



**HAL**  
open science

# Revisiting darwinian teleology: A case for inclusive fitness as design explanation

Philippe Huneman

## ► To cite this version:

Philippe Huneman. Revisiting darwinian teleology: A case for inclusive fitness as design explanation. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 2019, 76, pp.101188 - . <10.1016/j.shpsc.2019.101188>. <hal-03488290>

**HAL Id: hal-03488290**

**<https://hal.science/hal-03488290v1>**

Submitted on 20 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

Author: Philippe Huneman

Institut d'Histoire et de Philosophie des Sciences et des Techniques, CNRS/Université Paris I Sorbonne

13 rue du Four 75006 Paris

philippe.huneman@gmail.com

**Title: *Revisiting Darwinian teleology: a case for inclusive fitness as design explanation.***

### **Acknowledgements.**

I am hugely grateful to Andy Gardner, with whom I developed many of the ideas of this paper, and who provided me with invaluable comments, as well as materials for elaborating the present account. The present ideas couldn't have been elaborated without him.

I am indebted to Jonathan Birch, Arnaud Pocheville and Sébastien Dutreuil for many useful suggestions and criticism. I also thank two anonymous reviewers whose comments greatly helped to improve the paper.

This work has been funded by the ANR Grant "Explabio", #13 BSH3 0007, and by the Laboratoire International Associé CNRS Paris-Montréal ECIEB

## ***Revisiting Darwinian teleology: a case for inclusive fitness as design explanation.***

**Abstract.** This paper elaborates a general framework to make sense of teleological explanations in Darwinian evolutionary biology. It relies on an attempt to tie natural selection to a sense of optimization. First, after assessing the objections made by any attempt to view selection as a maximising process within population genetics, it understands Grafen's Formal Darwinism (FD) as a conceptual link established between population genetics and behavioral ecology's adaptationist framework (without any empirical commitments). Thus I suggest that this provides a way to make sense of teleological explanations in biology under their various modes. Then the paper criticizes two major ways of accounting for teleology: a Darwinian one, the etiological view of biological functions, and a non-Darwinian one, here labeled "intrinsic teleology" view, which covers several subtypes of accounts, including plasticity-oriented conceptions of evolution or organizational views of function. The former is centered on traits while the latter is centered on organisms; this is shown to imply that both accounts are unable to provide a systematic understanding of biological teleology. Finally the paper argues that viewing teleology as maximization of inclusive fitness along the FD lines as understood here allows one to make sense of both the design of organisms and the individual traits as adaptations. Such notion is thereby claimed to be the proper meaning of teleology in evolutionary biology, since it avoids the opposed pitfalls of etiological views and intrinsic-teleology view, while accounting for the same features as they do.

### **Introduction**

Darwinism is a theory of the process and purpose of adaptation, which is attested by the appearance of design in the living world (Maynard Smith 1968, Leigh 1971, Gardner 2009). In this view, adaptations are driven by cumulative natural selection, which has fitted organisms to their environments by a process of selection of slightly advantageous heritable variations. The concept of fitness intends to capture evolutionary success and, given that organisms feature a Mendelian inheritance, is often measured in terms of gene copies left to subsequent generations. In turn, these traits as adaptations can be understood as being there for the sake of maximising fitness, and, more precisely, "inclusive fitness" - that is, fitness defined by taking into account also the genes left by individuals that are genetically related to the focal actor (an inflexion of the concept unknown to Darwin and due to the fact that organisms often undergo social interactions; Hamilton 1964, Gardner and West 2015). The traditional notion of design, meaning

that organisms and their parts seem contrived toward a common purpose, is therefore explained by natural selection, which provides ways to explain all signs of design (cohesion, robustness, integration, etc.) without a designer. It is therefore legitimate to presuppose that biological traits, as functioning to maximise fitness, will appear as near-optimal answers to environmental demands, even if this assumption has to be a posteriori tested. Along those lines, behavioural ecology, the discipline that studies various traits of animals as ways to adapt to their environment, has since its inception been using this so-called adaptationist methodology, whose heart is the equation between optimisation and maximisation of fitness (Davies et al 2012). It is a massively successful body of science, which explained in detail e.g. the sex ratios of wasps, the foraging strategies of many animals in various environments, the relative investments of organisms in reproduction and survival, or the varieties of mutualistic relationships between plants and insects. As such, behavioral ecology can be seen as a modern science of biological design - even though this whole picture is quite contentious<sup>1</sup>, but the present paper will later on address those worries.

However, in contrast to how behavioral ecologists use and think about Darwinism, there is a persistent view that Darwin's theory abolished the notion of teleology and replaced it with a purely mechanistic study of the dynamics of heritable traits, illustrated by the successes of evolutionary genetics. Some authors who subscribe to this view but don't think teleology should be eliminated in biology - for example because functional statements are overwhelming in biology, and they are teleological - often argue that teleology has therefore to be

---

<sup>1</sup> Birch (2015) among others presents strong skeptical arguments against the role of maximization; some are considered below.

sought elsewhere than in natural selection. For example they would see phenotypic plasticity as the locus that underlies organismic teleology and finally drives evolution (West-Eberhardt 2005, Walsh 2015), or they would theorize the self-organising capacities of organisms as supporting functionality, adaptation and finally teleological aspects of life (Varela et al. 1974, Rosen, 1971, Mossio et al 2009, Shapiro 2011).

Here, I will argue that maximising inclusive fitness actually constitutes the essential meaning of biological teleology. I will defend it against views that oppose to Darwinian thinking another kind of teleology, by showing that all aspects of teleological judgement in biology rely on Darwinian teleology, which I'll capture by defining design on the basis of inclusive fitness; by so doing, I inherit from the view (originally sketched by Fisher) that an intrinsic connection exists between design and natural selection<sup>2</sup>.

I start by considering the status of optimisation in population genetics and evolutionary biology in general, and then by discussing a view of evolutionary biology – “Formal Darwinism” (Grafen 2002) – that proposes on the basis of an interpretation of natural selection some equivalences between population genetics and other approaches in evolutionary theory (section 1). Since it constitutes a framework to think about teleology in Darwinian biology, I'll turn to usual families of views of teleology (section 2), and then indicate some of their weaknesses, showing that neither captures all that is teleological in organisms. Especially, I will argue that the key concepts of *contrivance* and *design*, central in

---

<sup>2</sup> Birch (2015) contests this view and Edwards (1994) rejects this reading of Fisher; however regarding this latter point, even though it's not Fisher's exact meaning, the reading itself exists in the literature and is shared by people like Alan Grafen. For a defense of it see Gardner (2017).

Darwin's explanatory strategy, are not represented in these teleological conceptions (section 3). I will then confront them to the inclusive-fitness based teleology embedded within the Formal Darwinism, show that it recovers both the dimensions of design and contrivance proper to living organisms, and finally argue that it's the most complete account of biological teleology (section 4).

## ***1. Optimisation, design and teleology.***

### *1.1 Maximization and its critics.*

It is usually thought that where pre-darwinian biology saw teleological features, such as the fine adjustment of organisms to their milieu or their apparently designed character, Darwin (1859) proposed a novel account, where no intention or plan would support such features. Cumulative natural selection explains the complex adaptive organs such as the eye, which seemed to be unexplainable by the mere laws of physics, for the chances that these laws would provide such a sophisticated organ were too low<sup>3</sup>. Adaptations such as the beaks of finches, so well designed to seize their prey within the holes where they hide (Lack 1947), were explained by natural selection. In other words, even if organisms look designed, there is no designer: natural selection is enough to do the job. Finally, where people like Von Baer or Owen elaborated the most sophisticated version

---

<sup>3</sup> This argument of the too low chances for a physical explanation is traditional: it can be found in the §63 of Kant's *Critique of judgement*. In his *Only possible argument...* (Kant 1763) he indicates that the eye brings together so many different parts, acting under so many different and independent rules, and in such an adjusted manner, that there is no way to derive this adjustment from the laws of physics; and in the first Introduction to the *Critique of judgement* he sees the statement "the eyes are here to see" as the paradigm of a teleological judgement in biology – a judgement such that, if one does not subscribe to it, any explanation of the item, for instance, any explanation of the eye's structure, is precluded. On Kant's view see Huneman (2017b).

of the idea that all vertebrates, if not all animals, are molded on the same type – namely, the *Ur-typus*, or Owen’s vertebrate type - Darwin showed that the unity of type is due to common descent (Darwin 1859, ch.6) so, against authors from the “transcendental morphology” school (Rehbock, 1989, Balan, 1984), there is no need to conceive of a designer with a plan for organisms. This is the usual reading of Darwin’s take on teleology: while the biological design was related to a designer, and, on the basis of this idea, organisms were explained teleologically, Darwin explains the appearance of teleology by natural selection; the Modern Synthesis added the genetics, and the mathematical apparatus of population biology, but inherited this conception.

