Agency, goal-orientation and evolutionary explanations

Tobias Uller

Department of Biology

Lund University

Abbreviated title: Agential concepts and evolutionary explanations

Correspondence: Tobias Uller; Department of Biology, Lund University, Sölvegatan 37, 22362 Lund, Sweden; email: tobias.uller@biol.lu.se

Word count: 8375 including references (5954 main text)

Acknowledgements: I am grateful to the editors for the invitation to contribute to this volume, and to Nathalie Feiner, Juan Gefaell, Denis Walsh and the editors for feedback on the manuscript. I gratefully acknowledge the financial support of the John Templeton Foundation (#62220). The opinions expressed in this paper are those of the author and not those of the John Templeton Foundation.

To be published in *Evolution 'on Purpose': Teleonomy in Living Systems* (Peter Corning, Stuart Kauffman, Denis Noble, James Shapiro and Richard Vane-Wright, Eds.), MIT Press, Cambridge, Mass., forthcoming

Overview

The scepticism towards teleology in the natural sciences may give the impression that all reference to agency and goals in evolutionary explanations just is convenient short-hand. Here, I suggest that agential concepts may in fact serve several distinct epistemic functions. Firstly, the gene's-eye view demonstrates that agential concepts can promote the intelligibility of evolutionary theories, thereby facilitating the application of those theories to explain natural phenomena. Secondly, agential concepts can structure evolutionary investigation according to particular criteria of explanatory adequacy. These explanatory agendas admit non-selective causal influence on adaptive evolution, which begs the question how developmental and selective explanations should be integrated. Thirdly, a more radical proposal is that organismal goals themselves can be explanatory for evolutionary change. Such naturalistic teleological explanation is motivated by an explanatory gap left by causal explanations, and encourages development of theories and models that allow the principles of evolution to depend on organismal activities that originate as a result of the organisms' internal organisation.

1 Introduction

Anyone who's ever visited a rain forest or taken a virtual tour through a cell would surely agree that living systems are extraordinarily complex – far too complex to grasp in their entirety. Scientific explanation must therefore rely on simplified representations that leave out detail and even distort reality. A generic representation provides principles from which can be developed more specific representational theories and models that are applied to explain natural phenomena (Fig. 1). Such explanations commonly refer to variables, events, entities, or state-of-affairs that make a difference as to whether or not a phenomenon obtains. Explanations that concern humans and our societies may also refer to goals and purposes; perhaps your opening this book is explained by your goal to learn something new, or intention to see if your work was cited; the human strive for greater equality and justice can be invoked to explain the rise of modern democracy; and so on.

In contrast to the social sciences, the natural sciences consider goals and purposes unacceptable. Matter is not imbued with agency, and planets do not move around the sun because it is their function or purpose to do so. Such phenomena are supposed to be explained by mechanisms, causes and forces. Moreover, goals lie in the future, while causes must precede their effects. This is perhaps not so problematic for human affairs since we can think about the future and decide what to do on the basis of our mental representations, thereby keeping the cause-effect relation in the right order. But in the absence of the human capacity for deliberation, rational choice, and cumulative culture, any reference to goals may simply appear unscientific.

At first sight, evolutionary biology appears to adhere to the mechanistic ideal; mutations occur without regard to their effects, and selection favours current, not future, utility. Yet, the goal-oriented nature of development, physiology, and behaviour is hard to deny, and talk of agency, goals, and purpose is rife within biology (Box 1). One of the most famous perspectives on the evolutionary process – the gene's eye view – even seem to grant purposive agency to DNA; genes are said to have goals and interests, play strategies, and be in conflict with each other. Perhaps such use of 'agential concepts' is just convenient and innocuous short-hand. This chapter explores the alternative explanation that agential concepts really do serve important epistemic functions in evolutionary biology.[i]. Three broad possibilities will be considered (Fig. 1).

[Please insert Fig. 1 about here – now found at the end of this document]

Firstly, agential concepts may further epistemic goals by making evolutionary theories intelligible. Thinking of genes, biological processes, or organisms as goal-oriented can help scientists reason intuitively and make qualitative predictions, something that often is necessary to develop formal models. Secondly, agential concepts can set explanatory standards. By drawing attention to the adaptive biases imposed on evolution by development and behaviour, agential concepts structure scientific investigation around particular sets of problems (including agency itself) and with associated criteria of explanatory adequacy. Thirdly, if biological evolution is a consequence of the goal-oriented activities of individual organisms, naturalistic teleological explanation could perhaps be expanded beyond human cultural evolution, and make it scientifically legitimate to account for biological evolution by referring to organismal goals.

[Please insert Box 1 about here – now found at the end of this document]

2 Agential concepts can make evolutionary theories intelligible

The gene's-eye view is a useful starting place to explore the epistemic functions of agential concepts; if for no other reason then simply because a molecule seems a rather unlikely candidate for exhibiting goals and purposes. Proponents of the gene's-eye-view have indeed been quick to point out that ascribing agency to genes is merely a convenient short-hand, which they could translate 'into respectable terms if we wanted to' (Dawkins 1976. P. 88)[ii]. While there is some controversy over whether or not genes ever can be considered agents (Okasha 2018), it is widely accepted that thinking of genes as having goals, interests, intentions, or strategic repertoires can help scientists think clearly about evolution (Burt and Trivers 2006; Ågren 2021).

There is undoubtedly something right about this suggestion. But what exactly is it about the agential metaphors that make them helpful? If purposive agency does not correspond to an actual property of genes, the epistemic function of agential concepts seems unlikely to mediate the relationship between theory and phenomenon; it is not really the goal-oriented activities of genes that explain why ants cooperate, for example. A more plausible alternative is that applying agential concepts to genes makes evolutionary theory itself intelligible.

According to philosopher of science Henk De Regt, scientific understanding requires application of an appropriate explanatory model or theory (of which there are many) to natural phenomena (De Regt 2018; Fig. 1)[iii]. Scientists cannot apply a theory, or cannot apply it appropriately, unless the theory is intelligible to them. This makes the intelligibility of scientific theories central to understanding how scientist can deliver adequate explanations. De Regt defines scientific intelligibility as 'the value that scientists attribute to the cluster of qualities of a theory (in one or more of its representations) that facilitate the use of the theory' (De Regt 2018, p. 40). Drawing on physics, he further considers a theory to be (at least to a first approximation) intelligible to a scientist if that scientists can 'recognize qualitatively characteristic consequences of the theory without performing exact calculations' (De Regt 2018, p. 102).

