each ancestral species of animals and plants originally consisted of, and how the species gradually devolved from their ancestral species.' (Girtanner, 1796, p. 2) This was a new and specific science that would explore and explain how environmental changes on the earth—indeed 'violent revolutions in nature'—occasioned dramatic changes in life forms. Yet however dramatic, the point was that these were not *chaotic* changes; rather, the variation in observed traits in current species emerged always under the guidance of a 'natural law' requiring that 'in all of organic creation, species remain unaltered.' (Girtanner, 1796, p. 6) Kant's great achievement, in Girtanner's eyes, was his connection of this law to a more determinate 'natural law' (proposed by Buffon) to explain this process, namely that 'all animals or plants that produce fertile offspring belong to the same physical [i.e., real] species,' notwithstanding considerable observed variation in traits. (Girtanner, 1796, p. 4) That is, these organisms *must* have 'derived from one and the same stem [*Stamm*].' (Girtanner, 1796, p. 4) While there could be hereditary variations [*Abartungen*] *within the confines of the governing stem*, there could not be 'degenerations' [*Ausartungen*], that is, permanently heritable departures from the fundamental traits of the ancestral stem. *Races* constituted decisive evidence for this theory, because their crosses always showed perfect proportion in the offspring: *Halbschlachtigkeit* (half-breeding). To account for these internal variations within species, Kant had offered the view that 'the ancestral stem of each species of organic life contained a quantity of different germs [*Keime*] and natural potentialities [*natürliche Anlagen*].' (Girtanner, 1796, p. 11) Girtanner followed Kant literally in identifying *Keime* with the source of changes in the parts (organs) of an organic life form, while *natürliche Anlagen* occasioned changes only in the size or proportion of such parts. Kant used winter feathers in birds to exemplify the first, and thickness in the husk of grain to exemplify the second. Girtanner replicated these examples.[41]

To help explicate the *process* of variation, Girtanner turned to his teacher Blumenbach. It was 'through different directions of the *Bildungstrieb*, [that] now these and now those [germs or natural potentialities] developed, while the others remained inert.' (Girtanner, 1796, p. 11) Only climate acting on organisms over extended time could educe such variation, such shifts in the 'direction of the *Bildungstrieb*,' and thus permanently alter 'the primal forces of organic development and movement.' (Girtanner, 1796, p. 12) Moreover, once such shifts in direction took place, once certain germs or natural potentialities triggered into actualization, the rest atrophied and the process proved irreversible. (Girtanner, 1796, p. 27) This claim represented one of Kant's decisive interventions in the theory of race, separating him sharply from Buffon, for example. (Bernasconi, 2001a)

Girtanner was acutely aware of the way in which Kant's 'natural history' interpenetrated with his theory of organic form. Not only did Kant require a specific theory of generic transmission, but he needed a theory of organic life in which to cast it. The only form of generation that had been empirically observed, Girtanner noted, was *generatio homonyma*, the persistence of species, though *generatio heteronyma* [*Ausartung*] was not impossible (against reason), but only unheard of (against experience). The essential point was that these both contrasted with *generatio aequivoca* (spontaneous generation). 'That by mechanism organized beings should emerge from unorganized matter [...] contradicts reason as well as experience.' (Girtanner, 1796, p. 15)[42] That is, 'it contradicts all known laws of experience that matter which is not organized should have by itself, without the intervention of other, organized matter, organized itself.' (Girtanner, 1796, pp. 14–15) *Anti-hylozoism*, then, was the essential posit of Kant's theory of organic form. Girtanner stressed this about the idea of organism. Not only was it 'not a machine' in consequence of the mutuality of cause and effect, of parts and whole, but neither was it the 'analogue of art,' for 'organized Nature organizes itself.' (Girtanner, 1796, pp. 17–18) If Girtanner replicated Kant's presentation of the perplexity, he did nothing to advance its resolution. Certainly he did not find the regulative/constitutive distinction of any use in the science he proposed to elaborate.

Girtanner was clear that Blumenbach's *Bildungstrieb* was a *Lebenskraft*, namely 'that force by virtue of which the chemical and physical laws are subordinated under the laws of organization.' (Girtanner, 1796, p. 17) Because life forms showed characteristics—reproduction, growth through nourishment and assimilation, regeneration of lost organs and self-healing generally—which could not be assimilated to the mechanistic model of natural science, they represented anomalies requiring recourse to teleological judgment, the analogy of 'purposiveness.'

Girtanner, whom Lenoir is happy to identify as authentically Kantian in some places, clearly does not serve in that capacity for Lenoir here: 'Girtanner defended a view concerning Kant's *Stammgattung* which seems to run directly counter to the regulative function attributed to it in Kant's own works [...] Girtanner argued that the task of natural history was to delineate the original form (*Urbild*) of each *Stammgattung* and show how the present species were degenerated from these originals.' (Lenoir, 1978, p. 74) That was exactly what Girtanner endeavored, but—Lenoir notwithstanding—because of, not despite Kant's own statements regarding the *Stammgattungen* as actual ancestors. Though Lenoir seeks to exonerate Girtanner of 'sinn[ing] against a [...] sacred Kantian principle' and rescue him for authentic Kantian 'regulative' thinking and the 'ideal type' notion, he has to admit that Kielmeyer and Link—indeed the entire new generation of the 1790s—did go constitutive: 'For them the *Stamm* was not a regulative Ideal Type; it had a historical existence.' More, they believed 'a naturalistic explanation of organic form can be given.' (Lenoir, 1978, p. 92) Lenoir concludes for this generation of 1790: 'The *Urtyp*, transcendental ideal, or *Stamm* of the previous generation is no longer merely a regulative, necessary methodological tool of reason; it has become an actual historical entity shaped by the physical forces of nature.' (Lenoir, 1978, p. 98). I have established that this was always true for Blumenbach, even after his assimilation of Kant. Now Girtanner, Kielmeyer, Link—core members of the 'Göttingen School'—appear lost as well for any authentically Kantian 'transcendental philosophy of nature.' Lenoir's historical train of connections gets unhitched right at his locomotive!

Blumenbach and his school took the *Bildungstrieb* for actual, not regulative. Their project was to specify its *effects* through the mechanisms (*Bildungskräfte*) it set in motion. Kant's regulative/constitutive distinction proved useless for them in that pursuit. There is no doubt that the life scientists of Blumenbach's school did reach out to Kantian philosophy for legitimation of their methodology, as Lenoir contended. Kant's philosophical endorsement gave them some

[40] In a letter responding to the publisher Breitkopf's invitation to submit a more extended work on race in 1778, Kant, declining the invitation, explained: 'my frame of reference would need to be widely expanded and I would need to take fully into consideration the place of race among animal and plant species, which would occupy me too much and carry me into extensive new reading which in a measure lies outside my field, because natural history is not my study but only my game ...' (Kant to Breitkopf, April 1, 1778, in Kant, B, AA 10, pp. 227–230) The project of extending consideration of race to animals and plants took up the bulk of Girtanner's study.

[41] It is nor surprising, then, that Kant should have endorsed Girtanner's exposition of his theory of race. See Kant, 1798, p. 320.

[42] This is unquestionably a recapitulation of the argument in Kant, 1790, §§ 80–81.

epistemological and metaphysical comfort, especially given the thinness of their analogy to the Newtonian mysteriousness of gravity. (See Barnaby, 1988; Gregory, 1989; Huneman, 2006a, 2006b; Ingensiep, 1996; Larson, 1979; Lieber, 1950; Sloan, 2006; Williams, 1973) Yet the ultimate irony is that this affiliation went awry. There is perhaps no more widely accepted idea about the life sciences in the German 1790s, even—or especially—when they invoked Kantian critical terms, than that they slid one and all from a strictly regulative into an unmistakably constitutive use of natural teleology. (See, esp. Larson, 1979, 1994) This was a natural, indeed inevitable consequence of their commitment to the *empirical practice* of a life science, which Kant's philosophy of science in fact proscribed.[43] William Coleman demonstrates, for example, the way in which the crucial figure of Kielmeyer has to be read as having transgressed Kant's divide of constitutive and regulative. (Coleman, 1973, pp. 342–347) Leeann Hansen demonstrates, similarly, how J. C. Reil's *Von der Lebenskraft* (1795) 'included reason itself as an organic force; the highest force, it is true, but rooted in the chemical properties of matter like all the others.' (Hansen, 1993, p. 63) Indeed, as both Robert Richards and Frederick Beiser recognize, they are closer to Kant's disparaged former student, Johann Gottfried Herder, than they are to Kant. (Beiser, 2002; Richards, 2002; Zammito, 1998)

Daniel Kolb sums up the argument and then charges that the source of the 'border crossing' is in Kant's own ambiguity:

> Is the use of teleological explanations nothing more than a declaration of ignorance? [...] [Kant's] argument against reduction leaves open the question of the exact specification of organic teleology. His idea of teleology consequently proves to be frustratingly difficult to pin down. (Kolb, 1992, p. 13)

It is this irony that Clark Zumbach captures in his provocative title, *The Transcendent Science*. (Zumbach, 1984) Goethe even found Kant himself equivocating between constitutive and regulative uses of teleology in the *Critique of Judgment*. (Cited in Jardine, 1988, pp. 330–331) As Michael Friedman acutely notes, in his philosophy of science Kant was faced with a very uncomfortable question: 'how was [the] brilliantly successful Newtonian paradigm to be extended beyond astronomy and celestial mechanics?' (Friedman, 1992b, p. 240) Friedman elaborates: 'Kant's developing awareness (in 1785) of the new chemical developments and of the general importance of chemistry' made this problem of a unified 'order of nature' for natural science acute for Kant. (Friedman, 1992b, p. 285) Friedman establishes that Kant from this point onward saw himself caught in the toils of a 'gap in the critical system' which became the obsessive theme of the *Opus postumum*. (Friedman, 1992b, pp. 214–215)[44] Friedman's conclusion is grave:

> After the execution of the *Metaphysical Foundations* and the articulation of reflective judgment as an autonomous faculty, it becomes clear—from the point of view of the critical philosophy itself—that the absolute dichotomy between regulative and constitutive principles cannot be maintained. (Friedman, 1992b, p. 305)

Yet it was precisely in upholding that distinction that Kant sought to prescribe methodology to the emerging life sciences in Germany. (Zammito, 1998)

5. Conclusion

The failure of the regulative-constitutive barrier casts severe doubt on the adequacy of Kant's program, Lenoir notwithstanding. Lenoir recognizes the collapse of the constitutive-regulative distinction after 1790, although his commitment to a Kantian interpretation of the life sciences in this epoch simply presumes that 'teleomechanism' is unaffected by that collapse. (Lenoir, 1978)[45] My argument suggests that Lenoir's effort to construe a Kantian 'transcendental *Naturphilosophie*' as a coherent teleomechanist 'research programme' for the life sciences in the first half of the nineteenth century simply blurs too many categories on the one hand and introduces too many arbitrary distinctions on the other.[46]

The issue is what to make of vitalism in emergent life science at the end of the eighteenth century (which cannot set out from Kant's position that precisely vitalism *excluded* life from any valid *science*). Kant has a role in that historical constellation, but not as a coherent master model; rather, as a source of conflicting impulses that needed to be sorted out. I submit that Kant's language of *Keime* and *natürliche Anlagen* and his acceptance of the idea of a *Lebenskraft* as exemplified by Blumenbach's *Bildungstrieb* entangled him in a conception of science entailing the *objective reality* of forces which could not be reduced to those he admitted in the Newtonian order of physics. (Zammito, 2009) That was certainly where he ended up in the *Opus Postumum*. (Tuschling, 1991) If Kant himself could not hold this line, it can hardly be surprising when the leading biologists of his day, even in invoking his theory, found it impossible in practice to observe it.

Perhaps this helps explain why Kant's view came so swiftly to be overshadowed by Schelling. Frederick Gregory, no enthusiast for that development, identified three factors:that in Kant nature seemed somehow less real than mind, that Kant's scientific description of nature had to be restricted to mechanistic interaction alone, and the confusion that reigned about the status of scientific theory and the relation of science to religion (Gregory, 1989, p. 60)

Above all, what Kant refused to warrant was the overweening intuition of the epoch, that, in Gregory's formulation, 'Nature was not a timeless and immutable machine, but a temporal and developing organism.' (Gregory, 1989, p. 57) Goethe gave expression to this when he tried to explain how he reacted to Linnaeus: 'what he wanted to hold apart by force I had, according to the innermost need of my nature, to strive to bring together.' (Cited in Oppenheimer, 1967, p. 136) Robert Richards put it succinctly: 'The impact of Kant's *Kritik der Urteilskraft* on the disciplines of biology has, I believe, been radically misunderstood by many contemporary historians. [...] Those biologists who found something congenial in Kant's third *Critique* either misunderstood his project (Blumenbach and Goethe) or reconstructed certain ideas to have very different consequences from those Kant originally intended (Kielmeyer and Schelling).' (Richards, 2002, p. 229)

Lenoir was most concerned to establish that 'there were fundamental differences between Kant's teleology and that of the Naturphilosophen.' (Lenoir, 1989, p. 6) His aim was to rescue teleology from vitalism, but simultaneously to rescue biology from reductive mechanism. Kant's program for life science seemed to Lenoir to have been historically a viable path for one phase in the emergence

[43] Therefore, Frederick Beiser has it right: 'Kant's regulative doctrine was *not* the foundation of empirical science in the late eighteenth and early nineteenth century; rather it was completely at odds with it. It is striking that virtually all the notable German physiologists and biologists of the late eighteenth and early nineteenth centuries conceived of their vital powers as causal agents rather than regulative principles ...' (Beiser, 2002, p. 508).

[44] On this idea of the 'gap,' see Förster, 1987; and Tuschling (1971), Tuschling (1989, 1991).

[45] That one could still take oneself for a Kantian teleomechanist in the nineteenth century without subscribing to the regulative/constitutive distinction has been suggested in a private communication to me by Lenny Moss.

[46] Lenoir deserves the harsh judgment of Kenneth Caneva that he is guilty of 'errors, misrepresentations, inconsistences, unsupported claims, and plain unclear writing,' above all that 'he seems to loose [sic] sight of the fact that his categories are *his* categories, and not in any explicit sense also those of the scientists he studied.' (Caneva, 1990, p. 300)

of modern biology as a special science and a resource for its continued conceptualization in the present. There is much to appreciate in these ambitions. But if the notion of 'vital materialism' as it was developed by the Göttingen School is not quite the Kantian 'transcendental philosophy of nature' that Lenoir wants it to have been, then we in fact find the Göttingen School far closer to the *Naturphilosophen* than Lenoir would like.[47] Lenoir's 'third way' collapses back towards what has garnered historical opprobrium as 'vitalism,' and the only alternative seems the reductive mechanism Lenoir welcomes neither as a historical development nor as a current theory of life science.[48] My suggestion is that the historical question of 'vital materialism' needs to be reconsidered.[49] Instead of viewing the closeness of the Göttingen School to *Naturphilosophie* as a contamination, we might view it as historical evidence that something essential to the character of biology as a special science was at stake, and thus this episode in the history of biology might reopen issues in the contemporary philosophy of biology.[50] In such a scenario, however, I believe Kant's particular views on teleology constitute a hindrance, not an aid.

Acknowledgements

I would like to acknowledge the generosity of Mr. and Mrs. Bruce Dunlevie for their endowment of my John Antony Weir Professorship in History with a research account that made it possible to travel to the Brisbane conference where this paper was first delivered.

References

Adickes, E. (1924). *Kant als Naturforscher*. Berlin: Walter de Gruyter.
Allison, H. (1991). Kant's antinomy of teleological judgment. *Southern Journal of Philosophy*, 30 Supplement: System and teleology in Kant's *Critique of Judgment*, 25–42.
Allison, H. (1994). Causality and causal laws in Kant: A critique of Michael Friedman. In P. Parrini (Ed.), *Kant and contemporary epistemology* (pp. 291–308). Dordrecht: Kluwer.
Aulie, R. (1961). Caspar Friedrich Wolff and his 'Theoria Generationis', 1759. *Journal of the History of Medicine, 16*, 124–144.
Bach, T., & Breidbach, O. (Eds.). (2005). *Naturphilosophie nach schelling*. Stuttgart/Bad Cannstatt: Frommann-Hollzweg.
Barnaby, D. (1988). The early reception of Kant's *Metaphysical Foundations of Natural Science*. In R. Woolhouse (Ed.), *Metaphysics and philosophy of science in the Seventeenth and Eighteenth Centuries* (pp. 281–306). Dordrecht: Kluwer.
Baumanns, P. (1965). *Das problem der organischen Zweckmäßigkeit*. Bonn: Bouvier.
Beihart, C. (2009). Kant's characterization of natural ends, in *Kant Yearbook* I:2009: *Teleology*, 1–30.
Beiser, F. (2002). *German idealism: The struggle against subjectivism, 1781–1801*. Cambridge, MA/London: Harvard University Press.
Bernasconi, R. (2001a). Who invented the concept of race? Kant's role in the Enlightenment's construction of race. In R. Bernasconi (Ed.), *Race* (pp. 11–36). Oxford: Blackwell.
Bernasconi, R. (2001b). Kant and Blumenbach on the *Bildungstrieb*. Paper presented at the conference of the American Society of eighteenth-century studies, New Orleans, April, 2001.
Bernasconi, R. (Ed.). (2001c). *Introduction. Concepts of race in the eighteenth century. Volume IV: Blumenbach, De generis humani varietate native*. Bristol: Thoemmes.
Blumenbach, J. F. (1776, 1781, 1795). *De generis humani varietate nativa*. Reprint: in *Concepts of race in the eighteenth century*. Vol. IV. Bristol: Thoemmes, 2001. Translated in Bendysche, T., (trans.), (2005; reprint of 1865). *The anthropological treatises of Johann Friedrich Blumenbach*. London: Elibron.
Blumenbach, J. F. (1779). *Handbuch der Naturgeschichte*. Göttingen: Dieterich.
Blumenbach, J. F. (1781). *Über den Bildungstrieb und das Zeugungsgeschäfte*. Reprint: Stuttgart: G. Fischer (1971).
Blumenbach, J. F. (1782). *Handbuch der Naturgeschichte* (2nd ed.). Göttingen: Dieterich.
Blumenbach, J. F. (1785). *De nisu formativo et generationis negotio nuperae observationes*. N.p.
Blumenbach, J. F. (1787). *De nisu formativo et generationis negotio nuperae observationes*. Göttingen: Dieterich.
Blumenbach, J. F. (1788). *Handbuch der Naturgeschichte* (3rd ed.). Göttingen: Dieterich.
Blumenbach, J. F. (1789a). *Über den Bildungstrieb* (2nd ed.). Göttingen: Dieterich.
Blumenbach, J. F. (1789b). 'Versuch einer Beantwortung der von der kaiserlichen Akademie der Wissenschaften zu St. Petersburg, zum drittenmal aufgegebenen Preisfrage, *uti nutritio aequabilis, etc.*,' in *Zwo Abhandlungen über die Nutritionskraft [. . .] nebst einer fernen Erläuterung eben derselben Materie von C. F. Wolff*. St. Petersburg: Royal Academy of Sciences.
Blumenbach, J. F. (1791). *Handbuch der Naturgeschichte* (4th ed.). Göttingen: Dieterich.
Blumenbach, J. F. (1797). *Handbuch der Naturgeschichte* (5th ed.). Göttingen: Dieterich.
Bommersheim, P. (1919). Der Begriff der organischen Selbstregulation in Kants Kritik der Urteilskraft. *Kant-Studien, 23*, 209–220.
Bommersheim, P. (1927). Der vierfache Sinn der inneren Zweckmäßigkeit in Kants Philosophie des Organischen. *Kant-Studien, 32*, 290–309.
Breidbach, O. (1995). Die Geburt des Lebendigen—Embryogenese der Formen oder Embryologie der Natur?—Anmerkungen zum Bezug von Embryologie und Organismustheorien vor 1800. *Biologisches Zentralblatt, 114*, 191–199.
Breitenbach, A. (2009). Teleology and biology: A Kantian perspective. In *Kant Yearbook* I:2009: *Teleology*, 31–56.
Buchdahl, G. (1965). Causality, causal laws and scientific theory in the philosophy of Kant. *British Journal for Philosophy of Science, 16*, 187–208.
Buchdahl, G. (1967). The relation between 'understanding' and 'reason' in the architectonic of Kant's philosophy. *Proceedings of the Aristotelian Society, 67*, 209–226.
Buchdahl, G. (1969a). The Kantian 'dynamic of reason', with special reference to the place of causality in Kant's system. In L. W. Beck (Ed.), *Kant studies today* (pp. 341–374). La Salle, IL: Open Court.
Buchdahl, G. (1969b). *Metaphysics and the philosophy of science: The classical origins: Descartes to Kant*. Cambridge: MIT Press.
Buchdahl, G. (1971). The conception of lawlikeness in Kant's philosophy of science. *Synthese, 23*, 24–46.
Buchdahl, G. (1981). Zum Verhältnis von allgemeiner Metaphysic der Natur und besonderer metaphysischer Naturwissenschaft bei Kant. In B. Tuschling (Ed.), *Probleme der 'Kritik der reinen Vernunft'* (pp. 97–142). Berlin/NY: Walter de Gruyter.
Buchdahl, G. (1986). Kant's 'special metaphysics' and the *Metaphysical Foundations of Natural Science*. In R. E. Butts (Ed.), *Kant's philosophy of physical science* (pp. 121–161). Dordrecht: Reidel.
Buchdahl, G. (1991). Comments on Michael Friedman: 'Regulative and Constitutive.' *Southern Journal of Philosophy*, 30 Supplement: System and Teleology in Kant's *Critique of Judgment*, 103-108.
Burtt, E. A. (1954). *Metaphysical foundations of modern physical science*. Garden City: Doubleday.
Butts, R. E. (1990). Teleology and scientific method in Kant's *Critique of Judgment*. *Nous, 24*, 1–16.
Caneva, K. L. (1990). Teleology with regrets. *Annals of Science, 47*, 291–300.
Coleman, W. (1973). Limits of recapitulation theory: Carl Friedrich Kielmeyer's critique of the presumed parallelism of earth history, ontogeny, and the present order of organisms. *Isis, 64*, 341–350.
Daston, L., & Galison, P. (2007). *Objectivity*. NY: Zone.
Dörflinger, B. (2000). *Das Leben theoretischer Vernunft*. Berlin/NY: de Gruyter.
Duchesneau, F. (1979). Haller et les théories de Buffon et C. F. Wolff sur l'épigenèse. *History and Philosophy of the Life Sciences, 1*, 65–100.
Duchesneau, F. (1985). Vitalism in late eighteenth-century physiology: the cases of Barthez, Blumenbach and John Hunter. In W. F. Bynum & R. Porter (Eds.), *William Hunter and the eighteenth-century medical world* (pp. 259–295). Cambridge: Cambridge University Press.
Fambach, O. (Ed.). (1959). *Ein Jahrhundert deutscher Literaturkritik (1750–1850). Vol. III Der Aufstieg zur Klassik*. Berlin: Akademie (pp. 357–397).
Flasch, W. (1997). Kants empiriologie: Naturteleologie als wissenschaftstheorie. In P. Schmid & S. Zurbuchen (Eds.), *Grenzen der kritischen Vernunft* (pp. 273–289). Berlin: Schwabe.
Förster, E. (1987). Is there a 'gap' in Kant's critical system? *Journal of the History of Philosophy, 25*, 533–555.
Forster, G. (1786). Noch etwas über die Menschenraßen. *Teutsche Merkur, 57–86*, 150–166.
Fricke, C. (1990). Explaining the inexplicable: The hypotheses of the faculty of reflective judgment in Kant's third critique. *Nous, 24*, 45–62.
Friedman, M. (1986). The metaphysical foundations of Newtonian science. In R. E. Butts (Ed.), *Kant's philosophy of physical science* (pp. 25–60). Dordrecht: Reidel.
Friedman, M. (1990). Kant and Newton: Why gravity is essential to matter. In P. Bricker & R. I. G. Hughes (Eds.), *Philosophical perspectives on Newtonian science* (pp. 185–202). Cambridge, MA: MIT Press.
Friedman, M. (1991). Regulative and constitutive. *Southern Journal of Philosophy*, 30 Supplement: System and Teleology in Kant's *Critique of Judgment*, 73–102.

[47] For a view that seeks to come to better terms with the situation, see Huneman, 2006a, 2006b.
[48] For the most eminent statements, see Mayr (1982), Ghiselin (1997).
[49] For the most extensive and persuasive treatment of this question, see Reill (1989, 1992, 1998, and esp 2005).
[50] On the Göttingen School and emergent Naturphilosophie, see: Hansen (1993), Gloy and Burger, eds. (1993), Bach and Breidbach, eds. (2005), Breidbach (1995), Coleman (1973), Jahn (1995), Jardine (1988), Rheinberger (1981, 1986), Stein (2004).

Friedman, M. (1992a). Causal laws and the foundations of natural science. In P. Guyer (Ed.), *The Cambridge companion to Kant* (pp. 161–199). Cambridge: Cambridge University Press.

Friedman, M. (1992b). *Kant and the exact sciences*. Cambridge, MA: Harvard University Press.

Friedman, M., & Nordmann, A. (Eds.). (2006). *The Kantian legacy in nineteenth-century science*. Cambridge: MIT.

Gaissinovich, A.E. (1968). Le rôle du Newtonianisme dans la renaissance des idées épigénetiques en embryologie du XVIIIe siècle. In *Actes du XIe Congrès International d'Histoire des Sciences* (Vol. 5, pp. 105–110).

Gaissinovich, A. E. (1990). C. F. Wolff on variability and heredity. *History and Philosophy of the Life Sciences, 12*, 179–201.

Genova, A. C. (1974). Kant's epigenesis of pure reason. *Kant-Studien, 65*, 259–273.

Ghiselin, M. (1997). *Metaphysics and the origin of species*. Albany: SUNY.

Ginsborg, H. (1987). Kant on aesthetic and biological purposiveness. In A. Reath, B. Herman, & C. Korsgaard (Eds.), *Reclaiming the history of ethics* (pp. 329–360). Cambridge: Cambridge University Press.

Ginsborg, H. (2001). Kant on understanding organisms as natural purposes. In E. Watkins (Ed.), *Kant and the sciences* (pp. 231–258). Oxford & NY: Oxford University Press.

Ginsborg, H. (2004). Two kinds of mechanical inexplicability in Kant and aristotle. *Journal of the History of Philosophy, 42*, 33–65.

Girtanner, C. (1796). *Über das Kantische Prinzip für die Naturgeschichte*. Göttingen: Vandenhoeck & Ruprecht.

Gloy, K., & Burger, P. (Eds.). (1993). *Die Naturphilosophie im Deutschen Idealismus*. Stuttgart/Bad Cannstatt: Frommann-Hollzweg.

Gregory, F. (1989). Kant's influence on natural scientists in the German Romantic period. In R. P. W. Visser, H. J. M. Bos, L. C. Palm, & H. A. M. Snelders (Eds.), *New trends in the history of science* (pp. 53–66). Amsterdam/Atlanta: Rodopi.

Guyer, P. (2001). Organism and the unity of science. In Eric. Watkins (Ed.), *Kant and the sciences* (pp. 259–281). Oxford & NY: Oxford University Press.

Guyer, P. (2003). Kant and the systematicity of nature: Two puzzles. *History of Philosophy Quarterly, 20*, 277–295.

Guyer, P. (2005). *Kant's system of nature and freedom*. Oxford: Oxford University Press.

Hansen, L. (1993). From Enlightenment to *Naturphilosophie*: Marcus Herz, Johann Christian Reil, and the Problem of Border Crossings. *Journal of the History of Biology, 26*(1), 39–64.

Heidemann, D. (Ed.). (2009). *Kant Yearbook: I. Teleology*. Berlin: de Gruyter.

Herrlinger, R. (1959). C. F. Wolff's 'Theoria generationis' (1759). *Zeitschrift für Anatomie und Entwicklungsgeschichte, 121*, 245–270.

Herrlinger, R. (1966). *'Vorwort', to C. F. Wolff, Theorie von der Generation in zwei Abhandlungen erklärt und bewiesen (1764). Theoria generationis (1759)*. Reprint: Hildesheim, Olms. pp. 5–28.

Huneman, P. (2002). *Métaphysique et biologie: Kant et la constitution du concept d'organisme*. Villeneuve: Presses Universitaires du Septentrion.

Huneman, P. (2006a). From the *Critique of Judgment* to the hermeneutics of nature. *Continental Philosophy Review, 39*, 1–34.

Huneman, P. (2006b). Naturalising purpose: From comparative anatomy to the 'adventure of reason'. *Studies in History and Philosophy of the Biological and Biomedical Sciences, 37*, 649–674.

Huneman, P. (Ed.). (2007). *Understanding purpose: Collected essays on Kant and philosophy of biology*. University of Rochester Press/North American Kant Society Studies in Philosophy.

Ingensiep, H.-W. (1996). 'Die Welt ist ein Thier: aber die Seele desselben ist nicht Gott': Kant, das Organische und die Weltseele. In Ingensiep & R. Hoppe-Sailer (Eds.), *NaturStücke: Zur Kulturgeschichte der Natur* (pp. 101–120). Ostfildern [Germany]: Edition Tertium.

Ingensiep, H. W. (2004). Organismus und Leben bei Kant. In H. W. Ingensiep, H. Baranzke, & A. Eusterschulte (Eds.), *Kant-reader: Was kann ich wissen? Was soll ich tun? Was darf ich hoffen?* (pp 107–136). Würzburg: Königshausen & Neumann.

Jahn, I. (1995). Georg Forsters Lehrkonzeption für eine 'Allgemeine Naturgeschichte' (1786–1793) und seine Auseinandersetzung mit Caspar Friedrich Wolffs 'Epigenesis'-Theorie. *Biologisches Zentralblatt, 114*, 200–206.

Jardine, N. (1988). The significance of Schelling's 'epoch of a wholly new natural history': An essay on the realization of questions. In R. S. Woodhouse (Ed.), *Metaphysics and philosophy of science in the seventeenth and eighteenth centuries* (pp. 327–350). Dordrecht: Kluwer.

Jardine, N. (2000). *The scenes of inquiry: On the reality of questions in the sciences* (2nd ed.). Oxford: Clarendon.

Kant, I. (1775–1777). Gesammelte Schriften Herausgegeben von der Preussischen Akademie der Wissenschaften (Vols., 1–29 (to date). Berlin: de Gruyter (1901–Present) [hereafter noted as "Akademie-Ausgabe"], *2*, 427–444.

Kant, I. (1785a). Recensionen von J. G. Herders *Ideen zur Philosophie der Geschichte der Menschheit*. Theil 1. 2. *Akademie Ausgabe, 8*, 43–66.

Kant, I. (1785b). Bestimmung des Begriffs einer Menschenrace. *Akademie-Ausgabe, 8*, 89–106.

Kant, I. (1786). Metaphysische Anfangsgründe der Naturwissenschaft. *Akademie-Ausgabe, 4*, 465–566.

Kant, I. (1788). Über den Gebrauch teleologischer principien in der philosophie. *Akademie-Ausgabe, 8*, 157–184.

Kant, I. (1790). Kritik der Urteilskraft. *Akademie-Ausgabe, 5*, 165–486.

Kant, I. (1798). Anthropologie in pragmatischer Hinsicht. *Akademie-Ausgabe, 7*, 117–334.

Kant, I. (B) Briefwechsel. *Akademie-Ausgabe*, 10–13.

Karolyi, L. V. (1971). 'Vorwort' to Blumenbach. *Über den Bildungstrieb und das Zeugungsgeschäfte*, v–xx. Stuttgart: Fischer.

Kitcher, P. (1983). Kant's philosophy of science. *Midwest Studies in Philosophy, 8*, 387–407.

Kitcher, P. (1986). Projecting the order of nature. In R. E. Butts (Ed.), *Kant's philosophy of physical science* (pp. 201–235). Dordrecht: Reidel.

Kitcher, P. (1994). The unity of science and the unity of nature. In P. Parrini (Ed.), *Kant and contemporary epistemology* (pp. 253–272). Dordrecht: Kluwer.

Kolb, D. (1992). Kant, teleology, and evolution. *Synthese, 91*, 9–28.

Lagier, R. (2004). *Les races humaines selon Kant*. Paris: PUF.

Larson, J. (1979). Vital forces: Regulative principles or causal agents? *Isis, 70*, 235–249.

Larson, J. (1994). *Interpreting nature: The science of living form from Linnaeus to Kant*. Baltimore: Johns Hopkins University Press.

Lenoir, T. (1978). Generational Factors in the origin of *Romantische Naturphilosophie*. *Journal of the History of Biology, 11*, 57–100.

Lenoir, T. (1980). Kant, Blumenbach, and vital materialism in German biology. *Isis, 71*, 77–108.

Lenoir, T. (1981a). The Göttingen School and the development of transcendental Naturphilosophie in the Romantic Era. *Studies in History of Biology, 5*, 111–205.

Lenoir, T. (1981b). Teleology without regrets. The transformation of physiology in Germany: 1790–1847. *Studies in History and Philosophy of Science, 12*, 293–354.

Lenoir, T. (1988). Kant, Von Baer, and causal-historical thinking in biology. *Poetics Today, 9*, 103–115.

Lenoir, T. (1989). *The strategy of life: Teleology and mechanism in nineteenth-century biology*. Chicago/London: University of Chicago Press.

Lieber, H. J. (1950). Kants philosophie des organischen und die biologie seiner zeit. *Philosophia Naturalis, 1*, 553–570.

Locke, J. (1689). *Essay concerning human understanding*. Oxford: Clarendon. Reprint, 1988.

Löw, R. (1980). *Philosophie des lebendigen: Der begriff des organischen bei Kant, sein grund und seine aktualität*. Frankfurt: Suhrkamp.

Lovejoy, A. (1959). Kant and evolution. In B. Glass (Ed.), *Forerunners of Darwin 1745–1859* (pp. 173–206). Baltimore: Johns Hopkins University Press.

Lovejoy, A. (1936). *The great chain of being*. Cambridge: Harvard University Press.

Lüsebrink, H.-J. (1994). Aufgeklärtes humanismus: Philosophisches engagement am Beispiel der Kontroverse über die 'Menschenrassen'. In R. Reichardt & G. Roche (Eds.), *Weltbürger–Europäer–Deutscher–Franke: Georg Forster zum 200. Todestag1* (pp. 88–195). Mainz: Universitätsbibliothek Ausstellungskatalog.

Lukina, T. (1975). Caspar Friedrich Wolff und die Petersburger Akademie der Wissenschaften. *Acta Historica Leopoldina, 9*, 411–425.

Mayr, E. (1982). *The growth of biological thought*. Cambridge: Harvard.

McLaughlin, P. (1982). Blumenbach und der Bildungstrieb: Zum Verhältnis von epigenetischer Embryologie und typologischem Artbegriff. *Medizinhistorisches Journal, 17*, 357–372.

McLaughlin, P. (1990). *Kant's critique of teleology in biological explanation: Antinomy and teleology*. Lewiston: Mellen.

Menzer, P. (1911). *Kants theorie von der Entwicklung*. Reimer: Berlin.

Mocek, R. (1995). Caspar Friedrich Wolffs Epigenesis-Konzept—ein problem im Wandel der Zeit. *Biologisches Zentralblatt, 114*, 179–190.

Morrison, M. (1989). Methodological rules in Kant's philosophy of science. *Kant-Studien, 80*, 155–172.

Müller-Sievers, H. (1993). *Epigenesis: Naturphilosophie im Sprachdenken Wilhelm von Humboldts*. Paderborn: Schöningh.

Müller-Sievers, H. (1997). *Self-generation: Biology, philosophy and literature around 1800*. Stanford: Stanford University Press.

Okruhlik, K. (1983). Kant on the foundations of science. In W. Shea (Ed.), *Nature mathematicized* (pp. 251–268). Dordrecht: Reidel.

Oppenheimer, J. (1967). *Essays in the history of embryology and biology*. Cambridge: MIT Press.

Quarfood, M. (2004). *Transcendental idealism and the organism: Essays on Kant*. Almqvist & Wiksell: Stockholm.

Quarfood, M. (2006). Kant on biological teleology: Towards a two-level interpretation. *Studies in History and Philosophy of the Biological and Biomedical Sciences, 37*, 735–747.

Querner, H. (1990). Christoph Girtanner und die Anwendung des Kantischen Prinzips in der Bestimmung des Menschen. In G. Mann & F. Dumont (Eds.), *Die Natur des Menschen: Probleme der physischen anthropologie und rassenkunde (1750–1850)* (pp. 123–136). Stuttgart: G. Fischer.

Rang, B. (1998). Zweckmäßigkeit, Zweckursächlichkeit und Ganzheitlichkeit in der organischen Natur: Zum Problem einer teleologischen Naturauffassung in Kants 'Kritik der Urteilskraft'. *Philosophisches Jahrbuch, 100*, 39–71.

Reill, P. H. (1989). Anti-mechanism, vitalism and their political implications in late Enlightened scientific thought. *Francia, 16*, 195–212.

Reill, P. H. (1992). Between mechanism and hermeticism: nature and science in the late Enlightenment. In R. Vierhaus (Ed.), *Frühe Neuzeit—Frühe Moderne?* (pp 393–421). Göttingen: Vandenhoeck & Ruprecht.

Reill, P. H. (1998). Analogy, comparison, and active living forces: Late Enlightenment responses to the skeptical critique of causal analysis. In J. van der Zande & R. Popkin (Eds.), *The skeptical tradition around 1800* (pp. 203–211). Dordrecht: Kluwer.

Reill, P. H. (2005). *Vitalizing nature in the Enlightenment*. (Berkeley, etc.: University of California Press.

Rheinberger, H.-J. (1981). Über Formen und Gründe der Historisierung biologischer Modelle von Ordnung und Organisation am Ausgang des 18. Jahrhunderts. In M.

Hahn & H.-J. Sandkühler (Eds.), *Gesellschaftliche Bewegung und Naturprozeß* (pp. 71–81). Cologne: Paul-Rugenstein.
Rheinberger, H.-J. (1986). Aspekte des Bedeutungswandels im Begriff organismischer Ähnlichkeit vom 18. zum 19. Jahrhundert. *History and Philosophy of the Life Sciences, 8*, 237–250.
Richards, R. (2000). Kant and Blumenbach on the *Bildungstrieb*: A historical misunderstanding. *Studies in the History and Philosophy of Biology and the Biomedical Sciences, 31*, 11–32.
Richards, R. (2002). *The romantic conception of life*. Chicago/London: University of Chicago Press.
Riedel, M. (1980). Historizismus und Kritizismus: Kants Streit mit G. Forster und J.G. Herder. In B. Fabian & W. Schmid-Biggemann (Eds.), *Deutschlands kulturelle Entfaltung* (pp. 31–48). Munich: Kraus.
Roe, S. (1981). *Matter, life, and generation: 18th-Century embryology and the Haller-Wolff debate*. Cambridge: Cambridge University Press.
Roe, S. (1979). Rationalism and embryology: Caspar Friedrich Wolff's theory of epigenesis. *Journal of the History of Biology, 12*, 1–43.
Roger, J. (1963). *Les sciences de la vie dans la pensée française du XVIIIe si□cle; la génération des animaux de Descartes à l'encyclopédie*. Paris: Colin.
Roger, J. (1968). Leibniz et les sciences de la vie. *Studia Leibnitiana. Supplementa, 2*, 209–219.
Roger, J. (1980). The living world. In G. Rousseau & R. Porter (Eds.), *The ferment of knowledge: Studies in the historiography of eighteenth-century science* (pp. 255–284). Cambridge: Cambridge University Press.
Roretz, K. (1922). *Zur analyse von Kants philosophie des organischen*. Akademie der Wissenschaften: Vienna.
Schmied-Kowarzik, W. (1994). Der Streit um die Einheit des Menschengeschlechts. Gedanken zu Forster, Herder und Kant. In C.-V. Klenke, J. Garber, & D. Heintze (Eds.), *Georg Forster in interdisziplinärer perspektive* (pp. 115–132). Berlin: Akademie.
Schuster, J. (1941). Der Streit um die Erkenntnis des organischen Werdens im Lichte der Briefe C. F. Wolffs an A. von Haller. *Sudhoffs Archiv, 34*, 196–218.
Sloan, P. (1979). Buffon, German biology, and the historical interpretation of biological species. *British Journal for the History of Science, 12*, 109–153.
Sloan, P. (2006). Kant on the history of nature: The ambiguous heritage of the critical philosophy for natural history. *Studies in History and Philosophy of the Biological and Biomedical Sciences, 37*, 627–648.
Steigerwald, J. (2006a). Kant's concept of natural purpose and the reflecting power of judgement. *Studies in History and Philosophy of Biological and Biomedical Sciences, 37*(4), 712–734.
Steigerwald, J. (Ed.) (2006b). Special issue: Kant and biology. *Studies in History and Philosophy of Biological and Biomedical Sciences, 37*(4).
Stein, K. (2004). *Naturphilosophie der Frühromantik*. Paderborn: Schöningh.
Strack, T. (2001). Philosophical anthropology on the eve of biological determinism: Immanuel Kant and Georg Forster on the moral qualities and biological characteristics of the human race. *Central European History, 29*(3), 285–308.
Tuschling, B. (1971). *Metaphysische und transzendentale Dynamik in Kants Opus postumum*. Berlin: de Gruyter.
Tuschling, B. (1989). Apperception and ether: On the idea of a transcendental deduction of matter in Kant's opus postumum. In E. Förster (Ed.), *Kant's transcendental deductions* (pp. 193–216). Stanford: Stanford University Press.
Tuschling, B. (1991). The system of transcendental idealism: questions raised and left open in the *Kritik der Urteilskraft*. *Southern Journal of Philosophy*, 30 Supplement: System and Teleology in Kant's *Critique of Judgment*, 109–128.
Ungerer, E. (1922). *Die teleologie Kants und ihre bedeutung für die logik der biologie*. Berlin: Borntraeger.
Uschmann, G. (1955). *Caspar Friedrich Wolff: Ein Pionier der modernen embryologie*. Leipzig/Jena: Urania.
van Hoorn, T. (2004). *Dem Leibe abgelesen. Georg Forster im Kontext der physischen Anthropologie des 18. Jahrhunderts*. Tübingen: Niemeyer.
Warnke, C. (1992). 'Naturmechanismus' und 'Naturzweck': Bemerkungen zu Kants Organismus-Begriff. *Deutsche Zeitschrift fur Philosophie, 40*, 42–52.
Wegelin, C. (1957). Dr. Med. Christoph Girtanner (1760–1800). *Gesnerus, 14*, 141–163.
Weingarten, M. (1982). Menschenarten und Menschenrassen: Die Kontroverske zwischen Georg Forster und Immanuel Kant. In G. Pickerodt (Ed.), *Georg Forster in seiner Epoche* (pp. 117–148). Berlin: Argument (Sonderband 87).
Williams, L. P. (1973). Kant, Naturphilosophie and scientific method. In R. Giere & R. Westfall (Eds.), *Foundations of scientific method: The nineteenth century* (pp. 3–22). Bloomington: Indiana University Press.
Wolff, C. F. (1764). *Theorie von der Generation in zwei Abhandlungen erklärt und bewiesen (1764). Theoria generationis (1759).* Mit einer Einführung von Robert Herrlinger. Reprint: Stuttgart: G Fischer, 1966.
Wolff, C. F. (1789). *Von der eigenthümlichen und wesentlichen Kraft der vegetabilischen sowohl als auch der animalischen Substanz, in Zwo Abhandlungen über die Nutritionskraft [. . .] nebst einer fernen Erläuterung eben derselben Materie von C. F. Wolff.* St. Petersburg: Royal Academy of Sciences.
Wubnig, J. (1968/69). The epigenesis of pure reason. *Kant-Studien, 60*, 147–152.
Zammito, J. (1998). 'Method' vs 'Manner'?—Kant's critique of Herder's *Ideen* in light of the epoch of science, 1790–1820. *Herder Yearbook, 1998*, 1–25.
Zammito, J. (2003). 'This Inscrutable *Principle* of an Original *Organization*': Epigenesis and 'Looseness of Fit' in Kant's Philosophy of Science. *Studies in History and Philosophy of Science, 34*, 73–109.
Zammito, J. (2006a). Kant's early views on epigenesis: The role of maupertuis. In J. E. Smith (Ed.), *The problem of animal generation in modern philosophy* (pp. 317–354). Cambridge: Cambridge University Press.
Zammito, J. (2006b). 'Policing polygeneticism in Germany, 1775: (Kames,) Kant and Blumenbach'. In S. Eigen & M. Larrimore (Eds.), *The german invention of race* (pp. 35–54). Albany: State University Press of New York.
Zammito, J. (2006c). Teleology then and now: The question of Kant's relevance for contemporary controversies over function in biology. *Studies in History and Philosophy of Biological and Biomedical Sciences, 37*, 748–770.
Zammito, J. (2007). Kant's persistent ambivalence toward epigenesis, 1764–1790. In P. Hunemann (Ed.), *Understanding purpose: Collected essays on Kant and philosophy of biology* (pp. 51–74). University of Rochester Press/North American Kant Society Studies in Philosophy.
Zammito, J. (2009). 'Kant's Notion of Intrinsic Purposiveness in the *Critique of Judgment*: A Review Essay (and an Inversion) of Zuckert's *Kant on Beauty and Biology*,'*Kant Yearbook* I (2009): *Teleology* (Berlin/NY: de Gruyter), 223–247.
Zammito, J. (forthcoming). Kant and objective purposiveness: The problem of organisms. In W. Dudley, & K. Engelhard (Eds.), *Kant: Key Concepts.* London: Acumen.
Zammito, J. (2010). Kant, natural history, and the 'daring adventure of reason', Invited Lecture, North American Kant Society, Southern Section Inaugural Meeting, Texas A&M University, March 5–7, 2010 (unpublished).
Zuckert, R. (2007). *Kant on beauty and biology: An interpretation of the critique of judgment*. Cambridge: Cambridge University Press.
Zumbach, C. (1984). *The transcendent science: Kant's conception of biological methodology*. The Hague: Nijhoff.