We have to notice here a trend within Modern Synthesis thinking that pertains to some very general form of teleology, in the sense that the *end of a process appears explanatory for the process*. First, Wright’s adaptive landscapes are explanatory only insofar that their peaks are evolutionary attractors, so that natural selection drives population towards the most accessible peak. In a naïve language, this means that the fact that a gene combination is optimal in fitness – in some sense - is explanatory for the fact that it is there, whatever the process that led to this fact (in other words, trajectories on a fitness peak are not explanatorily relevant, but what is relevant is only the position of the summit). Second, Fisher’s (1930, 1941) Fundamental Theorem of Natural Selection (FTNS) intended to provide the basis of a teleological viewing, in the sense that, in a naïve reading of the theorem<sup>4</sup>, knowing that a combination of traits has the best fitness is

---

<sup>4</sup> Which is qualified below.

sufficient to predict that the population will reach this trait value, because according to it natural selection is supposed to maximize fitness<sup>5</sup>.

Those two major views of population genetics, Fisher's and Wright's, even if they are conflicting, are however in agreement on the fact that there is *some* optimization through natural selection - optimization corresponding to maximization in the sense that fitness maxima correspond to optimal phenotypes - ; and that optimization can play an explanatory role when one addresses evolutionary dynamics. But optimization is indeed a teleological explanation, in the sense that saying that the trait to be expected is in principle a trait close to the optimal trait means that I consider nature as directed towards reaching an optimum.

It is nevertheless not clear whether the theory of *evolution* by natural selection should incorporate some teleology. As Fisher famously remarks, evolution is *not* natural selection (Fisher 1930 p.1); the FTNS is about *natural selection*: hence it might be that optimization has no role to play in *evolution*. Therefore it is not obvious whether the occurrence of teleology within population genetics underwrites other views of biological teleology, or whether it is plainly orthogonal to them.

---

<sup>5</sup> Let us notice yet that the inspirations of Wright and Fisher about this optimization seem different (see also Winther, 2006; Frank 2012). Wright thinks in terms of optimization in physics: fitness landscapes show optimization in the same sense as wells of potential energy in physics, whereas Fisher's maximizing of fitness is much more complex. Fisher rejects explicitly the idea of potential wells. He indeed draws on physics, but rather on statistical mechanics and on the fact that the positive increase in fitness seems to analytically derive from the nature of the system (a collective of genes) exactly as the second principle of thermodynamics, which is also a necessary law stating a trend in systems, analytically derive from the system of molecules in motion in statistical mechanics.

But Fisher's idea of the FTNS also relies on a parallel with economics - Fisherian fitness is throughout the second chapter explained as analogue to a loan, that the parent would subscribe to, and whose interests are computed in representatives in the grand-children generation. Thus the 'maximization' idea in Fisher borrows from both statistical mechanics and economics (but from not potential wells, as in Wright's understanding).

Thus I will first consider the two aspects of maximization in population genetics as sketched above (Wright's climbing landscapes, Fisher's FTNS) and ask on what basis they help us to make sense of a specific kind of teleological explanation. The first issue one meets in this direction is that population genetics *per se* does not seem to be teleological at all. It is, mostly, a dynamics of gene frequencies. Wright's landscapes may help to visualize what is going on in some situations, but the idea of climbing is not, as such, explanatory - also because real landscapes are highly dimensional and here maximization is more difficult to make sense of (see Gavrilets 2004). On the other hand, Fisher's FTNS does not really state an *actual* maximization of anything, and some doubt it even shows a potential maximization, either because this is wrong (Birch 2015) or because this is not Fisher's intent (Edwards 1994). Granted, the first understanding of the theorem, by Wright and by population geneticists until the 70s, was that population mean fitness increases because its variation equals genetic additive variance, which is positive by definition. Actually since Fisher himself there has been a tradition of interpreting the FTNS as meaning population fitness maximization by organisms.

However, many counter-examples have then been found: for instance, in scenarios involving frequency dependent selection (e.g. hawk-dove dynamics), the mean fitness of the population may decrease over successive generations. Thus, a more accurate reading of the theorem has been proposed later, after Price (1972), then Edwards (1994), which states that the increase in fitness *directly due to natural selection* is positive: yet this says nothing about evolution as such and its internal tendencies. In this perspective, the positive contribution in

fitness change made by natural selection may, and often is, counterbalanced by what Fisher called the “deterioration of environment”, which can also be the change in fitness due for example by population structure change in frequency-dependent selection (which entails that purported counterexamples to the theorem indeed don’t invalidate it but challenge its empirical utility in biology). In this reading teleology, as maximization, plays *no role in explaining the actual dynamics of evolution* (recall Fisher's first line: “Natural Selection is not Evolution”... therefore analytic statements about natural selection may not explain actual evolution). Moreover, as Ewens (2014) indicates, the fact that the increase in mean fitness due to selection is always positive doesn’t imply that there is a maximum, not even a local maximum as in Wright’s fitness landscapes. In other words, the FTNS can’t show that selection intrinsically maximises anything; at least it may say that selection *tends* to maximise, which is very different.

However, when one considers *evolutionary biology in general*, one meets a striking contrast: some biologists, especially behavioral ecologists, make pervasive use of optimizing principles. They would explain traits by the fact that they maximize a proxy for individual fitness: clutch sizes are explained because they maximize a trade-off between investment in offspring number and life expectancy of eggs (Charnov and Krebs 1974), foraging theory is wholly based on the idea that foraging time and foraging behavior in general maximize energy intake (Krebs and Davis 1997), mating strategies of gorillas are explained by their maximizing the amount of offspring with high chances of survival (Dunbar et al. 2001). Overall, behavioral ecology is based on the idea that allelic frequency change underlies the change and fixation of traits, and that natural selection –

acting on pools of genes – optimizes traits with regard to the environment by retaining those alleles that better contribute to individual fitness. However, as indicated above, population geneticists have been reluctant to admit such notions of optimisation.

Hence there is a strange discrepancy within the theory of evolution: biological investigation of the origin or maintenance of traits by natural selection relies on *optimization*, whereas the theory supposed to capture the process of natural selection, focusing on gene frequencies change, is a *dynamical* theory and its practitioners do not support optimization (Gardner 2009).

### *1.2. Formal Darwinism as a conceptual analysis of maximization within natural selection.*

The Formal Darwinism is a framework developed by Alan Grafen to make sense of this difference between these types of theoretical understanding of evolution - optimization and dynamics – and justify the legitimacy of optimization reasoning in behavioral ecology (Grafen 2007, 2014). Grafen has shown that there is a mathematical equivalence between the population dynamics of allele frequencies (that correlate to phenotypes and distinct fitness values), and organisms “choosing” a strategy in a strategy set according to an optimization program, i.e. a function which aims at maximizing some maximand (Grafen 2002). He constructs an isomorphism between Price equation in population genetics and the optimization function whose arguments are strategies “chosen” by the organism. According to it, assuming some strong conditions (no casts, no muta-

tions, no frequency-dependence, etc.), when the derivative of the latter reaches 0 (hence an extremum is reached), the additive genetic variance nullifies and reciprocally.

The maximand of the optimization function that Grafen calls the organism's "objective" - is, most generally, the focal individual's inclusive fitness (Grafen 2006), which is her fitness *plus* the indirect fitness benefits, i.e. the contribution in fitness to the reproductive success of the individuals statistically correlated to the focal individual at a given locus. The measure of this correlation is called *relatedness*: if a trait provides a benefit  $b$  to another individual whose relatedness is  $r$  to the focal individual, then the indirect fitness benefit is  $br$  (Hamilton 1964; Frank 1998, West and Gardner 2013).

Stating this isomorphism does not claim that natural selection always maximizes, which is contrary to what is empirically known; it just means that the structure of both theories, dynamics and optimization, can be intertranslated, the maximizing strategy in the optimization program corresponding to the phenotype of the genotype fixed in the population genetics framework. Grafen establishes links between both - for instance, that when there is no "scope" or "potential" - as he calls possibilities for selection in the strategy set- for selection to optimize phenotypes, populations are at genetic equilibrium. Even though vagaries of the genetic make-up or of the development, which have an impact on the strategy set, may interfere with optimization, there is nevertheless a formal connection between the dynamics of natural selection and an optimisation view with its associated language of purposiveness. Unlike the initial understanding of the FTNS, or the hill-climbing in fitness landscapes, which stood within popula-

tion genetics and faced limits pointed by population geneticists, this equivalence provides a formal foundation for the use of optimization thinking and language in investigating the evolution of traits by natural selection (namely, what Grafen (1984) calls the “phenotypic gambit“ assumed by behavioral ecologists).

### *1.3. Formal Darwinism seen as a conceptual analysis of natural selection.*

A caveat is here necessary. I’m not at all claiming that optimisation is all over the place, or that selection produces optimisation all the time. Rather, the extent to which optimisation is realised in the world is an empirical issue, and selection may often produce suboptimal phenotypes, as it has been empirically established, and theoretically proven by population geneticists. The FD only entails that there is an intrinsic and mathematical relation between selection and optimality, which means that the population genetics and the behavioral ecology levels are not wholly orthogonal.