From these preliminaries, it follows that intelligibility cannot be intrinsic to a theory, but must depend on the cognitive ability and conceptual tools available to scientists. This, in turn, makes scientific intelligibility highly context-dependent. Scientists chose theories in part on their ability to put those theories to work, making intelligibility an important determinant of the success and propagation of scientific theories. The context-dependence of intelligibility explains why a theory that required great intellectual effort in the 1960's can be textbook material today[iv]. At the same time, a theory that is easy to grasp may continue to be applied even if a harder or unfamiliar theory would produce a deeper understanding of the phenomena to be explained.

De Regt illustrates the relationship between intelligibility of theories and understanding of phenomena with the kinetic theory of gases (De Regt 2018, p. 138). The kinetic theory represents real gases as an aggregation of particles that obey Newtonian mechanics. On its own, this generic representation merely provides the principles of the theory: even elementary gas laws (e.g., Boyle's law) can only be explained by developing more specific representational models that rely on further idealization, such as assuming that the particles are smooth and hard elastic spheres. To use the kinetic theory (i.e., to apply it to explain phenomena), scientists must make decisions that are tailored to their explanatory aims. These decisions would be difficult to make unless the theory was intelligible to the scientist in the sense described above; only by being able to "recognize qualitatively characteristic consequences of the theory" can the scientist make informed decisions about the idealizations and approximations that go into building a model.

Is this model of scientific understanding applicable to evolutionary biology? More specifically, does assigning agency and goals to genes facilitate the intelligibility of evolutionary theory; that is, does it help scientists grasp the consequences of the theory without writing down the maths? There seem to be good reasons to believe that it does.

One reason is that the traditional working horse of evolutionary explanation – theoretical population genetics – is something that many biologists find hard going. Dobzhansky, for example, apparently found it hard enough to hum through the equations as he went through Sewall Wright's papers (Provine 2003). Yet, population genetic models require nothing more than the algebra skills many biologists would have acquired before entering university. This suggests that it is not (only) the mathematical operations that make population genetics hard, but that the theory is difficult to understand. A skilled population geneticist intuitively 'knows' how to use the basic principles of population genetics in order to construct specific models, tailored to a particular explanation of a particular phenomenon. In contrast, biologists who lack this intuition[v] may not even know where to begin, even if they are perfectly apt at algebra.

Scientific theories that are difficult to grasp motivate the use of visual or conceptual tools to further their intelligibility (De Regt 2018). The 'adaptive landscape', introduced by Sewall Wright, is a good example of a visual tool: it makes it easier to infer consequences of evolutionary theory without calculations (Provine 1986). Thinking of genes as agents that pursue goals arguably serve a similar function; it makes it easier to intuitively grasp the possible consequences of an intervention on relevant variables, such as fitness or the rate of dispersal. Here is theoretical evolutionary biologist John Maynard-Smith speaking:

> I am prepared to think as loosely as necessary to give me an idea when I'm confronted with a new biological problem. If it helps me think to say, 'If I was a gene, I would do so-and-so' then I think that is OK. But when I've got an idea, I want to be able to write down the equation and show that the idea works. (…) I'm

> all for loose thinking. We all need ideas. (Maynard-Smith 1998, quoted in Ågren 2021)

Maynard-Smith here interprets his 'loose thinking' of what a gene should do as a way to generate possible solutions to a problem that can be formalized mathematically. This suggests that applying agential concepts to genes enabled Maynard-Smith to argue qualitatively and to recognize possible consequences of evolutionary theory without formal calculation.

If this interpretation is correct, agential concepts are not just a convenient short-hand, but rather conceptual tools that make evolutionary theory intelligible (at least to some scientists)[vi]. This conclusion is compatible with standard justification for agential thinking, which emphasizes its metaphorical nature (e.g., Haig 1997). However, it identifies a distinct epistemic function of agency or goal-directedness. This function makes no assumption about the metaphysical status of genes as agents; indeed, biologists could deny any biological entity agency and yet apply agential concepts to make biological theories intelligible. In practice, such a situation may be hard to sustain, as metaphors that make scientific theories intelligible have a tendency to shape metaphysical views and vice versa (Godfrey-Smith 2009; De Regt 2018)[vii].

3 Agential concepts can set explanatory standards

While genetic agency is widely acknowledged to be a metaphor, biologists really do consider individual organisms active, purposive agents (e.g., Dobzhansky 1968; Waddington 1969; Mayr 1974). Even a very demanding concept of agency will apply to humans, and less demanding concepts will grant agency to the smallest autonomous living system – the cell – and perhaps other organismal systems (Box 1). The dilemma is that, to be a respectable natural science, evolutionary biologists seemingly must deny the organism's goal-oriented activities any explanatory relevance for adaptation and diversification. That is, they must demonstrate that their evolutionary explanations not only are compatible with the causal-mechanistic ideal, but that the structure of evolutionary theory itself ensures 'an impermeable barrier between individual agency and evolutionary transformation' (Riskin 2020, p. 273).

Ernst Mayr's exemplar of how organismal agency can be at once accepted as factual but denied any explanatory relevance for evolution is an influential attempt to erect such a barrier (Laland et al. 2011; Corning 2019). Acknowledging that organisms appear as purposive agents, Mayr insisted that their goal-oriented activities imply the presence of a program, encoded by the inherited genome (Mayr 1961,1974,1988). Organisms are literally programmed to ensure that their development, physiology and behaviour are directed towards outcomes that serve their survival and reproductive interests (organismal activities and behaviours are *teleonomic* rather than *teleological*[viii]; Mayr 1988). In Mayr's view, developmental biologists and physiologists are concerned with explanations of how the genetic program is decoded, or how the program works. Such explanations rely on 'proximate' causes that readily fit within the mechanistic ideal. For example, a mechanistic explanation for a lizard escaping a predator may refer to the visual input stimulating the sensory system and the brain, which in turn triggers the muscles and tendons to cause movement of the limbs. Evolutionary biologists, on the other hand, are concerned with why particular programs exist and they rely on 'ultimate causes'. Lizards that do not run from predators fail to pass on their genes, making the survival difference between genes that code for skittish and docile lizards an explanation for why lizards run from predators, as well as an explanation for why the behaviour appears goal-oriented.