Studies in History and Philosophy of Biological and Biomedical Sciences 43 (2012) 133–139

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

Hegel's notion of natural purpose ☆

Francesca Michelini

Alexander von Humboldt Foundation, Humboldt-Universität zu Berlin, Philosophische Fakultät III, Institut für Kulturwissenschaft, Office: Mohrenstraße 40/41, Raum 519, 10117 Berlin, Germany

ARTICLE INFO

Article history:
Available online 12 July 2011

Keywords:
Hegel
Intrinsic purposiveness
Life
Contradiction
Deficiency
Steresis

ABSTRACT

This paper argues that the notion of natural purpose developed by Hegel can only be thoroughly grasped by considering its intimate connection with the idea of contradiction and, particularly, with what Hegel in his philosophy of nature called the 'activity of deficiency'. This expression is used by Hegel to denote the ontological situation of every living being, which is embodied most authentically in the concepts of need and drive. For Hegel, life itself is imbued with contradiction because it is inextricably bound up with what it lacks: its identity is at one with its negation. This paper defends the thesis that Hegel's philosophy—and not just his philosophy of nature—can be characterized as an 'ontology of life' (to use the same expression that Martin Heidegger applied to Aristotle's *De Anima*), or more precisely, as an ontology of living individuality.

> 'It is said that contradiction is unthinkable; but the fact is that in the pain of a living being it is even an actual existence' (Hegel, 1969, p. 770)

1. The project of a conjugation of Aristotle and Kant

In what follows I intend to draw attention to the model of natural purpose developed by Hegel. My thesis is that this notion can only be thoroughly grasped by considering its intimate connection with the notion of contradiction and, particularly, with what Hegel in his philosophy of nature called the 'activity of deficiency'. This is an expression that he used in particular contexts to denote the ontological situation of every living being. I maintain that Hegel's philosophy—and not just his philosophy of nature—can be characterized as an ontology of life, to use the same expression that Martin Heidegger applied to Aristotle's *De Anima*. Put better, it is an ontology of living individuality. Hegel drew on the two philosophers who, he believed, had discovered and developed an 'intrinsic' notion of purpose: Aristotle and Kant. In Hegel's words, Kant had the merit of having resuscitated 'the determination of life by *Aristotle*'. Because it already comprised internal purposiveness, the Aristotelian notion struck Hegel as 'infinitely' more advanced than the 'modern' concept of teleology, which only envisaged external purposiveness (Hegel, 1991, § 204r, p. 280).

From a certain point of view, Hegel's recognition granted to these two philosophers still maintains its validity today, given that Aristotle and Kant together constitute the primary model for those who wish to talk about the concept of intrinsic teleology in order to address the issues raised by contemporary biophilosophical research. Those who cite Aristotle stress that his notion of final cause by no means corresponds to the intention or the design of a mind. Nor did Aristotle envisage backward causation, this being one of the factors that most vex the opponents of finalism in nature. For example:

> The interpretation I have in mind recognized that the most important feature of Aristotelian teleology is that it presents an alternative to the anthropocentric, creationist, and providential schemes of teleology that were favored by Aristotle's predecessors, and were later popular in the commentarial tradition's

☆ Some reflections in this paper have already been developed in Michelini (2008).

E-mail address: francescamichelini@libero.it

1369-8486/$ - see front matter © 2011 Published by Elsevier Ltd.
doi:10.1016/j.shpsc.2011.05.012

appropriation of Aristotle, and in the early modern period's natural theology (Johnson, 2005, p. 3).[1]

Kant is recalled for manifold reasons—not least the 'productive' tension that he established between the mechanistic explanation of living organisms and the necessity of conceiving them as 'natural purposes'—in particular by those who stress the close conceptual connection between internal purposiveness and self-organization. By way of example, according to recent interpretations: 'It was Kant who elaborated for the first time the similarity of this *intrinsic teleology with a modern understanding of self-organization*. For Kant things that organize themselves are—in opposition to purposes of nature—called *natural purposes*' (Weber & Varela, 2002, p. 106); 'Kant's problematic may have been largely forgotten by contemporary biology, but it has strong resonances with issues that are only now beginning to attract biologists' attention—self-organization, the 'emergent' properties of organisms, their adaptability, their capacity to regulate their component parts and processes' (Walsh, 2006, p. 772).

I wish to suggest that Hegel's endeavour is particularly interesting in this context precisely because he stressed a close continuity between the two conceptions of Aristotle and Kant and tried to unify them. The purpose of this conjugation was not simply to blend the two conceptions together, but rather to remedy the limitations of what Hegel considered to be their 'unilateral' visions and develop the key theoretical aspects they did not elaborate in depth or to which they merely alluded. In a certain sense, their possible integration is subordinate to what Hegel calls true 'refutation'. The notion of immanent refutation is absolutely not to be confused with the simple rejection of particular doctrinal elements or with repudiation *in toto* of the philosophy being discussed. Nor does it consist in pointing out only the defects in the latter. In other words, refutation is not simply arguing against a particular doctrine in order to demonstrate its falsehood; in this way one would risk—from the Hegelian point of view—extrinsically opposing one unilateral doctrine with another unilaterality. This commonsense meaning of 'refutation' is firmly rejected by Hegel: 'Most commonly the refutation is taken in a purely negative sense to mean that the system refuted has ceased to count for anything, has been set aside and done for' (Hegel, 1991, § 86r, p. 2). On the contrary, true refutation, to be such, does not 'come from outside' but must penetrate the adversary's sinews and establish itself where his strength resides. It is necessary, that is to say, to start from assumptions immanent in the system itself, not from 'needs and suppositions' external to it. This, therefore, is not a matter of opposing the 'falsehood' of a doctrine, with a doctrine deemed 'true'; rather, it involves firstly recognizing the doctrine's perspective as 'essential and necessary', and secondly bringing to light the dialectic internal to the philosophy being refuted.

This therefore requires developing the essential elements of a philosophy—discarding those that are more contingent and tied to a particular historical period—so that they can be fruitfully preserved within one's own perspective. In the case in question, as we shall see, Hegel's intent is to recast the concept of intrinsic purpose that Aristotle first discovered and Kant subsequently developed. Hegel's operation can thus be summarized as follows: via Aristotle and Kant, beyond Aristotle and Kant.

However, unlike Aristotle and Kant, references to Hegel on this topic are rare in the contemporary philosophy of biology. Referring to Hegel may seem paradoxical, if not absurd, given that the 'scientific' critique of the second half of the nineteenth century seems to have definitively swept away the philosophy of nature, and with it the Hegelian notion of natural purposiveness. Hegelian philosophy did not significantly influence the subsequent debate in biology, whereas Kant's *Critique of the Power of Judgment* is considered a fundamental turning-point. According to widespread opinion, Hegel also epitomizes the fact that finalism is not considered in its specific (and therefore also natural) features but dissolved into a vast cosmic finalism, in which the individuality of the single living being is obfuscated to the advantage of an entirely general conception of life. But even the interpreters who have concentrated more closely on the purposiveness internal to the single living organism have nevertheless maintained that Hegel re-proposed a form of vitalism by considering the 'formative power' which Kant had held to be unknowable as intelligible.

What I shall instead seek to show is that the notion of intrinsic purpose is not understood by Hegel in either a 'cosmic' or a 'vitalist' sense; rather, he employs the notion to understand the fundamental structure of the living organism in a way which may be of interest to current bio-philosophy.

2. Hegel's critical appreciation and appropriation of Kant

Explanation of Hegel's notion of natural purposiveness can fruitfully start with a passage from the already-cited *Encyclopaedia Logic*:

> In dealing with the purpose, we must not think at once (or merely) of the form in which it occurs in consciousness as a determination that is present in representation. With his concept of *internal* purposiveness, Kant has resuscitated the Idea in general and especially the Idea of life. The determination of life by *Aristotle* already contains this internal purposiveness; hence, it stands infinitely far above the concept of modern teleology which had only *finite*, or *external*, purposiveness in view. Need and drive are the readiest examples of purpose. They are the *felt* contradiction, as it occurs *within* the living subject itself; and they lead into the activity of negating this negation (which is what mere subjectivity still is) (Hegel, 1991, § 204r, p. 280).

Let us concentrate on a crucial aspect: The 'readiest examples of purpose are need and drive'. What does this mean? Note that in this passage Hegel does not claim that the object of *need* and *drive* is the *purpose*, but that these are in themselves the *purpose*. In what sense? Understanding the point requires taking a small step backwards and considering the beginning of the passage: 'In dealing with the purpose, we must not think at once (or merely) of the form in which it occurs in consciousness as a determination that is present in representation'.

Implicit in this statement is Hegel's critical approach to Kant, which he developed at least from the *Phenomenology of Mind* (but also from his writings of the Jena period, such as *Faith and Knowledge*) onwards. Hegel's approach was centred on the work which he regarded as the most significant of all modernity, namely the *Critique of Judgment*. It consisted in acknowledgement of Kant's contribution—in having recovered the notion of internal purpose—but at the same time in criticism that he still adhered to the subjective perspective of consciousness.

In the *Phenomenology of Mind* Hegel equated Kantian consciousness with a state of 'being observation' separate from what it observes, organic nature. This separation prevented recognition of the objective existence of the end. According to Hegel, what the Kantian observative consciousness determines as constitutive of the concept of end is not 'the proper essence of the organic,' but

[1] Johnson quotes also Eduard Zeller's judgement: 'The most important feature of the Aristotelian teleology is the fact that it is neither anthropocentric, nor is it due to the actions of a creator existing outside the world or even of a mere arranger of the world, but it is always thought as immanent in nature. What Plato effected in the *Timaeus* by the introduction of the world-soul and the Demiurgus is here explained by the assumption of a teleological activity inherent in nature itself' (Johnson, 2005, p. 3r).

it occurs in the consciousness as the subjective mode of reflection on phenomena. And even when Kant seeks to save its objectivity by locating the end in an intellect ultimately responsible for the teleological ordering of the world, in this case, too, although objectivity is attributed to the end, nevertheless it is still not considered to be the essence of the organic. Hegel consequently concludes that:

> In this state of being, observation does not recognize the concept of purpose, or does not know that the notion of purpose is not in an intelligence anywhere else, but just exists here and in the form of a thing (Hegel, 1931, p. 299).

Hegel's criticism in the *Phenomenology* is based on ambiguities within the Kantian conception which become more evident in subsequent works like the *Science of Logic*. In his treatment of teleology, Hegel acknowledges Kant's achievements: 'One of Kant's great services to philosophy consists in the distinction he has made between relative or *external*, and *internal* purposiveness; in the latter he has opened up the Notion of life, the Idea' (Hegel, 1969, p. 737). This last line warrants attention: as in the above-quoted passage from the *Encyclopaedia*, here again Hegel describes Kant as the thinker who conceived internal purposiveness *as real*, connecting it with the more general topic of life.[2] Under the heading 'teleology' in the *Doctrine of the Concept*, Hegel only deals with 'external purposiveness'. He treats internal purposiveness in the section on life, which confirms that it is life and not teleology, Hegel maintains, which comprises the authentic concept of purpose. It is in the idea of life, moreover, that Hegel reconciles the contrast between mechanicism and teleology, which instead arises when the two determinations are considered in their total independence.

In (indirect) polemic with Kant, however, Hegel maintains that the relation of purpose is not a relation expressible through a judgement—which is tied to the subjective point of view of the consciousness—rather it is the syllogism that makes the relation fully such. External purposiveness is illustrated with a syllogism where the extreme terms cannot be exchanged with each other, and in which the middle term is not interchangeable with the extreme ones. In other words, the middle term introduces a unidirectional connection between the representation of the end and the means of achieving it. Natural things are considered only as means used or consumed to achieve an end that lies externally to them. This is the so-called 'utility point of view', which, if applied extensively, according to Hegel, can produce increasingly paltry reflections that he, like Kant, ridicules. For instance: it is already ridiculous to observe a grapevine from the point of view of its utility for man, but it is even more so to judge it in terms of the corks that can be cut from its bark to stopper bottles (Hegel, 1991, § 205r, p. 282).

Internal purposiveness is an entirely different concept, which envisages reversibility. To return to the above-mentioned syllogism, the extreme terms are exchanged: not only are the parts a function of the organism, but the organism is a function of its parts. This is what happens within a living organism: according to Kant's well-known definition in the *Critique of the Teleological Judgment* (which Hegel reiterated), '*an organised product of nature is that in which everything is an end and reciprocally a means*' (Kant, 2000, § 66, p. 247). Within an organism, each member is an integral part of the whole constituted by the same organism. It is the means for the organism's subsistence and life, but at the same time it is also its end. That is to say, the organism is a form of *Selbstorganisation* in which each part is thinkable 'only *through* all the others' and '*for the sake of the others* and on account of the whole'; in other words, every element produces the others and is reciprocally produced (Kant, 2000, § 65, p. 245). Whilst in a machine or an artifact, a part exists for the sake of the others, but not through them, in an organism, 'as an *organized* and *self-organizing* being', a part acquires sense only in its relation with the others and with the whole; at the same time the whole is such only in relation to its parts. Contrary to a purely mechanistic vision which holds that the parts have priority over the whole, 'an organism is an entity which has to be apprehended in such a way that the parts should presuppose the idea of a whole to be understood, and according to this idea that parts are reciprocally causes of their own production within this whole' (Huneman, 2006, pp. 9–10).

Hegel reformulates Kant's position in statements such as the following:

> The living being is the syllogism, whose very moments are inwardly systems and syllogisms. But they are active syllogisms, or processes; and within the subjective unity of the living being they are only *One* process (Hegel, 1991, § 217, p. 292).

Here Hegel's principal concern is to develop the type of purposiveness which comes into play in Kant's notion of the living organism; a purposiveness with a meaning other than that of the *mental* and *external* teleology mentioned earlier. Both the idea of the reciprocity of the parts within a whole, and the Kantian definition of *Naturzweck*, are diametrically opposed to the notion of a mind which plans, and in which external purposiveness is at work. The organism is a whole complete in itself, which is born complete, and which develops and grows in its completeness. No member is added subsequently; rather, as the organism develops, it realizes what it actually is. Implicit in the definition of *Naturzweck*[3] is the assumption that the organism is bound to realize nothing other than itself (in contrast to the case of external purposiveness). The organism is the origin of its own organization: it self-produces. However, precisely this connection between the concept of the organism and natural teleology constitutes one of the most problematic and controversial aspects of Kant's entire theory.

It is common knowledge that Kant considered natural purposiveness to be, not an objective principle but a merely regulative one, a subjective maxim of the reflecting power of judgment. Therefore it has a value that is not constitutive but simply heuristic. What exactly this means is contested in Kantian hermeneutics, where a range of different, if not opposed, interpretations have been put forward.[4]

The principal problem arises from the fact that according to Kant, even though we are unable to think of organized beings as anything but end-directed,[5] when we must represent this purposiveness to ourselves, we can only use an analogy with the human mode of operating: to think 'as if' beings have been planned is the

[2] It should noted that the expressions external and internal purposiveness are simplifications of a more complex Kantian conception. Within so-called *objective* purposiveness Kant distinguishes between a formal objective purposiveness and a material, or real, objective purposiveness. In turn, the latter (to which the *Critique of the Teleological Judgment* is entirely devoted) entails both internal and external purposiveness.

[3] 'A thing exists as a natural end *if it is cause and effect of itself*' (Kant 2000, § 64, p. 243).

[4] On the one hand some commentators view the principle as a subjective one resulting from the limited ability of our cognitive power to understand these unique natural objects, yet consider it indispensable in the investigation of organisms (Steigerwald, 2006, p. 718). Even if the principle does not have a constitutive value, nevertheless it is a transcendental rule necessary for our knowledge of certain natural objects as organized and self-organizing (Steigerwald, 2006, p. 718). At the opposite extreme other commentators doubt the explanatory power of teleological judgement and therefore regard it as largely useless when considering the self-organization present in nature (Zammito, 2006).

[5] 'It is merely a consequence of the particular constitution of our understanding that we represent products of nature as possible only in accordance with another kind of causality than that of the natural laws of matter, namely only in accordance with that of ends and final causes' (Kant, 2000, §77, p. 277).

only way to ground the purposiveness of nature which bears an analogy, however distant, with the purposiveness of the conscious operations of human beings (Kant, 2000, § 65, p. 247).

The difficulties that inhere in the reflecting judgement are due to what Kant sees as the human inability to conceive purposiveness independently from consciousness. This is also the ground for Kant's rather ambiguous notion of the 'technique of nature'. In fact, causality understood as technique cannot account for the process by which organisms are constituted. The term 'technique' refers to an operation based on a transitive causality and therefore to an external purposiveness (see Chiereghin, 1990; Illetterati, 2002, p. 37). Since 'strictly speaking, the organisation of nature is therefore not analogous with any causality that we know' (Kant, 2000, § 65, p. 246), in the end the option Kant chooses is a mental purposiveness still ultimately anchored to project and design. He does not find this option entirely satisfactory but it is the only way he can avoid the contradictions involved in other forms of purposiveness.[6] Nevertheless, stressing the heuristic value of this approach enables us to make the 'products and processes of nature far more intelligible than trying to express them purely in terms of mechanical laws' (Mayr, 1976, p. 402).[7]

Hegel's position on Kant can be summarized thus: Kant had the merit of showing that a merely mechanistic reading of life is not possible, and of emphasising that our intellect must necessarily explain organisms in teleological terms. Nevertheless, he was unable to provide a coherent explanation of the organism.[8]

3. Hegel's critical appreciation of Aristotle

I believe it possible to view Hegel's re-instatement of the constitutive character of natural purposiveness as an attempt to make sense of the non-conscious form of purposiveness which Kant had considered a contradiction in terms. His strategy consists in freeing the notion of purpose from analogy with the design—and thus in shaking off the Kantian constraint—and in separating the notion of purpose from the idea of its *representation.*

This, obviously, does not mean that the connection between the end and its representation does not exist at the conscious level: but it does mean that the purpose is already present at a more elementary level. It is precisely at this level that Hegel's revival of Aristotle takes place. But that does not mean—contrary to what some interpreters say—that he aims to merely recast the Kantian notion in the Aristotelian ontological form.

Hegel's references to Aristotle, and in particular to the *Physics* (II, 8), serve primarily this purpose: to uncouple the concept of end from that of awareness, but without the former ceasing to be effective. Exemplary in this regard is the following extract from an addition to the *Encyclopaedia*:

> The fundamental determination of living existence is that it is to be regarded as acting purposively. This has been grasped by *Aristotle*, but has been almost forgotten in more recent times. Kant revived the concept in his own way, however, with the doctrine of the *inner* purposiveness of living existence, which implies that this existence is to be regarded as an end in *itself*. [*Selbstzweck*]. The main sources of the difficulty here, [in Kant's case] are that the relation implied by purpose is usually imagined to be *external* and that purpose is generally thought to exist only in a *conscious* manner. Instinct is purposive activity operating in an unconscious manner (Hegel, 1970, § 360r, p. 145).

Again evident in this passage, as a real and proper leitmotiv, is reference to Kant, and with overtones very similar to those that we saw in § 204. Still to be verified, however, is the precise sense in which for Hegel the inner—unconscious—end in Aristotle's philosophy is effective, operating in nature and real. To understand this, it is necessary to recall Hegel's treatment of Aristotelian philosophy in his *Lectures on the History of Philosophy*. It is especially in these that Hegel conducts comparison with the Aristotelian notion of end. According to Hegel, the 'inner' end in modern philosophy has been lost because of two different tendencies: on the one hand, a mechanistic philosophy that has excluded the end and posited pressure, impact, and chemical reactions as the basis of nature; on the other hand, a theological physics which has conversely attributed real existence to purpose, but confined it to the realm of the thoughts of an other-worldly mind (God). It is evident, even if Hegel does not expressly say so, that these are two sides of the same coin. In fact, 'groping' between these two, apparently opposite, modes of viewing nature would produce the same result: 'dallying' in consideration of external teleology alone.

The Aristotelian concept of nature would thus be 'infinitely' more advanced than the modern concept of teleology, which only envisaged external purposiveness. Like Kant, Aristotle had grasped the idea of nature as life, or such that it is purpose in itself and unity with itself. Unlike Kant, however, he acknowledged a superior feature of nature to technology: the fact that it disposed of movement in itself. The authentic 'internal determination of the natural thing' consists for Aristotle (according to Hegel) in the fact that 'The natural is what as a principle within it, is active, and through its own activity attains its end.' In other words, it is in the principle of self-movement that the authentic nature of the end resides, or, put otherwise, 'the whole of the true profound Notion of living being' (Hegel, 1955, p. 159).

Hegel's resumption of the Aristotelian notion of self-movement, however, must not be taken in a vitalist sense. This interpretation might seem reasonable in light of the fact that, in the above Aristotelian citation, Hegel speaks of the natural as that which has an active principle within itself. It is in fact typical of vitalist approaches to maintain that in order to understand life, 'something further'—an immaterial entity, a force or a field, or a principle—must be added to the laws of physics and chemistry. Whilst this was the interpretation that some nineteenth-century thinkers notoriously gave to Aristotle's philosophy, it is not the case with Hegel, as evinced by this passage:

> In this expression of Aristotle's ["The natural is what as a principle within it, is active, and through its own activity attains its end"] we now find the whole of the true profound Notion of living being, which must be considered as an end in itself—a self-identity that independently impels itself on, and in its manifestation remains identical with its Notion [...] (Hegel, 1955, p. 159)

The living is not to be understood through the introduction of some 'additional' elements. The whole is not directed by a separate and superior entity, as the entelechy of Hans Driesch for example postulates. It should instead be recognized that an end-in-itself is 'a self-identity that independently impels itself on, and in its manifestation remains identical with its Notion'. Or, as will be argued in the next section, one must bring to light the intrinsically dialectical nature of the living.

[6] Non-intentional purposiveness and the intentional purposiveness immanent to nature itself –hylozoism (Kant, 2000, §§ 72–73, pp. 261–266).

[7] For Mayr, Kant still conceived purpose as analogous with design: 'Kant was unable to free himself from the design-designed analogy' (Mayr, 1976, p. 402).

[8] 'Our immodest conclusion is that Kant, though foreseeing the impossibility of a purely mechanical, Newtonian account of life, nonetheless was wrong in denying the possibility of a coherent explanation of the organism' (Weber & Varela, 2002, p. 120).

Accordingly, not only do we find an Aristotelian revival in Hegel, but what he reprises from Aristotle is an ontology of the living. But his is an attempt to escape from the dogmatic dichotomy which holds that there are only either vitalists or mechanicists in the life sciences (Lenoir, 1982, p. IX), and to stress that a 'third way' can be pursued. After all, it is precisely this that Hegel means when he states in the *Science of Logic* that the idea of life in itself surpasses both teleology and mechanicism.

4. Hegel's appropriation of Aristotle and the 'bio-philosophy formula': life is activity of deficiency

We may now seek to answer the question asked at the outset: in what manner are *need* and *drive* the readiest examples of purpose? (cf. § 2). Let us briefly recall the passage in § 204:

> Need and drive are the readiest example of purpose. They are the *felt* contradiction, as it occurs *within* the living subject itself; and they lead into the activity of negating this negation (which is what mere subjectivity still is).

Need and drive are for Hegel the readiest examples because they realise in the most evident way the dialectic characteristics of the idea of an 'end'. In a basic need like thirst or hunger the living being shows the unity of itself and its determined contrary, namely the deficiency of water and food. This kind of unity—Hegel calls it the 'unity of need'—is not something inert, it is not an 'empty' unity; rather, it is active, an activity which constantly distinguishes itself by two aspects, the subject and the negative of the subject. Although common thought has it that need indicates dependence on something else, in reality, in a paradoxical way, it is a manifestation of independence: in fact water and food would be totally indifferent to the living being and they would not be able to have a 'positive' relation with it if the living being was not, for Hegel, 'the possibility of this relation'. The need is not only a lack or a deficiency. If I remove a brick from the wall, a space is left open and this is a simple deficiency (see Chiereghin, 1990, p. 196). Need, however, is an 'active' deficiency: it must be fulfilled, even if, obviously, it may also not be fulfilled.

This is what Hegel once efficaciously termed (in a fragment entitled *Zum Mechanismus, Chemismus, Organismus und Erkennen* but also in his philosophy of nature) 'activity of deficiency' (*Thätigkeit des Mangels*). In this expression, the accent is not just on 'deficiency' but especially on 'activity': 'activity of deficiency' signifies that need and deficiency should not be viewed as defective moments to be eliminated in a movement back to some pre-existing unity. In fact, the movement of scission of self from self which distinguishes the living being when it opens to the external world through an instinctive urge or need does not consist in a distancing from the self in a process of progressive loss, but rather in the maintenance of the self also amid deficiency and want—as Hegel states in relation to Aristotelian self-movement. This is not the maintenance of a 'static' unity reaffirmed in unchanged form: rather, it is a ceaseless process in which the organism turns towards the outside in assimilation of organic nature; in this process the organism is primarily directed towards itself in its self-organization.

Living being is activity by deficiency. Deficiency is, so to speak, a constitutive part of living being itself. It is therefore in negativity and separation that life unfolds in its fullness and unity. Note, however, that in this movement deficiency and identity constitute an indivisible whole, one single thing. One should not think that there is some initial fixed identity to which one returns: Rather: life is inextricably bound up with what it lacks. Hence one can only speak of 'completion' on the basis of deficiency, or vice versa of deficiency only on the basis of completion (Illetterati, 1996, p. 64).

It is this activity of deficiency, according to Hegel, that distinguishes living beings from inorganic matter: it is the presence of this 'interior contradiction' which marks the transition to the living being and produces the most basic form of subjectivity:

> 'Only a living existence is aware of *deficiency*, for it alone in nature is the *Notion*, which is the unity *of itself* and its *specific antithesis*. Where there is a *limit* it is a negation, but only *for a third term*, an external comparative. However, the *limit* constitutes deficiency only in so far as the *contradiction* which is present in *one term* to the same extent as it is in the *being beyond* it, is as such immanent, and is posited within this term. The *subject* is a term such as this, which is able to contain and *support* its own contradiction; it is this which constitutes its *infinitude*' (Hegel, 1970, § 359r, p. 141).

Note firstly that in this passage Hegel distinguishes between the 'limit' which is proper to an external comparison and the 'limit' that constitutes the lack intrinsic to the living being. In the section on determined being in the *Encyclopedia Logic* he explains: 'If we take a closer look at what a limit implies, we see it involving a contradiction in itself, and thus evincing its dialectical nature' (Hegel, 1991, § 92r). On the one hand, the limit manifests the nature of every being as determined, it affirms its constitution and form; but, on the other, it is simultaneously also the negation of it. But as the negation of what is something, it is not an abstract negation; it does not simply stand before the other indifferently; it is—according to Hegel's famous expression—*the other of itself*. This 'contradiction' within the determined being—the fact that it is not statically identical to itself but constantly extends above and beyond itself—is fully manifest in change and mutation.

In the living being, the limit thus becomes dialectical, that is, the active deficiency which constitutes its essential property. Something is vital only in so far as it contains contradiction within itself and the 'strength' to contain it and to sustain it in itself.

But another point should be noted: this strength is what constitutes the infinitude of the subject. In the above-quoted passage Hegel writes: 'it is this which constitutes its *infinitude*'. This passage at first sight may appear more anachronistic, almost 'scandalous', in a context of scientific investigation into nature. But the infinity of the subject—in light of this passage—consists, not in its presumed supra-individuality, but in its ability to comprise the contradiction in itself. This is not something achieved externally to the subject which must be surpassed with an incessant effort in a 'bad' process to infinity but something which intimately pervades the subject, so interwoven with it that it distinguishes its very nature.

To summarize, therefore, it is now possible to see (in this last quotation) how closely connected the notion of natural purposiveness is with that of 'activity of deficiency': this latter notion in its turn must be set in strict relation to Hegel's concept of contradiction, this being the element, the 'root', which distinguishes the living from the inertness of the inorganic being. It is no coincidence that in the section of the *Science of Logic*, in which Hegel openly thematized what contradiction is, he again makes overt reference to the instinctive and natural sphere, and to the idea of deficiency:

> Similarly, internal self-movement proper, *instinctive urge* in general, [the entelechy of absolutely simple essence], is nothing else but the fact that something is, in one and the same respect, *self-contained and* deficient, *the negative of itself* (Hegel, 1969, p. 440).

This last quote also helps to show how Hegel ultimately draws on and transforms Aristotle. Here Hegel again refers to the Aristotelian notion of self-movement and to the 'entelechia'. However, he transforms it by 'injecting' into it the idea of a contradiction, in this way bringing it closer to his notion of activity of deficiency. 'Internal self movement—he says in the last quotation—is the fact that something

is 'self-contained and deficient in *one and the same respect*'. *In one and the same respect*: this expression is clearly used in opposition to Aristotle and his principle of non-contradiction. This has considerable implications. In the very notion of self-movement Hegel uncovers in Aristotle's philosophy something more profound and original than the principle of non-contradiction. In the Aristotelian idea of self-movement—as interpreted by Hegel—we can discern a 'germ' which revolutionizes the principle of non-contradiction itself: we can conceive of a level where contradiction turns out to be more profound than identity itself. Indeed, in his treatment of contradiction in the *Science of Logic*, Hegel clearly warns us: should we want to establish an order of precedence between identity and contradiction, the latter, *contradiction*, would have to be taken as the more profound and characteristic determination. But this, note, is only if we want to posit an order of priority: identity and contradiction are in reality so intimately bound up with each other that such an ordering is impossible.

5. Contradiction and *steresis*

On the basis of Hegel's notion of natural purpose as outlined above, we may now at least partially address the questions that have been raised. Firstly, characteristically, Hegel merges the Aristotelian and Kantian theories not by assuming one of the two points of view, but by transforming both. He maintains that the shortcoming of the Kantian perspective is that 'the relationship of purpose' is a *reflective* judgement *which considers external objects only 'according to a unity', 'as though* an intelligence has given this unity *for the convenience of our cognitive faculty*' (Hegel, 1969, p. 739). Nevertheless, although Hegel aims to restore purposiveness in the ontological sense of the term—and reprises the Aristotelian idea of the unconscious end—he does not merely reiterate Aristotelian teleology in the modern age. The aspect that he disputes in particular is the crucial relationship between identity and contradiction, with the latter radicalized to such an extent that it represents the essence of living beings, their most proper 'activity'.

Need and drive are 'the readiest examples of purpose' not so much because they are aimed, even though unconsciously, at a purpose. This would be anyway an 'external' form of teleology, situated only at one, namely unconscious, level. The essential difference does not lie in the intentionality or non-intentionality of aiming at the purpose, but rather in the fact that need and drive, in that they are the 'readiest' examples of purpose, are also the 'readiest' examples of contradiction. In other words, Hegel considers them to embody purposiveness precisely because they embody contradiction (Hegel, 1991, § 204r, p. 281). That a living being dies when it no longer has the strength to sustain contradiction indicates that this is posited by Hegel at a level for which, as the unity of life and death, it seems to be even more profound than the identity which contains it. Hence, contradiction is not understood as the contrast between our representation and the object, and between the object and our inner concept; rather, it intrinsically pertains to what Hegel calls 'notion', his idea of Life. The role of contradiction in Hegel's thought is much debated. The most controversial point is whether Hegel still adheres to the Aristotelian view of non-contradiction, or whether his principle of contradiction entails a revolution in traditional logic such as to throw into crisis the Aristotelian principle itself. As I have sought to show, I believe that Hegel 'subverts' Aristotelian logic (Lugarini, 2004, p. 9) in such a manner to involve the principle of non-contradiction as well. But this is a subversion that comes about from 'within'—that is, according to Hegel's method of authentic 'refutation'—through the development of those elements that would already lead the Aristotelian system beyond itself, as evidenced in the case of the self-movement and entelechy.

Nevertheless, almost paradoxically, Hegel makes no mention, either in the *Lectures on the History of Philosophy* or elsewhere, of an Aristotelian notion that, if developed and taken to its extreme consequences, could effectively constitute the model for his idea of activity of deficiency. I refer to the notion of *steresis* (usually translated into English as *privation* or *deprivation*). I shall not dwell here on why Hegel does not name it as such.[9] The importance of this notion—also for the Hegelian dialectic—has been emphasised with extreme acuteness by Martin Heidegger, albeit from a sharply critical perspective on the conception of negativity, which notoriously marginalizes its role in the history of metaphysics.

Steresis is one of the most complex concepts in Aristotle's philosophy. Even when Aristotle seeks to define it overtly, distinguishing among its different senses—as, for instance, in *Metaphysics* (V, 22, 1022b)—one gains the impression that it is an entirely elusive notion. It is no coincidence, in fact, that almost nothing has been written on the topic, even though *steresis* is commonly considered to be one of the most characteristic notions of Aristotle's thought (see Wieland, 1970, p. 131). *Steresis*, Aristotle states, has numerous meanings. *Steresis means lack, privation, deficiency*. Among the various meanings of the term distinguished by Aristotle, it can signify the loss or the deprivation of something that the thing (or its genus) should by its nature possess: as in the case of blindness, which is the lack of sight in an animal that by its nature should see. But deprivation also means forfeiture, the violent appropriation of a thing that belongs to another.

Moreover, *steresis*—and this is this the meaning of greatest interest here—is closely connected by Aristotle with potency (dunamis). In a famous passage in the *Metaphysics*, Aristotle writes:

> Sometimes [a thing] is thought to be of this sort [potent] because it has something, sometimes because it is deprived of something; but if privation is in a sense 'having' or 'habit', everything will be capable by having something, so that things are capable both by having a positive habit and principle, and by having the privation of this (1019 b 5–8).

Possession (having) here corresponds to the Greek term *hexis*, which can also be translated as 'habit' or 'faculty' (ability). When commenting on this passage, Giorgio Agamben stresses that Aristotle's principal interest is in potency (dunamis), such that it signifies 'privation'. 'In Aristotle's terminology, *steresis*, 'privation', stands in strategic relation to *hexis*, that is, to something that attests to the presence of what the action lacks. Having a potency, having a faculty means having a privation' (Agamben, 2005, p. 276). To explain the point, Agamben cites a passage from the *De Anima* in which Aristotle distinguishes between two different types of potency, and does so through an example. On the one hand, 'generic' potency is that possessed by a child in regard to knowledge. He is in fact potent in the sense that he must undergo alteration through learning. Instead, whoever already possesses a technique—for instance, grammar or arithmetic—need not undergo an alteration but is potent by virtue of a *hexis* which he may or may not actualize. For Agamben, the potency that really interests Aristotle is the one which stems from this *hexis*, and, therefore, essentially from the possibility of its non-exercise (Agamben, 2005, p. 277). This is a potency that is not lost in passage to actuality: 'The passage into act neither annuls nor exhausts the potency, but this conserves itself as such in actuality and, markedly, in its eminent form of power of not (being or doing)' (Agamben, 2005, pp. 285–6). Agamben's conclusions are that potency does not pass into actuality, suffering destruction or alteration, but, in the act, it grows and perfects itself.

[9] Presumably—given that he was certainly aware of the notion—because he considers it a corollary to the notions of power [dunamis] and action [entelechy].

According to him, our Western tradition must still take full account of all the consequences of this figure of potency, which does not disappear or become lost in the act. But taking account of this figure is essential for understanding the living in all its inexhaustible forms. Agamben's interpretation seems to be in (indirect) opposition to that of Heidegger. Heidegger has the indubitable merit of having been the first to underline the connection between *steresis* and potency in order to explain (and criticise) movement in Aristotle: 'Dynamis is in a pre-eminent sense exposed and bound to steresis' (Heidegger, 1995, p. 95). And it was on precisely this aspect that Heidegger centred his principal courses on Aristotle, besides his celebrated essay *On the Being and Concept of Physis*. Here *steresis* is considered to be the fundamental concept in Aristotle's *Physics*: 'the essence of *physis* reveals itself in *steresis*'. However, for Heidegger, lack is not understood by Aristotle in a radical way, but simply as a lack that can be overcome, or remedied. In a horizon like that of the history of metaphysics, of which Aristotle also is a protagonist, dominated by being as 'simple presence'[10], in fact, movement can be exclusively explained as a process of actualization in which the 'not' from which the movement has started is progressively left aside and lost.

In fact, against the background of Heidegger's interpretation and criticism of *steresis* and Aristotelian movement, the true object of his polemic, albeit indirectly, is Hegel. The latter had taken negativity to be the source of every movement, but only in order to deprive it of its authentic power, to subdue and tame it. In the same years when Heidegger wrote his essay on Aristotle's *Physics*, or in his sketches *Über die Negativität* of 1938–39 and 1941, he accuses Hegel of simply adhering to the traditional views of the *Metaphysics* regarding negativity. But already in some writings of the 1920s, and especially in his 1922 lectures, *Phenomenological Interpretations with respect to Aristotle*, Heidegger argued that the Hegelian dialectic has its roots in the concept of *steresis*: a point reiterated in his lecture of the 1931 summer semester, *Aristotle's Metaphysics Θ 1–3. On the Essence and Actuality of Force*, where he states that, although Aristotle and the ancients treated *steresis* to only a minor extent, we should not forget the movement in philosophy that, through such a concept, led up to Hegel (Heidegger, 1995, p. 110).

I believe that, in a certain sense, Heidegger is correct: that there is indeed a 'movement' in philosophising, the linking theme of *steresis*, that leads from antiquity to Hegel. Nevertheless, this line can be understood in a different sense from Heidegger's interpretation of it: that is, not in the sense of a 'taming' of negativity and contradiction, but, on the contrary, in the sense of its slow emergence—the gradual recognition of its essential role. Could not *steresis* therefore be interpreted, rather than as a 'remediable' deficiency, as a real activity of deficiency?

If, as Agamben maintains, potency can be recognized from its intrinsically privative nature, more than from its fulfilment and annulment in the act, can one not discern in this the forming of a 'logic of contradiction' divergent from and alternative to the 'logic of identity' to which the history of modern Western thought has often been schematically reduced? And may not 'internal purpose' not paradoxically have more to do with this logic of contradiction than with that of identity? And, finally, could it not be, almost paradoxically, Hegel himself (the philosopher *par excellence* of 'synthesis' and 'conciliation') who personifies—his intentions notwithstanding—this divergent tradition?

If this is so, one may better comprehend what Cusanus wrote in *De Berillo* at the beginning of the modern age:

> But if Aristotle had understood the beginning which he calls *privation*—understood it in such a way that privation is a beginning that posits a coincidence of contraries and that, therefore, (being 'deprived,' as it were, of every contrariety), precedes duality, which is necessary in the case of contraries—then he would have seen correctly (Cusa, 1998, p. 811).

Acknowledgements

I wish to thank Lenny Moss and Jonathan Davies for fruitful discussions on my paper, James Barham and Dan Nicholson for language help, and the Bruno Kessler Fondation Trento for having partially financed my stay at the 2009 conference of ISHPSSB (Brisbane, Australia).