Birch (2015) convincingly argued that neither Fisher’s theorem (at the population level) nor Grafen’s Formal Darwinism (at the organism level) establish that maximisation is at work with natural selection - either in the sense that population genetics equilibria would be fitness maximal, or in the sense that natural selection drives in principle towards maxima (but may not reach them). Regarding Formal Darwinism, he shows that cases like heterozygote superiority are problematic, since they present a situation where the absence of potential and scope for selection does not correlate with a genetic equilibrium. Indeed, if everyone is heterozygote (think of malaria resistance in sickle cell anemia), there is neither scope nor potential for selection (since all

are optimal) but no equilibrium (since reproduction will introduce novel homozygotes because of Mendel's laws), contrary to the link FD establishes.

I still think that the demonstration of Formal Darwinism is illuminating, in the sense that it sheds light on an conceptual intrinsic relation between natural selection and maximisation. To explicate it with an analogy, such intrinsic relation could be compared to some actions verbs, e.g. *hunting*, which are intrinsically teleological. The fact that these actions are directed towards a goal, e.g. killing preys, is built in the meaning of those verbs. This in turn doesn't entail that the goal is always reached, e.g. that the hunter always kills his targets. This might be very rare: it's an empirical issue; but the meaning of 'hunting' does not depend on those facts of the matter. It might even be a question of the structure of the world preventing hunters to kill. Suppose that we are in a world where preys run much, much faster than the velocity of the reflex reactions and cognitive apparatus of hunters: then, hunters would almost never kill preys, even though hunting would still be intrinsically related to the death of the prey. This latter case corresponds to the decoupling between optimisation and population genetic equilibria exemplified by heterozygote superiority. Therefore, Formal Darwinism does for natural selection the same thing as such grammatical analysis does for the verb 'hunting', namely establishing principled links between concepts.

There are two sets of reasons underlying cases of selection not optimizing phenotypes. The first, which can be extrinsic to population genetics, is about variation and its extent and biases: when variation is very much constrained, then selection has a very small scope<sup>6</sup>. Even though variation biases may be reflected within population genetics, the causes of the developmental constraints yielding such biases are not ex-

---

<sup>6</sup> See Huneman (2017) for a detailed analysis of those relations between direction of variation and scope for selection.

plained in this field, therefore they are exogenous parameters. The second is more intrinsic to population genetics: assortative mating, castes, and in general some particular structures of populations. Such population genetics facts indeed prevent the Formal Darwinism equivalences to hold (Grafen 2002, 2007), and support Birch's skeptical verdict. But as I said, Formal Darwinism informs us about the concept of natural selection as it is used in behavioral ecology and population genetics, thus about organisms and about gene frequencies : it only tells us that "natural selection" refers to a same thing, which in both fields under some conditions nullifies genetic additive variance while it makes phenotypes tend towards inclusive fitness maxima. (Notice that the case of heterozygotes superiority involves genetic variance due to dominance effects and is a more complex case where such nullification is precisely not relevant.)

To some extent, once the conceptual analysis of selection is done as it is, and this abstract link between behavioural ecology and population genetics is explicated, then one can make an inverse reasoning, starting from the fact that in this world much adaptation occurs, then infer that some of the conditions for the optimizing trend built in selection should be realised in this world, which in turn confers some further legitimacy to adaptationist assumptions instantiated by the phenotypic gambit. This is like inferring from the fact that a specific hunter indeed brings back killed animals, to his shooting skills, and then predict he'll bring back much more victims. Of course the prediction is fallible: the hunter may have been lucky; he can't plausibly be lucky everyday, so the prediction based on extended observation is more reliable; but then, he can also run out of preys precisely because he was such a successful hunter, and the prediction will fail. Our inverse reasoning regarding the conditions of optimization may be as fragile. However, it could also be strengthened by considering the amount of species known, the amount of living species hypothesized, and extrapolate

the extent of adaptation in the living world. The fragility of the whole inverse reasoning will therefore be the fragility of such extrapolation.

Notice then that the optimization approach embeds a clear teleological dimension: considering that a trait facilitates an organism's attempt to realize its "objective" - maximising the maximand - explains why it is present, so its objective-enhancing effect is indeed the cause of its being there. As in the classical example of teleology and efficient causes presented in the history of philosophy by Leibniz (*Discours de métaphysique*), there is a mechanism, a dynamics, underwriting the process which leads to an outcome (for example here a trait as an adaptation), but many dynamical trajectories are possible and the mere fact that the trait is maximizing a variable explains its presence (Rice 2012), whatever the dynamics through which it has increased in frequency. Echoing Leibniz, Formal Darwinism provides us with a conceptual and formal pluralist framework for thinking of teleology.

Therefore, from the consideration of population genetics as the science of evolution by natural selection we can, indeed, deduce a teleological explanatory framework, which is defined by the maximization of inclusive fitness: namely, X is teleologically explained means that X contributing to maximize inclusive fitness (of an organism of which X is a property<sup>7</sup>) explains X. This framework doesn't state that all biology is teleological, but suggests what teleology should be if indeed it plays an explanatory role in biology. My claim in the present paper here is that this is the genuine Darwinian teleology, and allows to make sense of many of the several dimensions of teleology, namely physiological functions,

---

<sup>7</sup> 'Property' is to be understood in the largest sense : a trait or an organ or a behaviour are properties, but so are the overall design of an organism, its form, its developmental patterns, etc.

goal-directedness of developmental process (which concerns the whole organism), adaptations of all sorts (Huneman 2006). In the next section I'll review major rival attempts to conceive of teleology in modern biology, point out their main flaws, and then in the end argue in favor of the present account based on inclusive fitness.

## ***2. Current approaches to teleology and functions: two families.***

### *2.1. Intrinsic-teleology views.*

The term "teleology" also gestures towards another very different view, which *opposes* the classical - Modern Synthesis centered - view of biology. Here, authors use another strategy in order to naturalize teleology : they think that organisms themselves (decoupled from natural selection) are *intrinsically teleological*. Though there is not a body of theory that one could summarize, all these views of teleology share a same reluctance for the Darwinian primacy of natural selection as the rationale for any teleology: I label them "intrinsic-teleology" view here, also to contrast them with the externalist character of explanations by natural selection (Godfrey-Smith 1996).

Along these lines, some are arguing that organisms are playing an important causal role in their evolution, because they are built to cope with their environment. That was the focus of West-Eberhardt's book on phenotypic plasticity (West-Eberhardt 2005). She contrasts classical "mutation first" scenarios, where evolution by natural selection acts on random mutations at the allelic level, with "phenotype first" scenarios where the organism phenotypically change

in relation to the environment, then initiating an evolutionary change that is registered (but not constituted, or caused) by the alleles' frequency change. Other authors focus on variation, arguing that it is not only genetic random variation, but often epigenetically transmitted hereditary variation, directed and not random, that fuels natural selection: in this case the organism somehow thereby directs its variations (Jablonka and Lamb 2005). This led Jablonka and Lamb to speak of a "Lamarckian dimension" (but see Merlin 2010 for a rebuttal of the term Lamarckian), referring to Lamarck's idea that organisms vary across evolutionary timescales because they strive to adapt to circumstances. Denis Walsh synthesized those views by arguing that the organismic teleology is the underlying rationale of the genetic evolution, which is a shadow of the real cause of evolution (esp. adaptation), that lies at the level of organism teleology (Walsh 2003, 2006, 2012). At the microbiological level, Shapiro (2011) develops an account of bacteria that is quite close in spirit to such views, since he argues that bacteria's features call for a teleological drive towards surviving and adapting built in themselves, and accounting for their capacity to produce adaptations via the biased production of mutations.

Evidence of transgenerational or intragenerational non-genetic inheritance (Jablonka and Raz 2009) plays an important role in those arguments, not exactly because the substrate of inheritance is not Mendelian, but because it allows variation to be non-random, hence possibly adaptively directed, and therefore allows for some purposiveness within the variation-inducing mechanisms. This is not crucial in Walsh's account (Walsh 2015), indeed, yet what's important for him is the fact that variation cannot be assumed to be blind variation occur-

ring in a population of organism, but is also emerging from organisms themselves in an adaptive way, and such variation plays a leading role in evolution. On Wash's view, many variation mechanisms - and not especially epigenetic or non-genetic inheritance - can support patterns of biased variation, and those patterns are crucial for establishing the teleological character of organisms as 'agents', as he calls them. Reciprocally, not all research on non-genetic inheritance is committed to intrinsic-teleology view (see Danchin et al. 2011), meaning that the project of extending the notion of inheritance is not logically leading to vindicating a non-Darwinian teleology.

Another variant of intrinsic-teleology accounts - not primarily interested in evolution - has been advocated by authors involved in the long-lasting philosophical controversies about understanding biological functions (the notion of function being an instance of teleological judgment, as emphasized above). Building on Varela and Maturana's theory of organisms as selforganising entities (known under the name "autopoiesis"), as well as on Rosen (1971) theoretical account of the notion of organism as closed circular system, Mossio et al. (2009) and Saborido et al. (2012) advocated an "organizational account" (OA) of functions and teleology, according to which, since organisms are self-maintaining entities in which the activities of parts achieve a constraint upon the whole and are in the same time maintained by such a whole, the parts are both causes and effects of themselves; therefore they can be teleologically understood. Here, teleology is a property of some natural entities (namely, organisms), therefore it is a realist concept, corresponding to objective features of these specific physical systems.