This 'neo-Darwinian' representation of evolution by natural selection solves the evolutionary biologist's dilemma. Firstly, it 'mechanizes' historical explanations by enabling more specific representational theories that explain adaptive (and non-adaptive) evolution by relying solely on causes and forces. This is well exemplified by the explanations delivered by theoretical population genetic models[ix]. Secondly, the structure of evolutionary theory seemingly rules out *a priori* goal-oriented processes – organismal development, physiology or behaviour – from evolutionary explanation (Mayr 1961).[x] Those processes do not directly change allele frequencies, and thus may seem unable to account for directional, adaptive change. Only one difference maker for adaptive bias remains: fitness differences between genotypes. Thus, according to this perspective, evolutionary biologists can consider organismal agency a real phenomenon, yet shrug it off as an intermediate, proximate, expression of a genetic program; an expression of past natural selection rather than a cause of future adaptation.[xi] Under the neo-Darwinian representation of the evolutionary process, evolutionary explanation is genes and natural selection all the way down.

The elimination of organismal agency and goal-oriented processes from evolutionary theory has successfully structured evolutionary inquiry around a set of problems that the theory is appropriate for. Yet, two limitations are difficult to avoid.

Firstly, some evolutionary problems inevitably fall outside the theory's domain. The origin of novelty is one familiar example. An explanation for a novel morphological feature, such as the flower, requires attention to the sequence of morphological transformation over evolutionary time and the genetic and developmental changes that were responsible for this change (Calcott 2009). This is not something that a population genetic model can explain: someone who explained the evolution of flowers in terms of fitness advantages and shifting allele frequencies would simply be off the mark. Selective explanations are valid explanations in their own right, but they are not adequate explanations for the flower as an evolutionary novelty (Love 2008).

Secondly, to account for adaptation in terms of fitness differences, the neo-Darwinian representation must make assumptions about the evolutionary process that are at odds with biological reality (Walsh 2015; Uller and Helanterä 2019; see Potochnik 2017 for a general discussion). The insistence that natural selection of genetic variation is the only legitimate difference maker for adaptive evolution (e.g., Charlesworth et al. 2017) is a consequence of these idealizations[xii], not a fundamental feature of the causal fabric of the world. Failure to appreciate that criteria of explanatory adequacy depend on how the evolutionary process is represented by theories and models can make it appear as if the privileged role of genes and natural selection is indispensable to *any* evolutionary explanation. This, in turn, may result in an over-reliance on fitness-based explanations and neglect of alternatives (e.g., Gould and Lewontin 1979; Lloyd 2005), slow acceptance of phenomena that do not fit assumptions (e.g., extra-genetic inheritance; Jablonka and Lamb 2014), and limited explanatory power as a result of failure to account for adaptive biases imposed by goal-oriented processes (e.g., development; West-Eberhard 2003; Kirschner and Gerhardt 2007). The genetic representation of development and evolution also carries several conceptual difficulties and inconsistencies (Keller 2000; Oyama 2000; Griffiths and Stotz 2007).

As these inconsistencies, explanatory gaps and deficits are built into the structure of evolutionary theory itself, they cannot easily be amended without alternative representations that pick out different causal patterns. Such attempts have often been motivated by understanding the evolutionary consequences of organismal agency and goal-oriented processes. For example, to Mary-Jane West-Eberhard, adaptive evolution begins with phenotypic accommodation – adaptive mutual adjustment among variable parts during development – in response to genetic or environmental perturbation (West-Eberhard 2003, p. 51, 140-141). To West-Eberhard and similar-minded biologists, development is not just a conservative force that constrains evolution, as the neo-Darwinian representations depicts it, but also what makes the generation of adaptive variation possible (e.g., Salazar-Ciudad 2007). Such ‘facilitated variation’ (Gerhardt and Kirschner 2007) contributes to evolution by providing natural selection with a source of putative adaptive phenotypes, which makes development a co-determinant of the rate and direction of adaptive change. Others (e.g., Lewontin 1983; Odling-Smee 1988; Edelaar & Bolnick 2019) have emphasized that organisms can influence their own individual fitness by modifying selective environments, which makes such 'niche construction' a co-determinant of adaptation (Odling-Smee et al. 2003).

Concepts like 'phenotypic accommodation', 'facilitated variation' and 'niche construction' capture aspects of the responsive, self-organising nature of development that allows organisms to maintain functional stability even when exposed to conditions that threaten their persistence. These concepts thus direct attention to 'internal' or 'agential' sources of consistent bias in evolution that may account for the evolution of particular adaptations, diversification or evolvability (Sultan et al. 2021)[xiii]. In so doing, those concepts set an alternative explanatory agenda; they structure scientific investigation of evolution according to criteria of explanatory adequacy that are different to those of the neo-Darwinian representation of evolution by natural selection (Love 2008; Brigandt 2011). These alternative criteria in turn determine what biological fields that are deemed relevant for a scientific understanding of evolution. A scientist that considers development a source of adaptive bias in evolution will draw on a different sets of knowledge, concepts and methods from other disciplines than will a scientist that considers natural selection alone responsible for the diversity and adaptive fit of organisms.

A characteristic feature of such evolutionary research is that it admits developmental and behavioural difference makers (e.g., developmental plasticity, habitat choice) alongside fitness differences as explanans for adaptation and diversification. A major challenge is therefore to distribute causal responsibility between transformational and selective processes, and among the different types of difference-makers involved (for a general discussion, see Love 2017) [xiv]. The developmental causes of adaptive bias that refer to, for example, exploratory processes of bone and tissue growth are not easily compared to the causes of fitness differences, which begs the question of how different causes should be integrated in an evolutionary explanation[xv]. Moreover, the processes that generate phenotypic variation and differential fitness can be intertwined or modify each other on the relevant time scales (Watson and Thies 2019), which makes it difficult to attribute adaptive change to either one or the other (Uller and Helanterä 2019).

Another challenge is to decide when the addition of particular difference makers results in better or worse explanations (Ylikoski and Kourikoski 2010). For example, the striking morphological diversity and convergence of cichlid fish in African lakes (Kocher et al. 1993) may be explained in part by the phylogenetically shared developmental biases of fish feeding on different diets, and in part by consistent fitness differences between fish with alternative morphologies (e.g., Muschick et al. 2011; Conith and Albertson 2021; review in Schneider and Meyer 2017). An explanation that refers jointly to developmental bias and natural selection can arguably be preferable over an explanation that refers to either natural selection or developmental bias. But exactly when is the first explanation better than the second or *vice versa*? Choosing between alternative explanations may be more difficult than it seems as it requires an evaluation of the explanatory power of different representations of the evolutionary process (Baedke et al. 2020; Uller et al. 2020).