References

Agamben, G. (2005). *La potenza del pensiero. Saggi e conferenze*. Vicenza: Neri Pozza.
Chiereghin, F. (1990). Finalità e idea della vita. La recezione hegeliana della teleologia in Kant. *Verifiche, 19*, 127–229.
Cusa, Nicholas (1998). *Metaphysical speculations: Six Latin texts* (J. Hopkins, trans.). Minneapolis: Minn. A.J. Banning Press.
Hegel, G. W. F. (1931). *The phenomenology of mind* (J.B. Baillie, trans.). London: Allen & Unwin.
Hegel, G. W. F. (1955). *Hegel's lectures on the history of philosophy* (Vol. II) (E. S Haldane & F. H. Simpson, trans.). New York: Humanities Press.
Hegel, G. W. F. (1969). *Science of logic* (A.V. Miller, trans.). New York: Humanities Press.
Hegel, G. W. F. (1970). *Hegel's philosophy of nature* (M. J. Petry, trans.). London: Allen & Unwin.
Hegel, G. W. F. (1991). *The encyclopedia logic. Part I of the encyclopaedia of philosophical sciences with the Zusätze* (T. E. Geraets, W. A. Suchting, H. S. Harris, trans.). Indianapolis-Cambridge: Hackett.
Heidegger, M. (1995). *Aristotle's metaphysics Θ 1–3: On the essence and actuality of force* (W. Brogan and P. Warneck, trans.). Bloomington-Indianapolis: Indiana University Press.
Heidegger, M. (1976). On the being and conception of *physis* in Aristotle's *Physics* B, 1. *Man and World, 9*(3), 219–270.
Huneman, P. (2006). From the critique of judgment to the hermeneutics of nature: sketching the fate of philosophy of nature after Kant. *Continental Philosophy Review, 39*, 1–34.
Illetterati, L. (1996). *Figure del limite. Esperienze e forme della finitezza*. Trento: Verifiche.
Illetterati, L. (2002). *Tra tecnica e natura. Problemi di ontologia del vivente in Heidegger*. Padova: Il poligrafo.
Johnson, M. R. (2005). *Aristotle on teleology*. New York: Oxford University Press.
Kant, I. (2000). *Critique of the power of judgment* (P. Guyer & E. Matthews, trans.). Cambridge: Cambridge University Press.
Lenoir, T. (1982). *The strategy of life. Teleology and mechanics in nineteenth century German biology*. Dordrecht: Reidel.
Lugarini, L. (2004). *Hegel e Heidegger. Divergenze e consonanze*. Napoli: Guerini e Associati.
Mayr, E. (1976). Teleological and teleonomic: A new analysis. In E. Mayr (Ed.), *Evolution and the diversity of life* (pp. 383–404). Cambridge-London: Belknap Press.
Michelini, F. (2008). Thinking life. Hegel's conceptualization of living being as an autopoietic theory of organized systems. In L. Illetterati & F. Michelini (Eds.), *Purposiveness. Teleology between nature and mind*. Frankfurt/Paris/Lancaster/New Brunswick: Ontos Verlag.
Steigerwald, J. (2006). Kant's concept of natural purpose and the reflecting power of judgement. *Studies in History and Philosophy of Biological and Biomedical Sciences, 37*, 712–734.
Walsh, D. M. (2006). Organisms as natural purposes: The contemporary evolutionary perspective. *Studies in History and Philosophy of Biological and Biomedical Sciences, 37*, 771–791.
Weber, A., & Varela, F. J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of individuality. *Phenomenology and the Cognitive Sciences, 1*, 97–125.
Wieland, W. (1970). *Die aristotelische Physik. Untersuchungen über die Grundlegung der Naturwissenschaft und die sprachlichen Bedingungen der Prinzipienforschung bei Aristotles*. Göttingen: Vandenhoeck & Ruprecht.
Zammito, J. (2006). Teleology then and now: The question of Kant's relevance for contemporary controversies over function in biology. *Studies in History and Philosophy of Biological and Biomedical Sciences, 37*, 748–779.

[10] 'Steresis as a becoming-present (*Abwesung*) is not simply absentness (*Abwesenheit*), but rather is a becoming-present, the kind in which the becoming-absent becomes present' (Heidegger 1976, p. 266).

Studies in History and Philosophy of Biological and Biomedical Sciences 43 (2012) 140–151

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

How was teleology eliminated in early molecular biology?

Phillip R. Sloan

Program of Liberal Studies and Program in History and Philosophy of Science, University of Notre Dame, Notre Dame, IN 46556, USA

ARTICLE INFO

Article history:
Available online 1 July 2011

Keywords:
Teleonomy
Niels Bohr
Complementarity
Max Delbrück
Erwin Schrödinger
Teleology

ABSTRACT

This paper approaches the issue of the status of teleological reasoning in contemporary biology through a historical examination of events of the 1930s that surrounded Niels Bohr's efforts to introduce 'complementarity' into biological discussions. The paper examines responses of three theoretical physicists who engaged boundary questions between the biological and physical sciences in this period in response to Bohr—Ernst Pascual Jordan (1902–80), Erwin Schrödinger (1887–1961), and Max Delbrück (1906–81). It is claimed that none of these physicists sufficiently understood Bohr's 'critical' teleological arguments, which are traced to the lineage of Kant and Harald Høffding and their respective resolutions of the Antinomy of Teleological Judgment. The positions of these four historical actors are discussed in terms of Ernst Mayr's distinction of 'teleological,' 'teleomatic,' and 'teleonomic' explanations. A return to some of the views articulated by Bohr, and behind him, to Høffding and Kant, is claimed to provide a framework for reintroducing a 'critical' teleology into biological discussions.

1. Introduction

It is common to encounter the claim in contemporary life science that modern biology has successfully eliminated traditional teleological explanations through the combination of natural selection theory and molecular biology. Furthermore, this is often seen as one of the consequences of the 'molecular' revolution in biology, since this has provided a way to give mechanistic 'bottom up' explanations of what may have seemed to be purpose-laden biological processes.[1]

I shall focus on the efforts to distinguish traditional teleological explanations from non-teleological accounts of biological function and process through their replacement by 'teleonomical' and 'teleomatic' accounts (Mayr, 1974, 1992).[2] Such explanations may acknowledge the obvious goal-directedness of organic life, but account for it by an eventual reduction to underlying material and efficient causation, denying the realism attributed to traditional teleological causes in the sense of formal and final principles of life, or even the necessity of reasoning about organisms in teleological ways. Ernst Mayr, whose discussions of teleonomic and teleomatic viewpoints have been influential, distinguished these forms of explanation from traditional teleological accounts, either in their external (i.e. Platonic, Stoic), or internal (Aristotelian) sense from the others. 'Teleomatic' explanations are in Mayr's definition the automatic achievements of changes of state in a passive, law-governed way in response to causal input from an *external* agency in accord with deterministic natural laws. His examples are such things as the cooling of hot iron, or radioactive decay. They 'may have an end point but they never have a goal' (Mayr, 1992, p. 125). 'Teleonomic' processes proper, in Mayr's restricted definition, are defined as follows: 'a teleonomic process or behavior is one that owes its goal-directedness to the operation of a program' (Mayr, 1992, p. 127). This implies the interaction of intrinsic properties within the organism with external laws and conditions. Mayr also sees a teleonomic process, in contrast to a teleomatic one, as clearly oriented to an end point: 'this end point might be a structure (in development), a physiological

E-mail address: Phillip.R.Sloan.1@nd.edu

[1] The broad discussion of functional explanation will not be directly at issue in this paper. For overviews see Zammito (2006), McLaughlin (2001).

[2] A wide range of meanings of 'teleonomy' is encountered in the current literature, and multiple definitions have been offered since the term was first introduced by Colin Pittendrigh (Pittendrigh, 1958, p. 394), with only some of these usages mapping on to Mayr's definition. Pross (2005), for example, interprets this to apply to systems which violate thermodynamic principles, without requiring the need to conform to an underlying 'program' in Mayr's sense. My usage will follow Mayr's definition.

1369-8486/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsc.2011.05.013

function, the attainment of a geographical position (in migration), or a 'consummatory act' in behavior' (ibid.). In organisms, this program is based on a genetic code, and in turn, this code is a product of evolution by natural selection. Mayr's definition also allows for the multiple pathways and complexities entailed in developmental biology, and allows for a complex understanding of gene action. None of this would be captured by rigidly deterministic teleomatic explanations. A teleonomic understanding of organisms in this sense does not claim that all teleological behavior is 'reducible to' causal mechanisms, and it interprets realistically the systemic and holistic aspects of organic life, while still acknowledging a causal closure on a naturalistic account. At the same time it does not involve commitment to realistic interpretations of formal and final causes, or to reverse causation, and in this sense denies traditional teleological realism. Controversies over Mayr's distinctions will not be explored here (McLaughlin, 2001, pp. 28–32; Nagel, 1977, pp. 267 ff). I will simply adopt Mayr's distinctions as a useful heuristic for this exposition, recognizing that his distinction between 'teleomatic' and 'teleonomic' may be unsustainable in final analysis. As Peter McLaughlin argues, Mayr's categories do not smoothly do away with the need for more traditional teleological accounts (McLaughlin, 2001, p. 32).

This paper will explore the question of teleology in biology through the examination of the historical interaction between 'critical' teleological perspectives (Roll-Hansen, 1976), against 'teleonomic' and 'teleomatic' alternatives as this occurred in the biophysics of the 1930s.[3] As a philosophical claim, this paper will argue for a return to some of the insights of critical teleology, exemplified imperfectly in Niels Bohr's arguments, that were lost in the subsequent historical discussion.

The discussions of the 1930s are particularly relevant. This era was formative in the development of modern biophysics in the form that emerged from the interactions of the new physics with traditional biology and genetics in this period, motivated in good part by the entry of prominent theoretical physicists into discussions of biology. These encounters shifted focus away from the longstanding Driesch-Loeb heritage of the 'mechanism-vitalism' debate of the early decades of the century to new territory defined by new theoretical developments in physics and chemistry that many sought to exploit as a way to redefine the relation of the physical and biological sciences. Out of this discussion emerged a more potent version of biological reductionism that has been influential since that date. As Francis Crick put this reductivist claim in an often-repeated statement:

> The ultimate aim of the modern movement in biology is in fact to explain *all* biology in terms of physics and chemistry. There is a very good reason for this. Since the revolution in physics in the mid-twenties, we have had a sound theoretical basis for chemistry and the relevant parts of physics [...] So far everything we have found can be explained without effort in terms of the standard bonds of chemistry—the homopolar chemical bond, the van der Waal attraction between non-bonded atoms, the all-important hydrogen bonds, and so on [...] Thus eventually one may hope to have the whole of biology 'explained' in terms of the level below it, and so on right down to the atomic level. And it is the realization that our knowledge on the atomic level is secure which has led to the great influx of physicists and chemists into biology. (Crick, 1966, pp. 10–11, 14)

Deeper exploration of these events referred to by Crick defines the problematic of this paper.

2. The Bohr debates

Niels Bohr's entry into the discussion of the relations of biology and physics in influential public statements he made between 1929 and 1937 sets the context for this discussion.[4] Because of his prestige in science generally, and also because of his dominant role in theoretical physics in the period, Bohr's decision to enter public discussions surrounding the implications of quantum physics for biology and psychology in the late 1920s generated attention that few other individuals could command in the scientific world of the 1930s.

Either directly or indirectly, Bohr's reflections formed the foil against which different conclusions on the relation of teleology and mechanism were formulated by the three theoretical physicists of interest here—Ernst Pascual Jordan (1902–80), Erwin Schrödinger (1887–1961), and Max Delbrück (1906–81). Their contrasting interpretation of these issues, and their different reactions to Bohr's views, had a considerable impact on the way in which teleological reasoning in biology was conceptualized within biophysical discussions that drew upon the heritage of this new form of biophysics. Two of these individuals—Schrödinger and Delbrück— had concrete impact on early theoretical interpretations of molecular biology (Sloan, 2012, forthcoming; Domondon, 2006; Yoxen, 1979; Olby, 1971).

This debate took shape—either directly or indirectly—against the backdrop of Bohr's efforts to develop his approach to biophysics through the extension of his philosophical program of 'complementarity,' first utilized publicly to describe the resolution of certain paradoxes in quantum physics in 1927. Bohr first extended this concept publicly outside physics to deal with issues in psychology in 1929, and then applied this to general biological questions in the early 1930s (Bohr, 1929, 1930, 1931). Since considerable scholarship, masterfully commanded by my colleague Don Howard, has addressed the wider dimensions of his concept of Bohr's 'complementarity,' and considerable discussion has also been given of Bohr's application of 'biological' complementarity in specific, (Domondon, 2006; Roll-Hansen, 2000, Roll-Hansen, 2012, forthcoming; McKaughan, 2005, McKaughan, 2012, forthcoming; Aaserud, 1990, chap. 2; Hoyningen-Huene, 1994; Folse, 1990; Kay, 1985), my effort here is limited to probing a few key aspects of these discussions beyond the analyses available, situating these more deeply in their historical context. This context illuminates several issues that otherwise are left unclear.

Several of the scholars cited above have drawn attention to the origins of Bohr's biophysical interests in the efforts of his physiologist-father Christian Bohr (1855–1911) to resolve the mechanism-vitalism dispute in the form he encountered it near the turn of the century. The importance of University of Copenhagen philosopher Harald Høffding (1843–1931), a close family friend and university colleague of the elder Bohr, and Niels Bohr's university instructor in philosophy, has also been a subject of substantial scholarship that has explored some of the importance of Høffding's philosophy for Niels Bohr's views on teleology, mechanism, and biophysics (Roll-Hansen, 2000; Roll-Hansen, 2012, forthcoming; Aaserud, 1990, pp. 68–109; Folse, 1990; Kaiser, 1992). I am extending this analysis by drawing particular attention to some additional

[3] In applying Mayr's categories to a debate of the 1930s, I acknowledge that 'teleonomy,' 'teleomatic,' and 'molecular biology' are neither actor's categories nor terminology in use in the historical period under examination. I will, however, argue that several ingredients that were later to characterize more recent accounts of these positions emerged from the context of the discussions I detail.

[4] My intention here is not to enter the vast area of general Bohr scholarship and the larger literature surrounding the debates over the Copenhagen interpretation and 'complementarity' in the philosophy of physics. I will simply tease out the strand which is of relevance to the biophysical debates in this specific historical context. On the more general issues see Howard (1994), Faye (2008).

details in Høffding's interpretation of Kant's solution to the 'antinomy' of teleological judgment beyond those highlighted by Roll-Hansen, in order to illuminate some additional aspects of Bohr's own positions. As Roll-Hansen (2012, forthcoming) has suggested, this philosophical framework was neither understood by those like Delbrück who claimed to be adopting Bohr's views, nor by those who were critical of Bohr. This framework is also not considered in some recent analyses of Bohr's approach to the issue of biological reductionism.

Kant's possible importance for Bohr's philosophy of science has been a subject of some disagreement (Folse, 1985; Kaiser, 1992), and the controversy bears in large measure on the specific texts and aspects of Kant's thought at issue. The text of relevance for the discussion of this paper is Kant's analysis of biological teleology in the second half of the *Critique of the Power of Judgment* of 1790. How Bohr developed his similar views—whether from readings and discussions with his philosophical mentor Høffding (Roll-Hansen, 2000; Roll-Hansen, 2012, forthcoming; Faye, 1979), or from his own reading of Kant, or from other intermediate sources, or simply from his own reflections on these matters—is not critical to resolve at this point. I am also not in any strong sense claiming that Bohr is a Kantian philosopher, a strict disciple of Høffding, or even a systematic philosopher on these issues. He departs from Kant on numerous technical issues, including those of relevance to this paper. Nonetheless, I will build here upon the line of scholarship that sees Bohr developing upon Høffding's reworking of Kantian perspectives in a form that rejected many aspects of Kant's technical philosophy (Christiansen, 2006).

3. Kant's resolution of the antinomy of teleological judgment

Kant's efforts in the late 1780s to supply a 'critical' solution to the conflict of mechanistic, vitalistic, and teleological explanations in biology in the form he encountered these in the late eighteenth-century—Cartesian-inspired biological mechanism, Stahlian vitalism, Leibniz-Wolffian teleomechanism, and Spinozistic hylozoism—were complex enough to have spawned a variety of interpretations of his views in recent decades (e.g. Watkins, 2009; Zammito, 2009, 2006; Ginsborg, 2006; Guyer, 2005, chap. 13; Quarfood, 2004, 2006; Allison, 2003 in Guyer, 2003; McLaughlin, 1990). Two fundamental alternatives have emerged in the interpretation of Kant's arguments of relevance to this essay. A common reading, the heritage of Marburg Neo-Kantianism, is a 'projectivist' reading of Kant's arguments, which interprets Kant to make the attribution of teleological purposiveness to organisms a subjective imposition on phenomena, non-constitutive in the Kantian sense, and subject to the restrictions of 'reflective' rather than determinative judgment. As such, teleological concepts lack explanatory function, with the latter reserved for mechanistic accounts that remain the ideal of Kantian science.

Opposing this reading are those who have seen a much deeper complexity in Kant's address of the issue of teleology, and reject attempts to press it onto the regulative-constitutive framework of the First Critique (Quarfood, 2004, 2006; Watkins, 2009; Zammito, 2009; Ginsborg, 2006; Guyer, 2005, chap. 13; Allison, 2003; McLaughlin, 1990, chap. 3). Some have even seen Kant as a defender of a form of critical teleological realism that places him closer to Aristotle than the projectionist reading would countenance (Quarfood, 2006, 2004; Ginsborg, 2006). Without pretending to resolve here this interpretive issue within Kant scholarship, the evidence provided by these discussions underlines the textual warrant for a complex, and even 'realistic' reading of Kant's views on teleological purposiveness sufficient to ground the interpretation that Bohr seems to have derived indirectly from Kant through his philosophical mentor, Harald Høffding. This is as much as I am claiming in this discussion.

I will presume that the main features of Kant's arguments in the *KU* do not need to be rehearsed here and I will move directly to the debate over Kant's resolution of the Antinomy of Teleological Judgment in the central sections of the Dialectic. This antinomy is based on the apparent opposition between a mechanistic and teleological explanation of organisms. The complexity that emerges from this discussion is that it does not bear a clear similarity to the antinomies of the First Critique (Watkins, 2009). Furthermore, and unlike what might be the expectations of those familiar with the arguments of the *CPR*, the resolution does not occur along a regulative-constitutive distinction. Instead, *both* mechanistic and teleological analyses are rendered 'regulative' in the *KU* discussion.[5] The privileging of mechanism that seems to be the argument of the Analytic of the *KU*, and the reading suggested by a text like Kant's *Metaphysical Foundations of Natural Science,* is considerably altered as the Dialectic of the *KU* is developed. Instead, in the crucial discussion of paragraph 78, we see that the status of *both* mechanism and teleology is essentially equalized. *Neither* can be constitutive of things in themselves. They stand instead as possibly reconcilable in an unknown ground.

> [I]f we are to have a principle that makes it possible to reconcile the mechanical and the teleological principles by which we judge nature, then we must posit this further principle in something that lies beyond both (and hence beyond any possible empirical presentation of nature), but that nonetheless contains the basis of nature, namely, we must posit it in the supersensible, to which we must refer both kinds of explanation [*von dieser aber doch den Grund enthält, d.i. im Übersinnlichen, gesetzt und eine jede beider Erklärungsarten darauf bezogen werden*]. (Kant, 1923, para. 78, 412. trans. Kant, 1987, p. 297)

Since we cannot have any cognitive knowledge of this supersensible domain, we cannot fully penetrate how this reconciliation comes about (ibid., para. 78). In spite of the 'seeming conflict that arises between the two principles for judging that product [...], we are assured that it is at least possible that objectively, too, both these principles might be reconcilable in one principle (since they concern appearances, which presuppose a supersensible basis)' (ibid). Fundamental to this resolution is therefore a claim that the conflict between teleological and mechanical explanations can be overcome, but only by appeal to a *transcendental* realm outside empirical nature. There is no solution within empirical nature itself.

Furthermore, the teleological and mechanical are not reconcilable in a single causal explanation. Similar to the point Bohr would later make with his notion of biological 'complementarity,' the teleological and mechanical are, instead, *mutually exclusive* accounts:

> For each mode of explanation excludes the other—even supposing that objectively both grounds of the possibility of such a product [of nature] rest on a single foundation. (Ibid.)

My point in drawing attention to this detail in Kant's unusual resolution of the antinomy, is that it is through this passageway that we can see with greater clarity some features of Bohr's arguments about the relations of teleology and mechanism in biology which are easily missed if one is not conversant with this philosophical background. More proximately this leads us to the importance of Harald Høffding's interpretation of these issues for Bohr's subsequent reflections.

[5] A deeper exploration of this issue would, I argue, require a detailed analysis of the resolution of the antinomy against Kant's complex philosophy of nature. See on this Zammito (2009) and Breitenbach (2008) for some examination of this issue.

4. Reworking Kant: Høffding on the antinomy

The evidence for the concrete impact of Harald Høffding on Bohr's philosophical and biological thinking leads us to attend less to the historical Kant, and more directly to Høffding's interpretation of the biological issues. Roll-Hansen's analysis of the importance of Høffding's development of Kant's insights for Bohr's views on the relations of mechanism, teleology, and vitalism are foundational for my analysis (Roll-Hansen, 2000, 2012, forthcoming). Extending his analyses, I focus on Høffding's revision of Kant's *transcendental* reconciliation of the antinomy of mechanism and teleology. As argued in a recent study (Christiansen, 2006, pp. 11–15), Høffding rejected Kant's distinction of the primary faculties of intuition, understanding, and reason, and reworked many other aspects of Kant's solutions to epistemological issues, with input from Heinrich Hertz and Hermann von Helmholtz. Instead of following Kant's transcendental Idealism, Høffding developed a metaphysical monism grounded in a theory of analogy, in which reality is given in experience, but always encountered as mediated by analogies and images (Høffding, 1905). This meant that our encounter with reality is based upon a 'likeness of the *relations of properties*, not identity of the individual properties' (Christiansen, 2006, pp. 11-15). This theory of analogy Høffding combined with his 'speculative hypothesis' of 'critical' monism, which 'presupposes unity and continuity in the real, and as reality is intelligible in a high degree' (Høffding, 1902, p. 150).

While Høffding was fully aware of Kant's attempt to resolve the antinomy of teleological judgment, like many after Kant who sought a naturalistic resolution of the antinomy (Zammito, 2009), he saw in this resolution defects that he sought to overcome by moving this unification from the supersensible to the real. As he explains in a discussion of 1926, the Kantian distinction of mechanism and teleology does not mean for him,

> a fundamental distinction between mechanism and organicism, but is based in the fact that our knowledge of these must utilize different methods, and must proceed sometimes from the parts to the whole and sometimes from the whole to the parts without being able to carry these through. It was Kant's belief that one and the same order of things lies at the foundation equally for the mechanical connection and for the formation and existence of organic totalities. We have therefore no warrant to consider the beauty and purposiveness of nature as due to chance. (Høffding 1926, p. 25)

Bohr's biological 'complementarity,' if not necessarily identical with his meaning of the term in physics, bears several resemblances to the resolution of the antinomy of teleological judgment offered by Høffding, and departs from that of Kant himself while still retaining some important similarities. For Kant, as we have seen, the teleological and mechanical understanding of living beings achieves unification in a *transcendental* realm, in keeping with Kant's transcendental Idealism. For Høffding, this unity is achieved within the *empirical* domain and is tied to his critical monism. To use terminology that has become current since Timothy Lenoir's important discussions (Lenoir, 1982), Høffding, and after him, Bohr, if not Kant, can be considered to be genuine ontological 'teleomechanists.' But for neither—and this is in accord with Kant's claim in the *KU* (para. 78)—does this mean a simple *union* of the teleological and mechanical in a single set of causes. Instead, there is a dual-aspect, and mutually exclusive, relation between the teleological and mechanical which requires alternative frameworks for the understanding of organisms. It is this critical point that I argue was not understood by those of his contemporaries who read Bohr and debated the meaning of his 'complementarity' in biology.

5. Misreading Bohr

Bohr's interest in extending his notion of 'complementarity' beyond physics to areas of biology and psychology was first carried out publicly in a lecture to a meeting of the Scandinavian Union of Natural Scientists in the summer of 1929. One historical detail that has not adequately been explored in relation to Bohr's engagement with these issues is the conversations with Jordan that began in the spring of 1929 at the annual meeting of theoretical physicists in Copenhagen with the young quantum physicist Ernst Pascual Jordan concerning the implications of quantum mechanics for issues outside physics (Bohr to Jordan, Jan. 25, 1930 in Favrholt, 1999b, p. 10, 515).[6] Already by 1927, Jordan had been interested in exploring the implications for biology of the new quantum mechanics in which he was a fundamental participant (Aaserud, 1990, pp. 82–92; Beyler 1994; Heisenberg, Born, & Jordan, 1925). More explicit conversations over psychological and biological topics between Bohr and Jordan were then continued at the April 1931 Copenhagen annual meeting of theoretical physicists, which generated further correspondence and a long praise-filled response by Bohr to the resultant draft of Jordan's paper on causality and freedom of the will sent him in May following this April conversation (Jordan to Bohr, May 20, 1931, Bohr to Jordan, June 5, Favrholdt, 1999b, p. 517, 520–23).

This generated a flurry of correspondence in June between Bohr and Jordan that illuminates the differences in their respective interpretations of the relation of the biological and physical, and the relevance of quantum mechanics for biological and psychological issues. Akin to the situation Roll-Hansen has illuminated in his analysis of the Bohr-Delbrück relationship, there is a significant failure of comprehension in evidence. While praising Jordan's 'beautiful essay' (*schönen Aufsatz*), Bohr also drew out some important points of disagreement with Jordan's views on the relations of biology, physics, and psychology.

Particularly relevant is Jordan's evident failure to understand the foundations of Bohr's views in Høffding's psycho-physical parallelism and his theory of analogy.[7] In a long letter of June 5, Bohr comments that 'your discussions of the parallelism of physical and psychical events could perhaps give rise to a misunderstanding of my point of view.' He discounts a strong analogy between wave-particle duality and issues in psychology, and worries about potential misunderstanding of his views:

> [...] I want to draw your attention to the fact that your discussions of the parallelism of physical and psychical events could perhaps give rise to a misunderstanding of my point of view. My emphasis on the formal similarity between the wave-particle problem and fundamental problems in psychology does of course not aim at a narrow [*enge*] analogy between psycho-physical parallelism and the wave-particle duality, but above all at the possibility of gaining mutual elucidation [*Belehrung*] from physical and psychological investigations. Parallelism itself involves of course a special complementarity that cannot

[6] Bohr speaks of being 'deeply engrossed' (*sehr erfüllt*) with issues of biology in this letter. The Jordan conversations provide an important contextualization that has not been discussed by others looking into this issue with the exception of Richard Beyler (1994, 1996). I acknowledge my deep debt to Beyler's studies on Jordan.

[7] In his unpublished lecture on Høffding's views on the relationship of physics and psychology, given in August of 1932 near the time of the 'Light and Life' lecture, Bohr speaks there about how Høffding saw the parallels between complementarity in physics and issues in psychology, and praises him for his caution in extending these analogies, while with possible reference to his conversations with Jordan, notes that 'in contrast to Høffding such caution is not always exerted neither [sic] by physicists nor by psychologists ....' As Bohr continues, '... as regards such problems as freedom of the will we cannot say anything else, that we here deal with forms of consistent life, the parallel of which in physical nature in the sense of Spinoza are not open to analysis by mechanical ideas' (Bohr MSS August, 1932.).

> be understood by means of laws of a one-sided physical or psychological kind. (Bohr to Jordan, June 5, 1931, Favrholdt, 1999b, pp. 521–22)

Bohr continues by noting that he is grounding his conclusions on the 'renunciation, emphasized in the mentioned articles [Bohr, 1929], of the concept of observation as regards living organisms, for which concept killing by the application of the means of observation sets a limit in principle' (Favrholdt, 1999b., 522 with slight revisions). Continuing, Bohr agrees with Jordan that 'acausality can be regarded as a characteristic of life [*die Akausalität als Merkmal des Lebens bezeichnen kann*],' but his reading of how this is manifest is made clearer as he addresses Jordan's arguments. For Bohr, 'the laws of biology [...] cannot be comprehended mechanically,' and as a result are in principle different 'from the technical amplification devices used in the study of fluctuation phenomena.' The analogy with quantum uncertainty is only an analogy:

> Just as the stability of the atomic phenomena is inseparably connected with the limitation of observation possibilities expressed by the uncertainty principle, so in my view the peculiarities of life phenomena are connected with the impossibility in principle of ascertaining the physical conditions under which life exists. Briefly, one could perhaps say that atomic statistics deals with the behaviour of atoms under well-defined external conditions, whereas we cannot define the state of the organism on an atomic scale. (Ibid)

The time-sequence within which these Bohr-Jordan discussions of spring and summer 1931 reached the public is important to follow carefully, since this sequence influenced the subsequent understanding of several issues. Except for the publication of his 1929 essays in German in 1931 (Bohr 1931, 1934), which did not reflect the extended discussions with Jordan, Bohr was the first to present his side of the conversation in his well-known 'Light and Life' lecture, delivered in English in August of 1932 to the International Congress on Light Therapy meeting in Copenhagen. From all reports, this lecture, given in a faint whispering presentation, was generally unintelligible to most of the audience.[8] The first half of the lecture summarized his general interpretation of complementarity in physics, and the second half applied this notion specifically to biology. In this second portion he offered his suggestions on the resolution of the conflict of mechanism and teleological purposiveness, and here we can see the connections with our preceding discussion. Rejecting the option of vitalism, he nonetheless discounts a purely reductive and mechanical analysis of life as adequate. Biological science requires *both* mechanism and teleology, and this is a 'complementary' perspective that means a dual description of a single reality, rather than a matter of alternative subjective 'perspectives' without purchase on an ontological given.[9] Similar to Høffding, and more remotely echoing Kant, these are parallel, but *mutually exclusive* descriptions of a single reality, with neither teleological nor mechanical descriptions reducible to the other. There is 'no well-defined limit [...] for the applicability of physical ideas to the phenomena of life' but this only displays one aspect of this reality. There is another 'complementary' analysis that requires the teleological. As this is put in the published English version of his lecture:

> [T]he concept of purpose, which is foreign to mechanical analysis, finds a certain field of application in problems where regard must be taken of the nature of life. In this respect, the role which teleological arguments play in biology reminds one of the endeavours, formulated in the correspondence argument, to take the quantum of action into account in a rational manner in atomic physics. (Bohr, 1933c, p. 458)

Bohr also develops the point he made in his letter to Jordan in June of 1931 concerning an in-principle barrier to reductionism, what David Favrholt has termed the 'thanatological' principle—namely the destruction of the coordination and interrelations essential to a biological system through a reductive-analytical approach (Favrholt, 1999a, p. 12). One must kill an organism in order to study it analytically.[10] But this barrier has to do with the conditions of living systems and the organization of life processes and not to any vitalistic properties.

Bohr's divergences from Jordan are revealed through a careful reading of 'Light and Life' against his June 1931 letter, even though there is no specific reference to Jordan contained in the published version of the lecture. Similar to the point he had discussed earlier with Jordan, but with greater emphasis, the issue is related to a correct understanding of psycho-physical parallelism, and cannot be viewed as an extension upward of quantum indeterminism at the micro-level that opens up a new domain for causes outside those encompassed by the 'complementary' relation of the biological and physical domains:

> I should like to emphasise that the considerations referred to here differ entirely from all attempts at viewing new possibilities for a direct spiritual influence on material phenomena in the limitation set for the causal mode of description in the analysis of atomic phenomena. For example, when it has been suggested that the will might have as its field of activity the regulation of certain atomic processes within the organism, for which on the atomic theory only probability calculations may be set up, we are dealing with a view that is incompatible with the interpretation of the psycho-physical parallelism here indicated. Indeed, from our point of view, the feeling of the freedom of the will must be considered as a trait peculiar to conscious life, the material parallel of which must be sought in organic functions, which permit neither a causal mechanical description nor a physical investigation sufficiently thoroughgoing for a well-defined application of the statistical laws of atomic mechanics. (Bohr, 1933c, pp. 458–459)

The sequence in which the *published* pronouncements appeared, however, considerably obscured these differences between Bohr and Jordan and introduced further complications. Jordan's controversial paper of November, 1932, entitled 'Quantum Mechanics and the Foundational Problem of Biology and Psychology,' was the first published outcome of the Bohr-Jordan conversations, appearing in the main German-language journal of general science, *Die Naturwissenschaften.* In this essay, Jordan put forth his novel claim that one could develop from the indeterminism of quantum physics the foundations for an argument for the freedom of the will and the autonomy of conscious phenomena (Jordan, 1932). In making this argument, Jordan relied once again on an amplifier analogy, in which the organism is conceived to be like a signal-multiplying

[8] There seems to be no manuscript version of this lecture in the Bohr archives. Max Delbrück later commented on how Bohr always talked without a manuscript or notes (Delbrück, 1981, MS, p. 29). The English text that was eventually published in March and April of 1933 (Bohr, 1933a, 1933c), bracketed the published German version of April (Bohr, 1933b). It then appeared in a revised form in Bohr (1934). I am using the original versions.

[9] This is to follow Folse and reject the 'two aspect' instrumentalist reading of Bohr. See Folse (1990).

[10] In his detailed analysis of Bohr's reductionism in this essay by Hoyningen-Huene (1994), Bohr is seen as an epistemological anti-reductionist, but an ontological reductionist because of his rejection of vitalism. This is not the position that would emerge from following the Kant-Høffding lineage I am developing. The resolution of the antinomy in terms of a unification in an ontological monism, described by alternative frameworks, would neither deny the possibility of a mechanistic analysis of living organisms, nor that of a teleological and non-mechanistic account. The important point is that these would be mutually- exclusive accounts of a single reality.

vacuum tube that receives a weak signal and then amplifies it to produce larger macroscopic effects (Beyler, 1994, 1996). Although Bohr had already explicitly rejected this analogy in his letter to Jordan of June, 1931, and implicitly in the August 1932 lecture, Jordan proceeds to exploit this similitude. Since the micro-events are causally indeterminate on Jordan's interpretation of the quantum theory, the amplification of this causal indeterminism underlies the phenomena of inner freedom, consciousness, and biological function, and in principle prevents their reduction to a deterministic and mechanistic explanation. In this 'inner zone' of freedom, there is also the basis for an affirmation of the unity and the purposive character of life.

Furthermore, Jordan essentially claimed the sanction of Bohr for these arguments. Nor was such an assumed endorsement immediately denied by Bohr. In his reply to the receipt of Jordan's reprint in late November, Bohr praises it as a 'beautiful article' (*schöner Artikel*), and expresses his 'great pleasure' in the review Jordan had also written in the *Zeitschrift für Physik* of Bohr's 1931 collection of essays, which included reprints of his 1929 and 1930 articles on quantum mechanics and psychology (Bohr to Jordan, 17 December, 1932 in Favrholdt, 1999b, p. 533). There are no substantial criticisms offered of anything Jordan had presented. Not surprisingly, many thereafter associated Bohr and Jordan together in the development of a common argument about irreducibility and a-causality in biology and psychology, derived from quantum physics. This presumed the autonomy of organic life, and supplied a foundation for the realistic teleological purposiveness of organisms. Even if one can see Bohr's implicit reservations about Jordan's program in the subsequent publications of 'Light and Life' in April of 1933 (Bohr, 1933b, 1933c), the points of difference were evidently not picked up by others at the time. Instead Bohr and Jordan became tightly linked.

6. Schrödinger's solution: drifting toward the teleomatic

Jordan's interpretation of the meaning of the new physics for biology, and his appeal to Bohr's own papers in support, forms the immediate context for Erwin Schrödinger's entry into these biophysical discussions. Schrödinger's concern with biophysics as early as 1932 has generally not been described in any detail, with the notable exception of Edward Yoxen's important work on the origins of Schrödinger's famous *What is Life?* lectures (Yoxen, 1978, 1979). Yoxen draws attention to three documents in the Schrödinger archive that are relevant to my topic. Two are letters from his acquaintance, the Austrian physicist Karl Przibram (1878–1973). The other is a notebook headed '*Warum*' that seems to be notes made in the fall of 1932 on background readings for a lecture on the relations of biology and physics which Schrödinger then delivered in early 1933 to the Prussian Academy of Science under the title 'Warum sind die Atome so Klein?,' prefiguring the theme of his opening 1943 Dublin lecture.

When Jordan's controversial paper appeared in early November of 1932, Schrödinger had already been independently exploring biological topics related to the causes of the Brownian motion of microbes and other unicellular organisms, and had written to Karl Przibram in September to discuss these issues. These inquiries led to the involvement in the conversation of Karl's brother Hans Leo Przibram (1874–1944), an experimental biologist, mentor of Paul Weiss and Paul Kammerer, and the Director of the biological research unit in Vienna (Coen, 2006). Following the publication of Jordan's paper in November, Schrödinger again wrote to Karl, apparently enclosing a reprint of Jordan's paper. The response by Karl reveals the sense of shock both brothers felt at Jordan's claims, noting that Hans was pleased to find in Schrödinger 'an ally in the war against all ''occult forces''.' and the two brothers called upon Schrödinger to respond publicly (letter of H. Przibram to Schrödinger, 15 November, 32, Schrödinger Correspondence, APS, Reel 37/11). Furthermore, Hans even volunteered to supply literature references and offered his professional advice as a biologist for such a response. The outcome of this exchange was an exploration by Schrödinger of a selection of papers on genetics that included note-taking on geneticist Herman J. Muller's important 1926 paper 'The Gene as the Basis of Life' (Muller, 1929) (Fig. 1).

This long paper, originally delivered at the International Congress on Genetics in Ithaca, New York in 1926, was Muller's most extended development of his arguments on the possible size of the material gene. It also was a vigorous defense of the concept of the gene as a unitary physical entity ordered linearly on the chromosome. This paper also contained Muller's most extreme statement of the program of strong genetic reductionism in which the material gene was conceived as the governing agency of life. As Muller argued:

> [...] the view that seems best to stand the tests of ultimate analysis, the great bulk, at least, of the protoplasm was, after all, only a by-product, originally, of the action of the gene material; its 'function' (its survival-value) lies only in its fostering the genes, and the primary secrets common to all life lie further back, in the gene material itself. (Muller, 1929, p. 918)

Schrödinger's close reading in late 1932 of Muller's essay in the context of the Jordan debates helps explain many issues in his later biophysical discussions of the 1940s. Although neither Bohr, Jordan, nor Muller are cited in the later *What is Life?* lectures, Muller's arguments of 1926 seem to be the source of Schrödinger's 'gene first' analysis of living phenomena in which the gene is put forth as a master molecule, the basis of heredity—the 'architect's plan and builder's craft—in one' as he would term it (Schrödinger, 2000, p. 22).

In the lecture to the Prussian Academy of Sciences on February 8 of 1933, Schrödinger directly engaged the issues that Jordan had raised the previous November, but did not mention him by name, at least in the published abstract.[11] Whereas Jordan denied, by appeal to quantum indeterminism, that there was a deterministic basis for the exact order-maintaining aspects of the organism through inheritance, Schrödinger argued to the contrary that the order and stability of the gene,

> is secured according to the law of quantum energy transfer, which Max Planck had developed a generation previously. The same quantum laws, which in recent times have led to strong doubts about whether the relation of individual atoms can be strictly causally determined, on the other hand furnishes the only foundation for the comprehension of the unprecedentedly precise lawfulness with which is disclosed with increasing clarity 'on the macroscopic [level]' [*im Großen*] not only for the physicist, but also for the biologist (Schrödinger, 1933, p. 126).

Rather than warranting the assumption of a-causal inner freedom in living beings, as Jordan was claiming, exactly the opposite conclusion could be drawn. In brief, this is the claim that would become one of the central themes of the later *What is Life?* lectures—quantum physics solves the order from order issues that might otherwise seem to require recourse to vitalism or to new and

[11] Since Bohr's published versions of his 'Light and Life' lecture had not yet appeared, and Schrödinger had not attended the August delivery, it is difficult to claim that this claim was in response to Bohr's lecture itself. However, Schrödinger's efforts to address the issue of atomic stability and the stability of organic systems over time through quantum mechanics, which Bohr's lecture had dealt with through his complementarity argument, would have made his opposition evident when 'Light and Life' was published (Bohr,1933a, p. 422).

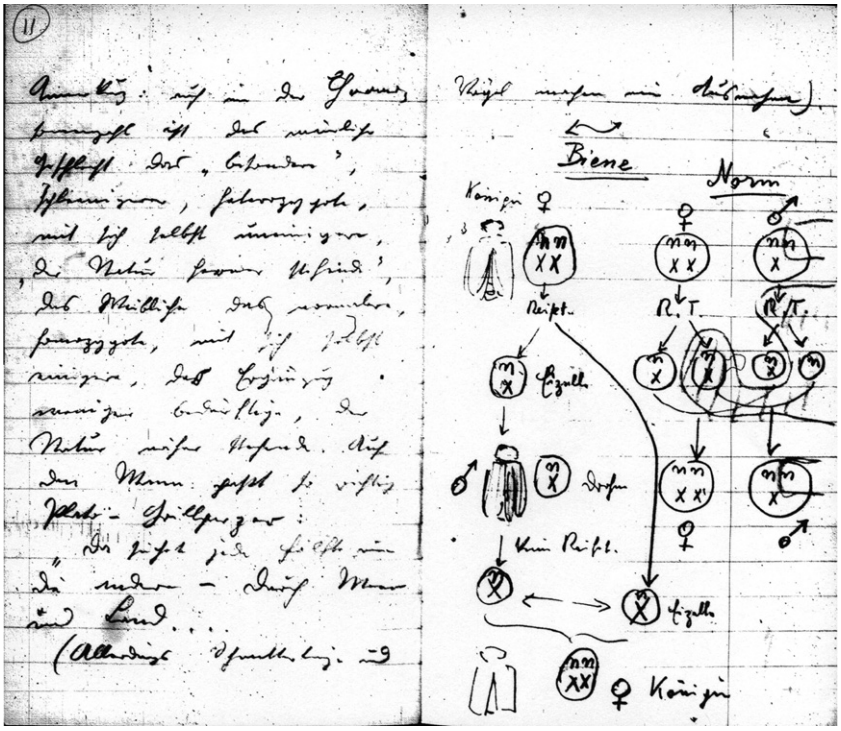

Fig. 1. A portion of Schrödinger's notes on Hermann Muller's 1929 paper in preparation for his 1933 lecture. (Source: E. Schrödinger, Notebook 'Warum,' Schrödinger Papers, Archives for the History of Quantum Physics, American Institute of Physics, M/f 43/Sect. 3. fol. 11. Used with permission of the AHQP and Ms. Ruth [Schrödinger] Braunziger.)

unknown laws in physics to deal with biology (McKaughan, 2005). Properly understood, contemporary physics grounds a causally deterministic interpretation of living processes. It explains stability, the transmission of order from order, and it also grounds the control of life by the gene.

Schrödinger's delivery of this lecture in February, and the printing of its brief abstract *prior to* the publication of Bohr's 'Light and Life' paper in March and April, presented an alternative view of the relation of organic life to physics to that defended by Jordan, and — so it seemed—, by Bohr. For Schrödinger, quantum physics supplies the basis for a deterministic account of life that explains apparent 'vital' properties of life from the 'bottom up' by the quantum interpretation of the chemical bond. Quantum theory grounds the discreteness of the gene and the strong discontinuity illustrated by mutation.[12] It also supplies a basis for the conception of the material gene as the main causal agency of life. Rather than supplying any basis for vital action or inner freedom, physics gives the basis for strong reductionism. It is to Schrödinger's reading of the contribution of the new physics to genetics to which Crick was evidently appealing in our opening quotation of this paper when he referred back to the developments in physics in the early decades of the century as securing the reductionist program. To apply categories borrowed from Mayr, Schrödinger is verging on Mayr's 'teleomatic' solution to organic teleology. All apparent teleological purposiveness would seem to be solely the product of deterministic causal forces that act from the bottom up. If his statements do indeed, at least by 1943, imply the notion of an inherent 'program' that verges on Mayr's concept of teleonomy, there is a determinism and a reduction of the holistic and multiple-pathway flexibility found in a properly teleonomic account in Mayr's sense that is lacking in Schrödinger's position. We see this difference by comparing Schrödinger's conclusions with the views of Max Delbrück two years later.