This reactivates a Kantian line of thinking, since Kant first argued that biology is intrinsically teleological (it explains things functionally, and it postulates norms for understanding development, see e.g. Ginsborg 2004, Huneman 2006), and for this reason should presuppose that organized beings are unfolded by a teleologically oriented formative force (the formative force, or *Bildungstrieb*, conceived of by biologists such as Blumenbach in the wake of Kant's texts). Kant was the first to argue that organisms have to be thought as self-organizing entities, meaning that we cannot understand and study them except if we presuppose that they are self-maintaining and self-producing according to an idea (that we think is) at the origin of their production (Huneman 2006). However, where Kant's analysis aimed at the necessary structures of our cognition, OA accounts argue that teleological features ontologically characterize organisms themselves<sup>8</sup>.

## 2.2. Etiological Darwinian views of teleology.

Philosophers of science have been keen on deflating in some ways the functional discourse that is so pervasive in biology, for the reason that such explanation, which cites the effects of a process as a cause of its presence, mainte-

---

<sup>8</sup> This OA of teleology that I am considering may also seem close to the well known "systemic" or "causal role" views of functional statements, initiated by Cummins (1975). Such accounts see functional traits as traits having a causal role in a focal system, a role defined by a regular connection between input and output – hence it ties teleology as an explanatory modality, to organization as a property of some systems, hence reminding the OA of teleology. Yet, being realist, the OA actually contrasts with the systemic view of functions, because for the latter the teleological aspect of the functional explanations is tied to a specific explanatory strategy, i.e. the choice of a system to investigate and the carving of it into parts, hence it is not realist regarding biological teleology. OA, on the other hand, emphasizes that the system is objectively defined by the set of constraining feedbacks, and the parts are objectively individualized by the circular processes that maintains the whole.

nance, form or effectuation, violates the requirements of correct science. Hempel (1955) and then Nagel (1961) famously tried to eliminate functional discourses in the context of their philosophy of science, quite unsuccessfully: for instance, if functional statements are merely stating a cause and effect relationship, one wonders why not all effects of a trait such as the kidney are not equally functional<sup>9</sup>... For these positivist authors, functions were mere ways of talking but don't capture something real in the world.

Yet, a Darwinian take on functions and then teleology has been precisely defended in this debate. Larry Wright (1973) initiated this approach, which considered that functional statements are legitimate statements that hold as such, with no need for any eliminative rephrasing. In this view, the "function of X is Y" is an explanation of the presence of Y, to the extent that it means that "X is here because it does Y". This can be explanatory when for instance doing Y (in the past) caused the selection of (past instances of bearers of) X. This so-called etiological approach (or "selected-effects theory") has then been tied to natural selection because the feedback by which the effects act upon the presence of their cause in biology is generally natural selection (Millikan 1984, Neander 1991)<sup>10</sup>. This theory of functions was called "Teleological" by Neander, since it naturalizes teleology via natural selection. Unlike positivist philosophers of science, for these philosophers the naturalization made possible by Darwin is such that a teleological thinking is legitimized because natural selection yields an explanatory scheme quite different from the mechanical explanations that are pervasive in

---

<sup>9</sup> See Mc Laughlin (2001, chap.1) for a thorough overview of these positions.

<sup>10</sup> Strictly speaking Wright's view differs from 'selected effect' theories since for him natural selection was not logically necessary to make sense of functions. But this is not relevant here.

physics. And since evolution by natural selection of a trait or a gene is something objective out there, the teleology resting on natural selection – which is mostly about functional statements – is not a mere way of speaking, a “stance” (as says Dennett (1995) about intentionality), but rather a real feature of *some* entities in this world, namely the organisms. The teleology proper to functional ascriptions is thereby naturalized in a realist way (i.e., it is not depending upon our explanatory strategy and interests). This conception of functions as teleology faces many problems and underwent successive refinements since its formulation (see e.g. Ariew et al. 2002, Godfrey Smith 1994, Buller 1999, Huneman 2013), but this is not the place to deal with that. I just emphasize that the focus of teleology here is the *trait itself*, explained in a teleological manner via the functional ascription. This focus on traits clearly contrasts with the focus on *organisms themselves* proper to the intrinsic-teleology accounts surveyed above.

Nevertheless, I will argue in the following that these two kinds of realist approaches to teleology – etiological views *sensu* Neander and others, and the intrinsic-teleology views - which are equally interested in giving a legitimate naturalized sense to teleology in life sciences, fail to capture the generality of the teleological explanatory practice in biology.

### **3. Critique of the most frequent theories of biological teleology.**

#### **3.1. Issue with the etiological theory**

The *etiological* theory focuses on traits as functional. “The function of kidneys is eliminating toxins”, means that kidneys, by eliminating toxins, provided

fitness advantages to those who had them, and therefore, present individuals are descendants of those ones and have kidneys. But notice first that this is a shortcut for a genuine biological explanation – in this case, like the case of the heart or other organs (often mentioned in literature about etiological views on functions, e.g. Millikan, 1984, Neander 1991, Wright 1973), selection was not acting in a population of kidneys-carrying vs. non-carrying individuals, but of course, the population was made up of individuals carrying slightly different kidneys, differing in terms of their toxin-elimination performances. And at each step of this process, the selection retained slightly better-performing variants, so that ultimately well-adapted kidneys became established in the population. This incremental dimension of the selective process is crucial, since it stems from the fact that most variations likely to be selected are generally small. The etiological view of functions refers to a notion of “selection for kidneys because of the toxin-eliminating effect that they achieve”, but this properly means that *better* toxin-elimination provides (*ceteris paribus*) a fitness advantage that yields what we call “selection for kidneys”. The idea of selection targeting kidneys against non-kidneys is of course too simplistic, and obscures the process of natural selection in favor of an emphasis on the parallel between natural selection and human intention (which often chooses A against non-A), that Wright (1973) considered as two aspects of a general concept of “function”.

Because of its focusing on *traits* themselves rather than *organisms*, the etiological theory of functions fails to make sense of the fact that traits are integrated within an organism, in a way which implies that their nature and form should somehow be adapted to other traits. The fact that toy examples (heart,

kidney...) favored by etiological theorists, as I just indicated, neglect the *incremental* nature of selection somehow illustrates this failure since, as Fisher's (1930) "geometric argument" made clear, the fact that selectable mutations are generally *small* derives precisely from the integrated character of organisms.

Intuitively, this echoes the longstanding critique of adaptationism (Gould and Lewontin 1978, Godfrey-Smith 2001), which argues that atomizing organisms into discrete traits and asking what adaptations-for they are misses the essence of the organism. More precisely, the fact that this account does not link functions of traits to any idea of system (unlike the other account) indicates that there is an issue with such idea of teleology, since it is unable to account for the fact that functional traits are also systematically related, and possibly related to non-functional traits. The basic issue is that even though etiological accounts may correctly make sense of what it is to ascribe a function to a trait or be a function of a trait, they are not capable of making sense of a whole organism as having an (apparent, at least) teleological nature. In other words, they can not see an organism as having parts collectively contrived to a purpose, because they only make sense of parts separately, each having its specific function in the environment. As an account of *biological teleology* in general, etiological views are wanting. Such account indeed, as many authors, tends to see functions as the sole dimension of teleological language in biology, while as Birch (2012) argues, "the reducibility of all teleological talk in biology to function ascription is not something we can safely take for granted"; many other instances of this language occur in biology, some of them implying the whole organism.

This doesn't mean that the etiological account is mistaken; it's rather incomplete as an account of teleology. Granted, making sense of the apparent teleological character of organisms may be seen as something extrinsic to the task of a theory of functions. But a complete theory of biological teleology should be able to do this<sup>11</sup>, at least because organisms seem *designed* in the sense that parts are adjusted to one another (without entailing that their function is to be adjusted to one another), and that *developing* organisms are seemingly teleological in the sense that they are oriented towards producing the typical design of the species. As Kant noticed, it's impossible to consider and understand embryogenesis without assuming that the system undergoing such process oriented towards a goal<sup>12</sup>. Thus the etiological theory is an incomplete theory of teleology.

Thinkers in the etiological tradition have sometimes diagnosed this problem: Kitcher's paper entitled "Function and design" (1995) indeed tries to plug into the etiological account the notion of a system, arguing that if X is here because of its doing Y, so that X has been designed to do Y, X is therefore meant to be integrated into a design. In such a design as a self-contained and orderly structure, not all parts should owe their presence to natural selection, but all may nevertheless be functional because other, naturally-selected, parts, have themselves adapted to their presence. The present paper takes a similar line of argument, but will rely on the view of maximisation sketched above<sup>13</sup>.

---

<sup>11</sup> Remember that Wright's etiological account was initially given as a general theory of teleology.

<sup>12</sup> Kant's solution is that this goal is a necessary assumption of our knowledge - see Ginsborg (2004); Breitenbach (2009); Huneman (2007).