4 Agential concepts can explain evolutionary change

The extensions to evolutionary theory discussed in the previous section have been accompanied by a lively debate about the re-interpretation of development and other 'intrinsic' or 'proximate' causes in evolution (Laland et al. 2014). Yet, these alternative explanatory agendas still explain in terms of causes and mechanisms: the goal-oriented developmental processes and feeding behaviours of cichlids may be sources of adaptively relevant causes, but it is not the goals themselves that account for the convergence in the cichlids' morphology. To some, this stance does not go far enough in how the agential perspective should influence evolutionary explanation. One critic is philosopher of science Denis Walsh. Walsh argues that, if we accept that adaptive evolution is a consequence of organisms' goal-oriented activities, there is a counterfactual relation between goals and organismal activities that can exploited to explain evolutionary events, and therefore to reinstall teleology as a mode of explanation in evolutionary biology (Walsh 2015,2018, see also Jaeger 2021).

To explain just how seriously he wishes us to take agency, Walsh distinguishes what he calls object theories from agent theories (Walsh 2015, ch. 10). Object theories are those that aim to explain what is happening to a set of objects within a system by externally imposing a set of rules or principles. Population genetics is a good example; the alleles are the objects of interest, and those alleles mutate and change in frequency according to principles that are external to the alleles and exist independently of them. The result is an explanatory asymmetry:

> The principles – e.g. laws of nature, initial conditions, and the space of possible configurations – explain the changes to the objects in the domain, but the objects do not explain the principles (Walsh 2015, p. 212).

Agent theories are different from object theories, because

> In an agent theory the entities in the domain include both agents and the principles we use to explain their dynamics. The agents' activities are generated endogenously; agents cause their own changes in state in response to the conditions they encounter. These conditions, in turn, are largely of the agent's making (Walsh 2015, p.212).

This explanatory symmetry between entities and principles means that the explanatory objective of an agent theory is different from that of an object theory: an agent theory explains the dynamics of a set of entities by accounting for the interplay between what these entities do and the principles used to explain their behaviour. Contemporary theory on the evolutionary causes and consequences of goal-oriented processes (e.g., developmental plasticity) do not meet these criteria; a tell-tale is that such models and explanations often rely on traditional tools and theories, such as population or quantitative genetics (e.g., Lande 2009; Chevin et al. 2010; Levis & Pfennig 2016).

Why would an understanding of biological evolution demand agent theories rather than object theories? One reason is that biological evolution appears to be an open-ended process. As organisms evolved from single-cell organisms into multicellular organisms, and eventually into organisms with symbolic means of transmitting information, they changed the evolutionary process itself. The evolution of multicellularity or symbolic communication are not 'just' the evolution of another unit of selection or an adaptation to transmit information; by changing the principles of the evolutionary dynamics these innovations opened up opportunities for evolution that were previously impossible.

Another reason for agent theories is even more fundamental, because it addresses why biological evolution can be open-ended (Jaeger 2021). Not all entities that can evolve by natural selection exhibit this open-endedness; algorithms in computer programs do not, for example. This suggests that biological evolution is possible because organisms are living beings, not because they fulfil abstract principles evolution by natural selection (e.g., as summarized by Lewontin 1970)[xvi]. On the agential view, evolution happens because organisms engage in goal-oriented activities. The organisational account of biological agency further tells us that these responses are initiated not by a program but 'from within' the organism itself, and hence cannot be predicted even with full knowledge of the population's selective history. For example, understanding why African cichlids evolved a similar suite of morphologies in different lakes would require both an understanding of the cichlids' developmental and behavioural repertoire (what they are capable of) and what their surroundings – rocks, algae, sand, snails, other fishes – offer or furnish to the cichlids as they pursue their goals[xvii]. In practice, a decent knowledge of cichlid biology may suffice. However, the philosophical point is that it is these repertoires and affordances that explain why the cichlids evolved similar adaptations, not 'natural selection' (Walsh 2015).

If organisms' goal-oriented engagement with their affordances is what enables adaptive evolution, would not citing those stable endpoints – or goals – also explain why evolution proceeded in one direction rather than in another (Walsh 2015; Jaeger 2021; Sultan et al. 2021)[xviii]? This is, perhaps, not so different from legitimate teleological explanations for adaptive cultural change in humans, a kind of explanation that can be preferred over cultural selection explanations (Chellappoo 2022). Jaeger goes as far as concluding that

> naturalistic teleological explanation is a necessary part of any agential theory of evolution, because of the immanence of rules which are generated by the agents themselves. (Jaeger 2021, p.31)

It is important to stress that none of this implies that the evolutionary process itself is goal-oriented; the goals and purposes in naturalistic teleological (or teleonomic) explanations would be those of organisms (and perhaps their parts or collectives of organisms that exhibit organisational closure; Box 1)[xix]. Neither does naturalistic teleology imply that goals somehow cause their means since teleological explanations are not causal (this is not as fatal as it may seem since natural sciences already do admit 'becauses' without 'causes'; Lange 2016).

Granting all this, what can an explanation that refers to organismal goals contribute to the scientific understanding of evolution?[xx]

One possibility is that naturalistic teleological explanations fill an explanatory gap that mechanistic explanations simply cannot fill. Consider again the evolutionary convergence of cichlid jaw morphologies. As mentioned above, an explanation for this convergence could cite a number of different sources of adaptive (and perhaps non-adaptive) bias. These include diet choices and preferences, the developmental genetics and plasticity of craniofacial development, and fitness differences between individuals with different morphologies. But the consistent biases imposed by the fishes' search for food or the development of their mouth parts may seem fortuitous – and hardly capable of promoting consistent adaptive bias in evolution – unless they are understood as means conducive to the goals that fish pursue. That is, a purely causal account of adaptive convergence seems to leave an explanatory gap; it refers to the adaptive biases on phenotypic evolution caused by development or differential fitness, but it struggles to make sense of why those biases (and not others) exist. Indeed, it is this explanatory gap that makes it tempting to explain any adaptive bias caused by developmental plasticity in terms of past natural selection on random genetic variation (e.g., Wray et al. 2014). However, if Walsh and others are right about agent theories, the regular attainment of particular phenotypes in cichlid evolution can neither be fully accounted for by chance and natural selection, nor by the addition of mechanistic developmental causation, but must credit the goals, affordances, and repertoires of individual fish. Goals can be legitimate difference makers (but, again, not in a *causal* explanation) because the presence of goals is a natural consequence of how organisms are organised, and those goals imposes a certain order on the world that, in the case of cichlids, resulted in the repeated evolution of a similar set of morphologies.