7. Transition to teleonomy: Delbrück and the three-man paper

Young Max Delbrück's entry into this conversation led it in yet another direction. As is well known, he was initially engaged with the issues of biophysics through conversations with Bohr, and he was in residence at Bohr's Institute on a Rockefeller Fellowship from late March to June of 1931, exactly in the time period when Bohr and Jordan were carrying on their conversations about the relations of quantum physics to biology. When he moved to study with Wolfgang Pauli in September of 1931, he recalled later being 'already [...] infected with a curiosity about the relation of physics to biology' (Delbrück, 1981 MS, p. 21).[13] His subsequent attendance at Bohr's 'Light and Life' lecture in August of 1932 is reported, in all his autobiographical recollections, to be the career transforming event that led him to accept a position as Lise Meitner's research assistant at the Kaiser-Wilhelm Institute for Chemistry in the fall of 1932, rather than returning to work with Pauli in Zurich as

[12] I will not attempt here to reconcile Schrödinger's strong claim in this paper, repeated in his *What is Life*? lectures, that quantum mechanics provides the basis for the stability of organic systems and the biophysical explanation of the discontinuity of mutation, with his more general wave-mechanical alternative to the Copenhagen interpretation of quantum mechanics expounded by Bohr and Heisenberg. His theory of mutation is developed on the basis of sharp discontinuity of energy levels between stable isomeric forms, a theory he also attributes to Delbrück. Hence he also gives strong endorsement to DeVries's mutation theory. For continued influence of Schrödinger's account of gene mutation, see Stamos (2001, p. 171). I thank Charles Pence for this reference.

[13] Delbrück interrupted his Göttingen training to do pre-doctoral work in physics at the University of Bristol with John Lennard-Jones from September 1929 to March 1931, with a return to Göttingen in December for his unsuccessful doctoral defense. He then received a Rockefeller Foundation fellowship for the calendar year of March 1931 to 1932, in which he spent the first six months with Bohr, and then the fall and winter of 1931-32 with Wolfgang Pauli in Zürich, followed by a return to Bristol from March to September of 1932. This was followed by his return to Berlin to work with Lise Meitner in September 1932. I have taken some of these details, which give more specificity than the general account in Fischer & Lipson (1988, pp. 50–61), from Delbrück's detailed curriculum vitae prepared for his admission to the Royal Society in 1967 (Delbruck Papers, Cal Tech, Box 40:3).

originally planned (Fischer & Lipson, 1988, p. 78 ff; Delbrück, MS. 1981, pp. 30–31; Harding, 1978, p. 42).

Delbrück's failure to connect with Bohr's complex, and admittedly opaque, views can be seen as we fast forward to Delbrück's contribution to the landmark paper on the nature of the gene and gene mutation that appeared in 1935, co-authored with Russian geneticist and evolution-theorist, Nikolai Timoféeff-Ressovsky (1900–1981), and radiation physicist Karl Günther Zimmer (1911–1988). Known in biological circles afterward as the 'Three Man Paper' (*Dreimännerwerk*), this co-authored paper marked Delbrück's first published entry into biology. Presented to the Göttingen Academy of Sciences in April of 1935 and then published in the short-lived *Nachrichten* of the Academy soon afterward, it appeared against the background of the Jordan-Bohr-Schrödinger debates (Timoféeff-Ressovsky, Zimmer, & Delbrück, 1935, hereafter *TZD*). Schrödinger would later draw upon this paper, and particularly on Delbrück's section of it, in support of his reductionist views of genetics in central chapters of *What is Life?* As a consequence, it has been primarily through Schrödinger's summary of the paper that it has been generally known and interpreted (Sloan and Fogel, 'Introduction,' forthcoming 2012; Perutz, 1987).

The origins of this paper, at least Delbrück's contribution to it, fall squarely into the context of the Bohr-Jordan debates. In the fall of 1934, Delbrück had assembled together at his family home in Grünewald near the Dahlem institutes of the Kaiser Wilhelm Gesellschaft a group of physicists and biologists to discuss theoretical issues in the relation of physics and biology, a meeting likely inspired by his enthusiasm for Bohr's views (Sloan, 2012, forthcoming, in Sloan & Fogel, 2012, forthcoming). Then in November of 1934, Jordan delivered a lecture at Dahlem, sponsored by the Society for Empirical Philosophy—the so-called 'Berlin Circle'—on the relation of physics and biology to a group of biologists and physicists attached to the K-W, evidently developing arguments similar to those he had published in his long *Erkenntnis* article of late summer 1934, in which he responded to the many criticisms of his 1932 *Die Naturwissenschaften* article and developed new arguments in support of his amplifier thesis (Jordan, 1934). Jordan's lecture proved to be deeply upsetting to the biologists in the audience, who jointly blamed Bohr and Jordan for introducing specious arguments into the literature on the relation of biology to physics (Delbrück to Bohr, November 30, 1934 in Favrholdt, 1999b, pp. 465–69).

We can see how Delbrück is diverging from Bohr in his summary statement of 'what we assert' (*was wir behaupten*) in a document appended to his letter to Bohr of November of 1934, reporting on Jordan's lecture. This was written to clarify for the Kaiser-Wilhelm biologist Max Hartmann the difference between the Bohr-Delbrück views and those of Jordan.

Delbrück opens his summary with an 'assertion' (*Behauptung*) which is elaborated in a set of four explanatory paragraphs. The assertion itself is:

> The assumptions having to do with the causal order of biological phenomena may in part stand in formal contradiction to the laws of physics and chemistry, because experiments on living organisms are *certainly* complementary to experiments establishing physical and chemical processes with *atomic* precision (Ibid, 468; italics denote underlined passages in original).

This claim is then elaborated as a set of statements about the completeness of the laws of atomic physics, and the inability to explain life on the basis of laws of physics and chemistry alone. He denies that he and Bohr clam that the 'laws of the atomic theory can explain *specific* life phenomena' (ibid., 468), while admitting that 'the laws of the *atomic theory* are the common root of *physics* and *chemistry*. Then as concerns the organism:

> Precisely *because* in a living organism physical and chemical phenomena are interwoven far into the atomic domain, the common root of biology *and* physics *and* chemistry *must* be found in the *atomic* domain. Just for this reason, however, a *causal* description of the relationship *cannot* be based on *physical and chemical concepts alone.* For in the atomic domain, *physics* and *chemistry* allow no common causal description. (Ibid, 469)[14]

As a consequence, he continues, 'for genetics *and* developmental mechanics *and* physiology *and* biochemistry *and* biophysics, it is characteristic and essential that they study processes in the *living* organism' (ibid.).

Although commenting in his reply that Delbrück's 'formulation is a very appropriate one (*Ihre Formulierung eine sehr zweckmässige ist*),' Bohr indicates some of the lines of disagreement that were already emerging between himself and his most illustrious protégé in biophysics. Observing that Delbrück has given 'not a comprehensive account of the viewpoints but only a correction of the misunderstandings that are unfortunately widespread among the biologists,' Bohr then comments that he has been spurred on by this exchange to work up for publication a short paper he wrote up in the summer of 1934 (Bohr to Delbruck, December 8 1934, Favrholdt, 1999b, p. 470). Although a published paper that directly matches this description does not seem to exist,[15] his short remarks on biophysics, delivered at the Galvani conference in October of 1937, seem to continue this conversation, and clarify some of the points of misunderstanding (Bohr, 1937, in Bohr 1987). In this paper, he rejects, again without mentioning him by name, Jordan's notion that we can find some 'direct correlation between life or free will and those features of atomic phenomena for the comprehension of which the frame of classical physics is obviously too narrow' (ibid., p. 20). Instead he emphasizes that the main implication of quantum mechanics for biology is the imposition of the requirement that 'the only way to reconcile the laws of physics with the concepts suited for a description of the phenomena of life is to examine the essential difference in the conditions of the observation of physical and biological phenomena' (ibid., p. 20). Similar to the claims of the 'Life and Light' lecture, 'the existence of life itself should be considered, both as regards its definition and observation, as a basic postulate of biology, not susceptible of further analysis' (ibid., p. 21).

Delbrück's response in the *TZD* paper to Jordan's 1934 lecture at the K-W, which preceded these remarks of Bohr in 1937, presented his own understanding of the relation of biology and physics, and in so doing moved in a direction that was neither that of Bohr, Schrödinger, nor of Jordan. These developments can be observed in sections of the *TZD* that seem to have been most clearly authored by Delbrück.

The *TZD* was a major theoretical synthesis that sought to incorporate the pre-existent work on radiation physiology and gene mutation theory, developing and deepening the earlier work of Friederich Dessauer, Richard Glocker, Hermann Muller, Fernand Holweck, F. B. Hanson and others who first explored the relations of radiation, cell damage, and mutation (Summers, 2012, forthcoming; Beyler, 2012, forthcoming; von Schwerin, 2004, pp. 119–36). It also went beyond these earlier explorations in its effort to synthesize this earlier medical and biological radiation research with some wider theoretical issues in physics.

[14] Emphasized statements are underlined in original MS.

[15] The first published paper on these issues was his lecture to the Second International Congress for the Unity of Science held in Copenhagen 21-26 June, 1936, and published as 'Causality and complementarity, Philosophy of Science 4 (1937): 289–98. In this he comments that his position 'stands far removed from every attempt to exploit in a spiritual sense the failure of causal description in atomic physics' (297).

To the first set of issues—the effects of radiation on genetic mutation—the *TZD* employed more exact methods of analysis than those employed by such pioneers of radiation genetics as Muller, utilizing the 'Target Theory' of radiation input developed in biological circles by Karl Zimmer on foundations laid by Dessauer and Holweck (Summers, 2012, forthcoming; Beyler, 2012, forthcoming). The more theoretical issues relating to physics were then developed largely by Delbrück in his sections of the paper. My attention will be on his section of this paper and the closing theoretical summary, attributed to all three authors, but from all signs written primarily by Delbrück.

Delbrück's discussion of radiation and genetic mutation engages subtly the theoretical issues we have previously examined in the context of the Jordan-Bohr-Schrödinger debates. In the background was Delbrück's own interpretation of Bohr, which he viewed, against Jordan, as the correct reading. This involved the claim that there was an in-principle barrier to a full reduction of biology to physics; but the nature of that barrier is essential to dissect out. As this is put in the concluding section of the *TZD*, repeating his claim of November of 1934: 'these domains are not causally reducible to one another, just as physics and chemistry are not causally reducible to one another.' (Timoféeff-Ressovsky et al., 1935, p. 469). This focus on causal reducibility does not, however, reflect what Bohr had defended as the teleological perspective within the context of biological complementarity. The views Delbrück presents are what Bohr would characterize subsequently, and possibly with reference to the issues of the *TZD*,[16] as those holding to 'an amplification of the effects of individual atomic processes.' This does not, in Bohr's view, achieve an explanation of the 'holistic and finalistic aspects of biological phenomena,' which 'can certainly not be immediately explained by the feature of individuality of atomic processes disclosed by the discovery of the quantum of action.' The only way 'to reconcile the laws of physics with the concepts suited for a description of the phenomena of life is to examine the essential difference in the conditions of the observation of physical and biological phenomena.' To achieve an understanding of biology, 'we are led to conceive the proper biological regularities as representing laws of nature complementary to those appropriate to the account of the properties of inanimate bodies' (Bohr 1937, in Bohr, 1987, pp. 20–21). The 'irreducibility' of biology again emerges not as a claim of an autonomous domain inaccessible to physical analysis, but rather something tied to the parallelism of descriptions of a single reality.

Delbrück's failure to understand exactly what Bohr was arguing is critically important. Bohr's positions are to be understood against Kant's resolution of the antinomy of teleological judgment along the lines of Höffding's empirical realism. The unification of the teleological and mechanical is indeed possible, with the teleological and mechanical accounts understood as descriptions of a single reality. But this is not an issue of *causality*. Rather it is a requirement that one hold both the teleological and mechanistic analyses together in a unity that is approached by alternative and mutually exclusive descriptions. This is not, however, a conclusion that can be gathered from Delbrück's conclusions.[17]

A careful reading of Delbrück's arguments in the *TZD* also allows us to see the nature of the differences that would separate Delbrück's interpretations from those extracted from the paper later by Schrödinger to support his reductionist positions. Delbrück *does* envision a strong relationship to hold between physics and biology, and he allows this to extend to genetics. He also makes none of the appeal to a-causality and indeterminism that Jordan had advocated as the 'secret' of life. Instead he proceeds to develop a causal explanation from theoretical physics for how the connection of mutation and radiation, developed by Timoféeff-Ressovsky and Zimmer in the first two sections of the paper, can be explained physically, and offers a theoretical model for how radiant energy can create a reversible rearrangement of the atomic structure of the material gene. It is this account that Schrödinger will later expound as 'the Delbrück Model of the Gene' in chapters four and five of *What is Life?* in support of the pre-existent reductive physicalism we have seen him express in 1933.

But Schrödinger's later reading was not the final conclusion of the *TZD*, as revealed clearly by the closing portions of the paper. Here there are two options posed, one which will essentially be that adopted by Schrödinger, and the other that put forth by the joint authors. The first option, probably pointing either to Muller or Schrödinger, reads as follows:

> According to the conception of many biologists, the genome is a highly complicated chemical-physical structure, consisting of a series of specific, chemical pieces of matter—the individual genes. Some attempts have been made to project back theoretically, by way of the hereditarily-modifiable, ontogenetic developmental sequences, from the organism to its individual genes. The genes are thus conceived as the immediate 'starting points' of the chains of reactions comprising the developmental processes. On the one hand, this conception requires that we assume a highly complicated structure and mode of operation for the genes, and that we deal with the gene problem from the standpoint of the requirements of developmental physiology. On the other hand, it leads to an explicit or implicit critique of [the] cell theory; the cell, thus far proving itself so magnificently as the unit of life, dissolves into the 'ultimate units of life,' the genes. (*TZD*, p. 240; as translated by Fogel in Sloan & Fogel, 2012, forthcoming, p. 270).

Read thus far, the argument supports a 'gene first' conception of biological order, with the 'ultimate units of life' the governors of the ontogenetic program. This does not imply any kind of goal-directedness of organic activity and verges on a complete genic reductionism. It seems clearly to be the view extracted from the paper by Schrödinger in *What is Life?*

But this is immediately followed with the second option which I read as primarily Delbrück's statement:

> Our ideas about the gene challenge this picture. Genes are physical-chemical units; perhaps the whole chromosome (to be sure the part containing genes) consists of such a unit, a large assemblage of atoms, with many individual, largely autonomous subgroups. Such genes are likely incapable of directly forming the morphogenic substances; they also can hardly be thought of as the 'starting points' of developmental sequences. Nevertheless, such a genome can be thought of as the foundation for specific, heredity-conditioned morphogenesis, by providing a steady, form- and function-determining framework for the cell . . .. Changes to its individual parts (gene mutation) would influence the overall functioning of the cell in specific ways and, thus, the individual development processes as well. Therefore, we need not dissolve the cell into genes, and the 'starting points' of the developmental sequences are not attributed to

[16] Bohr's comments are delivered after the September 27-29 1936 special meeting on the 'mechanism of mutation,' organized by Bohr in Copenhagen in the wake of the *TZD* to examine in more depth the relation of radiation, physics and genetics. This meeting was attended by Timoféeff-Ressovsky, Delbrück, and Herman Muller, and four other local biophysicists. I am indebted to Dr. Finn Aaserud and Felicity Pors of the Niels Bohr Archives in Copenhagen for information on this meeting (personal communication).

[17] As Daniel McKaughan has argued (McKaughan, 2012, forthcoming, and McKaughan, 2005), this does not preclude some important evolution in Delbrück's views over the years. My focus is limited to this initial mid-30s period. For alternative perspectives on Delbrück's 'complementarity,' see McKaughan (forthcoming), and Roll-Hansen (2012, forthcoming), both in Sloan & Fogel (2012, forthcoming).

> individual genes, but rather to operations of the cell, or even to intercellular processes (which are all eventually controlled by the genome) [*die alle letzten Endes vom Genom kontrolliert werden*]. (Timoféeff et al., 1935, p. 241, trans. Fogel in Sloan & Fogel, 2012, forthcoming, p. 270)

The concluding sentence requires careful analysis. It is the higher-level functions of the cell, and not the individual genes, that control development. There is no 'master molecule' concept in play here. But at the same time, it does not involve anything close to Bohr's own meaning of biological complementarity.[18] There is indeed no deeper teleological purposiveness of organic life implied, nor is there any notion of a union of the teleological and mechanical in the empirical order. The perspective is properly 'teleonomic,' to apply Mayr's distinctions, rather than 'teleological.' The apparent teleological aspects of living organisms are due to the action of an underlying genetic program. As Mayr argues, such a view involves a combined explanation of the directiveness of organic life through appeal to a pre-existent genome along with a natural selectionist account of origins. The authors of this final summary of the *TZD* employ both.[19] The genome as a whole, rather than an inner vital agency, or the atomic 'gene,' is behind living function. And this is the product of natural selection: As Delbrück writes in his own section of the paper:

> We have presented genes as well-defined molecules that do not generally change over the course of the development of individuals or of a population. This *stability* must have come about in some way through the conditions under which life evolves, where natural selection has surely played a decisive role as the controlling factor in the selection of especially stable formations. At the same time, we must expect that selection has driven this stability only so far as to exclude changes that emerge with appreciable frequency. There must remain, then, some rearrangements whose frequency is low relative to lifespan. We detect these in wild strains as mutations. (Ibid., p. 261)

Delbrück's maiden voyage into the world of biophysics begins from this point. If Delbrück does indeed recognize and explore the importance of organismal issues, as emphasized by McKaughan (2012, forthcoming), nonetheless, it is difficult to argue that he allows for a genuine teleological realism.

8. Conclusion

By situating this discussion of the elimination of teleology in a specific and local framework in the history of 1930s biophysics, my concern has been to illuminate contextually the issues in the controversy surrounding Bohr's philosophy of biology, and display how three alternative interpretations of Bohr's reflections on the relations of the biological and physical generated competing readings by physicists moving into biology who all rejected traditional vitalism, and who all sought to ground their arguments naturalistically in contemporary quantum physics.

The resultant elimination of a realistic view of teleological reasoning, and its replacement by teleonomic and teleomatic alternatives, does not actually address the root problem highlighted by Bohr. A 'critical' understanding of teleological purposiveness, whether read through the transcendental Idealism of Kant, or realistically, as Høffding and Bohr seem to allow, cannot be dissolved by the success of the reductionist program in biology. As argued by Matthew Ratcliffe (Ratcliffe, 2001) with reference to Kant's approach to biology, a teleological view of organisms forms the *framework within which* reductive biological analysis takes place. This is not simply old time ontological antireductionism in new clothes.[20] It is an epistemological precondition of life science. Because of the way Bohr's complex philosophical position was interpreted, and as I have argued, misinterpreted in part because of Bohr's own often confusing and unsystematic presentations, the Schrödinger interpretation of the relation of biology and physics was that which seems to have been embraced by several of the theoretical architects of molecular biology, at least those who reflected on these questions. Delbrück to be sure continued to reject Schrödinger's reading (McKaughan, 2012, forthcoming),[21] but his own philosophical program of 'complementarity' seems to have had few real disciples and in the end might seem to collapse into Schrödinger's view of the matter. The heritage of this discussion is still with us in contemporary debates about indeterminism, chance, and causation in natural selection theory (Brandon & Carson, 1996; Graves, Horan, & Rosenberg, 1999; Stamos, 2001).

To conclude on a philosophical note, I suggest that with the historical options of early molecular biology illuminated by these concrete historical examples, we can once again raise the issue of teleological realism. My suggestion is that a fruitful return can be made to some key aspects of Bohr's original argument as a way to clarify the role of the interrelations of the biological and the physical. With all the deficiencies in Bohr's actual formulations admitted, his arguments have drawn at least three things to our attention. First, the issue of the teleological purposiveness of organisms needs not be a question of causation in the way this is usually assumed—i.e. as a matter of backward causation or one involving special vital forces in matter or new laws of physics. Second, it does not mean a denial of the possibility of reductive biological explanations. Third, it is not an argument that attempts to ground the dynamism and purposive character of life on some kind of extension of quantum causality up the chain of organization. What I find of interest in Bohr's approach to these questions is that the issue is placed at the interface of epistemology and ontology rather than as a question of causation: the givenness of the teleological aspects of living things is a precondition for doing biological science as Kant saw, and furthermore, this can mean more than a subjective projection of human intentionality on a mechanistic universe as one might—incorrectly I feel—read Kant as arguing. Instead it presses us toward the recognition of a genuine union of the teleological and mechanical view of the organism that can, however, only be accessed non-simultaneously. To employ Michael Polanyi's useful distinction of the two forms of awareness—focal and subsidiary—the teleological account is rendered 'subsidiary' in analytic biological science when the explanation of organic process by efficient and material causation is 'focal' (Polanyi, 1964, chap. 4). But to pursue Polanyi's distinctions, it is also possible to shift focal attention in our biological inquiry to the teleological and purposive dimensions of life in order to deal with a wide range of other human and theoretical interests, in which the material and efficient causes of life become only subsidiary conditions underlying a robust experience of the organism. Both insights apply validly to a single reality, but in ways that cannot be held at the same time. Although more argument would be needed than can be given here

[18] Delbrück explicitly tells Bohr in sending him a reprint that the work 'contains no complementarity argument whatsoever' (*Die Arbeit enthalt keinerlei Komplementaritätsargumente*). Delbrück to Bohr, 5 April, 1935 (Favrholdt, 1999b, p. 471).

[19] Timoféeff-Ressovsky, in addition to his work on radiation genetics, was also a major theorist in the Russian school of population genetics. These comments likely represent his contribution to the discussion.

[20] Bohr himself later seemed to weaken his commitment to teleological realism in the face of the success of the Watson-Crick work. See McKaughan (2012, forthcoming.)

[21] Delbrück's review of Schrödinger's *What is Life*? displays some of this divergence. There he writes: 'Physicists and biologists who are not familiar with Bohr's subtle complementarity argument will be inclined to take the physical nature of the cellular processes for granted at the outset.' (Delbrück, 1945, p. 370).

to develop this point, this view bears some similarities to the view I see Bohr attempting to articulate with his notion of 'complementarity' in biology. We can, in this respect, acknowledge a realistic stance toward the teleological purposiveness of organisms without lapsing into vitalism of some traditional form.

Acknowledgements

Developed from a paper of a similar title delivered at the biennial meeting of the International Society for the History, Philosophy, and Social Studies of Biology in Brisbane, Australia, July 2009. Research for this paper was directly funded by National Science Foundation Grant #0646732 as part of a larger project. Travel support was through a grant from the Institute for Scholarship in the Liberal Arts of the University of Notre Dame. I wish to thank James Barham, Lenny Moss, Daniel McKaughan, Nils Roll-Hansen, and Charles Pence for valuable comments on earlier drafts. Crucial bibliographical assistance was rendered by James Barham, with some help from Charles Pence. I wish also to thank the archival staffs of the American Philosophical Library, Philadelphia, the Library of the California Institute of Technology, the Regenstein Library of the University of Chicago, Finn Aaserud of the Niels Bohr archives in Copenhagen, and Melanie Brown of the Niels Bohr Library and Archives at the American Institute of Physics for valuable assistance and for supplying necessary archival materials.

References

Aaserud, F. (1990). *Redirecting science: Niels Bohr, philanthropy, and the rise of nuclear physics*. Cambridge: Cambridge University Press.
Allison, H. E. (2003). Kant's antinomy of teleological judgment. In P. Guyer (Ed.), *Kant's critique of the power of judgment: Critical essays* (pp. 219–236). Lanham, MD: Rowman & Littlefield.
Beyler, R. H. (1994). *From positivism to organicism: Pascual Jordan's interpretations of modern physics in cultural context*. Unpublished Ph.D. dissertation, Department of the History of Science, Harvard University.
Beyler, R. H. (1996). Targeting the organism: The scientific and cultural context of Pascual Jordan's quantum biology, 1932–1947. *Isis, 87*, 248–273.
Beyler, R. H. (2012, forthcoming). Exhuming the three-man paper: Target-theoretical research in the 1930s and 1940s. In P. R. Sloan & B. F. Fogel (Eds.) (2012 forthcoming), infra.
Bohr, N. (1929). Wirkungsquantum und Naturbeschreibung. *Die Naturwissenschaften, 17*, 483–486.
Bohr, N. (1930). Die Atomtheorie und die Prinzipien der Naturbeschreibung. *Die Naturwissenschaften, 18*, 73–78.
Bohr, N. (1931). *Atomtheorie und Naturbeschreibung: Vier Aufsätze*. Berlin: Springer.
Bohr, N. (1932). Høffding's attitude towards physical science and its relationship to psychology. Niels Bohr Scientific Manuscripts, 1934–1934 Archives for the History of Quantum Physics, American Institute of Physics, MSS Mf no. 13. Quoted with Permission of the Niels Bohr Archives, Copenhagen.
Bohr, N. (1933a). Light and life I. *Nature, 131*, 421–423.
Bohr, N. (1933b). Licht und Leben. *Die Naturwissenschaften, 21*, 245–250.
Bohr, N. (1933c). Light and life II. *Nature, 131*, 457–459.
Bohr, N. (1934). *Atomic theory and the description of nature*. Cambridge: Cambridge University Press.
Bohr, N. (1937). Biology and atomic physics. In N. Bohr (Ed.), *Essays 1932-1957 on atomic physics and human knowledge. The philosophical writings of Niels Bohr* (Vol. 2, pp. 13–22). New York: Wiley. Reprinted Woodbridge, Conn: Oxbow (First published 1958).
Bohr, N. (1987). *Essays 1932–1957 on atomic physics and human knowledge. The philosophical writings of Niels Bohr* (Vol. 2). New York: Wiley. Reprinted Woodbridge, Conn: Oxbow (First published 1958).
Brandon, R. N., & Carson, S. (1996). The indeterministic character of evolutionary theory: No 'hidden variables proof' but no room for determinism either. *Philosophy of Science, 63*, 315–337.
Breitenbach, A. (2008). Two views on nature: a solution to Kant's antinomy of mechanism and teleology. *British Journal for the History of Philosophy, 16*, 351–369.
Christiansen, F. V. (2006). Henrich Hertz's neo-Kantian philosophy of science, and its development by Harald Høffding. *Journal for General Philosophy of Science, 37*, 1–20.
Coen, D. (2006). Living precisely in fin-de-siècle Vienna. *Journal of the History of Biology, 39*, 493–523.
Crick, F. (1966). *Of molecules and men*. Seattle: University of Washington Press.
Delbrück, M. (1945). What is life? and what is truth? *Quarterly Review of Biology, 20*, 370–372.
Delbruck, M. (1967?n.d.). Curriculum vitae prepared for admission to the Royal Society of London. Delbrück papers, California Institute of Technology, Box 40:3.
Delbrück, M. (1981 MS). The arrow of time. Unpublished autobiography, dictated January-March, 1981. Delbrück papers, California Institute of Technology, Box 45:7. Quoted with Permission of the Archives of the California Institute of Technology.
Domondon, A. T. (2006). Bringing physics to bear on the phenomenon of life: The divergent positions of Bohr, Delbrück, and Schrödinger. *Studies in History and Philosophy of Biological and Biomedical Sciences, 37*, 433–458.
Favrholdt, D. (1999a). General introduction: Complementarity beyond physics. In D. Favrholdt (Ed.). *Niels Bohr: Collected works* (Vol. 10, pp. xxiii–xlix). Amsterdam: Elsevier.
Favrholdt, D. (1999b). Complementarity beyond physics (1928–1962). In F. Aaserud (Ed.). *Neils Bohr: Collected works* (Vol. 10). Amsterdam: Elsevier.
Faye, J. (1979). The influence of Harald Høffding's philosophy on Niels Bohr's interpretation of quantum mechanics. *Danish Yearbook of Philosophy, 16*, 37–72.
Faye, J. (2008). Copenhagen interpretation of quantum mechanics. *Stanford on-line encyclopedia of philosophy*. Published May, 2005; rev. Jan. 2008. <http://plato.stanford.edu/entries/qm-copenhagen/>. (Accessed September 3, 2010).
Fischer, E. P., & Lipson, C. (1988). *Thinking about science: Max Delbrück and the origins of molecular biology*. New York & London: Norton.
Folse, H. J. (1985). *The philosophy of Niels Bohr*. Amsterdam: North Holland.
Folse, H. J. (1990). Complementarity and the description of nature in biological science. *Biology and Philosophy, 5*, 221–224.
Ginsborg, H. (2006). Kant's biological teleology and its philosophical significance. In Graham Bird (Ed.), *A companion to Kant* (pp. 455–469). Malden, MA: Blackwell Publishing.
Graves, L., Horan, B. L., & Rosenberg, A. (1999). Is indeterminism the source of the statistical character of evolutionary theory? *Philosophy of Science, 66*, 140–157.
Guyer, P. (Ed.). (2003). *Kant's critique of the power of judgment: Critical essays*. Lanham, MD: Rowman & Littlefield.
Guyer, P. (2005). *Kant's system of nature and freedom*. Oxford: Clarendon.
Harding, C. Interview with Max Delbrück, July-September 1978. <http://oralhistories.library.caltech.edu/16/01/OH_Delbruck_M.pdf>. (Accessed July 2007).
Heisenberg, W., Born, M., & Jordan, P. (1925). Zur Quantenmechanik. *Zeitschrift für Physik, 34*, 858–888. 35, 557–615.
Høffding, H. (1902). Philosophy and life. *International Journal of Ethics, 12*, 137–151.
Høffding, H. (1905). Analogy and its philosophical importance. *Mind, 14*, 199–209.
Høffding, H. (1926). *Erkenntnistheorie und Lebensauffassung*. Leipzig: Reisland.
Howard, D. (1994). What makes a classical concept classical? toward a reconstruction of Niels Bohr's philosophy of physics. In J. Faye & H. Folse (Eds.). *Niels Bohr and contemporary philosophy (Boston studies in the philosophy of science)* (Vol. 158, pp. 201–229). Dordrecht: Kluwer Academic Publisher.
Hoyningen-Huene, P. (1994). Niels Bohr's argument for the irreducibility of biology to physics. In J. Faye & H. Folse (Eds.). *Niels Bohr and contemporary philosophy (Boston studies in the philosophy of science)* (Vol. 158, pp. 231–255). Dordrecht: Kluwer Academic Publisher.
Jordan, P. (1932). Die Quantenmechanik und die Grundprobleme der Biologie und Psychologie. *Die Naturwissenschaften, 20*, 815–821.
Jordan, P. (1934). Quantenphysikalische Bemerkungen zur Biologie und Psychologie. *Erkenntnis, 4*, 215–252.
Kaiser, D. (1992). More roots of complementarity: Kantian aspects and influences. *Studies in History and Philosophy of Science A, 23*, 213–239.
Kant, I. (1923). *Kritik der Urteilskraft*, (1793). In *Koeniglich preussisichen Akademie der Wissenschaften* (Eds.). *Kants gesammelte Schriften* 25 vols. (2nd ed., Vol. 5, pp. 167-485). Berlin: de Gruyter.
Kant, I. (1987). *Critique of judgment*. Indianapolis: Hackett (W. Pluhar, Trans.).
Kay, L. (1985). The secret of life: Niels Bohr's influence on the biology program of Max Delbrück. *Rivista di storia della scienza, 2*, 487–510.
Lenoir, T. (1982). *The strategy of life: teleology and mechanics in nineteenth-century German biology*. Dordrecht and Boston: Reidel.
Mayr, E. (1974). Teleological and teleonomic, a new analysis. In R. S. Cohen & M. Wartofsky (Eds.), *Methodological and historical essays in the natural and social sciences* (pp. 91–117). Dordrecht and Boston: Reidel.
Mayr, E. (1992). The idea of teleology. *Journal of the History of Ideas, 53*, 117–135.
McKaughan, D. (2005). The influence of Niels Bohr on Max Delbrück: revisiting the hopes inspired by 'light and life'. *Isis, 96*, 507–529.
McKaughan, D. (forthcoming). Was Delbrück really a reductionist? In P. R. Sloan & B. F. Fogel (2012 forthcoming) infra.
McLaughlin, P. (1990). *Kant's critique of teleology in biological explanation: antinomy and teleology*. Mellon: Lewiston.
McLaughlin, P. (2001). *What functions explain: Functional explanation and self-reproducing systems*. Cambridge: Cambridge University Press.
Muller, H. J. (1929). The gene as the basis of life. In B. M. Duggar, (Ed.) , *Proceedings of the international congress of plant sciences, Ithaca, August 16-23, 1926* (pp. 897–921). Menasha, Wisc.: Banta.
Nagel, E. (1977). Teleology revisited: goal-directed processes in biology. *Journal of Philosophy, 74*, 261–279.
Niels Bohr Manuscripts, Archives for the History of Quantum Physics, American Institute of Physics (Microfilm).
Olby, R. (1971). Schrödinger's problem: what is life? *Journal of the History of Biology, 4*, 119–148.
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.

Polanyi, M. (1964). *Personal knowledge: Towards a post-critical philosophy (rev. ed.)*. New York: Harper (First published 1958).
Pross, A. (2005). On the chemical nature and origin of teleonomy. *Origins of Life and Evolution of Biospheres, 35*, 383–394.
Przibram, K. Letter to Erwin Schrödinger, 15 November,1932. Schrödinger Correspondence, Archives for the History of Quantum Physics, American Institute of Physics, 37/11 (microfilm). Cited with Permission of the AHQP and Ms. Ruth [Schrödinger] Braunziger.
Quarfood, M. (2004). *Transcendental idealism and the organism: Essays on Kant*. Stockholm: Alquist & Wiksell.
Quarfood, M. (2006). Kant on biological teleology: Towards a two-level interpretation. *Studies in History and Philosophy of Biological and Biomedical Sciences, 37*, 735–747.
Ratcliffe, M. (2001). A Kantian stance on the intentional stance. *Biology and Philosophy, 16*, 29–52.
Roll-Hansen, N. (1976). Critical teleology: Immanuel Kant and Claude Bernard on the limitations of experimental biology. *Journal of the History of Biology, 9*, 59–91.
Roll-Hansen, N. (2000). The application of complementarity to biology: From Niels Bohr to Max Delbrück. *Historical Studies in the Physical and Biological Sciences, 30*, 417–442.
Roll-Hansen, N. (2012 forthcoming). Niels Bohr and Max Delbrück; balancing autonomy and reductionism in biology. In P. R. Sloan & B. F. Fogel (Eds.) (2012 forthcoming), infra.
Schrödinger, E. (1933). Warum sind die Atome so klein? *Forschungen und Fortschritte, 9*, 125–126.
Schrödinger, E. (2000). *What is life? with mind and matter and autobiographical sketches*. Cambridge: Cambridge University Press (First published 1944).
Schrödinger Correspondence, Archives for the History of Quantum Physics, American Institute of Physics (Microfilm).
Schwerin, A. v. (2004). *Experimentalisierung des Menschen: der Genetiker Hans Nachtsheim und die vergleichende Erbpathologie 1920–1945*. Göttingen: Wallstein.
Sloan, P. R. (2012, forthcoming). Biophysics in Berlin: The Delbrück club. In Sloan & Fogel (2012 forthcoming), supra.
Sloan. P. R., & Fogel, B. F. (Eds.). (2012, forthcoming.). *Creating a biophysics of life: the three-man paper and early molecular biology*. In Press, University of Chicago Press.
Stamos, D. (2001). Quantum indeterminism and evolutionary biology. *Philosophy of Science, 68*, 164–184.
Summers, W. C. (2012 exp). Physics and genes: from Einstein to Delbrück. In Sloan & Fogel (2012 forthcoming), supra.
Timoféeff-Ressovsky, N. Zimmer, K. & Delbrück, M. (1935). Über die Natur der Genmutation und der Genstruktur. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, mathematisch-physikalische Klasse, Fachgruppe VI: Biologie 1*, 189–245. Full translation by Brandon Fogel to appear in Sloan and Fogel, 2012 forthcoming, supra.
Watkins, E. (2009). The antinomy of teleological judgment. In D. H. Heidemann (Ed.), *Kant yearbook, Vol. 1: teleology* (pp. 197–221). Berlin: Walter de Gruyter.
Yoxen, E. J. (1978). The social impact of molecular biology. Unpublished Ph.D. dissertation, Department of History and Philosophy of Science, Cambridge University.
Yoxen, E. J. (1979). Where does Schrödinger's 'What is life?' belong in the history of molecular biology? *History of Science, 17*, 17–52.
Zammito, J. (2006). Teleology then and now: the question of Kant's relevance for contemporary controversies over function in biology. *Studies in History and Philosophy of Biological and Biomedical Sciences, 37*, 748–770.
Zammito, J. (2009). Kant's notion of intrinsic purposiveness in the critique of judgment. In D. H. Heidemann (Ed.), *Kant yearbook, Vol. 1: teleology* (pp. 223–247). Berlin: Walter de Gruyter.

Studies in History and Philosophy of Biological and Biomedical Sciences 43 (2012) 152–163

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

The concept of mechanism in biology

Daniel J. Nicholson

Konrad Lorenz Institute for Evolution and Cognition Research, Adolf Lorenz Gasse, 2, Altenberg A-3422, Austria

ARTICLE INFO

Article history:
Available online 2 July 2011

Keywords:
Mechanism
Mechanicism
Machine
Causal explanation
Function
Organization

ABSTRACT

The concept of mechanism in biology has three distinct meanings. It may refer to a philosophical thesis about the nature of life and biology ('*mechanicism*'), to the internal workings of a machine-like structure ('*machine mechanism*'), or to the causal explanation of a particular phenomenon ('*causal mechanism*'). In this paper I trace the conceptual evolution of 'mechanism' in the history of biology, and I examine how the three meanings of this term have come to be featured in the philosophy of biology, situating the new 'mechanismic program' in this context. I argue that the leading advocates of the mechanismic program (i.e., Craver, Darden, Bechtel, etc.) inadvertently conflate the different senses of 'mechanism'. Specifically, they all inappropriately endow causal mechanisms with the ontic status of machine mechanisms, and this invariably results in problematic accounts of the role played by mechanism-talk in scientific practice. I suggest that for effective analyses of the concept of mechanism, causal mechanisms need to be distinguished from machine mechanisms, and the new mechanismic program in the philosophy of biology needs to be demarcated from the traditional concerns of mechanistic biology.

> 'Biological Mechanism is committed logically to a great deal more than is commonly supposed.' (Broad, 1925, pp. 91–92)

1. Introduction

The concept of mechanism has recently received a great deal of attention in the philosophy of science. The main catalyst for this new interest has been the realization that scientists, especially biologists, often refer to mechanisms in their inquiries into the phenomena they investigate. This has led to the development of a lively philosophical research program over the past decade that has attempted to make sense of scientists' 'mechanism-talk' and elucidate the role it plays in scientific practice. The standard philosophical strategy has been to begin by offering a general characterization of 'mechanism' that captures the way scientists use this word, and then show the ways in which mechanisms are involved in the explanation of phenomena. The mechanism account that has exerted the greatest influence in the development of this new discourse has been formulated by Machamer, Darden, and Craver (2000). Machamer et al. (MDC, hereafter) conceive mechanisms as 'entities and activities organized such that they are productive of regular changes from start or set-up conditions to finish or termination conditions' (MDC, 2000, p. 3). Glennan (2002) and Bechtel (2006) have also developed their own mechanism accounts. Glennan defines a mechanism for a behaviour as 'a complex system that produces that behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations' (Glennan, 2002, p. S344), whereas Bechtel characterizes a mechanism as 'a structure performing a function in virtue of its component parts, component operations, and their organization', adding that 'The orchestrated functioning of the mechanism is responsible for one or more phenomena' (Bechtel, 2006, p. 26).

This emerging mechanism movement aims to provide a new framework in which to tackle a number of classic problems in the philosophy of science. Central among them is the nature of explanation, in which a focus on mechanisms is deemed to constitute an effective antidote to the outmoded deductive-nomological conception of explanation inherited from logical empiricism. In addition, recent literature in the philosophy of science includes mechanism-based accounts of causation (Machamer, 2004),

E-mail address: dan.j.nicholson@gmail.com

1369-8486/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsc.2011.05.014

reduction (Craver, 2005; Darden, 2005), models (Craver, 2006; Darden, 2007; Glennan, 2005a), and reasoning in discovery (Bechtel, 2009; Darden, 2006). Moreover, it has been suggested that thinking about mechanisms may help resolve the problem of underdetermination (Glennan, 2005a, pp. 458–459), as well as render unnecessary discussions of laws (Glennan, 2002, p. S348; MDC, 2000, pp. 7–8) and theories (MDC, 2000, pp. 16–17). Nevertheless, despite the general applicability of mechanism-based philosophy of science, it is interesting to note that this research program has developed primarily within the philosophy of *biology*. Indeed, the most prominent defences and extensive elaborations of the mechanism approach have been advanced by philosophers interested in the life sciences, with book-length mechanism accounts now existing for several biological subdisciplines, including cell biology (Bechtel, 2006), molecular biology (Darden, 2006), and neuroscience (Craver, 2007). This partnership between mechanism-based philosophy and biology is no mere happenstance. In fact, I will show that attending to the role the concept of mechanism has played in the development of biological thought opens up a rich new perspective in which to effectively examine and critically evaluate the recent mechanism discourse.

In a nutshell, what a historically informed perspective reveals is that the term 'mechanism' has come to be used in biology in a number of different senses. As the new mechanism discourse proceeds with an almost complete disregard for how the concept of mechanism has been shaped by the history of its usage, current discussions frequently suffer from the inadvertent conflation of the different meanings of the term. Admittedly, philosophers are generally aware that 'mechanism' is a convoluted concept with a long history, as evidenced by MDC's assertion that 'What counts as a mechanism in science has developed over time and presumably will continue to do so' (MDC, 2000, p. 2). However, most of them deem the potential for semantic confusion minimal because they consider the various meanings of the concept to be neatly associated with discrete, non-overlapping historical periods. Craver (2007, p. 3), for instance, remarks: 'But what is a mechanism? History cannot answer this question. The term mechanism has been used in too many different ways, and most of those uses no longer have any application in biology'. This paper will demonstrate, in opposition to this claim, how an awareness of the semantic breadth of the concept of mechanism afforded by an examination of its history can help uncover a number of important tensions within the new mechanism discourse, as well as provide the necessary philosophical resources for resolving them.

I begin by distinguishing and characterizing the three meanings of the concept of mechanism in biology (Section 2). I then explore the way in which the different senses of 'mechanism' have been used in the history of biology (Section 3), and how they have come to be featured in the philosophical literature, situating the new mechanism discourse in this context (Section 4). Following this, I illustrate the various problems that arise in recent discussions from the inadvertent conflation of the different senses of 'mechanism' (Section 5). Finally, I show what amendments need to be made to current accounts of mechanism to effectively capture the way this concept is used by biologists in their research (Section 6).

2. The three meanings of 'mechanism' in biology

The term 'mechanism' is used to mean different things in different contexts. In biology, 'mechanism' has three distinct meanings, which can be distinguished and defined as follows:

(a) *Mechanicism*: The philosophical thesis that conceives living organisms as machines that can be completely explained in terms of the structure and interactions of their component parts.

(b) *Machine mechanism*: The internal workings of a machine-like structure.

(c) *Causal mechanism*: A step-by-step explanation of the mode of operation of a causal process that gives rise to a phenomenon of interest.