<sup>13</sup> This section on etiological account is comparatively short for several reasons. First, there has been many discussions about the problems and limits of this widely accepted account of functions - e.g. chapters in Ariew et al. (2001) or Huneman (2013); second, the present paper is much more discussing intrinsic-teleology accounts, which are not Darwinian. Etiological accounts are

### *3.2. Issue with the intrinsic-teleology accounts.*

On the other hand, the intrinsic-teleology accounts deal with the wholeness of organisms: such wholes have to be understood in relation to the environment within which they maintained themselves, and as oriented towards the maintenance of the parameter values that allow them to live (e.g., all physiological parameters), a maintenance often called homeostasis since Cannon. In Jablonka's, West-Eberhardt's, Shapiro's or Walsh's views, living beings (either bacteria, or multicellular organisms) are literally striving to adapt to their environment, and this striving is what allows us to understand their functioning and behavior.

But here the focus on organisms entails a drawback inverse to the former: it is not at all clear how to individuate functional traits in organismic accounts, and even worst, not clear how to ascribe function to particular traits. For instance West-Eberhardt's view of plasticity is about a general property of developmental systems, without making clear which traits could be so explained and why. So one cannot use plasticity to account for saying that a given trait is here to do such and such. Granted, plasticity may be something through which organisms cope with environmental changes or uncertainty; but this means that plasticity has no resources to explain traits of an organism in an unchanging environment, and yet these are still ascribed a role. While plasticity captures an as-

---

somehow Darwinian, and they are closer to the solution I will defend here, even though they are insufficient. Intrinsic-teleology views on the other hand are either connected to the biological theories that are currently challenging the Modern Synthesis, or to views that are not primarily evolutionary, such as Rosen's.

pect of teleology - that is, the ability to deal with environmental uncertainty, which may in turn be itself an adaptation and as such explainable by natural selection<sup>14</sup> – it is not embracing the generality of the teleological concepts in biology.

Someone could here object that etiological accounts of functions too do poorly in individualizing traits - and that's not their role, anyway. Generally they first assume individual traits, and interpret functional ascriptions to those traits. Moreover, using functions to individualize traits is very controversial, because, as made clear by Griffiths (2006), individuation of traits in evolutionary biology often relies on common descent, hence on homologies, and not on homoplasies and then adaptation or function. In effect, one should be able to say that the anterior limb of a vertebrate is the same thing, whether it's instantiated by a hand or by a whale's fin or a bird's wing, even though those three body parts have distinct functions. So one should not take the inability to *individualize traits* as an objection against intrinsic-teleology accounts such as West-Eberhardt's. However my objection here rather targets the *difficulty of ascribing functions to individual traits*.

Consider now the extravagant tail of the peacock: it is not maintaining the self-organisation of the organism, and indeed may hasten the demise of the individual by attracting the attention of a predator, so according to the organizational account OA it looks like non-teleological. But neither is it facultatively produced by the organism: peacocks does not grow tails as a plastic response,

---

<sup>14</sup> See Nicoglou (2015) on the controversies about the selection for plasticity.

thereby these are not produced by the organisms to cope with their current environment, as in principle plastic traits are.

So, few of the intrinsic-teleology accounts would apparently consider those adaptations such as peacock tail as teleologically explained. But the peacock tail *is* an adaptation (it enhances its bearer's fitness by enabling him greater mating success), and if we want to make sense of teleological explanations in a naturalized framework, this should count as an example of such explanation. Considering the rival account of teleology (etiologal accounts), peacock's tail is of course a paradigmatic case of a functional trait in etiologal accounts, since it is a classical case of selected trait.

Here, one could object that Walsh's account in fact sees functions as contributions to overall systems' goals, and thus, if the goals include reproduction, then surely the peacock's tail has a function. But now consider mixed strategies in behavioural ecology – such as “act as a hawk with probability 0.7, act as a dove with probability 0.4” – that are proven to be “evolutionary stable strategies” (Maynard-Smith 1982); they are adaptations, to the extent that they are the result of natural selection. However, the precise values of probabilities defining this strategy are in no way necessary or useful to the survival of the individual organism, and don't contribute to any specifiable goal better than any other probability value; therefore, such adaptations can't be accounted for in the intrinsic-teleology perspective.

And a more general criticism holds against OA accounts. Critically, some traits in organisms are not recruited to do self-maintenance of the organism: for instance, the altruistic behavior of sterile workers among the eusocial insects

does not do anything to its maintenance (and indeed they may be self-destructive) - yet we have explanations that are in a way teleological (choosing the best strategy, in a decision-theoretical model), and we definitively see them as adaptations (Krebs and Davies 2001). Saborido et al. (2012) tried to answer this concern about non-maintenance traits by arguing that one should consider another system, that defined by the lineage rather than the individual organism. In those cases of “cross-generational self-maintenance“, their answer here would probably be that the sterile casts contribute to the maintenance of the colony as a whole, which is therefore their function.

This faces a first objection : it seems that a price to pay is that some objectivity of teleology is given up, since deciding in favor of an organism or a lineage or a colony seems to rely on the modeler’s choices, hence we get closer to Cummins' systemic account, which makes teleology something so explanatory-strategy-dependent that it could be said dispensable. A second objection, derived from a slightly different case, is more radical. Acts of altruistic nepotism need not maximize the fitness of any group or lineage in order to be selectively favored. By definition, they will not contribute to the survival of the agent (since they are costly for her); but they are not promoting the survival of the group either, which includes various organisms with distinct genetic make-ups, and which may survive without them. What they increase is the amount of genes “for“ altruistic nepotism in the gene pool, but this is obviously not the kind of system that the OA account considers as a system likely to be autonomous and support functions.

Thus OA accounts and intrinsic-teleology views in general face difficulties in deriving from teleological organisms an account of those seemingly teleologi-

cal explanations of individual traits. This insufficiency is problematic for such accounts of teleology, because when asking the teleology question one wants to make sense of all dimensions of what before Darwin was supposed to pertain to teleology.

### 3.3. *Darwin's insight about design.*

We thereby think that neither of the here reviewed attempts of giving sense to teleology wholly succeeds, because either one of them – namely, adopting the etiological view of functions - loses the overall design of the organism, or the other – namely adopting any intrinsic-teleological account - loses the fact that adaptation sometimes differs from the maintenance of the organism or the strive to survive and adjust to the organism's environment, or in other words, that traits as adaptation can be decoupled from organism' maintenance. Etiological views of function and theories of intrinsic teleology both have a *realist* take on teleology, but it seems that they can't capture *at the same time* the functionality of parts *and* the design of organisms, and therefore are compelled to always exclude one aspect of the apparent teleology in biology.

Nevertheless, Darwin had a quite complex view on this matter. He thought that the design of organisms is explained by natural selection; and that *traits themselves are adaptations*: “natural selection acts by either now adapting the varying parts of each being to its organic and inorganic conditions of life; or by having adapted them during past periods of time” (1859, ch. 6); therefore an

etiological account of function (rephrasing functions of traits in terms of natural selection) would not be far from his view.

Yet he also *emphasized the coordinated contrivances of organisms*, for instance in the book on orchids, throughout which he explores the contrivances through which orchids can be fertilized: traits are contrived together, so that they somehow adjust to each other within an organism, and ultimately contrived to support reproduction<sup>15</sup>. This contrivance is in turn explained by natural selection. Contrived traits are not initially made for one specific purpose; at the contrary, various traits may have originated for various distinct purposes, yet by evolution they find themselves all together contrived for one specific purpose that is reproduction: “Although an organ may not have been originally formed for some special purpose, if it now serves for this end we are justified in saying that it is specially contrived for it. On the same principle, if a man were to make a machine for some special purpose, but were to use old wheels, springs, and pulleys, only slightly altered, the whole machine, with all its parts, might be said to be specially contrived for that purpose.” (Darwin 1862, 348) The very fact of contrivance as studied initially in the Orchids book, therefore, is about many traits of varied origins that conspire together into something that looks as if de-

---

<sup>15</sup> The following example from the chapter on *Spirantes autumnalis* exemplifies the point: “We thus see how beautifully everything is contrived that the pollinia should be withdrawn by insects visiting the flowers. They are already attached to the disc by their threads, and, from the early withering of the anther-cells, they hang loosely suspended but protected within the clinandrum. The touch of the proboscis causes the rostellum to split in front and behind, and frees the long, narrow, boat-formed disc, which is filled with extremely viscid matter, and is sure to adhere longitudinally to the proboscis. When the bee flies away, so surely will it carry away the pollinia. As the pollinia are attached parallel to the disc, they adhere parallel to the proboscis. When the flower first opens and is best adapted for the removal of the pollinia, the labellum lies so close to the rostellum, that the pollinia attached to the proboscis of an insect cannot possibly be forced into the passage so as to reach the stigma ; they would be either upturned or broken off: but we have seen that after two or three days the column becomes more reflexed and moves from the labellum, - a wider passage being thus left.” (Darwin 1862, 111)

signed to reproduce. Natural selection, for Darwin although in an implicit and weakly articulated way, accounts for *both* the *traits* as adaptations and the *overall design* of organisms. Of course Darwin's darwinism didn't include population genetics; thus, realizing such insight in current biology means implementing it in an evolutionary science structured into a darwinian biology of organisms (behavioral ecology) and a population genetics, and this is where FD should intervene.