Whatever one makes of this case for naturalistic teleological (or teleonomic) explanation, biologists do recognize that it can matter to evolutionary dynamics whether or not organismal activities are oriented towards goals. For example, the niche construction literature emphasizes that organismal activities that are goal-oriented result in more consistent selective pressures than do other sources of selection (Clark et al. 2020), and that goal-oriented modification of environments can result in highly regular sequences of adaptive change (Laland et al. 2017). Similarly, it has been suggested that the evolutionary consequences of developmental plasticity will depend on whether or not individual responses are directed towards goals, or what goals those responses actually serve (Feiner et al. 2019)[xxi]. Thus, insofar as biological systems exhibit the organisational closure that makes them goal-oriented, reference to those goals can perhaps help biologists understand patterns of evolution that may be difficult to grasp by explaining solely in terms of mechanisms and causes.

5 Conclusions

The goal-oriented processes that we observe as development and behaviour have proven difficult to fit within the explanatory standards of the dominant evolutionary theories. This has not prevented biologists from making liberal use of agential concepts, and those concepts can indeed play legitimate epistemic functions even when the objects, such as genes, do not fulfil criteria for biological agency. Biological agency is hard to deny for organisms, however, and some contemporary research on biological evolution is in fact organised around concepts that refer to organisms' ability to initiate activities from within their own boundaries, to sense and respond to the conditions they encounter, and to maintain their functional stability when perturbed. This suggests that biological agency can exercise a substantial influence on evolutionary biology by influencing the kinds of problems biologists address, what knowledge, concepts and methods they need to import from other disciplines, and what they consider a satisfactorily explanation.

References

Baedke, J., A. Fábregas-Tejeda, and F. Vergara-Silva. 2020. Does the extended evolutionary synthesis entail extended explanatory power? *Biology & Philosophy* 35:20.

Brigandt, I. 2010. Beyond reduction and pluralism: toward an epistemology of explanatory integration in biology. *Erkenntnis* 73: 295-311.

Calcott, B. 2009. Lineage explanations: explaining how biological mechanisms change. *British Journal for the Philosophy of Science* 60: 51-78.

Charlesworth, D., N. H. Barton, and B. Charlesworth. 2017. The sources of adaptive variation. *Proceedings of the Royal Society B-Biological Sciences* 284: 20162864.

Chellappoo, A. 2022. When can cultural selection explain adaptation? *Biology & Philosophy* 37: 2.

Chevin, L.-M., R. Lande, and G. M. Mace. 2010. Adaptation, plasticity, and extinction in a changing environment: Towards a predictive theory. *PLoS Biology* 8: e1000357.

Clark, A. D., D. Deffner, K. Laland, J. Odling-Smee, and J. Endler. 2020. Niche construction affects the variability and strength of natural selection. *American Naturalist* 195: 16-30.

Corning, P. A. 2019. Teleonomy and the proximate–ultimate distinction revisited. *Biological Journal of the Linnean Society* 127: 912-916.

Conith, A. J. and R. C. Albertson. 2021. The cichlid oral and pharyngeal jaws are evolutionarily and genetically coupled. *Nature Communications* 12: 5477.

Dawkins, R. 1976. *The selfish gene.* Oxford University Press.

De Regt, H.W. 2018. *Understanding scientific understanding.* Oxford University Press.

Dickins, T. E. and Q. Rahman. 2012. The extended evolutionary synthesis and the role of soft inheritance in evolution. *Proceedings of the Royal Society B-Biological Sciences* 279: 2913-2921.

Edelaar, P. and D. I. Bolnick. 2019. Appreciating the multiple processes increasing individual or population fitness. *Trends in Ecology & Evolution* 34: 435-446.

Feiner, N., I. S. C. Jackson, K. L. Munch, R. Radersma, and T. Uller. 2020. Plasticity and evolutionary convergence in the locomotor skeleton of Greater Antillean Anolis lizards. *Elife* 9: e57468.

Godfrey-Smith, P. *Darwinian populations and natural selection.* Oxford University Press

Gould, S. J. and R. C. Lewontin. 1979. Spandrels of San-Marco and the Panglossian paradigm - a critique of the adaptationist program. *Proceedings of the Royal Society of London Series B-Biological Sciences* 205: 581-598.

Griffiths. P. & K. Stotz 2013. *Genetics and philosophy: An introduction.* Cambridge University Press.

Haig, D. 1997. The social gene. In J.R. Krebs and N.B. Davies,eds. *Behavioral ecology, 4th edition,* pp.284-304. Blackwell Scientific.

Jaeger, J. 2021. The Fourth Perspective: Evolution and Organismal Agency. OSF Preprints. February 26. doi:10.31219/osf.io/2g7fh.

Keller, E.F. 2000. *The century of the gene.* Harvard University Press.

Gerhart, J. and M. Kirschner. 2007. The theory of facilitated variation. *Proceedings of the National Academy of Sciences of the United States of America* 104: 8582-8589.

Kocher, T. D., J. A. Conroy, K. R. McKaye, and J. R. Stauffer. 1993. Similar Morphologies of Cichlid Fish in Lakes Tanganyika and Malawi Are Due to Convergence. *Molecular Phylogenetics and Evolution* 2: 158-165.

Laland, K. N., K. Sterelny, J. Odling-Smee, W. Hoppitt, and T. Uller. 2011. Cause and effect in biology revisited: Is Mayr's proximate-ultimate dichotomy still useful? *Science* 334: 1512-1516.

Laland, K. N., J. Odling-Smee, W. Hoppitt, and T. Uller. 2013. More on how and why: cause and effect in biology revisited. *Biology & Philosophy* 28: 719-745.

Laland, K., T. Uller, M. Feldman, K. Sterelny, G. B. Muller, A. Moczek, E. Jablonka, and J. Odling-Smee. 2014. Does evolutionary theory need a rethink? *Nature* 514:161-164.

Laland, K. N., T. Uller, M. W. Fellman, K. Sterelny, G. B. Muller, A. Moczek, E. Jablonka, and J. Odling-Smee. 2015. The extended evolutionary synthesis: its structure, assumptions and predictions. *Proceedings of the Royal Society B-Biological Sciences* 282.