As this taxonomy illustrates, 'mechanism' may refer to (a) a philosophical thesis about the nature of life and biology, (b) the workings of a machine, and (c) a particular mode of explanation. In order to make the ensuing discussion as clear as possible, I will refrain from using the word 'mechanism' in favour of these three terms, employing it only when referring to the word itself and not to any of its meanings. Let us now examine each of the three senses of 'mechanism' in more detail.

Mechanicism (often called *mechanistic philosophy* or *mechanical philosophy*) has its roots in the natural philosophy that emerged from the work and ideas of Galileo Galilei, René Descartes, Pierre Gassendi, Robert Boyle, Isaac Newton and others during the Scientific Revolution. This philosophy is usually associated with a naturalistic, atomistic, and deterministic view of nature that tends to lend itself to mathematical characterization. However, biological mechanicism, or *mechanistic biology*, has a rather more specific meaning (cf. Allen, 2005; Bertalanffy, 1952; Broad, 1925; Dupré, 2007; Haldane, 1929; Lewontin, 2000; Loeb, 1912; Monod, 1977; Rosen, 1991; Woodger, 1929). It can be characterized in terms of the following key tenets:

1. The commitment to an ontological continuity between the living and the nonliving, exemplified by the quintessential mechanistic conception of organisms as machines, analogous and comparable to man-made artefacts
2. The view that biological wholes (i.e., organisms) are directly determined by the activities and interactions of their component parts, and that consequently all properties of organisms can be characterized from the bottom up in increasing levels of complexity
3. The focus on the efficient and material causes of organisms, and the unequivocal repudiation of final causes in biological explanation
4. The commitment to reductionism in the investigation and explanation of living systems

Mechanicism has been one of the most influential schools of biological thought since the late seventeenth century. It has its origins in the physiological writings of Descartes, though the doctrine has had numerous incarnations through the centuries. Some of the most illustrious biologists of the past three hundred and fifty years have developed their ideas within a mechanistic framework. Famous mechanistic biologists include Giovanni Borelli, Stephen Hales, Antoine Lavoisier, François Magendie, Emil du Bois-Reymond, Hermann von Hemholtz, Carl Ludwig, Wilhelm Roux, and Jacques Loeb. In modern times, the astounding successes of molecular biology have served to consolidate mechanicism as one of the central philosophies of life and biology. Most recently, the emerging field of synthetic biology, with its aim to apply engineering principles in order to design and manufacture living cells from scratch, constitutes the newest expression of the mechanistic research program in biology.

The *machine mechanism* sense of 'mechanism' is the closest to the etymological roots of the word, which can be traced to the Latin *machina* and the Greek *mechane*, terms meaning 'machine' or 'mechanical contrivance'. The notion of machine mechanism has traditionally been employed by biologists to describe machine-like systems, or rather, systems conceived in mechanical terms; that is, as stable assemblies of interacting parts arranged in such a way that their combined operation results in predetermined outcomes.

Since the time of Descartes, mechanistic biologists have conceived organisms in explicit analogy with the paradigmatic machine mechanism of the age, be it a seventeenth-century clock with its finely-tuned parts operating as a functionally-integrated whole, an eighteenth-century steam-engine consuming chemically-bound energy by combustion and performing work whilst releasing heat, or a twentieth-century computer with its inbuilt program capable of processing information about the environment and responding accordingly. Machine mechanisms, biological and technological, can be studied in isolation and are often decomposable into smaller machine mechanisms.

The *causal mechanism* sense of 'mechanism', in contrast to the first two, only acquired widespread currency in biology in the twentieth century, though it is the usage of the term that has become predominant today. Causal mechanisms are of fundamental importance in scientific practice because they enable the identification of causal relations. To inquire about the causal mechanism of P (where P is the phenomenon of interest) is to inquire about the causes that explain how P is brought about.[1] Although the majority of philosophers conceive causal mechanisms as real things in the world (akin to machine mechanisms), I will be arguing in this paper that they are actually better understood as heuristic models which target specific causal relations and thereby facilitate the explanation of the particular phenomena scientists investigate.

I am not, of course, the first to propose that the concept of mechanism needs to be terminologically fragmented to reflect its semantic breadth. In fact, the word 'mechanicism' as I have defined it above has had longstanding currency in the German ('*mechanizismus*'), French ('*mécanicisme*'), Italian ('*meccanicismo*'), and Spanish ('*mecanicismo*') scholarly literature, where it is commonly used to demarcate this sense of 'mechanism' from the machine mechanism and causal mechanism senses, but for some reason the term has not caught on in the English-speaking world. However, Allen (2005) has recently distinguished between the mechanicism sense (which he calls 'philosophical Mechanism') and the causal mechanism sense (which he calls 'explanatory mechanism'), though he does not discern the machine mechanism meaning of 'mechanism'. On the other hand, Ruse (2005) has distinguished between the machine mechanism and causal mechanism senses (designating the former 'mechanism in the specific sense' and the latter 'mechanism in the general sense'), but he fails to acknowledge the mechanicism meaning. So although previous attempts have been made to distinguish the various senses of 'mechanism', these efforts have tended to only discriminate two of the three meanings of the concept. Consequently, a tripartite distinction such as the one I have proposed in this section is needed to recognize the full semantic breadth of the concept of mechanism.

Proponents of the new mechanism movement may object that such convoluted distinctions are not really necessary, as at least in present philosophical discussions the term 'mechanism' is employed consistently. The reality, however, is that it is not uncommon to come across instances in the new mechanism discourse in which the concept is used in different senses, sometimes even in the same passage. For example, consider the following remark by Craver and Darden (2005, p. 234):

> From the perspective of biology [...] one might tell a triumphal story of the success of mechanism [i.e., *mechanicism*] over various forms of vitalism, as well as over biological theories appealing to intelligent design. Indeed, one cannot open a journal in any field of contemporary biology without encountering appeals to the mechanism [i.e., *causal mechanism*] for this or that phenomenon.[2]

One final terminological distinction is in order before moving on. It has become customary, following Skipper and Millstein's (2005) analysis, to refer to the recent mechanism discourse as 'the new mechanistic philosophy'. This is a very unfortunate and rather misleading designation, as it suggests that the new philosophical interest in the concept of mechanism represents some sort of continuation of mechanistic philosophy (i.e., mechanicism), which is not in fact the case. Mechanistic philosophy, both as a general doctrine and specifically as it applies to biology, is concerned with the characterization of machine mechanisms. The new mechanism discourse, in contrast, is devoted to examining the role played by causal mechanisms in scientific practice. The new mechanism discourse is not committed to a mechanistic worldview, nor does it prescribe a mechanistic approach in biology. In fact, there is nothing distinctively *mechanistic* about the new mechanism discourse, other than its focus on 'mechanisms'; and even this is not something it really shares with mechanicism given that each research program understands this concept in a different sense (see Fig. 1). Still, many contemporary philosophers of science routinely refer to explanations appealing to causal mechanisms as 'mechanistic', despite these generally having nothing to do with classic mechanistic explanations. Mechanistic explanations are ones in which wholes are accounted for in terms of the structure and interactions of their parts. Thus, to explain a system mechanistically is to explain it as one explains a machine mechanism; i.e., to explain the way in which the component parts of the system determine the properties and activities of the whole. However, it is increasingly the case that philosophers employ the term 'mechanistic' simply as a synonym for 'causal' when characterizing scientific explanations. This is regrettable because it blurs the longstanding tradition in biology of using 'mechanistic' to refer to the ontological and epistemological commitments of mechanicism (such as in the title of Jacques Loeb's seminal manifesto, *The mechanistic conception of life*), which remain at the heart of contemporary disciplines like molecular biology.[3] Consequently, for the sake of consistency it would be preferable to avoid the term 'mechanistic' altogether in discussions of causal mechanisms. In place of Skipper and Millstein's misleading banner, I will hereafter refer to the new mechanism movement in the philosophy of science as the *mechanismic program*, and to explanations given in terms of causal mechanisms as *mechanismic* explanations, retaining the word 'mechanistic' for discussions of mechanicism and machine mechanisms. This seems more appropriate, given that the term 'mechanismic' is already widely used in

[1] Interestingly, the machine mechanism and causal mechanism senses of 'mechanism' are sometimes invoked in the same context. For example, when biologists speak of 'the rotary mechanism of ATP synthase', 'mechanism' is used in the machine mechanism sense to draw attention to the engine-like structure of the enzyme. However, when biologists speak of 'the mechanism of ATP synthesis', 'mechanism' is employed in the causal mechanism sense to describe the sequence of steps involved in the chemiosmotic process responsible for the generation of ATP.

[2] In addition to conflating two senses of 'mechanism', this passage is historically inaccurate. Mechanicism cannot be contrasted historically with theories appealing to intelligent design because in its original formulation the mechanistic view of the world as a machine mechanism necessarily presupposed the existence of a Divine Creator. This has had important repercussions for biology. As Broad (1925, p. 91) recognized, 'Biological Mechanism [i.e., *mechanicism*] about the developed organism cannot consistently be held without an elaborate Deistic theory about the origin of organisms. This is because Biological Mechanism is a theory of the organism based on its analogy to self-acting and self-regulating machines. These, so far as we can see, neither do arise nor could have arisen without design and deliberate interference by someone with matter'. As I will discuss in the next section, it is only with the widespread acceptance of Darwin's theory of evolution that mechanistic biology became completely secularized.

[3] Jacques Monod, one of the founding fathers of molecular biology, captures the distinctively mechanistic mindset of this discipline in his characterization of the cell: 'By its properties, by the microscope clockwork function that establishes between DNA and protein, as between organism and medium, an entirely one-way relationship, this system obviously defies 'dialectical' description. It is not Hegelian at all, but thoroughly Cartesian: the cell is indeed a *machine*' (Monod, 1977, p. 108).

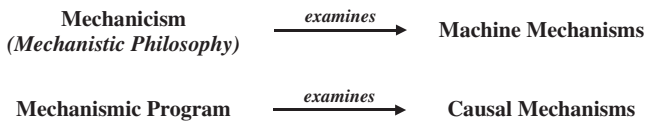

Fig. 1. Relationship between the different meanings of 'mechanism'

philosophical discussions of causal mechanisms in the social sciences (e.g., Bunge, 1997; Falleti & Lynch, 2009; Gerring, 2007; Norkus, 2005).[4]

3. The conceptual evolution of 'mechanism' in biology

Darden (2006, p. 289, fn. 5) has noted that 'The history of the usage of the concept of mechanism from the seventeenth century to molecular biology has yet to be written'. It would be impossible to provide a comprehensive account of this history in the present paper. Instead, I will restrict myself in this section to indicating what I take to be the critical episodes in that history which resulted in the semantic fragmentation of 'mechanism'.

The first two senses of 'mechanism' I distinguished, mechanicism and machine mechanism, can be traced back to the natural philosophy of the seventeenth century. Mechanicism in its first formulations was intertwined with natural theology, given that the mechanistic understanding of the universe as intricate clockwork (i.e., as a machine mechanism) necessarily implied a Divine Creator. As a result, all things in nature, including organisms, became conceived as complex assemblages of machinery created by an intelligent Designer. It is this mechanistic understanding of life which enabled the notion of machine mechanism to be employed beyond the realm of technological artefacts in explicitly biological contexts. For the mechanistic biologist, living systems are not just composed of machine mechanisms; they *are* themselves machine mechanisms. Indeed, allusions to the 'mechanism of the body' are commonplace throughout the history of physiology.

With Charles Darwin's theory of evolution by natural selection, it became possible to naturalistically explain the complex adaptations of organisms without needing to appeal to a Divine Creator. One of the implications of Darwin's theory was that its evolutionary understanding of organisms seemed to be at odds with the engineering-based conception of life of mechanicism, exemplified by its postulation of biological machine mechanisms. Therefore, to uncover the semantic evolution of the concept of mechanism, it is necessary to consider two key questions:

(a) What happened to the notion of machine mechanism in biology after Darwin?
(b) When and why did the notion of causal mechanism become pervasive in biology?

Ruse (2005) has actually provided answers to both of these questions, but his answers are problematic. In response to question (a), Ruse presents textual evidence which suggests that although Darwin did occasionally refer to biological machine mechanisms, unlike earlier biologists he *always* understood these machine mechanisms in a purely metaphorical sense. Ruse concludes from this that Darwin was responsible for demoting the notion of machine mechanism in biology to a heuristic status. With Darwin, machine mechanisms lost their ontic basis and became reconceptualized as heuristic tools that aid the investigation of adaptation. Darwin himself made use of the machine mechanism-heuristic in his inquiry into the workings of barnacles and orchids, and this remains a common practice in evolutionary biology, where it is known as 'reverse engineering'.

Although this account seems reasonable, a more careful examination reveals its problems. Despite the apparent incompatibility between the mechanistic conception of organisms as machines and a Darwinian understanding of organisms, what we actually find when we inspect modern evolutionary biology is that mechanistic language is *not* used exclusively at a heuristic level. Contrary to Ruse's expectations, Darwin did not strip the notion of machine mechanism of its ontic significance. Rather, it was evolutionary biology *itself* which adapted to accommodate mechanistic thinking about organisms, so that since Darwin, 'the idea that the world is full of *designed machines* has been replaced by the idea that it contains *evolved machines*' (Craver & Darden, 2005, p. 239, my emphasis). In fact, Gould and Lewontin's (1979) famous critique of adaptationism can be interpreted precisely as a reaction against this excessive reliance on mechanistic thinking in evolution, which all too often constitutes not just a heuristic tool but also a theoretical justification for understanding organisms as optimally-designed machines blindly engineered by natural selection (e.g., Dawkins, 1986; Dennett, 1995).

Moving to other areas of contemporary biology, it quickly becomes apparent that talk of machine mechanisms remains entrenched at an ontological level. In molecular and cellular biology, for instance, the standard conception of the organism is that of a machine programmed by its genes and decomposable into its component machine mechanisms. Subcellular protein complexes are frequently referred to as machines, and the cell itself is conceived as an assemblage of machine subunits (e.g., Alberts, 1998). An important point, however, is that despite the fact that machine mechanisms continue to play a fundamental role in many areas of biology, the term 'mechanism' is generally no longer used to designate them. Instead, biologists today tend to refer to machine mechanisms simply as 'machines', presumably to distinguish this notion from the sense in which 'mechanism' is now most commonly used in biology, namely causal mechanism.

Ruse's explanation for the displacement of machine mechanism by causal mechanism as the most widely used sense of 'mechanism', i.e., his answer to question (b), is also problematic. He suggests that Darwin's secularization of mechanicism enabled the concept of 'mechanism' to acquire widespread currency in the broader sense of causal mechanism. With Darwin, 'mechanism' came to be used to designate a much wider range of biological phenomena, including Darwin's own 'mechanism' of natural selection. However, after thoroughly searching through Darwin's works, Ruse actually discovers that Darwin 'simply does not speak of natural selection as a mechanism' (Ruse, 2005, p. 291). Darwin only uses 'mechanism' in the machine mechanism sense; the very idea of a causal mechanism is simply alien to him. As Ruse himself indicates, it is not until the late nineteen-thirties that natural selection came to be generally referred to as a 'mechanism'. Neither Fisher (1930) nor Haldane (1932) used this language, but Dobzhansky (1937) did, noting that 'the theory of natural selection is primarily an attempt to give an account of the probable mechanism [i.e., *causal mechanism*] of the origin of the adaptations of organisms to their

[4] Indeed, these authors have adopted this neologism precisely because they recognize the importance of distinguishing explanations based on causal mechanisms from mechanistic explanations of machine mechanisms. Gerring (2007, p. 163), for instance, remarks: 'It should be noted that this contemporary understanding of mechanism [i.e., *causal mechanism*] departs dramatically from common nineteenth-century and early twentieth-century understandings of the term, which invoked a *mechanistic* account of the world. In this context, mechanism [i.e., *mechanicism*] meant 'the theory that all phenomena can be explained in terms of the principles by which machines (mechanical systems) are explained without recourse to intelligence as an operating cause or principle' [...]. Evidently, to say 'mechanism' in a contemporary context does not mean that one is wedded to a *mechanistic* causal account modelled on Newtonian physics'.

environment' (Dobzhansky, 1937, p. 150). If Darwin's secularization of mechanicism truly brought about the widespread use of 'mechanism' in the causal mechanism sense, why is it that three-quarters of a century had to pass from the publication of Darwin's *Origin of species* for natural selection to be commonly referred to as a 'mechanism'?

In light of these difficulties, I want to suggest a rather different answer to question (b). When considering the factors that had the greatest impact on mechanicism in the late nineteenth and early twentieth centuries, far more important than the advent of Darwinism was the gradual erosion of vitalism. As the philosophical antithesis of mechanicism, vitalism can be characterized as the doctrine that upholds the direct inverse of the four core tenets of mechanicism outlined in Section 2. The heart of the vitalistic doctrine is the postulation of a vital principle (which, depending on the historical period, assumed the form of a soul, a force, or a mode of organization) that ontologically demarcates living from non-living systems. From the seventeenth century onwards, mechanicism and vitalism developed in parallel, with the mechanists continually disproving the claims of the vitalists, and the vitalists repeatedly re-emerging to pose new challenges to the mechanists. However, by the late nineteenth century the spectacular empirical success of mechanicism in disciplines as diverse as physiology, developmental biology, and biochemistry ultimately led to the marginalization of vitalism as a workable research program. No longer being confronted by serious opposition, the mechanistic conception of life became widely accepted as an elementary presupposition of biological research in the early decades of the twentieth century. 'At the present day', wrote the embryologist Joseph Needham in 1925, 'the situation is in effect the complete triumph of mechanistic biology. It is not alone in the field, because the neo-vitalists do exist as a small minority, but the vast preponderance of active biological workers are mechanists' (Needham, 1925, p. 235).

I want to argue that one of the key consequences of the consolidation of mechanicism was that it was no longer necessary to explicitly defend the core tenets of this doctrine. The view that living systems are machines did not need to be justified and could simply be taken as a given. As a result, mechanism-talk became applied to all kinds of biological phenomena, given the mechanistic confidence that everything would, in due course, be explained as effectively as engineers explain the operation of machines. This increasingly loose use of 'mechanism' caused the word to gradually lose its distinctive mechanistic connotations, becoming a 'dead metaphor' that could be readily applied beyond the realm of machine-like systems to any biological phenomenon in need of a causal explanation. It is this semantic shift, I suggest, which led the term 'mechanism', understood in the more general and inclusive sense of causal mechanism, to acquire such widespread currency in biology.

Evidence for this account can be found by inspecting the writings of the biologists of this period. For example, J. S. Haldane, one of the most influential physiologists of the early twentieth century, drew attention on several occasions to the increasing proliferation of mechanism-talk in biology, pointing out that using the term 'mechanism' with respect to a phenomenon no longer implied conceiving it mechanistically as a machine mechanism. In *The sciences and philosophy*, he observed that 'In current physiological literature it is still customary, in describing what is known as to different bodily activities, to refer to them as 'mechanisms'—for instance, the 'mechanisms' of reproduction, respiration, secretion, etc.' despite the fact that 'There are perhaps few physiologists who now consider that they have any real conception of these mechanisms [as *machine mechanisms*]'. The usage of 'mechanism', Haldane noted, has become 'a mere matter of custom' (Haldane, 1929, p. 59). In *The philosophical basis of biology*, Haldane reiterated these remarks, indicating that physiologists 'have acquired the habit, almost unconscious, of referring to the 'mechanisms' of various physiological activities, though they have not the remotest conception of what sort of mechanisms [i.e., *machine mechanisms*] these activities represent'. He concluded from this that 'the use of the word 'mechanism' is a mere empty formality' (Haldane, 1931, p. 11). Although Haldane openly voiced his concern regarding this looser use of 'mechanism' in the causal mechanism sense, warning that 'such a mode of expression is extremely misleading to that miscellaneous body which we call the public' (Haldane, 1929, p. 59), he clearly did not succeed in persuading his contemporaries against this usage of the term. Still, what is relevant in the present discussion is that his remarks lend credence to my proposed explanation of the supplantation of machine mechanism by causal mechanism as the most common meaning of the term in biology.[5]

4. The mechanismic program in relation to mechanicism

So far I have argued that due to the success of mechanicism in the early twentieth century, the causal mechanism sense of 'mechanism' became predominant in biology during this period, and remains so to this day. But how and when did the different senses of 'mechanism' come to be featured in the philosophy of biology? Exploring this question will help situate the recent mechanismic program in relation to mechanicism. This will be a key step in the development of my argument, as showing the fundamental differences between these two research programs will provide the basis for my critical engagement with the mechanismic program in Sections 5 and 6.

The longstanding conflict between mechanists on the one side and vitalists and organicists on the other, being in the final analysis a dispute concerning the very nature of life itself, constituted the central theme in the philosophy of biology during the first half of the twentieth century (see Bertalanffy, 1952; Johnstone, 1914; Woodger, 1929), even if by this time most experimental biologists (like Needham) considered that the dispute had already been resolved in favour of mechanicism. Mechanistic biology and machine mechanisms continued to be discussed in subsequent decades (e.g., Varela & Maturana, 1972), capturing even the attention of leading exponents of logical empiricism like Hempel (1966, ch. 8) and Nagel (1979, ch. 12). However, following the academic institutionalization of the philosophy of biology at the hands of David Hull, Michael Ruse and others, discussions of mechanistic biology came to an abrupt end as the new generation of philosophers of biology, influenced by philosophically-minded evolutionists like Ernst Mayr, turned its attention to theoretical issues in evolutionary biology, such as the levels of selection, the definition of fitness, and the nature of species. Nevertheless, critical examinations of mechanistic biology and machine mechanisms are still featured in the contemporary literature (e.g., Dupré, 2007; Lewens, 2004; Lewontin, 2000; McLaughlin, 2001; Rosen, 1991), although the terms in which the issues are discussed have changed somewhat.

What of the third sense of 'mechanism'? When did causal mechanisms enter into philosophical discussions of biology? Browsing the literature one finds references to the term 'mechanism' employed in the causal mechanism sense in articles by Kauffman (1970), Grene (1971) and Wimsatt (1972, 1974). However, Brandon (1985) appears to have been the first to provide a

[5] It is interesting to note that the causal mechanism sense of 'mechanism' first began to permeate the literature on natural selection only a few years after Haldane's warnings against this looser use of the term (see Ruse, 2005).

detailed analysis of the importance of causal mechanisms in biological research. Brandon's account is important for several reasons. For one thing, it is the first to explicitly recognize the semantic ambivalence inherent in the biological usage of 'mechanism', as well as the inevitable difficulties that arise when attempting to pin down this concept.[6] More crucially, it presents an understanding of the postulation of causal mechanisms in science that distinctly characterizes the mechanismic program today, namely that the appeal to causal mechanisms in scientific practice does *not* imply a commitment to the reductionistic agenda of mechanicism.[7] Indeed, whereas mechanicism, as Craver and Darden (2005, p. 235) note, is 'closely aligned with the spirit of reductionism and the unity of science', the mechanismic program focuses on multi-level explanations given in terms of causal mechanisms and with an explicitly *non*-reductive view of science (see Craver, 2005; Darden, 2005).

The mechanismic program, unlike mechanicism, is not primarily concerned with biological ontology, but with the nature of biological explanations. This is not surprising given that the postulation of causal mechanisms, having become a virtually ubiquitous practice in contemporary biology, discloses rather little about a biologist's ontological commitments. Physiologists, ecologists, neuroscientists, and cell biologists have different understandings of life, yet they all appeal to causal mechanisms in their explanations. Clearly, whatever ontological commitments they all share are likely to be very general in nature. This stands in contrast with molecular biologists' standard mechanistic conception of living systems as machine mechanisms, for which explanations are sought from the bottom up in increasing levels of complexity. In every respect, the appeal to machine mechanisms is indicative of far more substantive ontological commitments than the appeal to causal mechanisms. These ontological commitments (summarized in Section 2) are at the heart of the mechanistic conception of life that dominated biological thought for much of the twentieth century, but which today, with the growing emphasis on systemic thinking in biology, is increasingly viewed as simply one of several possible understandings of what living systems are and how they should be studied.

In the few occasions when mechanismic philosophers explicitly address matters of biological ontology, it is usually to distinguish the mechanists' appeal to machine mechanisms from their own concern with causal mechanisms (recall Fig. 1). By demarcating causal mechanisms from machine mechanisms, mechanismic philosophers distance their research program from the ontological commitments of mechanicism. Mechanismic philosophers distinguish causal mechanisms from machine mechanisms in two ways. The first strategy (which I already alluded to in the Introduction) is to focus on the way the term 'mechanism' is presently used in biology and disregard older uses of the term as irrelevant to current analyses of the concept (e.g., Craver, 2007, p. 3). What this does is minimize the scope for conflating the older biological usage of 'mechanism' in the machine mechanism sense (predominant in biology until the first third of the twentieth century) with the current biological usage of the term in the causal mechanism sense. The second strategy is to explicitly differentiate 'mechanisms' (i.e., causal mechanisms) from 'machines' (i.e., machine mechanisms), and both Darden (2006, pp. 280–281; 2007. p. 142) and Craver (2007, p. 4 and p. 140) do this on more than one occasion.

It is important to realize the extent to which MDC's (2000) account of causal mechanisms has marked a turning point in philosophical discussions of this concept. Before MDC's account, characterizations of 'mechanisms' routinely conflated the machine mechanism and causal mechanism senses. For instance, Thagard (1998) noticed that the term 'mechanism' is commonly featured in contemporary explanations of disease, but defined it in the machine mechanism sense as 'a system of parts that operate or interact *like those of a machine*' (p. 66, my emphasis). Similarly, when Glennan first defined 'mechanism', he indicated that his definition is meant to apply to 'complex systems *analogous to machines*' (Glennan, 1996, p. 51, my emphasis). In fact, Glennan has continued to heavily rely on the notion of machine mechanism in his account of 'mechanisms', going as far as to cite cells and organisms as prime examples of his conception of them (Glennan, 2002, p. S345). Although mechanistic biologists do indeed ontologically conceive cells and organisms as machine mechanisms, it makes little sense for *any* biologist to consider the causal mechanism of an entire cell or organism. Most mechanismic philosophers would disagree with Glennan's designation of cells and organisms as 'mechanisms', and the reason is clear. The new mechanismic program 'strives to characterize mechanism [...] in a manner faithful to biologists' own usages' (Darden, 2007, p.142) and causal mechanism *is* what most present-day biologists mean when they use the word 'mechanism'. This is why mechanismic philosophers focus exclusively on this sense of the term, and why most of them would not recognize supposed machine mechanisms like cells and organisms as 'mechanisms'.

The reason for Glennan's apparent unconcern regarding the lack of correlation between his mechanistically tinged understanding of the concept of 'mechanism' and the way the term is actually used by most contemporary biologists is that his account of mechanisms is not primarily motivated by an interest in scientific practice (like MDC and others), but by a concern with the nature of causation. Indeed, in his 1996 paper Glennan sets out to address Hume's sceptical challenge regarding the connection between cause and effect by suggesting that 'mechanisms' could provide a plausible metaphysics of causation. Glennan proposes that events are causally related *if* there is a 'mechanism' that connects them, and he uses this conception of 'mechanism' to develop a mechanical view of explanation (Glennan, 2002). In doing so, Glennan builds on Salmon's (1984) account of causal-mechanical explanation, which was itself an elaboration of Railton's (1978) deductive-nomological model of probabilistic explanation, in which the term 'mechanism' was introduced into the philosophical literature on scientific explanation (Glennan 2002, p. S343). Interestingly, this earlier work on 'mechanisms', unlike the more recent biologically-inspired mechanismic discourse, does actually show some clear links with mechanicism. Railton (1978) says the following regarding his *mechanistic* orientation:

> The goal of understanding the world is a theoretical goal, and if the world is a *machine* [...] then our theory ought to give us some insight into the structure and workings of the mechanism [i.e., *machine mechanism*], above and beyond the capability of *predicting* and *controlling* its outcomes. (Railton, 1978, p. 208, my emphasis)

This conception of the world as a machine mechanism, as well as the stated desire to understand, predict, and control it, are all

[6] Indeed, when Brandon asks what mechanisms are, he is unable to provide a precise definition. He notes that 'mechanism' may refer to 'spring-wound clocks and watches' (i.e., *machine mechanisms*) but also to 'small peripheral populations and geographic isolating barriers' (i.e., *causal mechanisms*). To make matters worse, Brandon observes, in the philosophy of biology 'mechanism' 'is typically used to designate the position opposing vitalism, holism, or organicism' (i.e., *mechanicism*). The semantic ambiguity is exacerbated by Brandon's proposal to use the term 'mechanism' in a fourth sense to refer to the practice of formulating causal mechanisms in science, stating confusingly that 'the position I call mechanism is given in terms of search of mechanisms' (Brandon, 1985, p. 346).

[7] Brandon further develops this important thesis in a more recent essay entitled 'Reductionism versus holism versus mechanism' (Brandon, 1996, ch. 11).

characteristic attributes of mechanistic philosophy. Along similar lines, Glennan points out that his account of 'mechanisms' is 'largely inspired by the insights of the Mechanical philosophers' of the seventeenth century' (Glennan, 1996, p. 51). Thus, Skipper and Millstein's (2005) banner of 'the new mechanistic philosophy' would have been far more appropriate if it had been used to refer to *this* literature on 'mechanisms', rather than to the more recent examinations of causal mechanisms in biology, which on the whole bear little connection to the original motivations of this earlier work in the philosophy of science. Darden's latest appraisal of the mechanismic program makes this explicit when she clarifies that 'work on mechanisms in biology originated (primarily) not as a response to past work in philosophy of science but from consideration of the work of biologists themselves' (Darden, 2008, p. 958).

Overall, it is clear that the mechanismic program must be regarded as being completely independent from mechanicism, both as a general doctrine and specifically as it applies to biology. Indeed, we have seen how leading proponents of the mechanismic program like Craver and Darden reject some of the core tenets of mechanicism, such as the reducibility of biology to physics and chemistry, and the exclusive reliance on reductionistic explanations. Demarcating the mechanismic program from mechanicism is crucial, as the failure to do so results in problematic analyses of causal mechanisms. The most glaring example of this, in my view, is found in some of Bechtel's recent work. While most mechanismic philosophers are rather cautious in their use of history when discussing causal mechanisms, drawing on relatively recent case studies when illustrating their claims, Bechtel traces the appeal to 'mechanisms' in scientific explanation not just to Descartes in the seventeenth century, but all the way back to the Ancient Greek atomists of the fifth century BCE (Bechtel, 2006, pp. 20–21; 2008, p. 10). But instead of considering how the meaning of 'mechanism' has developed over time (as Ruse (2005) does, and as I have attempted to do in Section 3), Bechtel just takes the modern sense of 'mechanism' as causal mechanism as his starting point and then simply projects it back in history. As a result, his historical discussions conflate the distinctive appeal to machine mechanisms by mechanistic biologists with the almost ubiquitous appeal to causal mechanisms by biologists today (e.g., Bechtel, 2006; 2007, ch. 2). Understanding the term 'mechanism' exclusively in the causal mechanism sense, Bechtel complains that critics of mechanistic biology commit a grave mistake in assimilating the notion of 'mechanism' to that of machine (Bechtel, 2008, p. 2), not realizing that the very reason for this is that when mechanists do speak of 'mechanisms', machine-like systems (i.e., machine mechanisms) is *precisely* what they have in mind.[8]

The striking thing is that Bechtel, just like Craver and Darden, actually *rejects* central tenets of mechanistic biology, such as the exclusive reliance on explanatory reductionism (Bechtel & Abrahamsen, 2008), and the privileging of the efficient and material causes of organisms over and above their systemic, self-organizing properties (Bechtel, 2007). But, again, instead of distancing himself from mechanicism, Bechtel seems to think that the only way to make sense of the pervasiveness of mechanism-talk in current biology is to broaden the doctrine of mechanistic biology accordingly, not realizing that the appeal to the term 'mechanism' in scientific practice today no longer commits one to mechanicism (as 'mechanism' is now generally employed in the causal mechanism sense). This leads Bechtel to formulate a very odd conception of mechanistic biology, so general in content and inclusive in its applicability that none of the distinctive ontological and epistemological commitments that tend to be associated with it (see Section 2) are relevant. Instead, all that qualifies a biologist as a 'mechanist' for Bechtel is that she appeals to 'mechanisms' in her research. Similarly, all that qualifies an explanation as 'mechanistic' is that a 'mechanism' is featured in it, regardless of the way in which this concept is used.[9]

I can think of two reasons for Bechtel's misrepresentation of mechanicism. The first is that some of his earlier work (e.g., Bechtel & Richardson, 1993) was in fact concerned with *mechanistic* explanations in biology, specifically with the strategies of decomposition and localization that are often featured in them.[10] So in the wake of the influence of MDC's (2000) account of causal mechanisms, Bechtel might have felt it natural to bridge his earlier discussion of machine mechanisms with an examination of causal mechanisms, since, after all, the term 'mechanism' is central to both discourses. Still, the main reason for Bechtel's misrepresentation is that he does not appear to recognize that the concept of 'mechanism' has more than one meaning. It is because he conflates the notions of machine mechanism and causal mechanism that he also conflates mechanicism with the mechanismic program (e.g., Bechtel, 2006, ch. 2; 2008, ch. 1).

Nevertheless, the mischaracterization of mechanicism is not the only, or even the main, problem that results from the conflation of causal mechanisms and machine mechanisms. The most serious consequence of not distinguishing these notions is that causal mechanisms become inappropriately endowed with the ontic status of machine mechanisms. This ontologization of causal mechanisms is very widespread in the philosophical literature, and in the next section I will discuss some of the problems that stem from it.

5. Problems resulting from the ontologization of causal mechanisms

Mechanismic philosophers tend to conceive causal mechanisms as real things in the world existing independently from our conceptualization of them. However, based on the role they play in scientific practice, I suggest that causal mechanisms are better understood as heuristic models that facilitate the explanation of phenomena. The fact that the overwhelming majority of mechanismic philosophers speak of them as 'real systems in nature' (Bechtel, 2006, p. 33) I attribute to an inadvertent transposition of the ontic status of machine mechanisms (the original sense in which 'mechanism' was used) onto the notion of causal mechanism (the standard meaning of 'mechanism' in biology today). This ontologization of causal mechanisms tends to result in a conception of them as autonomous complex systems (analogous to machine mechanisms), which constitute and operate *within* the organism (e.g., Bechtel, 2007; Glennan, 2002). I maintain that this ontic conception of causal mechanisms is problematic, and I will substantiate this claim by examining what are perhaps the two

[8] This is as true for seventeenth-century mechanists like Descartes as it is for twentieth-century mechanists like Loeb.

[9] Much more could be said regarding Bechtel's problematic reconstruction of mechanistic biology, such as the way in which he misappropriates classic vitalistic ideas like self-organization for the mechanists (Bechtel, 2007) despite the fact that the concept of self-organization was actually coined by Immanuel Kant in order to argue that organisms are fundamentally different from machines and thus *cannot* be explained in mechanistic terms, and that most vitalists after Kant took the distinctive self-organization of organisms to constitute the principal manifestation of the vital principle they postulated. However, elaborating these claims would take me beyond the scope of this paper.

[10] The fact that Bechtel & Richardson (1993) are interested in mechanistic explanations (pertaining to machine mechanisms) as opposed to mechanismic explanations (pertaining to causal mechanisms) is evidenced by their assertion that 'By calling the explanations *mechanistic*, we are highlighting the fact that they treat the systems as producing a certain behavior *in a manner analogous to that of machines developed through human technology*' (p. 17, my emphasis). Indeed, their analysis of *mechanistic* explanation begins with a characterization of *machines*, not of mechanisms. However, in his more recent work Bechtel readily describes as 'mechanistic' not just the distinctive appeal to machine mechanisms by mechanists, but also the widespread appeal to causal mechanisms in current scientific practice.

most distinctive features of causal mechanisms in biology: *function* and *organization* (cf. McKay & Williamson, 2010).

5.1. Function

The operation of a causal process described in a causal mechanism produces a particular phenomenon that serves to individuate and causally relate the entities and activities that are responsible for it. In biology, the phenomenon produced by the causal process described in a causal mechanism usually enables the fulfilment of a function, so that specifying the causal mechanism for a function explains how this function is causally brought about. The problem of conceiving causal mechanisms as autonomous complex systems is that it overlooks the conditions that actually enable the functions of these systems to be carried out, as well as the true biological significance of those functions.

A living organism is an organized network of processes of production, transformation, and regeneration of components that continuously realizes itself by means of the coordinated orchestration of the components that make it up (Maturana & Varela, 1980). In this way, the organism constitutes an integrated whole which maintains its identity through time by regulating, repairing, and reproducing its component parts. These parts stand in a relation of collective interdependence, as every one of them is necessary for the generation and operation of every other. Thus the attribution of functions to the parts of an organism is dictated by the means in which each part individually contributes to the maintenance and organization of all other parts and hence to the organism as a whole (see Edin, 2008; McLaughlin, 2001; Mossio, Saborido, & Moreno, 2009). This means that the function of all suborganismic systems and processes featured in causal mechanisms is ultimately that of preserving the autopoietic organization of the whole organism.

The idea of autonomous causal mechanisms operating within the organism is, I suggest, nothing more than a pragmatic idealization that biologists appeal to in order to narrow their focus on the particular parts of the organism they happen to be investigating. This heuristic fragmentation of the organism into causal mechanisms, despite being necessary for its investigation, often comes at the expense of neglecting the way in which the organism as a whole influences the behaviour of its parts. In current philosophical accounts, the ontic conception of causal mechanisms as real autonomous subsystems neglects the fact that in order to make appropriate biological sense of the subsystems' functions, these subsystems need to be framed within a set of background conditions, that is, the *organismic context* that enables them to carry out their functions in the first place.[11]

Craver (2007, p. 122) has indicated that 'The core normative requirement on mechanistic [i.e., *mechanismic*] explanations is that they must fully account for the *explanandum phenomena*'. That is, 'Good explanations account for all of the features of a phenomenon rather than a subset' (ibid., p. 161). This means that mechanismic explanations that do *not* include a full account of the organismic context that enables the production of the *explanandum* phenomenon (or function) are, on Craver's terms, necessarily incomplete. This is problematic as actual scientific practice demonstrates that mechanismic explanations are never exhaustive catalogues of *all* the causal relations necessary for the production of phenomena, such as the enabling conditions provided by the organism as a whole. Rather, mechanismic explanations specify only those features of the underlying causal networks that biologists deem *most relevant* for manipulating and controlling the phenomena whilst at the same time presupposing a great deal of the organismic context that makes them possible. For this reason, it makes more sense to view causal mechanisms as idealized spatiotemporal cross-sections of organisms that heuristically pick out certain causal features over others in order to account for how given functions within the organism are carried out, as these are generally the things that biologists describe when they use the term 'mechanism' in their explanations.

5.2. Organization

Mechanismic philosophers frequently emphasize the importance of organization for understanding how causal mechanisms account for functions or behaviours. MDC (2000, p. 3), for instance, state that 'The organization of entities and activities determines the ways in which they produce the phenomenon'. Bechtel (2006, p. 26) similarly notes that 'The orchestrated functioning of the mechanism is responsible for one or more phenomena'. The problem is that mechanismic philosophers do not actually explain *how* the entities and activities in a mechanism are organized, only *that* they are organized. MDC (2000, p. 3) point out that 'Entities often must be appropriately located, structured, and oriented, and the activities in which they engage must have a temporal order, rate, and duration' but say nothing about the means by which these crucial organizational requirements are actually met in living organisms. Instead, all that discussions of organization in the mechanismic literature amount to is the plain assertion that organization matters (e.g., Craver, 2007, pp. 134–139).

Still, if causal mechanisms are to be conceived ontically as real suborganismic systems (rather than epistemically as idealized models *of* those subsystems, as I suggest) then just paying lip service to the fact that these subsystems are organized is insufficient. To *fully* account for the *explanandum* phenomenon (Craver's normative requirement for a good mechanismic explanation) it becomes necessary not just to *specify*, but also to *explain* how this organization is generated and maintained. The problem is that this requires taking the description beyond the actual causal mechanism to the level of the organism as a whole, given that suborganismic parts do not organize themselves but rely on the action of the whole organism for their generation, organization, and maintenance. This is rarely understood in mechanismic accounts of organization. For example, when Craver (2007, p. 148) indicates that a 'mechanism might compensate for the loss of a part by recovering (healing the part), by making new use of other parts, or by reorganizing the remaining parts', he is inappropriately attributing actions to an ontologized causal mechanism that are actually performed by the organism which contains it.[12]

As I have argued in my discussion of function, one of the advantages of understanding causal mechanisms as idealized models of suborganismic causal processes rather than as real things is that a satisfactory mechanismic explanation need not include an account of how the target system is actually organized by the organism *even if* this organization is strictly speaking necessary for the system to causally bring about the phenomenon. This is more in

[11] The problematic transference of mechanistic thinking is particularly noticeable here. Whereas a machine mechanism can be broken down into discrete, self-contained parts with clearly-delineated output functions without the loss of information, the parts in an organism (ontologized in current accounts of causal mechanisms) stand in a relation of collective interdependence and are thus not autonomous in any important respect (even if they can be construed as such for the purposes of their investigation). Consequently, any explanation of the functions of parts in an organism needs to account not just for the parts themselves but also for the organismic context that makes their function possible.

[12] Not only does Craver not refer to the influence of the whole organism in explaining how the causal processes instantiated by causal mechanisms achieve and maintain their organization, but there is reason to believe that mechanismic explanations, by virtue of their nature, simply *cannot* accommodate organismic organization, given that mechanismic explanations are, in Craver's words, 'anchored in components' (Craver, 2007, p. 138), and an organism's autopoietic organization is a system-level phenomenon that is not explainable by attending exclusively to the properties of component parts.

accordance with actual scientific practice, in which causal mechanisms tend to pragmatically abstract away the organismic context and only specify the causal features that are taken to be most relevant for controlling and manipulating the phenomena being investigated. In the next section, I will elaborate and defend the epistemic account of causal mechanisms, indicating the further advantages of this view over the ontic conception that most mechanismic philosophers presently favour.

6. Defending an epistemic conception of causal mechanisms

It is important to keep in mind that the causal mechanism sense of 'mechanism' was not formulated *in abstracto* and then applied to scientific practice. Rather, it arose from scientific practice and it has only recently been philosophically reconstructed to make sense of how scientists explain phenomena. Consequently, the success of any given philosophical reconstruction of 'mechanism' must be measured in terms of how well it captures the way this term is used in scientific practice. The conception of causal mechanism that I argue best fits biologists' mechanism-talk is that of a contingent explanatory description which heuristically abstracts away the complexity of a living system sufficiently to describe some localized causal process within it which leads to the realization of some function of interest. That is, causal mechanisms are epistemic models that enable the explanation of how phenomena are causally brought about.

Interestingly, although most mechanismic philosophers claim to uphold an ontic view of causal mechanisms, much of what they say is actually perfectly compatible with an epistemic conception. In fact, it is not difficult to find instances in the philosophical literature in which ontically construed causal mechanisms are conflated with their epistemic representations, as I will show in a moment. This ambiguity, I suggest, is the result of the tension that inevitably arises from inappropriately transposing the ontic status of machine mechanisms onto causal mechanisms on the one hand, and paying close attention to the role that mechanism-talk actually plays in scientific practice on the other.