This Darwinian insight seems most present in Kitcher (1989). Kitcher held the view that functions of traits understood as "selected effects" fit in a general design, which also means that not all features of organisms are targeted by selection, but that many features are somehow entailed by the organism's design, which takes place as soon as some features are designed by natural selection<sup>16</sup>. His use of the word "design" intends to capture such a connection between selection and organismality. I think that this is close to the connection between adaptations, contrivances and design, which for Darwin defined the way in which teleological explanations could be reconceived in his framework. In the following section I propose an alternative view, sharing the concern for design present in Kitcher's approach, but relying on the Formal Darwinism account stated above.

---

<sup>16</sup> "It is enough (...) that genuine demands on the organism have been identified and that the entities to which they (biologists) attribute functions make causal contributions to the satisfaction of those demands. (...) To say that the function of X is F is to propose that a complete explanation of the presence of X (at the appropriate time) should be sought in terms of selection for F. Once we relax the demands on functional ascriptions, the role of selection is no longer clear; indeed, a biologist may explicitly allow that selection has not been responsible for maintaining X (or, at least, not completely responsible). But there is a different type of explanatory project to which the more lenient attributions contribute. They help us to understand the causal role that entities play in contributing to complex effects." (Kitcher 1993, 271)

#### ***4. Maximization of inclusive fitness: teleological explanation of design.***

##### *4.1. Inclusive fitness and design.*

In this last section, relying on the Formal Darwinism view stated above, I claim that the maximization of inclusive fitness underpins proper teleological explanations in biology (hereafter “IF teleology”), in a way faithful to Darwin’s sense of the unity between the contrivances and the traits as adaptations<sup>17</sup>. It is indeed teleological, since when a trait improves inclusive fitness it is for this precise reason that it is here<sup>18</sup>. So the inclusive-fitness *effect* itself explains, which indicates that this is a teleological explanation. Then the “IF teleology” account of teleology says that teleologically explaining X means showing that X contributes to maximising inclusive fitness by whatever means. It implies that teleology as an explanation is legitimate, because it is grounded in the dynamics of gene frequencies, according to the isomorphism stated by FD. Once again this does not mean that the whole evolutionary process has a goal; it just means that the evolutionary structure of biological reality makes some teleological reasoning legitimate, and therefore allows us to make on this basis predictions of which phenotypes will be found. I’ll argue now that, so defined, IF teleology accounts for most of the features which seem to need a teleological explanation, realizing in a Modern Synthesis setting Darwin’s intuition that explanations of design of organisms and of adaptive trait should go together.

---

<sup>17</sup> Birch (2012) also argues in favor of a key role of inclusive fitness to make sense of biological teleology - and not only of functions. His account relies on the idea of robust processes or attractors in phase spaces, and extends beyond biology, though the present view should be compatible with it.

<sup>18</sup> There might be an epistemic worry here, about which trait actually accounts for the improvement; see e.g. Huneman (2013) - but this is not the concern here

Indeed, IF teleology shares with “intrinsic-teleology” accounts *an emphasis on the role of organisms*, because within the optimization framework sketched above any (selected) trait can be understood as the ‘choice’ by the organism of a strategy which, through direct and/or indirect benefits, improves the net fitness result. So in this explanatory scheme the organism controls its inclusive fitness gains by ‘choosing’ its phenotypes. This use of the ‘maximizing agent analogy’ (Grafen 2002) means that organisms themselves are the locus of a teleological understanding, which acknowledges the point made by the supporters of intrinsic-teleology accounts in the wake of Kant’s views: as agents, organisms may be addressed as if acting on a purpose, or teleologically.

From this maximization principle follows thereby a concept of *design*: since evolutionary logics may be framed in terms of the *choice* of strategy, or phenotype, by an organism, then the set of actual phenotypes can be seen as in principle being likely to *manifest that it has been designed*. And since inclusive fitness is the maximand, then organisms can be conceived of as “designed to” maximize their inclusive fitness (Gardner 2009) - once again, even though the design fails, exactly as the hunting person can fail in killing preys.

To this extent organisms’ *overall design* is in accordance with this maximization principle - especially, the fact that the parts are contrived to a common purpose: this can be explained by the fact that too much independence between parts would affect the physiology and then the inclusive fitness, so that from this standpoint some coordination is necessary, from which a more or less achieved design results. The appearance of design – in the sense of “being designed” - stems from the teleological nature of organisms, as maximizing agents; and the

design itself, as contrivance and relations of parts, is a result of the fact that all the parts are oriented towards maximizing inclusive fitness. To sum up: since the genes-dynamics aspect of natural selection translates into optimization, the organisms can be said to be designed to achieve inclusive fitness, and for this reason the parts are more or less coordinated, giving for instance rise to those extremely complex and adapted structures such as the eye, whose design challenged causal explanations for decades<sup>19</sup>. (IF teleology can also, in principle, explain the evolution of genetic structures, so that one can say that these systems are designed to maximize inclusive fitness<sup>20</sup>.) What contrives all parts of an organism is thereby the value of inclusive fitness<sup>21</sup>.

So the IF teleology explains the coordinated design. Essentially, it explains the trade-offs between values of traits, in the sense that the value of each trait reachable by selection is contrived by the values of other traits and not only by its adjustment to

---

<sup>19</sup> It was presented by Kant as the paradigm of the set of coordinated action of myriads of parts seemingly contingent for physical laws (Kant 1763, Huneman 2015).

<sup>20</sup> Here I can mention Fisher's (1928) theory of the evolution of dominance, albeit with the qualifier that this is very far from being widely accepted. Fisher suggested this was an example of the inverse side of the usual research about the genetic basis of evolution, since usual evolutionary research intends to explain evolution on the basis of genetics, whereas here he investigated the evolutionary bases of genetic make up itself. Given that organisms are on the whole well designed, mutations in general will tend to reduce rather than improve (inclusive) fitness (Fisher 1932, 279). He wanted to understand why constant alterations by mutational pressure does not take place, thus he elaborated a theory according to which mutations, as often detrimental, will induce selection for modifier genes that bring about their recessiveness so that their chances to harm organisms in which they reside will be lower. So according to Fisher's theory, the very phenomenon of dominance results from a process through which the integrity of organisms is preserved in the face of mutational assault.

<sup>21</sup> As an alternative to inclusive fitness, during discussions about social evolution, another approach termed "personal fitness" or "neighbor modulated fitness", has been proposed (Hamilton, 1964 – recently revived by Taylor and Frank 1996). In the personal fitness approach, the fitness effect of a behavior is the addition of its effect on the fitness of the focal individual, plus the fitness benefits gained by this individual from the others, mitigated by relatedness (Taylor and Frank 1996). It has been claimed that personal fitness approach is equivalent to inclusive fitness in most situations (Taylor et al. 2007). However, the advantage of inclusive fitness, even though both approaches can be equivalent in describing the evolutionary dynamics, is that the process is similar to a rational process of deciding on the best options through maximizing the utility, as it is studied by microeconomics: it is as much a teleological process. It is to be noted that such teleology resembles much more to the one used in economics than the one used by physicists talking of entropy maximization or least action in Hamiltonian variation calculus.

environmental demands. For example, the running speed of a cheetah has some impact upon on metabolism and on the possible values of other traits, so that if selection were to maximize speed, it would impinge other functions and ultimately decrease inclusive fitness; hence the trait values that are predicted by IF maximization are optimal trade-offs, and adaptationism is always ‘trade-off adaptationism’ (Huneman 2017).

What was lacking in the etiological accounts of functional traits in general is the connection between trait-based teleology (a given trait is here for something) and the designedness of the overall organisms; in turn, the OA as well, as in general the intrinsic-teleology accounts, had a hard time making sense of *specific* functional traits and adaptations as functions and adaptations - especially regarding adaptations specifically at odds with the organism’s interests in survival. IF teleology by definition captures the link between these two things, as a principled link between the IF maximizing character of selected traits and the contrivances or contrived character of organisms themselves, established above; therefore I claim that it is the proper Darwinian teleology. Hence thinking teleological explanations in terms of inclusive fitness allows us to make sense of all teleologically-sounding explanations: the ones directed towards functional traits and the ones focusing on organisms apparently fulfilling some aims in their development.

While IF teleology does better than etiological accounts and intrinsic-teleology accounts because it allows making sense of a teleological understanding of both organisms and traits, it’s also more complete because some IF teleology cases can’t be accounted for by one or the other family of views, as I’ll show in this last section.

#### 4.2. IF teleology vs OA and vs SE views.

Regarding the traits themselves, according to IF teleology one explains their presence by the fact that they contribute directly to offspring or indirectly to offspring of the relatives of the organisms. Therefore, against the OA account - and intrinsic-teleology accounts more generally -, traits that don't help maintaining organisms into life can still be teleologically explained. It may be that traits fostering reproduction will seem to go against survival, which is *prima facie* hard for the OA to account for : for instance, the fragility of cottonwood branches seems less fit to their actual ecological location than a robust tree, but actually it increases the dispersal capacity because of their easiness to be broken combined with the fact that from the fragments trees easily sprout, so that those fragments can flourish by dispersing along streams even if the original cottonwood is undercut by the streams and dies (Endler 1986).