Laland, K. N., J. Odling-Smee, and J. Endler. 2017. Niche construction, sources of selection and trait coevolution. Interface Focus 7:20160147.

Lande, R. 2009. Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *Journal of Evolutionary Biology* 22: 1435–1446

Lange, M. 2016. *Because without cause: Non-causal explanations in science and mathematics.* Oxford University Press.

Levis, N. A. and D. W. Pfennig. 2016. Evaluating 'Plasticity-First' evolution in nature: key criteria and empirical approaches. *Trends in Ecology & Evolution* 31: 563-574.

Lewontin, R. C. 1970. The units of selection. *Annual Review of Ecology and Systematics* 1:1-18.

Lloyd, E.A. 2005. *The case of the female orgasm. Bias in the science of evolution.* Harvard University Press.

Love, A. C. 2008. Explaining evolutionary innovations and novelties: criteria of explanatory adequacy and epistemological prerequisites. *Philosophy of Science* 75: 874-886.

Love, A.C., T.A. Stewart, G.P. Wagner, and S.A. Newman. 2017. Perspectives on integrating genetic and physical explanations of evolution and development. *Integrative and Comparative Biology 57:* 1258–1268

Love, A.C. 2017. Building integrated explanatory models of complex biological phenomena: From Mill's methods to a causal mosaic. In M. Massimi and J.-W. Romeijn (eds), *EPSA Philosophy of Science: Düsseldorf 2015*. *The European Philosophy of Science Association Proceedings*, Vol. 5. Dordrecht: Springer, pp. 221–232.

Maynard-Smith, J. 1998. The units of selection. In G.R. Bock and J.A. Goode, eds. *The limits of reductionism in biology*, pp. 203-210. Novartis Foundation, Wiley.

Mayr, E. 1961. Cause and effect in biology – kinds of causes, predictability, and teleology are viewed by a practicing biologist. *Science* 134: 1501-1506.

Mayr 1974. Teleological and teleonomic: a new analysis. *Boston Studies in the Philosophy of Science* 14:91-117.

Mayr, E. 1988. *Toward a new philosophy of biology*. Observations of an evolutionist. Harvard University Press.

Mesoudi, A., S. Blanchet, A. Charmantier, E. Danchin, L. Fogarty, E. Jablonka, K. N. Laland, T. J. H. Morgan, G. B. Mueller, F. J. Odling-Smee, and B. Pujol. 2013. Is Non-genetic inheritance just a proximate mechanism? A corroboration of the extended evolutionary synthesis. *Biological Theory* 7: 189-195.

Montévil, M., and M. Mossio 2015. Biological organisation as closure of constraints. *Journal of Theoretical Biology* 372: 179–91.

Mossio, M., Saborido, C., and A. Moreno 2009. An organizational account of biological functions. *British Journal for the Philosophy of Science* 60: 813–41

Mossio, M. and L. Bich. 2017. What makes biological organisation teleological? *Synthese* 194: 1089-1114.

Muschick M, Barluenga M, Salzburger W, Meyer A (2011) Adaptive phenotypic plasticity in the Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. BMC Evolutionary Biology 11: 116.

Odling-Smee, F.J. 1988. Niche-constructing phenotypes. In H.C. Plotkin, ed. *The role of behaviour in evolution*, pp.73-132. MIT Press.

Odling-Smee, F.J., Laland, K.N., and Feldman, M. 2003. *Niche construction: the neglected process in evolution.* Princeton University Press.

Okasha, S. 2018. *Agents and goals in evolution.* Oxford University Press.

Otsuka, J. 2016. A critical review of the statisticalist debate. *Biology & Philosophy* 31: 459-482.

Oyama, S. 2000. *The ontogeny of information: Developmental systems and evolution.* Duke University Press.

Pigliucci, M. and R. Scholl, The proximate–ultimate distinction and evolutionary developmental biology: causal irrelevance versus explanatory abstraction. Biology and Philosophy 30: 653-670.

Potochnik, A. 2017. *Idealization and the aims of science.* University of Chicago Press.

Provine, W.B. 1986. *Sewall Wright and evolutionary biology.* Chicago University Press.

Provine, W.B. 2003. In R.C. Lewontin, J. Moore, W.B. Provine, B. Wallace, eds, *Dobzhansky's genetics of natural populations I-XLIII*. Columbia University Press.

Riskin, J. 2020. Biology's mistress, a brief history. *Interdisciplinary Science Reviews* 45: 268-298.

Schneider, R. F. and A. Meyer. 2017. How plasticity, genetic assimilation and cryptic genetic variation may contribute to adaptive radiations. *Molecular Ecology* 26: 330-350.

Scott-Phillips, T. C., T. E. Dickins, and S. A. West. 2011. Evolutionary theory and the ultimate-proximate distinction in the human behavioral sciences. *Perspectives on Psychological Science* 6: 38-47.

Stoltzfus, A. 2021. *Mutation, randomness, and evolution.* Oxford University Press.

Sultan, S. E., A. P. Moczek, and D. Walsh. 2022. Bridging the explanatory gaps: What can we learn from a biological agency perspective? *BioEssays* 44: 2100185.

Uller, T. and H. Helanterä. 2019. Niche construction and conceptual change in evolutionary biology. *The British Journal for the Philosophy of Science* 70: 351-375.

Uller, T., A. P. Moczek, R. A. Watson, P. M. Brakefield, and K. N. Laland. 2018. Developmental bias and evolution: a regulatory network perspective. *Genetics* 209: 949-966.

Uller, T., N. Feiner, R. Radersma, I. S. C. Jackson, and A. Rago. 2020. Developmental plasticity and evolutionary explanations. *Evolution & Development* 22: 47-55.

Waddington, C.H. 1968. *Towards a theoretical biology, Vol I.* Edinburgh University Press.

Walsh, D. 2012. Mechanism and purpose. A case for natural teleology. *Studies in the History and Philosophy of Biology and the Biomedical Sciences* 43: 173-181.

Walsh, D. 2015. *Organisms, agency and evolution*. Cambridge University Press

Walsh, D. 2018. Objectcy and agency: Towards a methodological vitalism. In D.J. Nicholson and J. Dupré, eds. *Everything flows. Towards a processual philosophy of biology*. Oxford University Press.