When scientists inquire about the causal mechanism of P (where P is the phenomenon of interest), the term 'mechanism' does not refer to that which is explained but rather to that which does the explaining. Craver (2007) acknowledges this when he asserts that 'The *explanans* is a mechanism' (p. 139) and the phenomenon of interest is the *explanandum* (p. 6).[13] In this way, specifying a causal mechanism for a phenomenon implies providing an explanation for it. As MDC indicate, 'Mechanisms are sought to *explain* how a phenomenon comes about or how some significant process works' (2000, p. 2, my emphasis). One of the advantages of the epistemic view of causal mechanisms is that it is no longer necessary to postulate additional epistemic notions like 'mechanism sketch' and 'mechanism schema' to make sense of mechanismic explanations. Depending on the degree of abstraction, causal mechanisms may constitute what mechanismic philosophers call 'sketches', 'schemas', or 'mechanisms'. Craver (2007, p. 114) tacitly admits the continuity between these notions when he indicates that progress in formulating a successful mechanismic explanation 'involves movement [...] along the sketch-schema-mechanism axis'.

Moreover, the very characterizations of causal mechanisms that mechanismic philosophers have proposed are in fact perfectly compatible with an epistemic understanding of them. According to the epistemic view, causal mechanisms constitute idealized representations of causal processes. These causal processes are abstracted *temporally* and *spatially*. Temporally, the causal mechanism delimits a particular causal process by specifying arbitrary beginning and end points that are selected on pragmatic grounds. MDC (2000, p. 11) explicitly recognize that the set-up and termination conditions of causal mechanisms are 'idealized states', and Darden has reiterated this point on several occasions, noting that the beginning and end points of causal mechanisms are 'more or less arbitrarily chosen' (Darden, 2007, p. 141; see also Torres, 2009, p. 240, fn. 10). So although MDC purport to uphold an ontic conception of causal mechanisms, they actually characterize them in terms of epistemically selected beginning and end points.

Causal mechanisms are also abstracted spatially, according to the epistemic view, as they can only capture certain ontic features of reality at the expense of neglecting others. What gets represented and what is omitted in a causal mechanism is dictated by the nature of the *explanandum* phenomenon. Craver (2007, pp. 139–160) reaches this same conclusion when he considers the normative requirements that determine whether or not something is included as part of a causal mechanism, asserting repeatedly that the delimitation of causal mechanisms can only occur *in the context of explanation*. That is, entities, activities, and organizational features are part of the causal mechanism for P (where P is the phenomenon of interest) if and only if they are relevant to the explanation of P. The act of individuating the causal mechanism for P is thus the act of determining what aspects are causally relevant to the explanation of P. The delimitation of causal mechanisms hence 'depend[s] on the *epistemologically prior* delineation of relevance boundaries' (Craver, 2007, p. 144, my emphasis).

This view of causal mechanisms significantly departs from the ontic conception of them as autonomous systems akin to machine mechanisms (defended by Glennan, Bechtel, and at times by Craver himself, as shown in Section 5), given that the parts of a causal mechanism do not even need to be structurally demarcated.[14] All that matters is that they are causally relevant to the production of the *explanandum* phenomenon. Craver fleshes out this notion of causal relevance by appealing to Woodward's (2003) manipulability theory of causation. In this way, a part is causally relevant to the phenomenon produced by a causal mechanism if one can modify the production of this phenomenon by manipulating the behaviour of the part, and one can modify the behaviour of the part by manipulating the production of the phenomenon by the causal mechanism.

Although Craver's account of explanatory relevance is compatible with both an ontic and an epistemic conception of causal mechanisms, there do not appear to be any obvious reasons for favouring the former over the latter view; if anything, the latter view seems more plausible. Explanations always presuppose a context that specifies what is to be explained and how much detail will suffice for a satisfying answer, and Craver recognizes that it is this very epistemic context that determines how causal mechanisms are individuated and what details are featured in them. The crucial requirement of any causal mechanism, according to Craver, is that it must capture the underlying causal relationships of the target system in such a way that it exhibits the necessary resources for explaining how the target system will behave as a result of interventions and manipulations of its parts. An epistemic view of causal mechanisms fulfils this requirement.

It may be helpful to illustrate these claims with an example. Consider the causal mechanism for the membrane trafficking of the delta-opioid receptor (DOR) induced by pain stimulation, shown in Fig. 2 (adapted from Bie & Pan, 2007). This causal mechanism exhibits all of the features I have discussed. It is a step-by-step explanation of the mode of operation of the signal transduction pathway induced by pain stimulation that triggers

[13] Thus, mechanismic explanations should be understood not as explanations *of* causal mechanisms, but as explanations *given in terms of* causal mechanisms.

[14] Indeed, on several occasions Craver (2007, p. 141) acknowledges that causal mechanisms 'frequently transgress compartmental boundaries', and 'are often spatially quite distributed' or 'tightly interwoven into their systematic context' (ibid., p. 143, fn. 23).

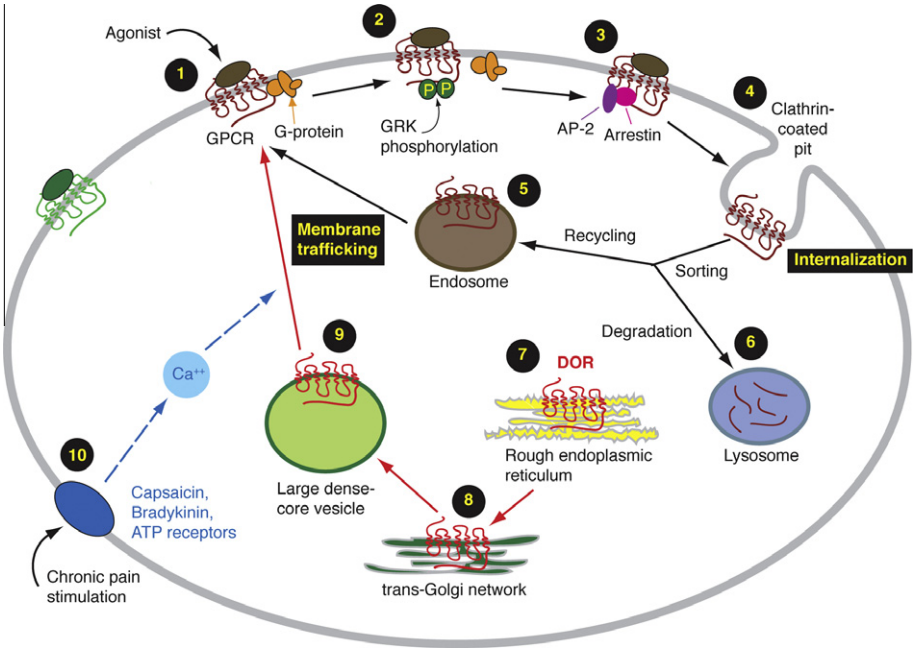

Fig. 2. Causal mechanism for the membrane trafficking of the Delta-Opioid Receptor (DOR). Upon agonist binding (1), DOR is phosphorylated by GRK (2). It then binds to proteins AP-2 and arrestin (3), and undergoes a process of internalization via endocytosis (4). Once internalized, the receptor is sorted and targeted either to endosomes via the recycling pathway (5) for membrane insertion, or to lysosomes for degradation via the degradation pathway (6). DOR is synthesized in the endoplasmic reticulum (7), and transported to the trans-Golgi network (8), becoming a mature receptor which is targeted in dense-core vesicles (9), ready for membrane trafficking and insertion. Chronic pain stimulation activates receptors (10) and increases intracellular calcium concentration, inducing the membrane trafficking of DOR.

the intracellular activation of DOR, which results in effective pain relief. The causal mechanism is abstracted both temporally and spatially. Temporally, it abstracts the continuous life-cycle of DOR into a series of discrete idealized stages, which are numbered in the causal mechanism. Spatially, although the whole cell is depicted, only the features that are causally relevant to the membrane trafficking of DOR (i.e., the *explanandum* phenomenon) are featured in the causal mechanism. What is represented in the causal mechanism is contingent on epistemic considerations given that if we happened to be interested in explaining any other cellular phenomenon, a different yet partially-overlapping set of features would be included in the causal mechanism. Moreover, the organismic context (in this case, the cell) is almost completely abstracted away and yet it is heavily presupposed, as it provides the enabling conditions that are ultimately necessary for the membrane trafficking of DOR. Finally, the causal mechanism constitutes an explanatory model of a particular cross-section of the cell that provides the necessary resources for anticipating how interventions and manipulations of any of the causally relevant parts within the cell and any of the successive stages of the described process will affect the membrane trafficking of DOR. In this way, this causal mechanism serves the heuristic purpose of aiding the physiological and pharmacological investigation of pain relief.

So far in this section, I have advanced my defence of an epistemic view of causal mechanisms by showing how the key features of causal mechanisms that mechanismic philosophers deem most important for understanding them are not only not incompatible with the epistemic account I propose, but actually provide strong support for it. Nevertheless, the compatibility of the central claims of MDC and others with an epistemic conception of causal mechanisms does not constitute the main incentive for adopting it. The major reason for defending an epistemic account, as I will argue in the remainder of this section, is that it captures the meaning of biologists' mechanism-talk in ways that are simply beyond the reach of any single ontic conception of causal mechanisms.

Causal mechanisms are invoked to explain an extremely wide range of phenomena. As Allen (2005, p. 264) indicates, causal mechanism 'can refer to very specific processes, such as the nucleophilic attack by the reactive group of an enzyme on an exposed covalent bond of its substrate, or to a whole category of reactions such as cell signal responses due to protein kinase A (PKA) second messengers'. As the postulation of causal mechanisms has become a virtually ubiquitous practice in biological research, it is practically impossible to define what a causal mechanism is in a way that meaningfully captures all the different uses of this notion, given that the conditions of satisfaction for what counts as a causal mechanism are entirely determined by the context in which it is postulated and on the kind of questions that are asked of the *explanandum* phenomenon. If, as I suggest, the notion of causal mechanism is understood epistemically, then it can be characterized as an explanation where the *explanans* and *explanandum* are sorted out from the context of its formulation. However, if causal mechanisms keep being conceived as 'real systems in nature' (Bechtel, 2006, p. 33), it becomes exceedingly difficult to specify exactly what these 'systems' actually are, not to mention what they all have in common.

Paradoxically, this problem stems from the mechanismic program's desire to closely adhere to scientific practice, given that as long as it remains 'faithful to biologists' own usages' of 'mechanism' (Darden, 2007, p. 142), it cannot fulfil its objective of ontically characterizing this notion in a concrete and unified manner. The reason for this is that there is an unavoidable trade-off between the degree of concreteness of any given ontic characterization of causal mechanisms and the breadth of its applicability. In other words, an ontic characterization of causal mechanisms can only increase its domain of applicability at the expense of sacrificing the concreteness of its formulation. Consequently, the only way mechanismic philosophers could encompass all the different ways in which the notion of causal mechanism is employed in scientific practice would be to propose an ontic characterization so general and so abstract that it would be effectively vacuous.

The recent debate concerning the nature of the causal mechanism of natural selection provides an instructive illustration of this dilemma. Skipper and Millstein (2005) have convincingly argued that none of the major ontic conceptions of causal mechanism successfully captures 'the mechanism of natural selection'. The causal mechanism of natural selection is *not* composed of entities and activities organized to produce regular changes (á la MDC), *nor* is it a series of parts in a complex system interacting to produce a behaviour (á la Glennan), *nor* is it a structure performing a function in virtue of its component parts (á la Bechtel). The different ways in which mechanismic philosophers have dealt with this incompatibility is quite revealing. Glennan (2005b) bites the bullet and concludes that 'there is no such thing as the mechanism of natural selection'. This strategy is problematic because it is at odds with the mechanismic commitment to the 'details of scientific practice' (MDC, 2000, p. 2), given that evolutionary biologists *do* routinely refer to natural selection as a 'mechanism'. Craver and Darden (2005, p. 240) instead contemplate 'whether the account of mechanism should be broadened to allow for stochastic processes and other forms of organization'. Skipper and Millstein (2005, p. 344) also consider this option but decide against it because postulating such a broad conception of causal mechanism 'may not be desirable if it means sacrificing an understanding of the things that make mechanisms distinctive in particular fields, such as molecular biology'. This concern aptly illustrates the danger of vacuity that arises from formulating exceedingly broad ontic characterizations of causal mechanisms. Barros (2008) proposes a third solution, which is to formulate various ontic characterizations of causal mechanism, among them one which can effectively capture the causal mechanism of natural selection. The problem with this strategy is that it means giving up the objective of having a unified conception of causal mechanisms that can be used to make generalizations regarding the nature of mechanismic explanations across biology. In this way, all three proposed solutions are unsatisfactory. However, when we adopt an epistemic view of causal mechanisms, the tensions generated by the efforts to ontically reconstruct this causal mechanism disappear.[15]

Some mechanismic philosophers may object that the thesis that causal mechanisms are epistemic rather than ontic can be refuted on the grounds that biologists often use 'mechanism' to refer to the causal process itself and not (just) to the explanation of it. In response, I would argue that it is very important to understand why biologists use the term 'mechanism' in their research in the first place. The inadvertent conflation of the machine mechanism and causal mechanism senses is once again at the heart of the matter. Mechanismic philosophers tend to assume that using the term 'mechanism' in relation to P (where P is the phenomenon of interest) indicates something distinctive about the nature of P that motivates and legitimates the use of the word 'mechanism' in the context of its explanation. Although this has indeed been the case in the past when mechanists conceived organisms and their parts as machine mechanisms, the ubiquitous appeal to 'mechanisms' by the majority of biologists today is no longer determined by the prescriptive ontological commitments of mechanicism, as I showed in Section 4. Mechanism-talk in contemporary biology is simply a contingent product of history, or as Haldane put it, 'a mere matter of custom'. Consequently, the use of the word 'mechanism' in an ontic sense by some biologists does not demonstrate that causal mechanisms need to be understood as real things. The ontic-epistemic dispute concerning the nature of causal mechanisms will not be settled by simply listing examples of the usage of 'mechanism' in the scientific literature, but by considering how best to make philosophical sense of the role played by mechanism-talk in scientific reasoning and explanation.

7. Conclusion

In this paper I have attempted to clarify the semantic confusion surrounding the concept of 'mechanism' as it is used in biology. I have argued that causal mechanisms—the targets of the new mechanismic program in the philosophy of biology—owe their ubiquity in contemporary biological explanations to the stunning successes of mechanistic investigations in the late nineteenth and early twentieth centuries. Historically, I have claimed that the mechanistic confidence during this period that all phenomena would ultimately be explained in terms of machine mechanisms caused the term 'mechanism' to gradually lose its distinctive mechanistic connotations, becoming a 'dead metaphor' that came to informally signify a commitment to *causal* explanation—no more and no less. Philosophically, I have argued that judging by the way biologists today use this notion, causal mechanisms are better understood as heuristic explanatory devices than as real things in nature, and that the reason why most mechanismic philosophers think otherwise is because they inadvertently transpose the ontic status of machine mechanisms onto their analyses of causal mechanisms. I have shown that by conceiving causal mechanisms epistemically it is possible to come to terms with the multitude of different biological contexts in which they are featured. My examination has also revealed that biologists today who habitually resort to the concept of 'mechanism' in their explanations are not necessarily mechanists, as the contemporary appeal to mechanism-talk neither entails nor derives from the ontological and epistemological commitments of mechanicism. Mechanismic explanations (i.e., explanations given in terms of causal mechanisms) need not be mechanistic; in fact they often deal with population-level phenomena, such as the causal mechanism of natural selection.

As my historico-philosophical analysis of the concept of 'mechanism' has been restricted to biology, it would be interesting to see whether similar analyses in other sciences support or conflict with the conclusions arrived at here for biology, such as the thesis that causal mechanisms are explanations rather than real things. Ramsey (2008) has recently examined the role of mechanisms in organic chemistry, and one of his main findings is that 'Organic chemists take mechanisms to be explanations' (Ramsey, 2008, p. 976) in the form of 'inferences based on observational data' (ibid., p. 972).[16] This suggests that the epistemic account of causal mechanisms that I have defended is probably applicable to other areas of science outside of biology. Expanding the range of perspectives on scientific practice should help provide further insight into the role played by the concept of 'mechanism' across the sciences.

Acknowledgements

I thank Lenny Moss, John Dupré, Paul Griffiths, Staffan Müller-Wille, Maureen O'Malley, Sabina Leonelli, and three anonymous

[15] Kuorikoski (2009) has recently proposed a sort of compromise between ontic and epistemic conceptions of causal mechanisms by proposing two concepts of 'mechanism': an ontic one referring to componential causal systems (like the causal mechanisms of cell biology), and an epistemic one referring to abstract forms of interaction (like the causal mechanism of natural selection). Although I am sympathetic towards this sort of reconstruction, I believe that the inherent problems of the ontic account discussed in Section 5, together with the broad applicability of an epistemic view, justifies defending a general epistemic conception of causal mechanisms.

[16] In fact, the textbook definition of 'mechanism' that Ramsey cites in his analysis closely resembles the epistemic definition of causal mechanism I offered in Section 2. In organic chemistry, a mechanism 'is a *specification*, by means of a sequence of elementary chemical steps, of the detailed process by which a chemical change occurs' (Lowry & Richardson, 1981, p. 174, my emphasis).

reviewers for helpful comments on earlier versions of this paper. I also thank audiences at the ESRC Centre for Genomics in Society at Exeter (UK), the biennial meeting of the International Society for the History, Philosophy, and Social Studies of Biology held in Brisbane (Australia), and the Konrad Lorenz Institute for Evolution and Cognition Research in Altenberg (Austria), for feedback on presentations on this topic. Finally, I gratefully acknowledge financial support from the University of Exeter, where much of the work for this paper was carried out.

References

Alberts, B. (1998). The cell as a collection of protein machines: Preparing the next generation of molecular biologists. *Cell, 92*, 291–294.

Allen, G. (2005). Mechanism, vitalism, and organicism in late nineteenth and twentieth-century biology: The importance of historical context. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 261–283.

Barros, D. B. (2008). Natural selection as a mechanism. *Philosophy of Science, 75*, 306–322.

Bechtel, W. (2006). *Discovering cell mechanisms: The creation of modern cell biology*. Cambridge: Cambridge University Press.

Bechtel, W. (2007). Biological mechanisms: Organized to maintain autonomy. In F. C. Boogerd, F. J. Bruggerman, J. S. Hofmeyr, & H. V. Westerhoff, et al. (Eds.), *Systems biology: Philosophical foundations* (pp. 269–302). New York: Elsevier.

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. London: Routledge.

Bechtel, W. (2009). Generalization and discovery by assuming conserved mechanisms: Cross-species research on circadian oscillators. *Philosophy of Science, 76*, 762–773.

Bechtel, W., & Abrahamsen, A. (2008). From reduction back to higher levels. In *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*.

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton: Princeton University Press.

Bertalanffy, L. v. (1952). *Problems of life: An evaluation of modern biological and scientific thought*. New York: Harper & Brothers.

Bie, B., & Pan, Z. (2007). Trafficking of central opioid receptors and descending pain inhibition. *Molecular Pain, 3*, 1–7.

Brandon, R. N. (1985). Grene on mechanism and reductionism: More than just a side issue. In P. Asquith & P. Kitcher (Eds.). *PSA 1984* (Vol. 2, pp. 345–353). EastLansing, MI: Philosophy of Science Association.

Brandon, R. N. (1996). *Concepts and methods in evolutionary biology*. Cambridge: Cambridge University Press.

Broad, C. D. (1925). *The mind and its place in nature*. London: Kegan Paul.

Bunge, M. (1997). Mechanism and explanation. *Philosophy of the Social Sciences, 27*, 410–465.

Craver, C. F. (2005). Beyond reduction: Mechanisms, multifield integration, and the unity of neuroscience. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 373–397.

Craver, C. F. (2006). When mechanistic models explain. *Synthese, 153*, 355–376.

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. New York: Oxford University Press.

Craver, C. F., & Darden, L. (2005). Introduction. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 233–244.

Darden, L. (2005). Relations among fields: Mendelian, cytological and molecular mechanisms. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 357–371.

Darden, L. (2006). *Reasoning in biological discoveries: Essays on mechanisms, interfield relations, and anomaly resolution*. Cambridge: Cambridge University Press.

Darden, L. (2007). Mechanisms and models. In D. L. Hull & M. Ruse (Eds.), *The Cambridge companion to philosophy of biology* (pp. 139–159). Cambridge: Cambridge University Press.

Darden, L. (2008). Thinking again about biological mechanisms. *Philosophy of Science, 75*, 958–969.

Dawkins, R. (1986). *The blind watchmaker*. New York: W. W. Norton.

Dennett, D. C. (1995). *Darwin's dangerous idea*. New York: Simon & Schuster.

Dobzhansky, T. (1937). *Genetics and the origin of species*. New York: Columbia University Press.

Dupré, J. (2007). *The constituents of life*. Assen: Van Gorcum.

Edin, B. B. (2008). Assigning biological functions: Making sense of causal chains. *Synthese, 161*, 203–218.

Falleti, T. G., & Lynch, J. F. (2009). Context and causal mechanisms in political analysis. *Comparative Political Studies, 42*, 1143–1166.

Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford: Oxford University Press.

Gerring, J. (2007). The mechanismic world view: Thinking inside the box. *British Journal of Political Science, 38*, 161–179.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*, 49–71.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*(1), S342–S353.

Glennan, S. (2005a). Modeling mechanisms. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 443–464.

Glennan, S. (2005b). Is there a mechanism of natural selection? Paper presented at the *International Society for the History, Philosophy, and Social Studies of Biology* meeting in Guelph, Canada, July 2005. (Abstract: <http://www.ishpssb.org/ocs/viewabstract.php?id=215>).

Grene, M. (1971). Reducibility: Another side issue? In M. Grene (Ed.), *Interpretations of life and mind* (pp. 14–37). New York: Humanities Press.

Gould, S. J., & Lewontin, R. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London, 205*, 581–598.

Haldane, J. S. (1929). *The sciences and philosophy*. London: Hazell, Watson & Viney, Ltd..

Haldane, J. S. (1931). *The philosophical basis of biology*. London: Hodder & Stoughton Ltd.

Haldane, J. B. S. (1932). *The causes of evolution*. London: Longmans, Green.

Hempel, C. G. (1966). *Philosophy of natural science*. London: Prentice-Hall International, Inc..

Johnstone, J. (1914). *The philosophy of biology*. Cambridge: Cambridge University Press.

Kauffman, S. A. (1970). Articulation of parts explanation in biology and the rational search for them. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, 1970*, 257–272.

Kuorikoski, J. (2009). Two concepts of mechanism: Componential causal system and abstract form of interaction. *International Studies in the Philosophy of Science, 23*, 143–160.

Lewens, T. (2004). *Organisms and artifacts*. Cambridge, MA.: MIT Press.

Lewontin, R. C. (2000). *The triple helix: Gene, organism, and environment*. Cambridge, MA.: Harvard University Press.

Loeb, J. (1912). *The mechanistic conception of life*. Chicago: Chicago University Press.

Lowry, T., & Richardson, K. (1981). *Mechanism and theory in organic chemistry*. New York: Harper & Row.

Machamer, P. (2004). Activities and causation: The metaphysics and epistemology of mechanisms. *International Studies in the Philosophy of Science, 18*, 27–39.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living*. Boston: Reidel.

McKay, P., & Williamson, J. (2010). Function and organization: Comparing the mechanisms of protein synthesis and natural selection. *Studies in History and Philosophy of Biological and Biomedical Sciences, 41*, 279–291.

McLaughlin, P. (2001). *What functions explain: Functional explanation and self-reproducing systems*. Cambridge: Cambridge University Press.

Monod, J. (1977). *Chance and necessity: An essay on the natural philosophy of modern biology*. Glasgow: Williams Collins Sons & Co Ltd. (First published 1970).

Mossio, M., Saborido, C., & Moreno, A. (2009). An organizational account of biological functions. *British Journal for the Philosophy of Science, 60*, 813–841.

Nagel, E. (1979). *The structure of science*. Indianapolis: Hackett Publishing Company.

Needham, J. (1925). Mechanistic biology and the religious consequences. In J. Needham (Ed.), *Science, religion and reality* (pp. 219–258). New York: The Macmillan Company.

Norkus, Z. (2005). Mechanisms as miracle makers? The rise and inconsistencies of the 'mechanismic approach' in social science and history. *History and Theory, 44*, 348–372.

Railton, P. (1978). A deductive-nomological model of probabilistic explanation. *Philosophy of Science, 45*, 206–226.

Ramsey, J. (2008). Mechanisms and their explanatory challenges in organic chemistry. *Philosophy of Science, 75*, 970–982.

Rosen, R. (1991). *Life itself: A comprehensive inquiry into the nature, origin, and fabrication of life*. New York: Columbia University Press.

Ruse, M. (2005). Darwinism and mechanism: Metaphor in science. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 285–302.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.

Skipper, R., & Millstein, R. (2005). Thinking about evolutionary mechanisms: Natural selection. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 327–347.

Thagard, P. (1998). Explaining disease: Causes, correlations, and mechanisms. *Minds and Machines, 8*, 61–78.

Torres, P. J. (2009). A modified conception of mechanisms. *Erkenntnis, 71*, 233–251.

Varela, F. J., & Maturana, H. R. (1972). Mechanism and biological explanation. *Philosophy of Science, 39*, 378–382.

Wimsatt, W. C. (1972). Complexity and organization. In *PSA: 1972, Proceedings of the Philosophy of Science Association*, pp. 67–86.

Wimsatt, W. C. (1974). Reductive explanation: A functional account. In *PSA: 1974, Proceedings of the Philosophy of Science Association*, pp. 671–710.

Woodger, J. H. (1929). *Biological principles*. London: Routledge & Kegan Paul.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

Studies in History and Philosophy of Biological and Biomedical Sciences 43 (2012) 164–172

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

Is the philosophy of mechanism philosophy enough? ☆

Lenny Moss

Department of Sociology and Philosophy, University of Exeter, Exeter, EX44RJ, UK

ARTICLE INFO

Article history:
Available online 2 July 2011

Keywords:
Mechanism
Carl Craver
Lifeworld
Pleiomorphic ensembles
IUPs

ABSTRACT

Recognition of the widespread use of the word 'mechanism' in bio-molecular research has resulted in the concept of 'mechanism' becoming a focal point for a highly visible group of philosophers of biology. Rather, however, than grasping and elucidating the situated aims and practices of biologists themselves, the philosophical investigation of the contemporary meaning of mechanism in biology has been commandeered by the needs of 'hard naturalists' to replace the old deductive-nomological model of the 'received view' with a new normative-explanatory gold-standard. It is argued that rather than an orientation toward an increasingly precise characterization of mechanisms as being an ultimate end in biological research, in actual biological practice 'mechanism' means different things in different contexts, pragmatically draws on our embodied know-how in the use of machines and is not, nor should be, an ultimate end of biological research. Further, it is argued, that classic work on low-level mechanisms became taken up *qualitatively* as parts of the scaffolding for investigating higher level regulatory processes and that in so doing, and in light of new findings such as that of the regulatory significance of 'pleiomorphic ensembles' and 'intrinsically unstructured proteins' the explanatory limits of the mechanism image have already come into view.

1. Introduction

During the past decade an interest in the practice of identifying mechanisms in the cell and molecular biomedical sciences has emerged (one might say with a vengeance) in the philosophy of biology. This new turn, which has frequently been contrasted with the practice of subsuming phenomena under covering laws, reflects no new developments within the bio-cellular and bio-molecular sciences (where talk of mechanisms is hardly novel) but rather a new perspective amongst what has become a growing number of *philosophers* of science.[1] For many of the latter who have long been seeking adequate epistemic grounds to warrant a philosophical identity structured by way of a non-anthropocentric 'hard' naturalism, the philosophical 'discovery' of biological mechanism provides a new hope. The philosophy of the 'new mechanism' must be understood both in terms of its claims about the life sciences *but also* in terms of how it anchors, or aspires to anchor, a certain philosophical-disciplinary self-conception. More than in any other discipline, a piece of work in philosophy, whatever its specific focus may happen to be, simultaneously and performatively, also constitutes a claim about what philosophy is and ought to be. Work in the philosophy of science is surely no exception. Where a work in the philosophy of science may appear to present itself as little more than a reflective extension of the science itself, one must be especially mindful of its own disciplinary agenda with the possibility, indeed likelihood, of deep-seated divergences in disciplinary aims. One ignores the interests and agonistics of disciplinary self-identification and self-assertion only at the risk of much conceptual confusion. I will be arguing that this divergence of basic aims has played a decisive, even if ostensibly subtle, role in the advancement of a misleading philosophical view about the status of 'mechanisms' in the world of cell and molecular biomedical research. The bulk of the present paper however will not be focused on questions of disciplinary self-understanding (performative or otherwise) but rather

☆ Many readers of this article will recall Van Orman Quine's pithy phrase 'philosophy of science is philosophy enough' (Quine, 1953, p. 446).

E-mail address: Lenny.Moss@exeter.ac.uk

[1] Beginning with the now 'classical' MDC paper (Machamer, Darden, & Craver, 2000) this has become a growing list of 'usual suspects' including but not limited to Glennan (1996), Glennan (2002), Bechtel & Abrahmsen (2005), Bechtel (2006), Craver & Darden (2001), Craver (2006, 2007), Darden (2008) and Woodward (2002).

1369-8486/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsc.2011.05.015

upon critically clarifying the role (or roles) and status of the concept of 'mechanism' in contemporary biomedical research. Issues of philosophical disciplinary identity and self-justification will be revisited in the concluding remarks.

2. The (situated) meanings of 'mechanism'

While a great deal of the philosophy of mechanism literature of the last decade has been devoted to proposing (and debating) the best and most precise definition of what constitutes a mechanism, nothing of this sort is to be found in the scientific literature. Nowhere in the textbooks, in the pedagogy or in the published research literature of the life sciences is there a place where efforts are made to define the necessary and/or sufficient conditions for what counts as a 'mechanism' and yet it is a term that is used freely in the biological research environment. This is not by any means the case for all keywords in the biological-research lexicon. There are concepts in the biomedical sciences with which one gains familiarity through explicit definition and explication and then there are other concepts which one becomes acquainted with through situated use in skilled activity. Examples of the former would include the concepts of a protein or that of the processes of cancer metastasis or of DNA replication. A layperson could well gain relevant 'textbook' understandings of such concepts without ever stepping foot inside of a research environment. The concept of 'mechanism' is an example of the latter. There is no place in an undergraduate course curriculum where, for example, a student is instructed as to the ways in which the term 'mechanism' means something different in the context of a chemistry laboratory versus that of a biology laboratory, nor would a full grasp of this distinction be easily conveyed to a lay person. And yet a young researcher in a biomedical laboratory who displayed a practical failure to grasp an appropriate understanding of references to 'mechanism' would not just be perceived as lacking merely a discrete piece of knowledge, but rather of lacking the kind of basic know-how that is requisite to any competent performance in the research environment. In any living human practical context there are 'background skills' that are so deeply embedded in the fabric of a particular everyday 'lifeworld' that they have become invisible (or 'atmospheric') to the denizens of that world. Where an understanding is so pervasively embedded in the background know-how of a community, and reproduced and inculcated accordingly, its explication requires a kind of 'phenomenological' reconstruction. By a phenomenological reconstruction I mean here an ability to in effect take up the performative stance of a competent practitioner and then make explicit to oneself how one would otherwise unreflectively understand, use, and act upon the concept of 'mechanism' in relevant practical contexts. On the basis of such a phenomenological reconstruction, I suggest that there are three basic senses of the term 'mechanism' that are embedded in the know-how understandings of any competent research practitioner of the cell and molecular biomedical sciences, and further that the wherewithal for combining or separating these three senses is also part of the invisible context-specific capabilities of the skilled researcher. These three senses of mechanism are as follows:

2.1. Mechanism or artifact?

A mechanism is a mechanism *for* something and in a biological context the 'for' must be determined by the living system of interest itself. To count as a biological mechanism the phenomenon in question thus must be perceived as being an expression of the ostensible 'purposiveness' of the living cell or organism. The contrast category, in the context of the biomedical research environment, would be that of an 'artifact'. Should fibrillar structures form in the course of a chemical fixation process (in preparation for electron microscopy, for example) these structures would not be an expression of the purposiveness of the cell and they would thus be classified as *artifactual*. The competent biomedical research scientist intuitively understands this distinction and does not inquire after the 'mechanism' of formation of an artifact (although a chemist certainly might). Likewise if cells stick to tissue culture plastic because of a chemical reaction with the plastic that resulted in the happenstance chemical production of an epoxy resin this too would be registered as an artifact and not as a 'mechanism of adhesion.' Such artifacts are no less chemical, physical or beholding to the laws of nature than are proper biological mechanisms; it is strictly the teleological aspect which makes the difference. Nor, to be clear, does this distinction necessarily rely upon, or assume anything about, the ancestral origins of the function. The differentiae that distinguish between processes that are or are not suitable for investigation as mechanisms are synchronic and are not beholding to any adaptive evolutionary 'seal of approval'. An oncologist seeking to understand the emergent 'mechanisms' of tumor immune evasion certainly does not seek to account for its evolutionary origins. Accordingly, the mechanism/artifact distinction does not rule out the possibility of unearthing a novel mechanism, even one that may have first been set into motion by an experimenter's intervention. If a living system under study responds adaptively to some experimentally induced perturbation then one certainly can speak of the 'mechanisms' of said response. The key is the ability of the investigator to provide evidence of an activity as being functionally embedded within the self-purposive (i.e., 'teleological') totality of the living system (as opposed to just being an 'artifact').

2.2. Mechanism as susceptibility to explanation in a general sense

To refer to a biological event or phenomenon in term of its 'mechanism' is also, however loosely, to categorize it as being physically plausible, that is, as being within the ambit of empirical investigation. There are times when the 'logical grammar' of use of the word mechanism in the biomedical research environment means this and nothing more. Should the claim be made that praying results in a demonstrable regression of malignant tumors (I have heard such claims), the research scientist will, in just this sense, inquire as to the relevant mechanism. What degree of detail and predictive precision, what hierarchical level, and what configuration of components and processes are needed and germane, is entirely contingent upon particulars, and largely beside the point in this case, just so long as a descriptive explanatory framework is provided that meets the practical intuitions of being within the realm of a legitimate (i.e., *empirically plausible*) account. As will be discussed in more depth below, biomedical researchers refer to many things other than mechanisms as the object of their investigative enterprise (e.g., processes, networks, mediators, regulators, signals, etc.), even while assuming that these phenomena are still 'mechanisms' in this relatively vague second sense. By no means are all explanatory objectives specifically identified as 'mechanisms', but when they are it is for a reason (generally having to do with problematizing this second or 'weak' sense of mechanism). It takes the background skill of the competent practitioner to know just what sense of mechanism is being intended in a particular context.

2.3. Mechanism, machines and the projection of embodied know-how

Beyond this weak meaning of 'mechanism' described in Subsection 2.2, the crux of a stronger more specific explanatory use of the concept of a bio-molecular mechanism is that of an

attempt to assimilate something that we cannot directly see, feel, or manipulate with that which we intuitively know best. Biomedical scientists are not referencing Newton's three laws of motion when they describe sub-cellular mechanisms (indeed the vast majority of bio-molecular 'mechanisms' take place in the non-Newtonian environment of the aqueous phase). Unlike for physicists, reference to mechanism in cellular and sub-cellular biology also does not assume a move in the direction of mathematization, which for biologists is by no means the *sine qua non* of something to count as a mechanism. More often than not bio-medical researchers when talking about mechanisms are drawing metaphorically on our everyday intuitions about the workings of machines and our instrumental encounters with medium sized objects in general. The binding of an enzyme to a substrate, or an antibody to an antigen, has long since been assimilated to a 'lock and key' or 'hand in glove' model. What could be closer to our deep-seated everyday practical know-how than 'lock and key' and 'hand in glove' experience? When cell biologists identify as the 'mechanism' of adhesion the binding of cell surface receptors to binding sites on collagen or other 'ECM' (extracellular-matrix) components they are not doing physics; they are drawing on and extending a metaphor that allows a microscopic world to become graspable in terms of our everyday instrumental intuitions. What we know by virtue of embodied, practical know-how, *we know best*.[2] Where and when we can put to good cognitive use the deep reserves of our everyday, embodied know-how we are both intuitively inclined, as well as epistemically well advised, to do so. To assimilate a process or phenomenon to such a schema is to open up a rich reserve of everyday embodied, physical intuitions. On the basis of such practical intuitions one can virtually project one's hands into a model and 'feel' the consequences of acting upon it in some way. The virtuoso investigator can then devise an ingenious way to experimentally turn the physical intuition and dextrous projection into a testable proposition. There is not, *nor can there be*, any universal normative gold standard for the right form of a mechanism-based explanatory model anymore than there can be a plausible hope of exhausting the entire possibility space of embodied human physical intuitions, which will vary both historically with different cultural lifeworld experience and individually with variant personal developmental patterns even within the same cultural space. Further, when we turn our normative focus away from a new Holy Grail quest for a mechanism-archetype and towards an appreciation of the open-ended cognitive and practical benefits of drawing on background know-how, a sobering consequence also comes into view. As explanatory devices, mechanisms are no better than the wealth of projective intuitions that come with them. When, through the concatenation of mechanisms or other complexity expanding means, a schema comes to overwhelm the reservoir of practical intuitions from which it was originally derived, its use-value begins to quickly deteriorate. The flip-side, good news realization then is that we can also come to discover that other kinds of background intuitions may very well similarly become serviceable for our explanatory missions. So too do new embodied practices come to inform new explanatory metaphors and intuitions. These ideas will be elaborated upon below.

3. On the ontogeny (and ontology) of a mechanism

Where does the characterization of a mechanism begin? If students of the 'new mechanism' philosophy imagine that a scientist can just look at a cell and start picking out 'entities' and 'activities' they are badly mistaken. The very idea of a mechanism in biology begins with a holistic pre-conception of a living system as a functional end-in-itself that sustains itself through functional-physical means. Functions are always functions *for* something and cannot veer off into a life of their own (and still be functional). However tacit it may become in practice, the point of departure of any functional, let alone mechanism-based, analysis is the holistic assumption of a unified entity that acts flexibly and contingently to sustain its own existence. Implicit in the very meaning of biological mechanism (as discussed above), the presupposition of a self-sustaining entity (only in relation to which some activity can count as a biological mechanism) is not a contingent feature of a mechanism-based account but is rather an *a priori* feature of any such possible investigation. Whether, and in what way, some process and the material components of said process can be reckoned as a 'mechanism' is never an intrinsic feature of any such process and its components, but is rather a function of its relationship to the living (i.e., self-purposive) system. There are no entities or processes which, taken in isolation, can necessarily be determined to be mechanisms (or parts of mechanisms) nor to be outside of the realm of mechanism (i.e., not the 'stuff' of mechanisms). So, for example, spontaneous decay of various entities exists on many levels, are generally stochastic in nature, and are expressed in terms of an average 'half-life'. On the face of it, neither the spontaneous decay of atoms nor the spontaneous hydrolysis of either small or large biomolecules would appear to count as a mechanism (biologically speaking) for anything. Yet, in the context of, for example, microtubule mediated cell-shape formation, the spontaneous and stochastic hydrolytic decay of high-energy GTP molecules associated with newly polymerized tubulin monomers has been shown to be central to the mechanism of the randomization of tubule outgrowth that ultimately allows a cell to respond selectively to environmental cues in the morphogenesis of its differentiated shape. In other words, in the right context, the stochastic hydrolysis of GTP into GDP would be the correct and appropriate answer to a 'what is the mechanism?' question (Kirschner & Mitchison, 1986). Similarly, as where random Brownian motion is not in an obvious answer to a 'What is the mechanism?' question, in fact, wherever a functional outcome is achieved in a particular context by means of passive diffusion, Brownian motion is very much part and parcel (if not the entirety) of the appropriate mechanism-based account. Random diffusion of receptors in the two-dimensional plane of a lipid-bilayer membrane, for example, is part of the mechanism of a great many receptor-mediated processes.

Methodologically, a transition occurs in the constituting and characterization of a mechanism. Given the holistic context of a living system, which is flexibly responding to perturbations and sustaining its form of existence, one may begin to ask about the ways and means that such ongoing achievements are accomplished. As an analytical grasp of the dynamics of even the simplest living system taken *in toto* is beyond our conceptual ken we attempt to home-in and focus on something simpler and more cognitively manageable. With the identification of a process that is functioning in some way on behalf of the self-sustaining system, some putatively tractable subset of the system is empirically circumscribed while the larger complex whole is methodologically bracketed. In so doing, in essentially freezing a piece of the action for analysis, we proceed provisionally and heuristically to hold some components and activities of an organism as if they were constants. We borrow the purposiveness we identified as intrinsic and provisionally appropriate it as our own—we tell the cell or organism for that moment what its business is and grant ourselves authority to freeze this piece of contingent functionality. We then

[2] This discussion of embodied know-how has been strongly influenced by the various expositions of the phenomenology of expert skill acquisition advanced by Hubert Dreyfus, including his recent debate with John McDowell (see, e.g., Dreyfus, 2002, 2006).

proceed to assimilate an abstracted, pragmatically isolated, zone of activity to the workings of a machine, i.e., to the kind of reality with which we have hands-on, practical familiarity. In this act of practical abstraction and assimilation to that which has a machine-like nature we are treating this piece of nature mechanistically as if it were now, in the manner of a machine, serving *our* purpose. In this act of reification a 'mechanism' is born and baptized.

New knowledge is built on old knowledge. Knowledge is most secure and most supple in its application where it is built upon deeply seated know-how. Assuming a 'mechanistic stance' in the cell and molecular life sciences has surely been extremely productive in opening up new windows and doors onto hitherto inaccessible phenomena but it also comes with some necessary provisos without which methodological gains and epistemic inroads can quickly lead to ontological fallacies. The mechanisms that our hands know so well did not invent themselves, are not part and parcel, both subject and object if you will, of a self-sustaining form of existence and do not have the wherewithal to transform themselves (without notice) into entirely different functionalities. The *stuff* of living things is, at least in potential, *all of that*. When we take a mechanistic stance we are comporting ourselves toward an ontological moving target. We have no warrant for proceeding as if 'our mechanism' is necessarily the use to which these cellular constituents will always be put nor the way in which this function will always be realized in all circumstances and for all time. Recent work on heritable phenotypic plasticity and facilitated variation lend strong credence to the significance of these provisos (see, e.g., Kirschner & Gerhart, 2005; West-Eberhrad, 2003). Many questions follow. How reliable and relevant will the characterization of lower-level mechanisms be in our ability to grapple with the complexity of higher-level systems? Is the characterization of mechanisms largely a transient episode in the unfolding of biomedical science or a permanent fixture? Where does the characterization of mechanisms stand in relation to the leading challenges faced by contemporary biomedical research?