Granted, Saborido et al. (2012) attempted to extend the OA account by including reproduction in their definition of the organismic constraints that underlie functions in their sense, so that the lineage itself may in the end become the subject of teleology. However, some traits can still be teleologically explained by their contribution to *indirect* fitness benefits, so not by direct offspring benefit, and here, even this extended OA would fail. For example, there is a teleological explanation of the phenomenon of human menopause<sup>22</sup>: grandmothers help raising children, which increases the reproductive success of their offspring (Foster

---

<sup>22</sup> Birch (2012) gives the same example; he indicates that this teleological explanation would hardly be expressed as a "function of menopause", because many would be reluctant to the normative connotations carried in this case by the concept of function.

and Ratnieks 2005); menopause in this example evolves by increasing the indirect benefit side of the inclusive fitness magnitude, because it proves more beneficial for the inclusive fitness than no menopause, which would privilege the direct benefit side. Alarm calls in vervet monkey are of the same type: though costly for the agent, who puts herself at risk by yelling, they evolve by increasing the indirect fitness benefits, which has been established by considering the family structures of tribes in which they live, since they are actually made of highly genetically related individuals (Cheney and Seyfarth 1985).

Even though beyond the organism itself the lineage becomes the subject of the teleology, as the OA tends to do, it won't account for traits owing their existence to very indirect benefits. Otherwise, given that all the degrees of relatedness may be found in the population, the teleological entity would become the population itself, which puts a very high cost onto the account, since in this case, teleology becomes something very far from what it was supposed to capture, mostly related to an individual organism.

Summing up, OA may account for traits beneficial to the offspring, through taking reproduction into account - but it can't account for traits owing their existence to the 'indirect benefit' side of the inclusive fitness equation, since no reproduction is involved in the relatedness coefficient. But IF teleology does that, by definition.

And now, against the etiological account, the trait does not need to be teleologically explained only in the sense that it would have been *selected for a specific effect*; it might just be that the trait contributes to increasing both direct and indirect fitness, in a way greater than other alternative traits within the same

system – a plurality of effects can be the reason of this increase of inclusive fitness (one effect causing direct benefits, the other causing indirect benefits). Thus it appears that the etiological view of functions only captures one subcase of what is teleological regarding biological traits.

Think of the classic example of kidneys, whose function is to eliminate toxins since our ancestors, carrying properly functioning kidneys, have been favored by selection because of the toxin-eliminating activity of kidneys. In this case, the trait and its payoffs only concern the focal actor: there is no mention of any other beneficiary of the trait. Thus the only question is: “which activity of the trait was causing a net benefit for the actor?” Yet in the most general cases, not only the benefiting effect is relevant, but also the relatedness to other potential beneficiaries, and traits beneficial for the focal actor can be counterselected if less beneficial traits come with higher benefits for related actors (namely, benefits that, mitigated by the relatedness value and added together, overcome the direct benefit).

For instance, Ratcliff et al. (2013) consider the function of microbial dormancy. Many microbes indeed form “dormant resting stages (e.g. spores and cysts)”. When environment is stressful this may help survival, however genotypes forming dormant microbes pay an “opportunity cost” since their delayed reproduction may cause them to have on the average less offspring than their non-dormant competitors - thus the survival enhancement effect is not enough to make dormancy into a selected trait. Yet genotypes with dormant microbes spare resources, which can be in turn consumed by surrounding non-dormant microbes. If those are enough related to the dormant microbes, then the genes of

the dormant microbes will increase in frequency and the indirect fitness benefits explain the evolution and maintenance of the trait “microbial dormancy”. Clearly, this happens, as Ratcliff et al. (2013) have shown, only when the population is structured in the proper way, namely its structure gathers dormant microbes with mostly their kin.

So here the indirect benefits are key to the teleological explanation of the microbial dormancy. But one may think that extending the etiological account by changing ‘fitness’ into ‘inclusive fitness’ in its formulation may allow it to account for such a case. This was already suggested by Neander (1991). She writes:

“It is a/the proper function of an item (X) of an organism (O) to do that which items of X’s type did to contribute to the inclusive fitness of O’s ancestors and which caused the genotype of which is X is the phenotypic expression (or which may be X itself where X is the genotype) to increase proportionally in the gene pool.”

Thus, an SE account could perfectly make sense of the explanation of microbial dormancy as teleological. However, in details it’s not at all obvious. Inclusive fitness here is composed of several things : 1. a positive impact on direct fitness in periods of environmental stress; 2. a negative impact because of reproductive delay; and 3. a positive impact via resources for kins, when the population structure is favorable. Moreover, this latter point is not independent from the microbial dormancy itself as a possibly cooperative behavior, since as Power et al. (2012) have shown “the opportunity for cooperation can in turn drive the creation of population structures that support it”. Thus, it’s impossible to state (following Neander’s definition) “that, which things of type “microbial dormancy” did to contribute to inclusive fitness”. Even though the inclusive fitness explanation is ultimately teleological, as I argued, it’s not likely to allow us to dis-

tinguish a “proper function“ in Neander’s sense<sup>23</sup>, and therefore etiological accounts of teleology can’t cover all cases deemed teleological according to IF teleology.

### ***Conclusions.***

In this paper I presented a view of evolutionary theory that confers an overarching role to a specific teleological explanation understood in terms of the maximization of inclusive fitness. I argued that this view could, better than many classical rephrasing of teleology by philosophers, account for the teleological features of explanation present in Darwinian theory, especially the use by Darwin of the concepts of design and contrivances. This paper therefore offers two general claims:

a) that the genuine and legitimate meaning of teleology in biology is maximization of inclusive fitness – because, given the isomorphism established by FD, this accounts for and justifies the explanatory power of teleological explanations in evolutionary biology;

b) that extant accounts of teleology captures aspect of teleology but aren’t general enough to make sense of the whole Darwinian teleology.

---

<sup>23</sup> Actually Griffiths (1993) have noticed the difficulty of introducing a reference to “inclusive fitness“ in a definition of a proper function and revise Neander’s definition in a way that gets rid of this phrase.

## Acknowledgement.

I am hugely grateful to Andy Gardner, with whom I developed many of the ideas of this paper, and who provided me with invaluable comments, as well as materials for elaborating the present account. The present ideas couldn't have been elaborated without him.

I am indebted to Jonathan Birch, Arnaud Pocheville and Sébastien Dutreuil for many useful suggestions and criticism. I also thank two anonymous reviewers whose comments greatly helped to improve the paper.

## References.

- Amundson, R. (2005) *The Changing Role of the Embryo in Evolutionary Thought*. Cambridge: Cambridge University Press.
- Ariew, A. R., Cummins, R. and Perlman, M. (eds) (2002) *Functions*, Oxford: Oxford University Press.
- Balan, B. (1980) *L'ordre et le temps*. Paris: Vrin.
- Birch J. (2012) "Robust processes and teleological language" *European Journal for Philosophy of Science*
- Birch J. (2015) "Natural selection and the maximization of fitness." *Biological Review*, 91, 3: 712-727.
- Buller, D. J. (ed.) (1999) *Function, Selection, and Design*, Albany, NY: SUNY Press.
- Charnov, E. L. Krebs, J. R. (1974) "On clutch-size and fitness". *Ibis* 116, 217-219.
- Charnov, E. L., Krebs, J. R. (1975) "The evolution of alarm calls: altruism and manipulation?" *Am.Nat.* 109: 107-112.
- Cheney D. L., Seyfarth R. M. (1985) "Vervet monkey alarm calls: Manipulation through shared information?" *Behaviour*, 94, 1:150-166.
- Cummins, R. (1975) "Functional Analysis", *Journal of Philosophy*, 72: 741-65.
- Cummins, R.(2002) "Neo-Teleology", in Ariew *et al.* (2002), pp. 157-72.
- Darwin, C. (1862) *On the various contrivances by which British and foreign orchids are fertilised by insects, and on the good effects of intercrossing*. London: John Murray.
- Dowe, P. (2000) *Physical Causation* New York: Cambridge University Press,
- Dunbar R. (2001) "The economics of male mating strategies among primates". In: van Hooff J, Noë R, Hammerstein P, eds. *Economic Models of Animal and Human Behaviour*. Cambridge: Cambridge University Press, pp.245-269.
- Edwards AWF. (1994) "The fundamental theorem of natural selection". *Biological Reviews*, 69:443-474.
- Endler J. (1986) *Natural Selection in the wild*. Princeton: Princeton University Press.