Watson, R. A and C. Thies. 2019. Are developmental plasticity, niche construction and extended inheritance necessary for evolution by natural selection? The role of active phenotypes in the minimal criteria for Darwinian individuality. Pp. 197-226 in T. Uller and K. N. Laland (eds) *Evolutionary Causation*. Cambridge, MA: MIT Press.

Welch J.J. 2017. What's wrong with evolutionary biology? *Biology & Philosophy* 32: 263–279

West-Eberhard, M.J. 2003. *Developmental plasticity and evolution.* Oxford University Press

Wray, G. A., H. E. Hoekstra, D. J. Futuyma, R. E. Lenski, T. F. C. Mackay, D. Schluter, and J. E. Strassmann. 2014. Does evolutionary theory need a rethink? - No, all is well. *Nature* 514: 161-164.

Ylikoski, P. and J. Kuorikoski. 2010. Dissecting explanatory power. *Philosophical Studies* 148: 201-219.

Ågren, A. 2021. *The gene's-eye view of evolution.* Oxford University Press

Box 1. What is biological agency?

In daily speech, reference to agency, goals and purpose often implies intentions or desires. The biological agency concept(s) of this chapter is much broader than that, although it of course includes the sophisticated cognitive abilities of humans. An inclusive concept of agency considers it a dynamical property of a system (e.g., an organism) that makes the system able to 'transduce, configure, and respond to the conditions it encounters', and to maintain 'functional stability in response to conditions that would otherwise compromise their viability' (Sultan et al. 2021). Agential systems are characterized by their ability to initiate activity from within their own boundaries, to sustain and transform themselves through novel structures, functions and activities, often in ways that ensure their continued existence (Walsh 2015). In contrast to a storm or a biogeochemical cycle, an organism can change what it does to navigate obstacles and overcome challenges that threaten its survival. We observe this as goal-oriented activities or behaviours, which are characteristic features of all organisms; the *E. coli* as well as the elephant.

These behavioural repertoires are often considered to be encoded in a genetic program, which makes organisms appear goal-oriented (or *teleonomic;* Mayr 1974) without them really being agents by their own making. That is, agency is not a property of organisms as much as a property of their genomes, much like how a piece of hardware can respond to commands and carry out its functions only because of its software. But in contrast to the programs written by human software developers, the presence of organismal goal-orientation and the organisms' pursuit of particular goals are not designed, but externally provided by natural selection on random genetic variation (more on that in the main part of the chapter).

An alternative way to think about organismal agency is to attribute it, not to a program, but to a particular kind of closed organisation where 'the processes and constraints…logically and materially entail each other' (Jaeger 2021, p.8; Mossio et al. 2009; Montevil & Mossio

2015). A system in which processes that are essential for the organism’s continued existence regulate and sustain each other can be considered to demonstrate an intrinsic orientation towards goals (or an *internal teleology*) because the organism must act upon the world to stay alive (Mossio & Bich 2017). Organisms are agents because of what they do, and they must continue to be agents to stay alive. Since the closed organisation of living systems are not reducible to genes, or even gene regulatory networks, it seems problematic to explain organismal agency by natural selection of random genetic variation alone, and thus this perspective tends to direct attention to the transformation of biological processes during evolution.

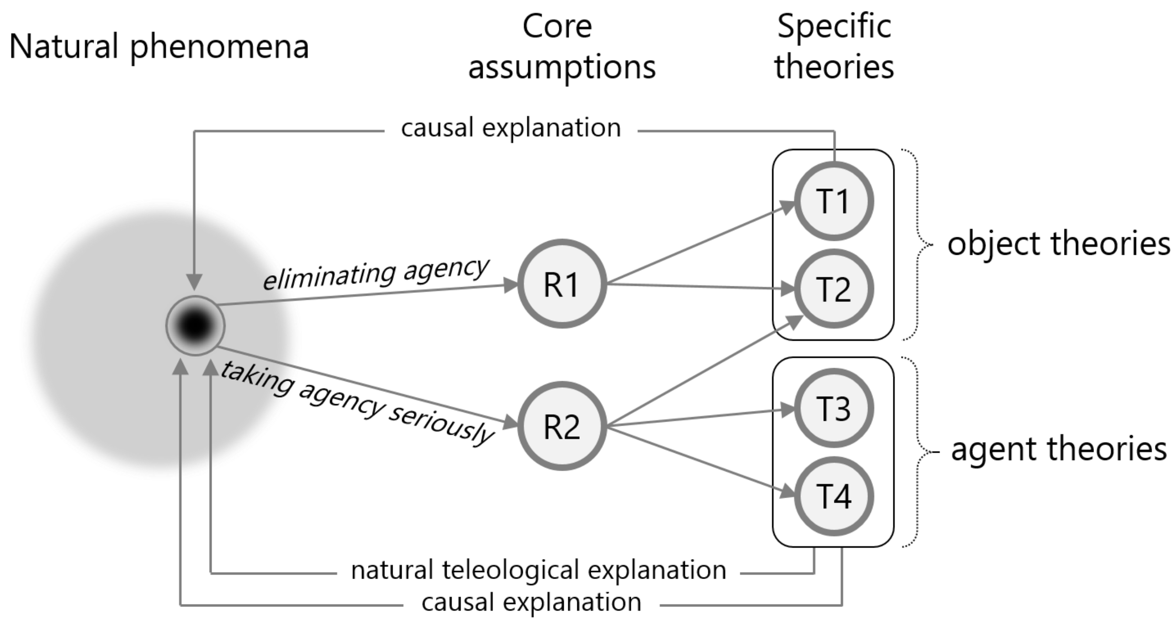

Figure 1. Scientific understanding requires application of an appropriate explanatory model or theory to natural phenomena. The complexity of biological evolution means that scientists must rely on simplified generic representations (R1-R2), with particular core assumptions, from which more specific scientific theories (T1-T4) or models can be developed. The standard theories in evolutionary biology are object theories: they explain by externally imposing a set of rules or principles on the objects or entities of the theory (e.g., individuals or alleles). An alternative set of theories are agent theories, which explain by accounting for the interplay between what entities do and the principles used to explain their behaviour (this is explained more fully in the main chapter).