4. What is the *terminus ad quem* of biomedical research?

Since its inception over 35 years ago, the journal *Cell* has been widely regarded as the flagship publication of molecular (and sub-cellular) biology and a bellwether of the guiding trends and insights of the discipline. As one way of gaining a perspective on the centrality of the concept of 'mechanism' in current molecular biomedical research, I carried out a qualitative survey of the key words to be found in the titles of just over a recent year's worth of articles. From the first week of January 2009 through the first week of 2010 there were 27 issues published of the biweekly publication with approximately 15 research and review articles in each issue (and so roughly a total of 400 articles during this span of time). The word 'mechanism' itself was found to be used fairly rarely in the titles of these articles. In only 18 cases, or less than five per cent of the research and review articles during this recent year, was the word 'mechanism' present in the title of an article. The most common keyword used for indicating the nature of the explanatory objective of the article was 'regulation' which appeared in title headings 48 times. The next most frequent key-concept word was 'mediate' (or 'mediation') which appeared 20 times. All told the number of keywords used to capture the explanatory mission of an article was quite extensive and involved both verb and nominal forms. On the face of it, and without helping oneself to any hasty conclusions, the words chosen tend to convey a sense of agency that the word 'mechanism' lacks. The inventory of additional keywords found, expressed in verb form, includes: 'controls', 'stimulates', 'silences', 'governs', 'assists', 'affects', 'suppresses', 'contributes', 'orchestrates', 'induces', 'restores', 'amplifies', 'targets', 'represses', 'organizes', 'coordinates', 'activates', 'signals', 'modulates', 'modifies', 'interacts', and 'allows'. What are we to make of this? Does the teleological flavor of these words refer to the self-sustaining, purposive character of the living system? Might it reflect an attempt to link the activities of parts to the 'goods' of the whole? Just what role should the concept of 'mechanism' play in achieving such an end? Will the concept of 'mechanism' in the strong sense, that is not merely the second sense described above, always have a central role to play in biological explanation or might it be more transient and just reflect a phase in the history of bio-molecular science? Alternatively might the 'discovery' of mechanisms be inseparable from anything we would want to call a '*terminus ad quem*' of biology as such?

4.1. Craver's normative inference

Of those working on the philosophy of the 'new mechanism', Carl Craver in particular has gone to some length to offer an unabashedly normative account of biological explanation. Craver's point of departure (Craver, 2006, 2007[3]) and motivational exemplar has been Hodgkin and Huxley's mid-twentieth century work on the action potential of neurons, and in particular their distinction between a merely predictive model and what they held to be a bona fide *explanatory* account. Hodgkin and Huxley provided a mathematical, biophysical model of the propagation of an action potential in terms of changes in membrane potential associated with changes in membrane permeability to sodium and potassium over a time course, but they insisted that in so doing 'the success of the equations is no evidence in favor of the mechanism of permeability change that we tentatively had in mind when formulating them' (quoted in Craver, 2006, p. 364). Drawing on this exemplar, Craver has set out upon an understanding of the very *terminus ad quem* of biomedical research as such in terms of strongly normative distinctions between 'possibly-how', 'plausibly-how' and 'actually-how' accounts of biological mechanisms. Just as for Hodgkin and Huxley, whose mathematical model only lent itself to a 'possibly-how' account of the biochemical (as opposed to purely formal biophysical) aspects of the propagation of an action potential, so too Craver wants to claim that all explanatory work in the biomedical sciences can be judged along similar, and strongly normative lines. Craver tells us that

> Models are explanatory when they describe mechanisms. Perhaps not all explanations are mechanistic. In many cases, however, the distinction between explanatory and non-explanatory models is that the latter, and not the former, describe mechanisms. It is for this reason that models are useful tools for controlling and manipulating phenomena in the world. (Craver, 2006, p. 367)

Further in his line of argument, 'The core normative requirement on mechanistic explanations is that they must account fully for the explanandum phenomenon. As such, a mechanistic explanation must begin with an accurate and complete characterization of the phenomenon to be explained' (Craver, 2006, p. 368). Curiously, early in the very same paper, Craver acknowledged that 'Few if any mechanistic models provide ideally complete descriptions of a mechanism. In fact, such descriptions would include so many potential factors that they would be unwieldy for the

[3] While Craver's 2007 book elaborates on various aspects of his larger view, such as the 'constitutive' nature of 'mechanistic' explanation, it is entirely in accord with the normative thrust of his 2006 account, which remains a fully adequate and more parsimonious presentation of his core position.

purpose of prediction and control and utterly unilluminating to human beings' (Craver, 2006, p. 360). Why this apt realization of the *de facto* limits of the kind of explanatory model he advocates does not impact upon the steadfastness of his normative claims is somewhat perplexing but suggestive of the possible force of a divergent disciplinary agenda. Be that as it may, Craver proceeds to set out in no uncertain terms a normative framework, a veritable *terminus ad quem* for biomedical research, in terms of the intentions of bringing 'possibly-how' and 'plausibly-how' accounts of biological mechanisms to full and complete 'actually-how' status. Pace Craver, I will suggest that such normative admonitions are actually at variance with the immanent aims of contemporary biomedical research, are inexorably locked into the fate of committing an ontological fallacy, and are ultimately driven not by the goods of biological science at all but by the vicissitudes of a particular disciplinary self-understanding within the world of philosophy. Nor is Craver in any sense an exception or an outlier amongst the philosophical champions of the 'new mechanism.' Rather, and entirely to his credit, it is Craver who has most clearly placed his *normative* cards on the table.

Craver and other advocates of the 'new mechanism' philosophy would be quick to claim that the explanatory objectives of the *Cell* research and review papers, even while expressed in the language of intentional actions such as regulation, mediation, silencing, coordinating, orchestrating, assisting, etc., are still pursued in terms of 'discovering' and characterizing mechanisms. Of course this is certainly the case with respect to the second (weak) sense of mechanism already indicated above. But why the use of such ostensibly colorful language and the absence of explicit reference to 'mechanisms' if anything like Craver's norms of explanation were in force? What the *Cell* title keywords indicate, beyond the background sense of the larger purposive-living context of the inquiries, is an explicit interest in questions of complex *relationality*. Talk of regulation, mediation, orchestration, coordinating, silencing, assisting and the like is about interaction and the interrelatedness of function. Far from describing research programs that are about digging deeper and deeper into increasingly airtight, physically-complete, actually-how accounts of 'mechanisms', the hallmark of leading edge research in the biomedical sciences has become that of seeking to understand how the very many low-level basic pieces of chemistry are responsively, flexibly and contingently weaved together (*orchestrated, mediated, regulated*, etc.) into coherently global responses to developmental and environmental cues, internal and external perturbations. Let us consider another example.

4.2. Plausibly-how mechanisms as constituents of explanatory scaffolding

Far more fundamental to basic biology than even the propagation of an action potential (which is limited to 'excitable' cells) is the process of membrane fusion, which is fundamental to each and every eukaryotic cell. While one may readily think of membrane fusion in the context of synaptic activation and the exocytosis of synaptic vesicle contents at the pre-syntaptic membrane, a membrane fusion event is involved in every step of intracellular transit from the endoplasmic reticulum, through each sector of the Golgi apparatus to the plasma membrane. So too are membrane fusion events at play between endosomes, between endosomes and lysosomes, between mitochondria and between whole cells for example in the cases of muscle syncytium formation or of fertilization of egg and sperm. It should not be difficult to see that what is of paramount biological interest and significance are not the exact details of the fusion event itself so much as the vast complexity of organization and regulation involved in maintaining the specificity of fusion events and of temporal coordination. To be able to achieve this, fusion events are 'mediated by a bewildering number of unrelated molecules' (Martens & McMahon, 2008, p. 554). What is principally understood about the fusion of lipid bilayer membranes is two things: that it requires bringing the membranes into close proximity, and that it involves creating physical stress (to overcome the negentropic resistance to exposing hydrophobic lipids to an aqueous environment) such as through a transient acute curvature that can then be relieved through fusion. Südhof and Rothman (2009) offered a plausible model of how a complex of SNARE and SM proteins may provide the wherewithal, along with synaptotagmin and other auxiliary factors, for the regulated fusion of the synaptic vesicle membrane and the plasma membranes on the pre-synaptic side of the synapse.

A plausibly-how model accomplishes several things. First of all it satisfies the second (weak) criterion of a mechanism-based explanation. It shows that that the kind of phenomenon in question can be accommodated within the general framework of physical/empirical thinking. Second, it provides a kind of conceptual and practical scaffolding for thinking roughly about how various basic categories of biological actions are accomplished. With a plausibly-how model in hand, biologists quickly become interested in the differences that make a difference in real biology. Biological interest in moving from plausibly-how to actually-how characterizations of each and every sub-class of membrane fusion event (mediated as they are by a 'bewildering number of unrelated molecules'), let alone each actual fusion event, is basically zero. What real investigators are seeking, using a plausibly-how model for general scaffolding, are clues to grasping the complexity of overall regulation and hence the use of words like 'orchestration', 'mediation', etc., as opposed to 'mechanism'. When cell biologists first attempted to culture primary mammalian cells *in vitro* they hit a brick wall until they began to explore the ways of better simulating an *in vivo* environment through the use of collagen and other components of a natural, stromal extra-cellular matrix (ECM). The culturing of cells on ECM led to the discovery of whole classes of cell-surface ECM binding cell-adhesion molecules 'CAMs' (such as the 'Integrins') that both mediate cell-surface adhesion and the conveyance of signals across the plasma membrane to within the cell that contribute to subsequent behavioral responses. There are hundreds if not thousands of ways that cells bind to ECM; hammering out the actually-how details of each and every mechanism is of interest to precisely no one. There are tens and probably hundreds of ways that membrane fusion is accomplished by eukaryotic cells. Beyond establishing plausibly-how schemas for however many basic categories of fusion events there may be (the fusion of mitochondrial membranes may well constitute an entirely different category of basic membrane-fusion mechanism, for example) the interest in hammering out all of the actually-how details will approach nil. For both cell adhesion and membrane fusion, both absolutely fundamental categories of living processes, the ongoing interest is and will be in the direction of understanding those highly complex, contextually sensitive and contingent, ostensibly adaptive and 'purposeful' behaviorial outcomes that stretch and challenge the mechanistic paradigm.

5. Infelicities in the philosophy of mechanism

The Craver program (and its extended family of closely related kin) of mechanism-based analysis radically diverges from the practice of contemporary cell and molecular biomedical research along several lines. I will enumerate these divergences as follows:

5.1. On the stratigraphy of bio-medical models

The Huxley-Hodgkin work on the action potential which Craver has used as his exemplar for the right way to conduct biomedical

research was done over 50 years ago. The 2006 book which fellow 'new mechanism' advocate William Bechtel wrote on the history and philosophy of mechanism-based analysis in cell biology likewise ended in work done over 40 years ago. What Craver, Bechtel and others have failed to consider is the possibility that early investigations in cell and molecular biology represented only a phase during which baseline models of basic mechanisms were required in order to provide a scaffolding composed of 'plausibly-how' models that higher level work could then rely upon. A baseline stock of plausible house-keeping mechanisms that every cell or almost every cell relies upon satisfies the second (weak) sense of mechanism (i.e., it supports the plausibility of carrying out an empirical program of analysis in general), it provides a platform for subsequent studies but it tells us nothing about the difference between a fungus and an elephant. The very idea that there could be stages in the unfolding of mechanism-based research in which the founding exemplars, and normative touchpoints of explanatory mechanism undergo transformation and replacement as biological/biomedical research agendas develop and advance is as yet virgin territory within the ambit of the philosophical discourse of the 'new mechanism' community.

5.2. Explanation or description? A question of perspective

Craver's (et. al.) idea that there can be a general criterion for what counts as a *complete* (normatively upstanding), even ideal, 'how-actually' account of a mechanism is a fiction. There are no general criteria for what would count as a complete 'actually-how' explanation, nor even what ever and always distinguishes explanation from description. What counts as an explanation as opposed to 'merely a description' depends on the particulars of an investigative moment. Yesterday's explanation can become today's mere description depending upon needs and interests.[4] Does the binding of a receptor to a ligand explain or describe? To an immunologist it looks a lot like an explanation but to an organic chemist it looks like a description. An organic chemist's model of protein binding steeped in electron orbital theory looks like an explanation to a chemist but only a pragmatically useful description to a quantum physicist. If where the explanatory buck stops when it comes to even a simple ligand binding is already a perpetually moving target, consider how much less warranted it is to speak of a definitive 'actually-how' explanation in the case of a membrane fusion event where even the most basic physics of hydrophobic chemistry is far from secure, let alone the micro-physical implications of each of very many variations in the actual composition of the effective membrane fusion apparatus.

5.3. Molecular machines, pleiomorphic ensembles and counter-factual conditionals

As the research agenda of the biomedical sciences progresses from that of characterizing basic processes that are fundamental to all cells to that of the complex regulatory interactions that determine the differences between one cell type and another, or the same cell in different physiological or developmental states, the very meaning of what counts as a mechanism-based explanation is going to change. Recognition of the complexity of signaling phenomena only expands and never contracts. As more and more factors become seen as irreducibly relevant to any regulatory event our simple mechanistic intuitions become less and less suitable for modeling biological phenomena and we become challenged to cultivate new resources of possible understanding. The hallmark for Craver et al. of the explanatory worth of a mechanism, even one which has been knowingly pared down to basics in order to avoid intractable complexity, is that it can still be shown to support counterfactual conditionals. We can say that mechanism M results in outcome O on the basis of the interaction of components x, y and z when organized in organization Q because if we were to omit any of these three components or reorganize them into organization P outcome O would no longer come to pass but could be fixed by restoring x, y, z or Q. But such simple schemas are no longer even close to being equal to the task of modeling the kinds of interaction systems that have come into view. What has taken the center stage of the state of the art drama of complex signaling systems is performed in a theatre haunted by the specter of everything 'cross-talking' to everything else. Where there are, for all intents and purposes, an innumerable number of different ways for an outcome to be brought about, the aspirations for a causal model that robustly upholds counterfactual conditionals, as championed by Woodward and adopted by Craver and the 'new mechanism' community, becomes utopian. An insightful recent review, by Mayer, Blinov, and Loew (2009) contrasts two kinds of molecular assemblies: the old standard bearer of the classic mechanism metaphor that they refer to as 'molecular machines', examples of which include ribosomes, molecular motors, the nuclear pore complex, flagella and proteasomes, with the emerging conception of 'pleiomorphic ensembles', examples of which include receptor complexes, adhesion complexes, rRNA splicing complexes and trafficking intermediates. Point by point, all the features of the 'molecular machine' model that lent itself so nicely to the projective, know-how intuitions of the workshop (or the kitchen) and to the kindred Woodward counterfactual conditionals are absent or radically transfigured in the case of pleiomorphic ensembles. Where, for example, the former assemblies consist of regular stoichiometric proportions of specific components (with the likely counterfactual conditionals that follow thereof) the latter are non-stoichiometric with proportions variable with circumstance. Where the former engage in specific, discrete interactions, the latter engage in dispersed combinatorial interactions. When, for example, tyrosine residues of membrane-bound platelet derived growth factor (PDGF) receptor dimers become autophosphorylated, the phosphorylated sites become available for binding to and activating 100 different cytosolic effector proteins with phospho-tyrosine binding sites. Autophosphorylation is stochastic, resulting in 500,000 different tyrosine-phosphate configurations of the PDGF receptor dimer, each of which can bind, in theory, to each of 100 different effector proteins. This results in the possibility of 2 billion different activation states for each and every dimerized PDGF receptor complex (Mayer et al., 2009, p. 82). Is this really a venue for strict compositional analyses with counterfactual implications? Where the 'molecular machine' model displays discrete molecular states, the pleiomorphic ensemble displays a full spectrum of molecular states; again where the former requires a complete set of subunits to be functional, the latter is again more complex with subunits often competing with each other and the complex not identified with any particular composition. Those who would like to turn back the clock or dismiss the pleiomorphic phenomena as aberrant or underdeveloped need to be reminded that it is at this level of complexity that creatures become distinct as species and as individuals, endure, perdure and respond to the contingencies of existence on the ground and in real time.

5.4. Minding your IUPs

Hans Jonas had long since pointed out to us that we have neglected the dialectical flip side of the oft-referenced Darwinian

[4] This point is made abundantly clear in an incisive and highly critical editorial by the editors of *Infection and Immunity* (Casadevall & Fang, 2009).

lesson that the theory of natural selection brought us closer to animals, i.e., that animals were likewise brought closer us. We are now poised to embark upon yet another new turn in our dialectical learning process. Where molecular biology once taught us that life is more about the interplay of molecules than we might had previously imagined, molecular biology is now beginning to reveal the extent to which macromolecules, with their surprisingly flexible and adaptive complex behavior, turn out to be more *life-like* than we had previously imagined. As the central focus of bio-molecular investigation has moved from that of those basic processes that are nearly universal in their eukaryotic phylogenetic distribution to that of the signaling and regulatory phenomena that separate and distinguish the developmental pathways of fungi and elephants and enable all organisms to have adaptive, spontaneous and often unpredictable responses to their environments and to each other, a startling new finding has come into view. Over thirty percent of key functional proteins of eukaryotic cells and organisms lack any unique native three-dimensional structure. Referred to as both 'IUPs' (intrinsically unstructured proteins [Gsponer & Babu, 2009] or as 'IDPs' (intrinsically disordered proteins [Uversky, 2010], it is precisely the *lack* of pre-ordained, machine-type, structure that enables these proteins to play key roles in the ongoing real-time construction of global regulatory and response states. This surprising departure from one of our most cherished dogmas about the structure-function determination of proteins now beckons us to rethink our most fundamental assumptions about the nature of on-going regulatory interactions. As with the hominid achievement of the upright posture, IUPs provide a full surface exposure that maximizes possibilities for multiple communicative contact with regulatory and effector binding partners and thus makes them ideal candidates to serve as the platform 'hub' units in scale-free regulatory networks. A protein termed 'p300', for example, involved in chromatin remodeling and transcription regulation is estimated (Gsponer & Babu, 2009, p. 97) to bind with up to 400 different 'partners'. Consider for a moment what it would mean to characterize a single regulatory event involving p300 in 'Craveresque' terms. It would already take a universe of investigators just to elucidate the combinatoric possibility space of n-many different effector proteins being able to bind to subsets of 400 different locations on the protein, and even that would only begin to give us a well-articulated 'possibly-how' depiction of the event. But here the plot only *begins* to thicken. The regulatory proclivities of p300 are an on-going function of a highly contingent history of binding *these* ligands as opposed to *those*. The trajectory of an IUP involved in complex signaling and regulatory dynamics bears a closer resemblance to the vagaries of social group dynamics than it does to the predictable course of a canonically well-oiled machine. Its flexibility and its open exposure make it a centerpiece of ongoing conversations of post-translational modification and de-modifications which in turn set-up and shape the direction of subsequent conversational interactions. The probability space for subsequent regulatory events is an ongoing constructive achievement of a highly contingent prior history.

> Overall the inherent flexibility of unstructured segments in proteins facilitates binding of different enzymes such as kinases, phosphatases, acetyltransferases, deacetyltransferases, methylases, ubiquitin ligases and others to specific post-translational modification sites that reside in these unstructured protein segments. As it is highly likely that many of the PTMS are used in a combinatorial manner, a plethora of effectors may be necessary to read, write or erase the PTM 'code' and mediate the specific biological responses. (Gsponer & Babu, 2009, pp. 97–98)

The shape and exposure and chemical specificities and affinities of IUPs change over time, contingent upon ambient circumstance and upon their prior histories, they quickly sample alternative conformational states and respond flexibly and adaptively to intra-cellular and extra-cellular perturbations. Far from further extending the scope of our machine metaphor into the realm of systematic cell signaling and regulation, we have rather rediscovered, even in the life-history of a single IUP molecule, the complex reciprocal causalities and ostensible purposiveness of Kant's famous blade of grass.[5]

6. The knower and the known

All knowledge may ultimately be rooted in the experience of a 'lifeworld', but lifeworld knowledge is not limited to mechanical intuitions or even to the routine 'everyday'. Where intuitions steeped in everyday know-how are no longer adequate, specialized (sub-)lifeworlds, (be they called schools of thought, paradigms, research styles, research programmes, etc.) emerge and become socially-organized cauldrons of newly cultivated, embedded and embodied skills and know-how. The 'worlds' of the mathematician and the quantum physicist are worlds that are distinctively removed from that of routine everyday intuitions, even if they are unto themselves 'worlds' nonetheless. But even the embodied know-how intuitions of our everyday lifeworld are by no means exhausted by our instrumental, dextrous familiarity with the workings of machines and mechanisms. Most of what we know in an everyday way we know in a very different way. Between mechanical know-how and the explicit give-and-take of reasons there is another sense in which we 'get-to-know' each other, we get-to-know pre-linguistic children and animals, pre-linguistic children and animals get-to-know us and get-to-know each other, and we get-to-know even our own bodies as they change over time and temporarily become foreign to us. Unlike mechanistic intuition, there is an empathetic element to this kind of understanding, and unlike the assumption of philosophers or scientists wedded to a Cartesian epistemology, it is not a kind of knowing restricted to explicit representations.

By means of what kind of knowing can and should the biomedical scientist and biologist approach the kinds of strikingly non-machine like empirical realities revealed in studies of the pleiomorphic ensembles and intrinsically unstructured proteins (IUPs) that constitute the woof and warp of systematic cellular signaling and adaptive regulation? Is it possible that where the 'rods and pistons' of the mechanistic metaphor have run out of steam that a transition to alternative forms of understanding will be in order? Where philosophers of mechanism, largely focused on the science of a bygone era, have yet to recognize this as a problem, it has not escaped the notice of contemporary biomedical investigators.

> Perhaps the most significant barrier to appreciating the dynamic, heterogeneous aspect of signaling complexes is the lack of a good analogy from our daily experience. This contributes to a second related problem, our inability to depict such interactions diagrammatically. Indeed, the typical 'cartoon' of signaling pathways, with their reassuring arrows and limited number of states could be the real villain. (Mayer et al., 2009, p. 816)

Craver, as quoted above, anticipated the problem of signaling complexity in suggesting that a truly complete account of a mechanism 'would include so many potential factors that they would be unwieldy for the purpose of prediction and control and utterly unilluminating to human beings'. Craver's solution and normative

[5] Many thanks to Steve Talbot for introducing me to much of this new literature.

prescription is to squint and see only that which can be sufficiently isolated as to be assimilatable to a toy mechanism. What is lost in the squint are the very interconnections that constitute the flexible, adaptable, breathing, pulsing fabric of a living being. Implicit in Craver's program is both the mechanistic assumption that 'prediction and control' is the bottom line, and that somehow squinting out excess complexity will still get us there. Little bits of machine-like predictability and control can be isolated out of a complex interactive system, but to what end? As contributions to plausibility maps of possible pathways it may have heuristic value, although it may also, as Mayer et al. suggest, mislead with its 'reassuring arrows and limited number of states'. When taken in the aspirational spirit of some *aufbau* of 'actually-how' mechanisms, it becomes a recipe for ontological fool's gold.

Is it really the case that the *terminus ad quem* of the biological sciences just is the ongoing mechanistic struggle, come hell or high water, to gain that predictive power and control over living systems that comes with explanatory accounts that support counterfactual conditionals? Or might it be the case that ultimately biology is about an understanding of life on behalf of the living and in the interest of a co-flourishing of life with life (and non-life)? And if the latter then might it not be the case that while we may have reached some of the limits of the applications of our everyday mechanical know-how, that we have only begun to see the larger relevance of the kind of everyday quasi-empathic, neither mechanical nor strictly rational, know-how that acquaints us with the other life-forms in our life (human and otherwise). Nor would this kind of knowing really be novel in the life sciences (as opposed to just being officially or at least philosophically unacknowledged). Scratch the surface of a hard-nosed ethologist, zoologist, cell biologist, or even botanist or mycologist and you will find someone who has feelings and intuitions for 'their organism' that they cannot account for in rational cum mechanical terms. And if Andrew Pickering is right, than this kind of knowing might even be germane not only to complex living systems (and subsystems) but even to complex machines! In his recent book *The Cybernetic Brain: Sketches of Another Future* (2010), Pickering finds that British cyberneticians such W. Ross Ashby, W. Grey Walter, Stafford Beer and Gordon Pask, each in different ways came to understand our relationship with even machine complexity, not as one even oriented toward prediction and control, but rather toward mutual interactive performances with outcomes that only emerge during the course of an encounter. If indeed the mechanistic paradigm is nothing but the projection and extension of one kind of practical lifeworld know-how into the realm of explicit knowledge and explanation (i.e., natural science), then it would surely follow that where and when machines surpass a certain level of complexity then even non-living entities will exhaust the utility of at least the 'classical' mechanistic viewpoint.

7. Coda: What is at stake for philosophers?

The late twentieth century philosopher Hans Blumenberg (1989) characterized the Copernican Revolution as a form of 'reoccupation'. Through uniting the heavens and the earth and bringing all of nature under the province of mathematics, the Copernican Revolution allowed the human mind to reoccupy that cognitive domain which had become relegated to God alone under the reign of medieval theological absolutism. The cognitive austerity of theological absolutism left a deficiency in compensatory human understanding that could not remain indefinitely stable. The compensatory provisioning afforded by the Copernican Revolution, the very keynote of modern science, was not predicated upon 'mechanism' but upon the in-principle comprehensibility to the mind of Man of a Nature written in the logos of mathematics. If mathematizability had become an onto-epistemic gold-standard, the living organism was hardly in the position, as it had once been for Aristotle, to serve as the exemplar of natural being - but not so that creature of human know-how: the machine. Not only had, by the late middle ages, clock-work mechanism long exemplified the complexity of what machines could do, but astronomical clocks, which date back to the second century BCE, had provided a palpable assimilation of cosmic movement to terrestrial machine mechanism for nearly one and a half millennia. This mechanical bridge between the cosmos and terrestrial physics had only long awaited its epistemic, metaphysical and anthropological moment to be crossed. The further Galilean mathematization of mechanical know-how as the basis of the new terrestrial physics provided the scaffolding, and Newton's achievement in hurling the new terrestrial physics back into the cosmos was the culminating *pièce de résistance*.

If the Copernican Revolution of astronomy set the metaphysical stage for mathematizable mechanical know-how to become the context of objective knowledge, it was Kant's Copernican Revolution of philosophy that provided it with its explicit epistemological warrant albeit on the basis of the 'Critical Philosophy's' transcendental anthropology of the conditions of possible human understanding. But while Kant's philosophical intentions included the securing, on a priori grounds, the warranted basis of the objectivity of mechanistic explanation it was *by no means* limited to such. Indeed, for philosophy to recognize that objectivity has anthropological grounds is at once to unavoidably recognize that philosophy has *not just* the question of knowledge to contend with, but all else that comes with the meaning of being human. It is thus no anomaly that Kant, in a mature work, would pithily aver that all the proper questions of philosophy ('What can we know?', 'How ought we to act?', 'What may we hope for?') can be reckoned under a fourth: 'What is it to be Man?'. If the cognitive 'reoccupation' of the Scientific Revolution gave us a Godly sense of knowledge as objective in its coming from nowhere and everywhere, for Kant objectivity was possible precisely in its coming from *somewhere*. But with that somewhere comes the normative entailments of *subjective* being; entailments that not all subsequent philosophers proved keen to cotton up with. For an intellectual trajectory that included the likes of Boltzano, Frege, Russell, and Vienna Circle thinkers (including Hempel, an important point of reference for contemporary philosophers of the 'new mechanism'), even transcendental subjectivity is *subjectivity too much*. Beginning with the stripping away of Kant's synthetic a priori's, the desubjectivization of objective knowledge, and *not* Kant's own quest for the achievement of the *summum bonum* of human existence, became the *terminus ad quem* of their philosophical enterprise. What has since become known simply as the 'Philosophy of Science' has never been merely some form of circumspect speciality that minds its 'Ps and Qs' when it comes to other 'specialities'; far from it. The 'Philosophy of Science', beginning with the particular bent and intentions of its departure from Kant, has always been about what the nature and aims and methods of philosophy *as such* should (and should *not*) be.

For most contemporary philosophers of science, the covering law or D-N model of explanation was presented early in one's philosophical education as the 'received view' and the standard bearer of any normative philosophy of science. In the D-N/Covering Law model one can still hear the echoes of both the 're-occupational' intentions of modern science to perceive a unifiable rational nature that bears an elective affinity to the mind of 'Man', as well as the efforts of over two centuries of post-Kantian logicians to save and secure the methods of science from the taint of human subjectivity. While the D-N/Covering law model, along with the other pillars of the Vienna Circle legacy, has surely been beaten and battered from many directions, yet as the lynchpin of a

normative theory of scientific explanation, it has been a hard act to follow. For philosophical 'naturalists', or at least philosophical 'hard naturalists', the ability to find warrant for privileging the natural sciences in matters of knowledge, explanation and understanding is of no small moment. Philosophically, there are also *other* games in town. There are, for example, also those descendants of Kant for whom questions of the ultimate goals and norms of human life have remained paramount. There are those for whom phenomenological insight, performative reconstruction, or other sources of knowledge obtained reflectively from within the ambit of the human lifeworld, rightly claim epistemic parity (if not superiority) with knowledge obtained by way of the subject-distantiated posture of a purely empirical approach. As philosophers of science most, if not all, of the leading expositors of the 'new mechanism' are *anti*-reductionist in their specific orientation yet they are *hard naturalists* to the bone.[6] The recognition that biological and biomedical sciences, especially at the investigative level of molecular, cellular and sub-cellular research, make frequent references to 'mechanisms' while very rarely invoking a covering law, presented them with two very different kinds of challenges. As suggested earlier, reference to 'mechanisms' and the lack of reliance upon covering laws is nothing at all new in the life and biomedical sciences, and this surely could not have escaped all previous philosophical notice. Philosophers of science had merely hitherto sidelined biology as a presumptively immature or cognitively under-resourced enterprise. What changed toward the end of the twentieth century was not the epistemology of biologists but the overall status of the biological and biomedical sciences. With its increasing molecular and subcellular prowess, biomedical and biological research had become the behemoth of the scientific research world; it had come to capture the greatest public notice and to boast the presence and participation of many who had become the most celebrated scientists of their generation. Biology could just no longer be ignored... not even by philosophers of science.

For those thinkers who made the turn into philosophically thematizing biological mechanisms, the double-edged sword they faced was this. On the one hand, a whole new horizon of philosophically un- or under-explored material—the reconstruction of the methodology and logos of 'causal-mechanistic explanation'—became sumptuously available. On the other hand, the manner in which this cornucopia of scientific practice had to be elucidated and reconstructed was taken to bear the burden for renewing the warrant of a strong naturalist identity and standpoint no longer provided by the covering law model of explanation. An enormous amount of hard work, scholarship, and intelligence has gone into the 'new mechanism' enterprise and yet I have claimed in this paper that they have, both in seeking to establish a concrete, clearly delimited canonical model of 'a mechanism' and in further using this to clarify and specify the normative basis of biological explanation, misconstrued the nature and status of the 'mechanism' stance and metaphor. But more damagingly, in so doing they actually serve to inhibit, rather than promote, philosophical recognition of the best conceptual, empirical and reflective efforts scientists are currently making for moving ahead. Surely there is much great work that the industry and intelligence of the purveyors of the new mechanism research program could provide if only they could loosen their allegiance to the *terminus ad quem* of the hard-naturalist creed.

Acknowledgements

I want to acknowledge and thank my former student and now co-editor and collaborator Dan Nicholson for both the intellectual drive and appetite that drove me to write what eventually became this paper as part of a triple session on Nature and Normativity we co-organized for the 2009 meeting of the International Society of the History, Philosophy and Social Studies of Biology, as well as for his critical reading and suggestions. I would also acknowledge the benefit of comments offered on the manuscript from Stuart Newman, Anya Plutynski, Justin Garson, and Carl Craver.

References

Bechtel, W. (2006). *Discovering cell mechanisms*. Cambridge: Cambridge University Press.
Bechtel, W., & Abrahmsen, A. (2005). Explanation: A mechanistic alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 421–441.
Blumenberg, H. (1989). *The genesis of the Copernican world (R. Wallace, Trans.)*. Cambridge, MA: The MIT Press.
Casadevall, A., & Fang, F. (2009). Editorial: Mechanistic science. *Infection and Immunity, 77*, 3517–3519.
Craver, C. (2006). When mechanistic models explain. *Synthese, 153*, 355–376.
Craver, C. (2007). *Explaining the brain: Mechanisms and the Mosaic unity of neuroscience*. Oxford: Oxford University Press.
Craver, C., & Darden, L. (2001). Discovering mechanisms in neurobiology: The case of spatial memory. In P. Machamer, R. Grush, & P. McLaughlin (Eds.), *Theory and method in the neurosciences* (pp. 112–137). Pittsburgh, PA: University of Pittsburgh Press.
Darden, L. (2008). Thinking again about mechanisms. *Philosophy of Science, 75*, 958–969.
Dreyfus, H. (2002). Intelligence without representation—Merleau Ponty's critique of mental representation. The relevance of phenomenology to scientific explanation. *Phenomenology and The Cognitive Sciences, 1*, 367–383.
Dreyfus, H. (2006). Overcoming the myth of the mental. *Topoi, 23*, 43–49.
Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*, 49–71.
Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*(Suppl.), S342–S353.
Gsponer, J., & Babu, M. (2009). The rules of disorder or why disorder rules. *Progress In Biophysics and Molecular Biology, 99*, 94–103.
Kirschner, M., & Gerhart, J. (2005). *The plausibility of life*. New Haven: Yale University Press.
Kirschner, M., & Mitchison, T. (1986). Beyond self-assembly: From microtubules to morphogenesis. *Cell, 45*, 329–342.
Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.
Martens, S., & McMahon, H. (2008). Mechanisms of membrane fusion: Disparate players and common principles. *Nature Reviews Molecular Cell Biology, 9*, 543–556.
Mayer, B., Blinov, M., & Loew, L. (2009). Molecular machines or pleiomorphic ensembles: Signaling complexes revisited. *Journal of Biology, 8*, 81.
Pickering, A. (2010). *The cybernetic brain: Sketches of another future*. Cambridge MA: Harvard University Press.
Quine, W. (1953). Mr. Strawson on logical theory. *Mind, 62*, 433–451.
Südhof, T., & Rothman, J. (2009). Membrane fusion: Grappling with SNARE and SM proteins. *Science, 323*, 474–477.
Uversky, V. (2010). The mysterious unfoldome: Structureless, underappreciated yet vital part of any given proteome. *Journal of Biomedicine and Biotechnology, 2010*, 1–14.
West-Eberhrad, M. (2003). *Developmental plasticity and evolution*. Oxford: Oxford University Press.
Woodward, J. (2002). What is a mechanism? A counterfactual account. *Philosophy of Science, 69*(Suppl.), S366–S377.

[6] For example, Darden is well known for her earlier arguments in favor of an 'interfield theory' as opposed to reductionist account of scientific unification, and Craver and Bechtel distance themselves from purely reductive bottom-up accounts of causation in complex systems.

Studies in History and Philosophy of Biological and Biomedical Sciences 43 (2012) 173–181

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

Mechanism and purpose: A case for natural teleology

Denis Walsh

Department of Philosophy, Institute for the History and Philosophy of Science and Technology, The University of Toronto, Toronto, Canada
Department of Ecology and Evolutionary Biology, The University of Toronto, Toronto, Canada

ARTICLE INFO

Article history:
Available online 2 July 2011

Keywords:
Teleological explanation
Mechanism
Emergence
Explanatory exclusion
Explanatory autonomy
Jaegwon Kim

ABSTRACT

This paper develops a naturalised account of teleological explanation. Any such account must be consistent with the completeness of mechanism. The account I offer of teleological explanation is adapted from a prominent approach to the understanding of causal-mechanistic explanations. In order to give a scientific explanation—whether it be mechanistic or teleological—one must provide two things: (i) an invariance relation and (ii) and an elucidating description. In a causal-mechanistic explanation there is an invariance relation between the mechanism and the effect. The elucidating description illustrates how the mechanism produces the effect. In a teleological explanation the invariance holds between the goal (effect) and the means (cause). The elucidating description illustrates the way the means conduces to the goal. Mechanistic and teleological explanations are both complete in their own right, and they are mutually autonomous. One cannot replace the other without explanatory loss. It follows, then, that some natural phenomena—those that contribute to goals—are susceptible of more than one complete, autonomous explanation. Teleology thus demonstrates the falsity of Jaegwon Kim's Explanatory Exclusion Principle.

1. Introduction

Mechanism is a doctrine about the explanation of natural phenomena. It holds that every event has a productive cause, and that every event can be explained by citing that cause. One fundamental commitment of mechanism is that such explanations are exhaustive and complete: exhaustive in the sense that the causal-mechanistic explanation applies to the occurrence of every event, and complete in the sense that a good causal-mechanistic explanation leaves no unexplained residuum. For these reasons, and perhaps more, mechanism is the methodological triumph of modern science. So dominant is it in the natural sciences, so successful has it been, that it seems to exclude the possibility of, or the need for, any other mode of scientific explanation.

In particular, mechanism seems to foreclose on the possibility of teleological explanation. A teleological explanation, in contrast to a causal-mechanistic one, explains the occurrence of an event, or the nature of some entity, by appeal to the goal or purpose that it subserves, and *not* to the mechanism that caused it. Untutored intuition accepts a variety of teleological explanations at face value. They have a particularly important place in folk psychology and what, analogously, we might call 'folk biology'. We readily accept the reason-giving explanations in which the occurrence of an action is explained by appeal to the agent's goal in performing the action. Similarly, it seems uncontroversially explanatory to account for the occurrence of many a biological event by appeal to an organism's goals: thermoregulatory systems make their specific responses in order to maintain proper body temperature; immune systems respond in order to restore the body to health. Citing the way that these physiological responses conduce to the fulfillment of an organism's goals appears to explain their occurrence. But, if mechanism is right, untutored intuition about purposive explanation is wrong. There is no need for purposive explanation because there are no phenomena left unexplained by mechanism. So goes the prevalent opinion in the philosophy of science.

The question whether the existence of genuine purposive explanations is compatible with the completeness of mechanism is an old one. It is the motivating question behind Aristotle's *Physics*

E-mail address: denis.walsh@utoronto.ca
[1] But see Simmons (2001) on Descartes' evident teleological commitments.

1369-8486/$ - see front matter
doi:10.1016/j.shpsc.2011.05.016

Book II. It is the methodological issue informing much of Descartes' biological works (Grene & Depew, 2005).[1] Boyle's mechanism is explicitly formulated to obviate the need for positing goals, or Aristotelian 'forms' in nature (Pyle, 2002). More recently, it has dominated much of the discussion in naturalised philosophy of mind and action. In an influential discussion of the compatibility of purpose and mechanism, Malcolm (1967) asks whether there is anything more to be explained, that requires an appeal to an agent's purposes, once her actions have been given a complete physiological causal-mechanistic account. Finding no unexplained residual effects left over by physiological explanation, Malcolm concludes that there is not. Kim (1989) endorses Malcolm's conclusion that purposive explanations are otiose given the completeness of mechanism. To insist on a role for purposive explanation, in the light of the completeness of mechanism, would be to insist that for some events, their occurrence needs to be explained twice over: once from the perspective of the mechanisms that cause them and once from the perspective of the purposes they serve. But there is no need for explanation twice over: 'say something once, why say it again?'[2]

The commitment to mechanism seems to render teleological explanation worse than otiose; it looks to be incoherent. Mechanism gives us a model of scientific explanation—to explain is to cite causes—which no teleological explanation could instantiate. Teleological explanations explain by citing goals. Goals do not cause the occurrence of their means. For the most part, at the time of the occurrence of means, *C*, to goals *E*, *E* is an unactualised, future state of affairs. Arguably, such states cannot cause anything.

These impediments notwithstanding, I wish to make an appeal for teleological (i.e. 'purposive') explanations. Any attempt to do so, it might be thought, would have to take either one of two tactical approaches. Either it must address the 'otiose' charge; it must deny the explanatory completeness of mechanism, establishing a gap, an unexplained residuum, to be filled in by teleological explanation. Failing that, it might address the coherence charge: if to explain is to cite a causal mechanism, it must show that teleological explanations really are just a cryptic form of causal-mechanistic explanations. Recent naturalism is full of such strategies.[3] I find them unsatisfactory and implausible. Denying the completeness of mechanism, it seems to me, is bootless, while assimilating purposive explanations to causal-mechanistic ones, mischaracterises them. Instead, I attempt to offer an account of teleological explanation according to which teleological explanations are autonomous from causal-mechanical explanations, non-redundant, and yet compatible with the completeness of mechanism. Teleological explanations are autonomous from causal-mechanistic explanations in the sense that we cannot reduce one to the other without explanatory loss. But the completeness of mechanism does not render purposive explanation redundant. For some events an adequate understanding of their place in the natural world requires that their occurrence really is explained twice over: once by appeal to the mechanisms that cause them and once by appeal to the purposes they subserve. One of the consequences of modern science's embrace of mechanism is that the possibility of, indeed the need for, this explanation twice over has been obscured.

I proceed in the following way. In Section 2 I outline what I take to be an uncontroversial analysis of causal-mechanistic explanation. I abstract from recent discussion of mechanism what might be thought of as a generalised form for a legitimate scientific explanation. In Section 3, I develop an account of teleological explanations—explanations that advert to goals—that fits this generalised form just as well as causal-mechanistic explanations do. This form of explanation has the virtue of demonstrating how some events may be susceptible to both causal-mechanistic and teleological explanations. Causal-mechanistic and teleological explanations of the same event have a special relation; they are mutually autonomous. One cannot replace or supplant the other without explanatory loss.

2. Mechanism

Mechanism is the thesis that to explain the occurrence of an event, or the properties of a complex entity, one must cite the mechanisms that cause it. It appears to have two corollaries: (i) causal closure—every event or entity has a complete causal history—and (ii) causal inheritance—the causal capacities of a complex entity are a consequence of the capacities of its parts.[4,5] Together, these entail that the occurrence of every event has a causal explanation, and that the capacities and activities of complex entities can be explained mechanistically in terms of the activities of their parts.

Mechanism is an old doctrine. It finds its roots in the pre-Socratic Atomists, particularly Democritus, according to whom '[...] the properties of the macroscopic world are to be explained on the basis of the micro-properties of the fundamental components of the universe, the atoms and the void in which they move [...] the apparent macroproperties are simply emergent upon suitable arrangements and configurations of atomic qualities' (Hankinson, 1998, p. 203). But mechanism is no mere vestige of early classical science; it has an impressive modern pedigree too. Its most prominent early modern expositors include Descartes, Newton and Boyle. It has even received a more recent upgrade (Bechtel & Abrahamsen, 2005; Craver, 2007; Glennan, 1996; Glennan, 2002; Machamer, Darden, & Craver, 2000).[6] According to this new version of mechanism: 'Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions' (Machamer et al., 2000, pp. 2–3). A causal-mechanistic explanation demonstrates the way in which the entity's engagement in the activity produces the phenomenon to be explained.

> [E]xplanation involves revealing the productive relation. It is the unwinding, bonding, and breaking that explain protein synthesis; it is the binding, bending, and opening that explain the activity of Na+ channels. It is not the regularities that explain but the activities that sustain the regularities (Machamer et al., 2000, pp. 21–22).

The emphasis on activities is of crucial importance here. The understanding that is the hallmark of an explanation is secured not simply by identifying the right entity, or causal capacity, but by specifying the *activity* of that entity that is responsible for producing the effect to be explained. The activities in question are referred to as 'bottom out' activities. A bottom out activity is said to be a 'relatively unproblematic' behaviour of a fundamental entity. The purpose of citing a bottom out activity is to produce an explanation that is 'descriptively adequate'. An explanation is descriptively adequate only if it illuminates the relation between the explanans and the explanandum. That

[2] With acknowledgements to David Byrne.