- Ewens, W.J. (2014) "Grafen, the Price equations, fitness maximization, optimisation and the fundamental theorem of natural selection" *Biology & Philosophy*, 29: 197.
- Fisher R. (1930) *The genetical theory of natural selection*. London
- Fisher R. (1932) "The bearing of genetics on theories of evolution." *Science progress*, 27: 273-287.
- Frank S. (1998) *Foundations of social evolution*. Princeton: Princeton University Press.
- Frank S. (2012) "Wright's adaptive landscapes versus Fisher's fundamental theorem". In *The adaptive landscape in evolutionary biology*, Svensson E, Calsbeek R, New-York: Oxford University Press, pp.41-73
- Frank SA, Slatkin M. (1992) "Fisher's fundamental theorem of natural selection". *Trends in Ecology and Evolution*. 1992;7:92-95.
- Frank SA. (2006) "Social selection". In: Fox CW, Wolf JB, eds. *Evolutionary Genetics: Concepts and Case Studies*. Oxford: Oxford University Press, pp.350-336.
- Gardner A. (2009) "Adaptation as organism design". *Biology Letters*, 5:861-864.
- Gardner A (2017) "The purpose of adaptation". *Interface Focus* 7, 20170005.
- Gardner A, West S. (2010) "Greenbeards". *Evolution*, 64:25-38.
- Gardner A, West S. (2014) "Inclusive fitness: 50 years on". *Phil.Trans. R. Soc. B* 369: 20130356.
- Gardner A, West SA, Wild G. (2011) "The genetical theory of kin selection." *Journal of Evolutionary Biology*, 24:1020-1043.
- Gavrilets S. (2004) *Fitness Landscapes and the Origin of Species*. Princeton: Princeton University Press.
- Gilbert, S, Sarkar, S. (2000) "Embracing complexity: Organicism for the Twenty-first Century". *Developmental Dynamics*, 219: 1 – 9
- Ginsborg, H. (2004) "Two kinds of mechanical inexplicability in Kant and Aristotle". *Journal of the History of Philosophy*, 42, 33–65.
- Godfrey-Smith, P. (1994) "A Modern History Theory of Functions", *Noûs*, 28: 344–62.
- Gould SJ, Lewontin R. (1978) "The spandrels of san Marco and the panglossian paradigm: A critique of the adaptationist programme". *Proc. Roy. Soc. Lond.* B205: 581-598
- Gould SJ (2002) *The structure of evolutionary theory*. Cambridge: Belknap.
- Grafen A. (1984) "Natural selection, kin selection and group selection". In: Krebs J, Davies N, eds. *Behavioural Ecology: An Evolutionary Approach*. Oxford: Blackwell, pp. 62-84.
- Grafen, A. (2002) "A first formal link between the Price equation and an optimisation program." *J. Theor. Biol.* 217: 75–91.
- Grafen, A. (2006) "Optimisation of inclusive fitness." *J. Theor. Biol.* 238: 541–563.
- Grafen, A. (2007) "The formal Darwinism project: a mid-term report." *Journal of Evolutionary Biology*, 20: 1243–1254.
- Griffiths P. (2006). "Function, Homology, and Character Individuation". *Philosophy of Science*, 73: 1–25.
- Griffiths P.(1993). "Functional Analysis and Proper Functions". *British Journal for the Philosophy of Science* 44: 409–422 .
- Hamilton W. (1964) "The genetic evolution of social behavior". *Journal of Theoretical Biology*, 7:1-16.

- Hempel, C. G. (1959) 'The Logic of Functional Analysis', in L. Gross (ed.), *Symposium on Sociological Theory*, New York: Harper and Row, pp. 271–87.
- Huneman, P. (2006) "From Comparative Anatomy to the 'Adventures of Reason.'" *Studies in History and Philosophy of Biological and Biomedical Sciences*, 37, 4: 649-674.
- Huneman P. (2013) "Weak realism in the etiological theory of functions", *Functions: selection and mechanisms*, Huneman (ed.), Dordrecht: Springer, pp.105-13.
- Huneman P. (2015) "Redesigning the argument from design." *Paradigmi*, 2: 105-132.
- Huneman P. (2017) "Variation, extension and selection: a synthesis of the reasons for a new evolutionary synthesis." in *Challenging the Modern Synthesis. Development, adaptation and inheritance*, P Huneman & D Walsh (eds.), New-York: Oxford UP, pp. 68-110.
- Huneman P. (2017b) "Kant's Concept of Organism Revisited: A Framework for a Possible Synthesis between Developmentalism and Adaptationism?" *The Monist*, in press
- Huxley J. (1942) *Evolution: The Modern Synthesis*. New York: Harper.
- Jablonka E., Lamb M. (2005) *Evolution in four dimensions*. Cambridge : MIT Press
- Jablonka, E., G. Raz (2009) "Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution." *Quarterly Review of Biology* 84: 131-176.
- Kant, I. [1790] (1987) *Critique of Judgment*. Trans. Werner S. Pluhar. Indianapolis: Hackett. (often translation revised) (noted CJ)
- Kant I. [1763] (1992) "The Only Possible Argument in Support of a Demonstration of the Existence of God." In Walford D. , Meerbote R., *The Cambridge Edition of the Works of Immanuel Kant. Theoretical Philosophy, 1755—1770*. Cambridge: Cambridge University Press, pp. 111-201.
- Kitcher, P. (1993) 'Function and Design', *Midwest Studies in Philosophy*, 18: 379–97.
- Lack D. (1947) "The significance of clutch-size," *Ibis*, 89: 302-352
- Leigh E. (1971) *Adaptation and diversity : natural history and the mathematics of evolution*. San Francisco: Freeman Cooper.
- Maynard Smith, J. (1958). *The Theory of Evolution*. London: Penguin Books
- Maynard-Smith J. (1982) *Evolution and the theory of games*. New-York: Oxford University Press.
- McLaughlin, P. (2001)] *What Functions Explain. Functional Explanation and Self-Reproducing Systems*, Cambridge: Cambridge University Press.
- Merlin F. (2010) 'Evolutionary Chance Mutation: A Defense of the Modern Synthesis' Consensus View', *Philosophy & Theory in Biology*, 2,
- Millikan, R. G. (1984) *Language, Thought, and Other Biological Categories*, Cambridge, MA: MIT Press.
- Moran P. (1964) "On the non-existence of adaptive topographies." *Ann. Hum. Genet.* 27: 383-393.
- Mossio, M., Saborido, C. and Moreno, A. (2009) 'An Organizational Account of Biological Functions', *British Journal for the Philosophy of Science*, 60: 813–41.
- Nagel, E. (1961) *The Structure of Science*, London: Routledge & Kegan Paul.
- Neander, K. (1991) 'Function as Selected Effects: The Conceptual Analyst's Defense', *Philosophy of Science*, 58:168–84.

- Nicoglou A. (2015) "The evolution of phenotypic plasticity: **genealogy** of a debate in genetics", *Studies in the History and Philosophy of Science Part C*. 50: 67-76
- Power S., Penn A., Watson R. (2012) "The concurrent evolution of cooperation and the population structures that support it." *Evolution* 65-6: 1527–1543
- Ratcliff W., Hoverman M., Travisano M., Ford Denison R. (2013) "Disentangling Direct and Indirect Fitness Effects of Microbial Dormancy." *The American Naturalist* 182, 2: 147-156
- Rehbock, P. (1983). *The philosophical naturalist: Themes in early nineteenth century British biology*. Madison: University of Wisconsin Press.
- Saborido C, Mossio M, Moreno A (2011) Biological Organization and Cross-Generation Functions. *British Journal Philosophy of Science*, 62 (3): 583-606.
- Shapiro J. (2011) *Evolution: A View from the 21st Century* San Francisco: FT Press Science.
- Sober E. (1984) *The Nature of Selection*. Cambridge: MIT Press.
- Varela, F. J., Maturana, H., Uribe, R. (1974) "Autopoiesis: The Organisation of Living Systems, its Characterization and a Model", *BioSystems*, 5:187–96.
- Walsh D. (2003) "Fit and diversity: explaining adaptive evolution". *Philosophy of Science*, 70:280-301.
- Walsh D. (2010) "Two neo-Darwinisms", *History and Philosophy of the Life Sciences*, 32(2-3): 317-339.
- Walsh, D. (2006) "Organisms as natural purposes: The contemporary evolutionary perspective" *Studies in History and Philosophy of Biological and Biomedical Sciences*, 37, 4: 771-791
- Walsh, D. M. (2015). *Organisms, agency, and evolution*. Cambridge, UK: Cambridge University Press.
- West-Eberhard, M.J. (2003) *Developmental Plasticity and Evolution*. Oxford: Oxford University Press.
- West SA, Gardner A (2013) "Adaptation and inclusive fitness". *Current Biology* 23, R557-R584
- West SA, Gardner A, Griffin AS (2006) "Altruism". *Current Biology* 16: R482-R483.
- West SA, Griffin AS, Gardner A. (2007) „Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection". *Journal of Evolutionary Biology*, 20:415-432.
- Williams GC. (1966) *Adaptation and Natural Selection*. Princeton, NJ: Princeton University Press.
- Winther RG (2006) "Fisherian and Wrightian perspectives in evolutionary genetics and model-mediated imposition of theoretical assumptions". *Journal of Theoretical Biology* 240:218–232.
- Woodward, J. (2003): *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
- Wright, L. (1973) "Functions", *Philosophical Review*, 82: 139–68.
- Birch J. (2012) "Robust processes and teleological language" *European Journal for Philosophy of Science*