The three broad categories of epistemic functions of agential concepts discussed in this chapter can be illustrated in this figure. As discussed in section 2, applying agential concepts to living systems or their parts (e.g., genes) can facilitate evolutionary explanation by making evolutionary theory (e.g., T1) intelligible to the scientists who wish to use it to explain natural phenomena. Section 3 demonstrates how agential concepts (e.g., phenotypic accommodation) can structure investigation of biological evolution by influencing how scientists chose to represent the evolutionary process. These representations can admit, e.g., development, extra-genetic inheritance, and niche construction to explain (alone or jointly with natural selection) adaptive evolution. Such theories may be developed from existing object theories (e.g., T2), which means that the interpretative understanding of a single representative model can differ depending on which core assumptions particular scientists hold on to. Section 4 discusses how representations of evolution that are structured around organismal agency also may motivate the development of agent theories, which are incompatible with the core assumptions of traditional representations of the evolutionary process (arrow connecting R2 and T3). As agent theories are fundamentally different from object theories, they may admit both causal explanation and naturalistic teleological explanations, where the latter involves the demonstration that the attainment of an evolutionary event was conducive to organismal goals ('naturalistic teleological explanation' arrow from agent theories to natural phenomena).

---

[i] This chapter is concerned with the explanatory roles played by the goal-oriented activities of organisms; not with the potential utility of ascribing agency to natural selection or attributing the evolutionary process itself with goal-oriented properties.

[ii] This does not mean it is just loose talk; to Dawkins (1976), the agency metaphor appears to communicate a deep metaphysical commitment about the world.

[iii] Scientific understanding also requires that the explanation based on the application of a theory fulfils the fundamental scientific values of empirical adequacy and internal consistency (De Regt, 2018: 93).

[iv] Conversely, a scientific theory that was widely used in the past, often to good effect, can seem entirely unintelligible to the scientists of today.

[v] The use of 'intuition' here follows De Regt (2018: 109-113).

[vi] Not all scientists will find the same conceptual tools useful; thus, we may expect differences in opinion regarding the scientific value of agential thinking to reflect the background, cognitive ability, and skill sets of individual scientists.

[vii] It has been pointed out there are significant risks associated with applying agential concepts to genes or other entities that do not have those properties (see e.g. Godfrey-Smith, 2009: ch.7). Perhaps the success of agential concepts in making evolutionary theory intelligible have contributed to illegitimate metaphysical views of genes among both scientists and the general public. These views may in turn have influenced what biologists consider fundamental or indispensable to evolutionary explanation.

[viii] On the teleonomy concept, see the chapter by Corning and Vane-Wright.

[ix] Not unlike statistical mechanics in physics, population geneticists represent evolutionary change in terms of forces affecting the spread and maintenance of alleles coding for alternative versions of a trait. This representation of evolution can answer a range of 'what-if-things-were-different' questions about adaptive and non-adaptive change; it is quantitative, predictive and empirically testable, all hallmarks of good natural science. For example, by assigning fitness values to genotypes, population genetic models demonstrate under what conditions natural selection will maintain more than one genotype. Models like these bring understanding because they help us grasp why different genotypes (and hence phenotypes) can co-exist by demonstrating how interventions on variables like fitness or population size influence the composition of genotypes within a population.

[x] Explanations that cite organismal development, physiology or behaviour are considered to violate the distinction between proximate and ultimate causation (Mayr, 1961). For more recent examples of how this distinction has been used to identify 'inadequate' evolutionary explanations, see Scott-Phillips et al., 2011; Dickins & Rahman, 2012; For counter-points, see Mesoudi et al., 2013; Laland et al., 2015. Different interpretations of the status of the proximate ultimate distinction in evolutionary biology are discussed in Laland et al., 2011; Laland et al., 2013 and the commentaries and author response to this paper in the same issue of *Biology & Philosophy*, and by Pigliucci & Scholl, 2015.

[xi] Biases imposed by mutation, development or inheritance that are not fitness-enhancing are also easily neglected under this representation as they will appear inconsequential for the 'interesting bits' of evolution (e.g., adaptation; see Stolzfus, 2021).

[xii] Key assumptions that underlie the explanatory standards of the neo-Darwinian representation of the evolutionary process concern the nature of genes (Oyama, 2000) and the autonomy of variation, fitness, and inheritance (Walsh, 2015; Uller & Helanterä, 2019).

[xiii] To those who emphasize the role of the organism in evolution, metaphors like genetic 'programs' and 'blueprints' discourage such work because they explain *away* agency rather than single out agency as a fundamental property of living systems.

[xiv] Love et al. (2017) discuss these challenges from the perspective of integrating genetic and physical explanations for the origins of novelties.

[xv] There is some debate over whether or not explanations that refer to natural selection are causal (Otsuka, 2016 ;Walsh et al., 2018), but this does not deny that there are challenges of explanatory integration (on the contrary, it would arguably make matters worse).

[xvi] While abstract criteria for evolution by natural selection are helpful to understand how evolution works, it makes a difference to evolution how those principles are instantiated by the evolving entities.

[xvii] *Affordance* is a concept used to describe this complementarity of organism and environment that are salient to an agent's pursuit of its goal.

[xviii] 'Because an agent is capable of attaining and maintaining stable endpoints that reliably secure its stability, one can cite the stable endpoint to which the system tends in explaining its activities' (Sultan et al., 2021). For Walsh's defence of naturalistic teleological explanations, see Walsh (2012;2015;2018). See also Mossio & Bich, 2017 and Jaeger, 2021.

[xix] Organisms, parts of organisms (e.g., metabolic processes within cells) and collectives (e.g., social insect colonies) are candidates for exhibiting closure (Mossio et al. 2009; Montévil & Mossio 2015). Whether or not they do is an empirical issue, and an important one to understand the role of agency in evolution. On this organisation account of biological agency, organismal agency can impose regularities in evolution and may therefore be responsible for macroevolutionary trends, but for the evolutionary process itself to be goal-oriented it too should exhibit closure.

[xx] To even grant the possibility of naturalistic teleological explanation in science may seem heretical. Some evolutionary biologists are understandably concerned that it blurs the distinction between scientific and unscientific explanation, and thus can be exploited by creationists, for example. Others may welcome naturalistic teleological explanation exactly because it addresses *scientifically* features of the living world that have been left unexplained by past scientific theories, and hence left vulnerable to exploitation by those hostile to science and scientific knowledge in general, and evolution in particular.

[xxi] Of course, there is nothing that prevent biologists to reject all these arguments for taking agency seriously, but still agree that goal-oriented activities have special evolutionary consequences: in the neo-Darwinian explanatory framework, the evolutionary consequences of goal-oriented activities would be fully accounted for by random genetic variation and natural selection, while the proximate development and behaviour of organisms are optional causal detail (Wray et al. 2014).