[3] Much of contemporary action theory, for example, follows the second tack, according to which propositional attitude explanations are recast as a form of causal explanation (see Velleman, 2000). Teleosemantics seeks to recast intentional function as the result of selectional history (Millikan, 1989). Etiological theories of biological function recast the apparent biological teleological explanations as implicitly referring to past episodes of selection (e.g. Godfrey-Smith, 1994).

[4] See Kim (1992).

[5] I do not intend to contest these here. Indeed, I intend to show that these metaphysical commitments notwithstanding, there is still a coherent account of teleological explanation to be given.

[6] My use of 'mechanism' and 'causal-mechanistic explanations' refer to the recent upgrade.

is to say, a descriptively adequate explanation is one that enhances our understanding of the phenomenon to be explained. As I read it, the importance of 'bottom out' activities for causal-mechanistic explanations lies not so much in the fact that they identify the causal relation between explanans and explanandum, but that they 'elucidate' that relation. They provide understanding.

Not just any old description of the activity will do. An activity can only be elucidative if it is described in the right way. It may be correct, but entirely unilluminating to say, for example, that your favourite bottom out activity produces my favourite effect. Without a proper description of the activity and the effect (and the productive relation between them) we are left without an explanation. I suggest, then, that a bottom out activity, is an activity *under a particular kind of description*, namely an 'elucidative' one. An elucidative description is one such that if one has a grasp of the concept by which the activity is picked out, then one understands that the activity *produces* the effect to be explained. The 'bottom out activities' cited by the new mechanists, like binding, bending, and opening are activities of this kind. So, I suppose, are twisting, attracting, and repelling. Grasping the concepts of *twisting*, *attracting*, *repelling*, *binding*, *bending* involves knowing what effect *x* has on *y*, if *x* twists, attracts, repels, binds or bends *y*.[7] Cartwright (2004) calls concepts of this sort 'thick causal concepts' and she notes the central role they play in the new mechanism's conception of explanation. I suggest that what makes a concept a thick causal concept is precisely that it elucidates a productive relation. Taking an activity to be a bottom out activity simply implies that once we understand the activity, we also see that it is productive of the kind of effect to be explained.

The New Mechanism, then, brings to prominence a fairly obvious, but little appreciated, feature of explanation, namely that explanation is *description dependent*.[8] In order to explain the occurrence of an event, it does not suffice just to identify the cause. One must also provide the proper description of the cause that elucidates the relation between cause and effect. This is the import of Machamer, Darden and Craver's claim: 'Intelligibility arises not from an explanation's correctness, but rather from an elucidative relation between the explanans [...] and the explanandum' (2000, p. 22).

Description dependence is important because explanation has a dual nature: it is both factive and informative. An explanation must meet two more or less distinct demands: a metaphysical one and a cognitive one. A successful causal explanation must identify a feature in the world, the explanans event or phenomenon, that causes the effect to be explained. This is the metaphysical demand. But it must do more; it must also enhance our understanding of the occurrence of the explanandum event. This is the cognitive demand. Providing the appropriate description, a bottom out activity picked out by a thick causal concept, discharges the cognitive role.

One important question about causal-mechanistic explanations is what is there about the relation between mechanisms and their effects that suits mechanisms to filling the metaphysical role. One appealing, and popular, response is that mechanisms explain their effects in virtue of the fact that they cause their effects. It is a widespread and popular belief that only causes can do this; causation is necessary for explanation (Salmon, 1984).

Woodward (2002, 2003) proposes an intriguing variant on this answer that will be of particular importance to my project. The relation between a mechanism and its effect is one of *change-relating invariance*. Change-relating invariance is a kind of robust counterfactual relation. Woodward characterises it in terms of interventions.

> [...] what counts as regular "productive" behavior in a part or component of a mechanism? I understand this in terms of the notion of invariance under interventions. Suppose that X and Y are variables that can take at least two values. [...] The intuitive idea is that an intervention on X with respect to Y is a change in the value of X that changes Y, if at all, only via a route that goes through X and not in some other way (Woodward, 2002, pp. S369–S370)

This sort of counterfactual dependence is required for explanation.

> [...] the sorts of counterfactuals that matter for purposes of causation and explanation are just such counterfactuals that describe how the value of one variable would change under interventions that change the value of another. Thus, as a rough approximation, a necessary and sufficient condition for X to cause Y or to figure in a causal explanation of Y is that the value of X would change under some intervention on X in some background circumstances [...] (Woodward, 2003, p. 15).[9]

Intervention is offered by Woodward as a diagnostic mark of invariance, but not as a definition. Whereas intervention is a distinctly causal notion, invariance is not. Invariance is a kind of robust counterfactual regularity. If this is right, then, the relation between a mechanism and its effect, by dint of which mechanisms are suited to explaining their effects, is not essentially causal.[10] There remains at least the possibility that non-causal relations might have the same explanatory property.

Our analysis of the new mechanism has got us this far. A causal-mechanistic explanation comprises two things (i) a change-relating invariance relation between explanans, *C*, and explanandum, *E*, (to be written as '$\langle C, E\rangle$')and (ii) a description, *d*, of the relation between *C* and *E* such that understanding *d* provides an understanding (elucidation) of why, under the circumstances, the occurrence of *c* produces *e*. The mark of causal-mechanistic explanation is invariance plus description. My strategy, below, will be to argue that teleological explanations have the same invariance-plus-description form.

2.1. Explanatory exclusion

One question arising from this analysis of causal-mechanistic explanation is whether it leaves room for any other mode of explanation. Kim (1989) has offered an influential argument to the effect that it does not. If Kim is right, there is little point in pursuing an account of naturalised teleology.

Causal-mechanistic explanation is, according to Kim, both exhaustive and complete. There are no lacunae, or unexplained residua, left by causal-mechanistic explanation. So, in order for an event to have multiple complete causal explanations, it must have multiple complete causes. But, Kim insists, the world is not multiply causally determined. Hence the occurrences of the world are not susceptible to multiple explanations. His considerations lead him to an articulation of the *Explanatory Exclusion Principle* (EEP), which may be roughly stated as follows:

> EEP: No event can have more than one complete and independent explanation

Kim expresses concern that his principle will be seen as 'absurdly strong and unacceptable' (1989, p. 78). Indeed, it has not met with

[7] It's probably no coincidence that these concepts are dispositional ones.

[8] Davidson (1967) makes much of this feature of explanation. According to Davidson the appropriate description is one that represents the event as falling under the covering law that governs the event's occurrence.

[9] It will be evident that unlike Woodward I do not think that such an invariance relation *alone* is sufficient for either causation or explanation.

[10] There are theories of causation, most notably Lewis' (1973) that attempt to reduce causation to counterfactual dependence. We shall see in the next section that there are robust counterfactual dependencies that are nevertheless not causal relations.

general acceptance. Nevertheless, there is an important insight embodied in the EEP, and, interpreted correctly, it does catch onto a widespread view in philosophy of science, and in action theory. That view is that the completeness of causal-mechanistic explanation precludes a role for non-reduced teleological (i.e. purposive) explanation.

As the EEP demonstrates, the concepts of the completeness and the independence (autonomy) of an explanation figure crucially in Kim's treatment for explanatory exclusion. The analysis I have given of causal-mechanistic explanation suggests an account of each. They are much different from those implicit by Kim's treatment. Nevertheless, they suggest a liberalised interpretation of the EEP; one that is more plausible and appealing than Kim's own. The appeal is that this liberalised version of the EEP, while allowing that an event may have multiple causal-mechanistic explanations, appears to preserve the intuition that mechanism excludes teleology.

2.2. Completeness

A causal-mechanistic explanation is complete, on this account, if it is true and descriptively adequate. Such an explanation identifies a cause or mechanism that, under the circumstances, is sufficient to bring about the effect to be explained so that no further causal facts need to be adduced to explain the effect. Furthermore, the adequate description suffices to furnish us with an understanding of how the mechanism (the entity undergoing its activity) produces the effect. When an explanation is complete in this sense it needs no augmentation by further causes or extra information to demonstrate how the mechanism produces the effect. It is reasonable to accept such an explanation without having to look for further causes or elucidative description.[11] Admittedly, this is a weaker conception of completeness than Kim has in mind.[12]

Kim often uses his stringent conception of completeness to argue for a robust kind of reductionism. His argument also requires the *Causal Inheritance Principle* (CIP):

> CIP: If [C] is instantiated on a given occasion by being realized by [$c_1, \ldots, c_n$], then the causal powers of *this instance of* [C] are identical with [...] the causal powers of [$c_1, \ldots, c_n$].[13]
> In other words, higher states are to inherit their causal powers from the underlying states that realize them. (Kim, 1993, p. 355; emphasis in original)

Given CIP, then for some macrostates C and E, if C causes E and $c_1, \ldots, c_n$ are the microstate realisers of C, then C inherits its causal capacities from $c_1, \ldots, c_n$. In this case $c_1, \ldots, c_n$ both cause E and conjointly determine the capacity of C to bring about E. Kim maintains, somewhat controversially, that the real causal work is being done *exclusively* by the microstate properties $c_1, \ldots, c_n$. In that event, he avers, the explanatory work should also be done exclusively by $c_1, \ldots, c_n$. The microstate realisers, $c_1, \ldots, c_n$, and their activities are sufficient to cause and hence to explain E. Thus, the system-level capacities of C are causally and explanatorily otiose. To insist that the macrostate, C, also fully explains E would be to insist on complete explanation twice over, which in turn, according to Kim, entails causation twice over.

The new mechanists typically eschew this brand of mereological reductionism. They point out that scientific explanations legitimately invoke invariance relations at a variety of levels of organisation (Craver, in press). The resistance is easily motivated. If some effect, E, is caused by the activities of some complex entity, C, then typically E can be explained *either* by citing the activities of C, *or* by citing the concerted activities of the parts of C, $c_1, \ldots, c_n$. After all, a complex entity is constituted of its parts, and the activities of a complex entity are constituted of the activities of the parts. So if $c_1, \ldots, c_n$ causes E, and constitutes the causal powers of C, then plausibly C causes E too. In that event, if there is an invariance relation between the capacities and activities of a system's parts $c_i, \ldots, c_n$ and effect E, there is an invariance relation C and E. Each meets the conditions outlined above for figuring in genuine explanation. Each relation may be susceptible to a different elucidating description, one that adverts to the bottom out activities of the system as a whole (in the case of $\langle C, E\rangle$), and another that adverts to the bottom out activities of the parts (in the case of $\langle c_i, \ldots, c_n, E\rangle$). If so, we have two distinct explanations of the same phenomenon. We might, for example, legitimately explain the popping of a balloon by adverting to the increase pressure of the gas inside, or by the bombardment of the membrane by gas particles (Jackson & Pettit, 2004). Each of these explanations (properly filled out) is complete, in that it is true and descriptively adequate. Each identifies a bottom out activity—pressure in one case, impact of molecules in the other—that elucidates the relation between the explanans event and the explanandum effect. Neither leaves an unexplained residuum. Mechanism is thus consistent with a form of explanatory pluralism, or 'explanatory ecumenism' (Jackson & Pettit, 2004).

The important point, for our purposes, is that as long as we are willing to interpret explanatory completeness as truth invariance plus descriptive adequacy, mechanism is compatible with explanatory pluralism.

Explanatory pluralism looks to be a straightforward violation of Kim's EEP. Here the phenomena in question have at least two complete causal explanations. Mechanistic explanatory pluralism is generally thought to be incompatible with explanatory exclusion (Shapiro & Sober, 2007). Sober (2003), for example, contrasts his version of pluralism with the explanatory exclusion of Putnam (1975) and Kim:

> Putnam, like Jaegwon Kim, thinks one has to choose between the micro- and macro-accounts. Kim argues that the causal action is to be found solely at the micro-level; Putnam contends, as we have seen, that the explanatory action is at the macro-level alone. In fact, both of these monolithic positions are mistaken; there is no need to choose. Both micro and macro provide true descriptions of the causal facts, and both thereby provide true causal explanations. (Sober, 2003, p. 210)

It would be a mistake to conclude from this that mechanism is antithetical to the Explanatory Exclusion Principle. The EEP proscribes multiple complete *autonomous* explanations of the same occurrence. It follows that the existence of multiple complete causal-mechanistic explanations of the same phenomena is inconsistent with exclusion, only if these explanations are autonomous. Again, the new mechanism suggests a plausible, if liberalized, account of what makes explanations autonomous.

2.3. Autonomy

Recall that a causal-mechanistic explanation must provide both an invariance relation and an elucidating description. In the cases discussed above, where there are two distinct explanations, one at the macro-level and another at the micro, each has its own invariance relation: $\langle C, E\rangle$ in the macro case and $\langle c_1, \ldots, c_n, E\rangle$ in the micro explanation case. Even though these are distinct explanations, there is a relation of mutual dependence between them.

[11] But it may also be reasonable to seek further explanations.
[12] In places, Kim suggests that a causal explanation is complete when it cites *all* the causal antecedents of the event to be explained (Kim, 1993).
[13] I have taken the liberty of changing the variables to conform to the conventions used throughout this paper.

One is simply the macro-state realisation ($\langle C, E\rangle$) of the microstate relation ($\langle c_1, \ldots, c_n, e\rangle$); the other the is micro-state realiser of the macro-relation. The metaphysical conditions for a successful explanation are met in the micro case if and only if they are met in the macro-case. In this sense the invariance relations are mutually non-autonomous. We could not intervene on one without also intervening on the other. Because of this, it is fair to say that the respective explanations in which they figure are also mutually non-autonomous.

In virtue of being mutually non-autonomous, macro and micro-level explanations of the same phenomena have a special feature: one can replace the other without explanatory loss. That is to say, for every complete explanation at one level, there is a corresponding complete explanation at the other level. Either is sufficient to render a complete (as in true, descriptively adequate) account of the explanandum event.

This relation is nicely underscored by a thought experiment due to Nozick (1980) and elaborated by Dennett (1987). These authors ask us to imagine that there is a race of smart aliens who know the physical principles governing our world, but lack our concepts of macro-level phenomena. Every causal phenomenon we describe as, say, one billiard ball impacting another, or as a volume of gas exerting a pressure, they describe in terms of the activities of sub-atomic particles. They could explain all the phenomena we explain by use of our macro-level concepts by adverting only to the activities of micro-level causes, with no unexplained residuum. They would have no need to couch their explanations in terms of the macro-level relations and activities that we commonly employ. Their understanding of our world would not be augmented by learning our macro-level concepts. The content of every one of our macro-mechanistic explanations could be captured by their micro-mechanistic explanations. Everything that we can explain by citing macro-level mechanisms, smart aliens could explain by adverting to the micro-state realisers of those mechanisms. A smart alien could replace all our macro-mechanistic explanations with micro-mechanistic explanations and incur no explanatory loss.

This is not to pronounce on the desirability of *our* doing so. Any argument for or against mereological reduction of this kind must be made on the grounds of some explanatory *gain* to be had by choosing one level over the other (not merely on the lack of explanatory loss). There may pragmatic reasons for choosing one level over another, but, in my view, there are no arguments for replacing macro-explanations with micro-explanations, or for declining to do so, that appeal to their comparative status as explanations in good standing.

My intention here is simply to demonstrate that the causal-mechanistic explanations of the same effect that advert to mechanisms at different levels of organisation are compatible because they are mutually non-autonomous. Because of this, the existence of multiple causal-mechanistic explanations of the same effect is consistent with a sensibly liberalised version of Kim's EEP. Explanatory pluralists need not resist the EEP, when properly interpreted.

2.4. Emergentism

Liberal it may be, but the revised interpretation of the EEP still appears to foreclose on any form of robust emergentism. Emergentism is the view that the properties of complex entities have an indispensable, irreducible role to play in the explanation of natural phenomena (Walsh, in press-a). According to emergentists, there are phenomena that can be explained by adverting to the properties of complex entities that cannot be explained by adverting to the activities of their parts.[14] In our terms, emergentism is the thesis that the properties of complex systems figure in explanations that are complete in their own terms, and wholly *autonomous* from explanations that advert to those systems' parts.[15] Emergentism entails the negation of the EEP, even in its liberalised form.

Kim suggests that for there to autonomous emergent explanations, it would have to be the case that a complex system had the capacity to bring about effects that the concerted activities of its parts could not. But, according to an influential argument, again from Kim (1989, 2006), this is incoherent. Kim's argument relies on the CIP, outlined above. It goes as follows: Suppose that some complex entity has autonomous causal powers, that is to say it that C has the capacity to bring about some effect E, and yet its parts $c_1, \ldots, c_n$, do not. C would then have to have the capacity to change the causal capacities of the parts, $c_1, \ldots, c_n$, because, by the CIP, C could not bring about E unless the parts, $c_1, \ldots, c_n$ could too. Kim calls this capacity of a complex system to confer causal powers on its parts 'reflexive downward causation'. Reflexive downward causation is incoherent, according to Kim, because if C has this capacity, then, again, by the CIP, the parts, $c_1, \ldots, c_n$, must too. So, the causal autonomy of complex entities requires that the entities' parts both have and do not have the causal powers of the complex entity.

This may be an effective argument against emergent causation—complex wholes do not have causal autonomy over their parts—but it is not effective against emergent *explanation*. In order to be effective as an argument against *explanatory* autonomy, it needs the further lemma—*viz.* that causal autonomy is necessary for explanatory autonomy. But, as we shall see below, this lemma is not true. Explanatory autonomy entails causal autonomy only if the emergent properties complex entities enter into emergent explanations *as causes*. To date, no argument has been offered to the effect that emergent properties of systems could not enter into explanations as *non-causal* explanantia. The principal weakness in Kim's argument is that it relies too heavily on a causal-mechanistic conception of explanation. It presupposes that any autonomous emergent explanation must be a causal-mechanistic one. There may be no autonomous, emergent causal explanations, but there are, I contend, autonomous, emergent non-causal explanations. Goals-directed systems provide us with an example.

3. Teleology

A teleological explanation is one that explains the nature or activities of an entity, or the occurrence of an event, by citing the goal that it subserves.[16] A system has goal, E, just in case it exhibits goal-directed behaviour toward E. Goal-directed behaviour is a gross property of a system as a whole. A system is goal-directed just if it approaches and maintains its end-state in a particular way. Typically goal-directed systems exhibit three kinds of properties: persistence, plasticity (Sommerhof, 1950), and, (I would recommend adding), repertoire. By 'persistence' Sommerhof means that a goal-directed system will persist in the pursuit of its goal across a wide range of perturbations. By 'plasticity' he means that a system has the capacity to respond to its circumstances in a manner appropriate to the attainment of those ends. By 'repertoire' I simply mean that on any occasion, in any given circumstance, the system has the capacity to produce an array of responses to occurrent conditions. Some, typically small, subset of those responses comprises those that are

[14] This is a minimal commitment of 'weak emergence' (Clayton, 2006).

[15] Walsh (in press-a) calls this 'explanatory emergence' and contrasts in with various brands of ontological emergence.

[16] In what follows I shall concentrate on the explanation of occurrences, with the understanding that these considerations adduced here apply equally *mutatis mutandis* to the teleological explanation of the properties and activities of entities.

conducive to the attainment of the system's goals. Goal-directed systems exhibit a bias toward the goal-conducive elements of their repertoire. A goal-directed system has the capacity to marshal the causal capacities and activities of its component parts in such a way that it is capable of producing its goals, through an array of means, across a range of circumstances. That is to say that given a goal, *E*, of a system, and a set of background conditions, the system reliably produces from its repertoire those activities, *C*, that bring about *E*. Having a goal is a system-level, emergent, property in that a system may have goals whether or not its parts do. These goals, I shall argue, underwrite emergent explanations of events that are complete in their own right and autonomous from the causal-mechanistic explanations of these same events.

Organisms and rational agents are the very paradigms of goal-directed systems: 'you cannot even think of an organism [...] without taking into account what variously and rather loosely is called adaptiveness, purposiveness, goal seeking and the like' (Von Bertalanffy, 1969, p. 45). Organisms are self-building, self-organising, highly plastic entities. They have enormously rich phenotypic (West Eberhard, 2003), developmental (Kirschner & Gerhart, 2005) and behavioural repertoires. Similarly, to be a rational agent a system must have a wide behavioural repertoire, and the capacity to marshal that repertoire in pursuit of one's cognitive and conative goals. Organisms and agents are the systems for which teleological explanation is appropriate, and, I maintain, indispensable.

The case for teleology as an autonomous, non-mechanistic mode of explanation begins from the account of causal-mechanistic explanation outlined above. There we saw that an explanation comprises two parts: a robust, counterfactually supporting invariance relation and an elucidating description of that relation. The first step in rehabilitating teleology involves demonstrating that teleological explanations can have this very structure.

The relation between a goal and the means to its attainment has the same invariant structure as the relation between a mechanism and its effect, but with a twist. In a mechanistic system the effect counterfactually depends upon the cause; in a teleological system, the cause (or means) counterfactually depends upon the effect (goal). If we intervene on a mechanism, while holding the background conditions constant, the effect changes in a systematic way. Similarly, if we intervene to change the goal of a system, under constant background conditions, the means toward its attainment would differ in a systematic way. The means produced would be those that contribute to the newly altered goal. In each case—i.e. in mechanistic and teleological invariances—the counterfactual relation is robust. This means that holding a mechanism constant, and intervening on the initial or background conditions, produces a systematic difference in the effect. Similarly, holding the goal of a goal-directed system constant and varying the initial or background conditions, produces a systematic change in the means. The change is systematic in that the different means, under the changed circumstances, will lead to the attainment of the goal.[17] So mechanistic systems and goal-directed systems each exhibit a robust invariance relation. The principal difference is that the direction of counterfactual dependence is reversed. In a mechanistic system effects counterfactually depend on causes. In a goal-directed system, the causes counterfactually depend on the goals.[18]

Now we begin to see the relevance of the preceding discussion of the structure of causal-mechanistic explanation. The metaphysical requirement of an explanation—robust invariance—is met by the relation between a goal and its means, in the same way that it is met by the relation between a mechanism and its effects. If goals instantiate a relation with their means of the same form as the relation that mechanisms instantiate with their effects, and it is by virtue of this relation that mechanisms explain their effects, then there is at least the prospect that goals might explain their means.

But an explanation needs more than an invariance relation. It also needs an elucidating description. In the case of causal-mechanistic explanations, the description invokes a 'bottom out' activity that uses a thick causal concept; such concepts disclose the way that the mechanism *produces* the effect to be explained. In the case of a teleological explanation, the description does not specify the way that the explanandum—the goal—*produces* the explanans. Goals do not produce their means. Goals *require* their means.

This idea that the teleological explanations account for the occurrence of some phenomenon, *C*, by citing the fact that it was *required* for the occurrence of *E*, seems to put teleological explanation beyond the pale of naturalism. They look to be irreducibly normative (Bedau, 1998). However, the sense in which goals require their means has a perfectly non-normative description. Means are hypothetically necessary for their goals in the Aristotelian sense (Walsh, 2007). Activity *C* is hypothetically necessary for goal *E*, under the circumstances, just if, under those circumstance, *C* is the only element (or one of a few elements) of the systems repertoire that conduces to *E*. One thing we know about goal-directed systems is that they are capable of bringing about those activities that conduce to their goals. When we offer a teleological explanation, we describe the way that the mechanism or cause in question *conduces* to the goal.

Conducing, like producing, is a causal relation; means conduce to the attainment of goals only when means cause the attainment of goals. Nevertheless, the descriptions of causal relations that invoke conducing are importantly different from those descriptions that invoke producing. Crucially, to describe how some cause *C* conduces to the attainment of goal *E*, *E* must be designated (implicitly or explicitly) as a goal in the description. Locutions such as 'in order to', 'for the sake of', 'so that', 'for the purpose of', among others, signify that the effect is a goal.[19] Contrast this with the case of the elucidating description in a causal-mechanistic explanation. There the description is strictly neutral on the question whether or not the effect is a goal. Furthermore, conducing descriptions differ from producing descriptions in that while producing descriptions elucidate *how* the effect comes about, conducing descriptions elucidate the fact that *given* that the effect is a goal, *C* is an appropriate way to achieve it.

Demonstrating that *C conduces to* the attainment of *E*, involves demonstrating that, under the circumstances, *C* is an effective way of producing *E*. Once we understand that the system in question can marshal its causal capacities in such a way as to bring about those states and processes that are *conducive* to its goals, then being told that *E* is the goal, and *C* conduces to *E*, is genuinely informative about the occurrence of *C*. It tells us, for one thing, why *C*, occurred rather than any of the other states or processes in the system's repertoire that don't conduce to *E*.

The important differences between the causal-mechanistic and purposive explanations of the dynamics of goal-directed systems are these: (i) the mechanism that produces the effect, *e*, appears essentially in the description of the relation between the cause, *C*, and *E* in the mechanical explanation, but not in the purposive explanation, and (ii) the fact that *E* is a goal appears essentially

[17] This is the sense in which self-organising systems are said to be 'insensitive' to initial conditions. They are not, strictly-speaking insensitive to initial conditions, but robust across a range of such conditions (see Kirschner & Gerhart, 2005).

[18] I do not intend these to be read as exclusive categories. Goal-directed systems are mechanistic systems.

[19] *Cf* Bedau's (1998) insightful claim that the mark of a teleological explanation is that the goal's being a goal appears within the scope of the explanans.

in the description of the relation between *C* and *E* in the teleological explanation, but not in the mechanical explanation.

Consider, for example, the causal-mechanistic and teleological explanations of some particular event of human thermoregulation, say, vasodilation. The causal-mechanistic description must specify that vasodilation produces an increase in the surface of the blood vessel which in turn increases the rate of heat exchange between the blood, the effect of which is to lower the body temperature to 37 °C. This explanation (properly filled out) is complete, in the sense that it needs no further augmentation by causal facts or descriptive content for us to understand that vasodilation produces an increased rate of heat exchange, and a decrease in body temperature. It certainly needs no invocation of goals. Once we know this it is a further and quite different question *why* the thermoregulatory system does this. The conducing description of the same event specifies that vasodilation, is required, under the circumstances, to return the body to its optimum temperature of 37 °C. This explanation, too, is complete in its own right It needs no augmentation *qua* teleological explanation by a specification of the mechanism. Taken by itself, it elucidates the fact that the event occurs because it achieves the system's goals. Once we know this, it is a further, independent and quite legitimate question, *how* it occurs.

We have got this far. Teleological explanations have the same form as causal-mechanistic explanations. They comprise two features: (i) a counterfactually robust invariance relation and (ii) an elucidating description. If causal-mechanistic explanations explain in virtue of the fact that they have this structure, then, by parity, teleological explanations do too. In light of this, it is worth revisiting the question whether the standard arguments against the coherence of teleological explanation are sound. The account I have offered should also occasion a reassessment of the claim motivated by the revised EEP that causal-mechanistic explanation leaves no role for unreduced teleology.

3.1. Coherence

Teleological explanation is a kind of emergent explanation. It adverts to properties (goals) of an entire system. One of Kim's arguments against the coherence of emergence derives from the claim that it requires reflexive downward causation, which I have conceded really is incoherent. So if teleological explanation requires reflexive downward causation, any attempt to justify it is futile. But it is clear that teleological explanation carries no such commitment. In a teleological explanation, goals *explain* the activities of a system's parts. In order for them to do so, there must be a relation of counterfactual dependence between a system's goal and the activities of the parts: the activities counterfactually depend on the goal. Formally all that is required for this relation to hold is that (i) were the goal to be different, the activities of the parts would be too, and (ii) were the goal to be the same, and the initial or background conditions changed, the activities of the parts would differ systematically—they would undertake *other* activities conducive to the attainment of the goal. This relation does not require that the system's goals *cause* the activities of the parts.[20]

They couldn't. Typically, at the time that the parts of a system undertake their activities, the system's goal is an unactualised, future event. Unactualised, future events, as far as I'm aware, don't cause anything. So no downward causation of activities of parts by a system's goals is required.

There is an important relation, however, between the *goal-directed capacity* of a system and the activities of the parts that is distinctively characteristic of goal-directed systems. This relation I call 'reflexive downward regulation' (Walsh, in press-a). It is easily conflated with Kim's reflexive downward causation, but it is importantly different.[21] A goal-directed system typically has a broad repertoire; each of its parts, on each occasion, can undertake a range of activities. On an occasion, the parts of the system undertake the activities they do—rather than other activities that they might—because the ones they adopt are conducive to the attainment of the goal. This is possible because the activities of the parts are regulated by the capacities of the system as a whole. This relation I call 'reflexive downward regulation'. Reflexive downward causation is extremely important. It is this capacity that underwrites the counterfactual dependence of means on goals. But reflexive downward regulation is not the reflexive downward causation that worries Kim so much. The capacity for reflexive downward causation is inherited (in Kim's sense) from the causal capacities of the system's parts, but it does not require that the system as a whole *causes* the parts to have causal properties or dispositions that they would not otherwise have. It simply requires that the *system as a whole introduces a bias into the parts' already existing causal repertoire*. This is entirely coherent.

3.2. Autonomy

The other argument proffered by Kim against emergence is that the autonomy of emergent explanations requires that system-level (emergent) properties are causally autonomous—that is to say, they must have causal properties that the concerted activities of the system's parts do not. That, Kim correctly avers, is inconsistent with the CIP. So, if the autonomy of teleological explanations requires that the goal-directedness of a system is causally autonomous from the activities of its parts, it too is incoherent. Nevertheless, it can be demonstrated that teleological explanation do not require that system's goals are causally autonomous from the activities of its parts.

The case for the autonomy of teleological explanations has two phases. The first demonstrates, *contra* Kim, that properties of entire systems can enter into genuine teleological explanations even if they are not causally autonomous from the capacities of the system's parts. The second demonstrates that teleological explanations are genuinely autonomous over causal-mechanistic explanations in the sense of being irreplaceable without explanatory loss, as I have argued for above.

As we saw above, teleological explanation simply requires an invariance relation between the goal of the system and the activities of the parts; the activities of the parts counterfactually depend upon the goal. This, of course, is consistent with the system's capacity to attain its goal being inherited from the capacities and concerted activities of the system's parts. It is more than merely consistent. It is a requirement on the very possibility of a goal's explaining its means that the concerted activities of the system's parts are capable of bringing about the goal.

All that this establishes, of course, is that goals do not need to be *causally* autonomous in order for teleological explanations to be explanatorily autonomous. More is needed to establish that teleological explanations genuinely are autonomous. As argued above, we need further to demonstrate that the teleological—'*conducing*'—description of the relation between means and goal cannot be replaced, without loss of understanding, by a causal-mechanistic—'*producing*'—description. For any sequence of events *C* and *E*, where *C* causes *E* and *E* is a goal, the relation $\langle C, E \rangle$ will be

[20] Walsh (2010, in press-b) discusses a variety of non-causal, explanatory counterfactual dependence relations.

[21] One possible instance of this conflation is found in Mitchell: 'Furthermore, a type of downward causation is in evidence when higher-level properties initially emerge by means of self-organization then place constraints on the behavior of their constituent parts' (Mitchell, in press, p. 7). I suspect that by 'a type of downward causation' Mitchell merely means what I mean by the term 'reflexive downward regulation'.

susceptible of *both* a teleological explanation and a causal-mechanistic one. So there will be a causal-mechanistic—'producing'—description, and a teleological—'conducing'—description. These descriptions of the same relation are not-intersubstitutable without loss of explanatory content, because they are different kinds of elucidating descriptions. 'Conducing' descriptions cite goals, but do not describe the mechanisms that produce the goals. Conversely, 'producing' descriptions cite mechanisms (entities and activities) that produce their effects; they do not cite goals. Consequently, for any causal relation among means and end $\langle C, E\rangle$, and teleological description of it d_1, there is no causal-mechanistic description, d_2, of $\langle C, E\rangle$ such that d_2 can replace d_1 without explanatory loss. There is an important, and explanatorily relevant fact about the relation $\langle C, E\rangle$—*viz.* that it is a relation that holds between a goal, *E*, and the means for attaining it, *C*, that is lost unless the description invokes the system's goal. A mere causal-mechanistic explanation will fail to account for this important feature of the relation. Similarly, there is an important feature of the relation between *C* and *E*, *viz.* the way that *C produces E*, that is lost by describing that *C* is conducive to *E*. Merely citing the fact that *C* contributes to *E* does not identify the mechanism by which it does so. The causal-mechanistic and teleological explanations of the relation, $\langle C, E\rangle$, each have their own proprietary sort of description, each of which provides explanatorily relevant 'elucidative' information that the other does not. Consequently, these descriptions are not intersubstitutable without loss of explanatory content.

It is instructive to compare the relation between causal-mechanistic and teleological explanations of the same event to that which holds between macro and micro-mechanistic explanations of the same event. Macro and micro-mechanistic explanations of the same event are mutually *non*-autonomous. The reason, as we saw, is that the respective descriptions are interchangeable without explanatory loss. The micro-level description does just as good a job of elucidating the way the cause produces the effect as the macro-level description does. Knowing either the micro-mechanistic or the macro-mechanistic explanation of event *E*, it would make little sense to ask 'yes, but how did *E* happen?' On the other hand, being in possession of a causal-mechanistic explanation of *how C* causes *E*, it might make perfect sense to ask 'Yes, but *why* did *C* occur?' or '*What purpose* did *C* fulfil?'. Conversely, knowing that *C* conduces to *E*, it makes perfect sense to ask *how* it does so. The difference has to do with the kinds of descriptions involved. Replacing a micro-mechanistic explanation with a macro-mechanistic explanation involves substituting one *producing* description with another, each of which is sufficient to elucidate the way the cause produces the effect. Replacing a teleological explanation with its causal-mechanistic counterpart involves substituting a *producing* description for a *conducing* one. The producing description tells us *how E* occurred (it was *produced* by *C*). The conducing description tells us *why C* occurred (it was *conducive* to *E*).

3.3. Exclusion

Where does this leave the Explanatory Exclusion Principle? We saw that it has an implausible strict reading and a plausible liberalised reading. According to the latter, no event has more than one complete, autonomous explanation, but one and the same event may have multiple complete, mutually *non*-autonomous explanations. This liberalised explanatory exclusion is consistent with various extant forms of explanatory pluralism, but is inconsistent with any form of emergent explanation.

If the account of teleological explanation I have offered is correct, then the EEP, even on this weakened reading, is false. There are events—those that contribute to the attainment of goals—for which it is appropriate to offer *both* causal-mechanistic and teleological explanations. Each of these modes of explanation is complete—in the sense of descriptively adequate—in its own right. And they are autonomous. The explanations are non-intersubstitutable. One cannot replace the other without explanatory loss. In a goal directed, purposive system, causal-mechanistic and teleological explanations of the same event are severally complete and mutually autonomous.

If this is right, then we have provided the outline of a case for a natural teleology that is consistent with the new mechanism. One and the same event may be explained in either of at least two distinct ways. But we have also provided some considerations to the effect that the EEP is false. There are many phenomena that need to be explained twice over: once by appeal to their causes (or mechanism) and once by appeal to the goals that they subserve. There are events such that a complete understanding of them requires that we explain how they occur—by citing the mechanisms that cause them—and why they occur, by citing the goals to which they contribute.

Every event in the world has a cause (let us suppose). So, the occurrence of every event can be adequately explained—mechanistically—by citing those causes and describing the relation between the cause and the effect. Some of these events also occur because they contribute to the attainment of goals. The capacity of goal directed entities to bring about the means to the attainment of their goal underwrites an extremely important set of regularities in the world. These facts about the world—*viz. that things happen because they contribute to goals*—are not captured by causal-mechanistic explanations. In order to capture these events as instances of these regularities we must represent them explicitly as instances of the relation between means and goals, and describe the way in which the means conduce to the attainment of the goals.

4. Conclusion

I have attempted to offer an account of naturalized teleological explanation that is consistent with the completeness of the new mechanism. The leading idea arises from an insight that I see implicit in much of the recent literature of causal-mechanistic explanation. On that view, to give a scientific explanation of a natural phenomenon one must do two things: (i) identify the phenomenon to be explained as an instance of an invariance relation, and (ii) provide an elucidating description that provides an understanding of how (or why) instances of the invariance hold.

An invariance relation is a robust counterfactual dependence. I have attempted to show that just as a relation of counterfactual dependence holds between a cause (or mechanism) and its effect, so a relation of the same form holds between a goal and the means to its attainment. These are distinct relations. So for any relation between cause and effect, where *C* is a goal and *E* is a means to its attainment, $\langle C, E\rangle$ will instantiate two distinct invariance relations: one that holds of it as an instance of a mechanical relation, the other that holds it if as an instance of a teleological relation. Each of these invariance relations, in turn, receives a distinct elucidating description. The causal-mechanistic description cites the cause *qua* mechanism and demonstrates how the mechanism *produces* the effect to be explained. The teleological explanation cites the effect *qua* goal and demonstrates that the cause is a means to the attainment of the goal. If this is correct, then causal-mechanistic and teleological explanations share the same form. In virtue of doing so, each ought to count as a *bona fide* mode of scientific explanation.

These two modes are mutually autonomous in the sense that a causal-mechanistic explanation cannot replace a teleological explanation of the same event without explanatory loss—and *vice versa*. The reason lies in the respective contents of their elucidating descriptions. The causal-mechanistic explanation essentially

describes the '*producing*' relation between mechanism and effect. The teleological explanation essentially describes the *conducing* relation between means and goals. These are different forms of description; they convey different information. Consequently, one cannot replace the other without explanatory loss.

The upshot, then, is that despite the completeness of mechanism, scientific practice needs to countenance two non-exclusive kinds of regular occurrences: those that are the consequence of causes and those that are conducive to goals. Each of these sorts of regularities has its own proprietary mode of explanation. Mechanism applies completely and exhaustively to the first. Teleology applies exclusively and exhaustively to the second.

Acknowledgements

I would like to thank Dominic Alford-Duguid and Zachary Irving, and Fermin Fulda for discussions on explanation and Kim's Exclusion Principle. I am pleased to acknowledge the assistance of Lenny Moss and Dan Nicholson in the preparation of this paper.

References

Aristotle. (2007). *Physics* (Translated by R.P. Hardie and R.K. Gray). <http://www.classics.mit.edu/Aristotle/physics.2.ii.html>.
Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in the History and Philosophy of Biology and Biomedical Sciences, 36*, 421–441.
Bedau, M. (1998). Where's the good in teleology? In C. Allen, M. Bekoff, & G. Lauder (Eds.), *Nature's purposes: Analyses of function and design in biology* (pp. 261–291). Cambridge, MA: MIT Press (Reprinted).
Cartwright, N. (2004). Causation: One word, many things. *Philosophy of Science, 71*, 805–819.
Clayton, P. (2006). Conceptual foundations of emergence theory. In P. Clayton & P. Davies (Eds.), *The re-emergence of emergence* (pp. 1–31). Oxford: Oxford University Press.
Craver, C. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford University Press.
Craver, C. (in press). Functions and mechanisms in contemporary neuroscience. *Synthese*.
Davidson, D. (1967). Causal relations. *Journal of Philosophy, 64*, 691–703.
Dennett, D. (1987). Evolution, error and intentionality. In D. Dennett (Ed.), *The intentional stance* (pp. 287–321). Cambridge, MA: MIT Press.
Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*, 49–71.
Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 64*, 605–626.
Godfrey-Smith, P. (1994). A modern history theory of functions. *Nous, 28*, 344–362.
Grene, M., & Depew, D. (2005). *The philosophy of biology: An episodic history*. Cambridge: Cambridge University Press.
Hankinson, J. (1998). *Cause and explanation in ancient Greek thought*. Oxford: Oxford University Press.
Jackson, F., & Pettit, P. (2004). In defence of explanatory ecumenism. In F. Jackson, P. Pettit, & M. Smith (Eds.), *Mind, morality and explanation* (pp. 163–185). Oxford: Oxford University Press (Reprinted).
Kim, J. (1989). Mechanism, purpose and explanatory exclusion. In J. Tomberlin (Ed.), *Philosophical perspectives 3: Philosophy of mind and action theory* (pp. 77–108). Atascadero, CA: Ridgeview.
Kim, J. (1992). Multiple realization and the metaphysics of reduction. In J. Kim (Ed.), *Supervenience and mind* (pp. 309–335). New York: Cambridge University Press (Reprinted).
Kim, J. (1993). The nonreductivist's troubles with mental causation. In J. Kim (Ed.), *Supervenience and mind* (pp. 336–357). New York: Cambridge University Press (Reprinted).
Kim, J. (2006). Emergence: Core ideas and issues. *Synthese, 151*, 547–559.
Kirschner, M., & Gerhart, J. (2005). *The plausibility of life: Resolving Darwin's dilemma*. New Haven: Yale University Press.
Lewis, D. (1973). Causation. *Journal of Philosophy, 70*, 556–567.
Machamer, P. L., Darden, N., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 57*, 1–25.
Malcolm, N. (1967). The conceivability of physicalism. *Philosophical Review, 77*, 45–72.
Millikan, R. G. (1989). Biosemantics. *Journal of Philosophy, 86*, 281–297.
Mitchell, S. (in press). Emergence: Logical, functional, dynamic. *Synthese*. doi:10.1007/s11229-010-9719-1.
Nozick, R. (1980). *Philosophical explanations*. Oxford: Oxford University Press.
Putnam, H. (1975). Philosophy and our mental life. In H. Putnam (Ed.), *Mind, language, and reality* (pp. 291–303). Cambridge: Cambridge University Press.
Pyle, A. (2002). Boyle on science and the mechanical philosophy: A reply to Chalmers. *Studies in the History and Philosophy of Science, 33*, 175–190.
Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.
Shapiro, L., & Sober, E. (2007). Epiphenomenalism—The do's and don'ts. In G. Wolters & P. Machamer (Eds.), *Studies in causality: Historical and contemporary* (pp. 235–264). Pittsburgh: University of Pittsburgh Press.
Simmons, A. (2001). Sensible ends: Latent teleology in Descartes' account of sensation. *Journal of the History of Philosophy, 39*, 49–76.
Sober, E. (2003). Two uses of unification. In F. Stadler (Ed.), *The Vienna circle and logical empiricism: Re-evaluation and future perspectives* (pp. 205–216). Dordrecht: Kluwer.
Sommerhof, G. (1950). *Analytic biology*. Oxford: Oxford University Press.
Velleman, D. (2000). *The possibility of practical reason*. Oxford: Oxford University Press (Reprinted, Velleman, D. (1992). The guise of the good. *Nous, 26*, 3–26).
Von Bertalanffy, L. (1969). *General systems theory*. New York: George Barziller.
Walsh, D. (2007). Teleology. In M. Ruse (Ed.), *The Oxford handbook of philosophy of biology*. Oxford: Oxford University Press.
Walsh, D. (2010). Not a sure thing: Fitness, probability and causation. *Philosophy of Science, 77*, 147–171.
Walsh, D. (in press-a). Teleological emergence. *Synthese*.
Walsh, D. (in press-b). Variance, invariance and statistical explanation. *Erkenntnis*.
West Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford: Oxford University Press.
Woodward, J. (2002). What is a mechanism: A counterfactual account. *Philosophy of Science, 69*, S366–S377.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.