Integrating proximate and ultimate causation: Just one more go!

B. Thierry

Centre d'Ecologie, Physiologie et Ethologie, Centre National de la Recherche Scientifique, UPR 9010, Strasbourg, France

Parting between proximate and ultimate causation is questioned. Mayr elaborated this dual scheme to distinguish the making up and functioning of the phenotype from the genesis of the genotype. But evolution cannot reduce to a uniformitarian progress resulting from the regular action of natural selection, it also includes such processes as random sorting and species selection, which occur at different time scales. In addition, the epigenetic processes responsible for the development of organisms generate multiple constraints that direct the action of selection. If we want to integrate mechanistic and adaptive functions, we should drop the very formulation that perpetuates their separation.

Keywords: Causality, constraint, epigenesis, evolution, finality.

The classical parting between proximate and ultimate causation was elaborated by Mayr[1]. Proximate causes deal with the mechanisms responsible for the making up and functioning of the individual phenotype. Ultimate causes refer to the past conditions having led to the information encoded in DNA. According to Mayr[1], proximate causation takes place once the genetic program encoded is actualised in the individual, whereas ultimate causation determines the shaping of the program itself. This dual scheme may be viewed as a logical consequence of the Weismannian separation between soma and germ line. It assumes that we need a different perspective to understand the phenotype and the genotype. Biologists studying proximate causes ask *how* questions about mechanisms whereas those studying ultimate causes ask *why* questions about evolutionary finality. The phenomena involved in these different levels of causation occur on different time scales.

There is some mystery about the origin of the dual scheme of biological causality. Baker[2] is usually credited for having introduced the proximate/ultimate formulation in biology, but he does not make any attempt to define it, he only mentions the terms in passing[3]. The first occurrences of the dual scheme in biology predate Baker's paper by more than twenty years. Huxley[4] explicitly opposes 'immediate' to 'ultimate' causes as early as 1916. Ultimate causation also appears from the sceptical pen of D'Arcy Thompson[5] in his book *On Growth and Form*, initially published in 1917. In an enlightening introduction about the use of the concepts of efficient and final causes in science and philosophy, he states that the search for final causes had for long overshadowed the study of mechanisms. He observes the 'insuperable difficulty of giving any just and tenable account of the relation of cause and effect from the empirical point of view', emphasising 'the difficulties which surround the concept of ultimate or "real" causation' (p. 8). D'Arcy Thompson also writes: 'mechanism and teleology are interwoven together, and we must not cleave to the one nor despise the other; for their union is rooted in the very nature of totality' (p. 7).

The search for ultimate causes proved to be heuristic in the second half of the last century. By pointing at the adaptive function of biological characters, the finalist perspective allowed the formulation of hypotheses about their evolutionary origins. Today there is general agreement that reconciling our views about mechanisms and functions is a necessary step to deepen the theory of evolution. The reservations of D'Arcy Thompson[5] remain topical in this respect. In what follows, I will examine the difficulties raised by the dual scheme regarding the time scales of evolution and the role of epigenetic processes.

The different time scales of the evolutionary process

Modern science was founded on a banning of occult qualities. The final causes of Aristotle were rejected as explanations of the physical world and replaced by the action of mechanical causes, that is, a mechanism was required to account for any effect. Living beings and their purposive actions were left apart for a while. As stated by Mayr[1], 'The Aristotelians and their successors asked themselves what goal-directed process could have produced such a well-ordered design in nature' (p. 1504). The problem was eventually solved by the mechanism of natural selection. If a character confers some advantage to its bearers, the latter will survive and reproduce at higher rates than others, increasing the representation of the character in the population. The reintroduction of finality in science by the way of natural selection has represented a main achievement of the Darwinian thought. Mayr[1] advocates this teleological stance. The origins of a biological character lie in its adaptive function, hence we

e-mail: thierry@neurochem.u-stras.fr

must find the ultimate causes of a character within the ecological environment where a given population evolved.

But are ultimate causes really causal and really ultimate? As stressed by previous authors, the concept of ultimate cause is questionable in its very label[5–8]. When we study the function of a character, we focus on its effects rather than on its causes. But the cause must precede the effect. To ascribe a causal role to the effects, we commonly assume that the effects observed today are similar to those having produced the character in the past, hence they become antecedent and by the way causal. The Darwinian thought is based on such an extrapolation. At this point comes the qualifying 'ultimate'. The opposite of 'proximal' should be 'distal', not 'ultimate'. Whereas 'distal' is a relative concept, 'ultimate' is an absolute one, there is nothing after the end. An absolute concept is necessary to mean the reversal of the causal arrow between present and past causation. It allows us to oppose two symmetric realms; one belongs to the ecological time of proximate causes, the other to the evolutionary time of ultimate causes. Both involve the action of the same factors. The issue is whether the ultimate realm is unidimensional. We cannot take for granted that selection processes may be extrapolated at all time scales from present-day events[9].

The Darwinian thought envisions evolution as a cumulative process that results from the continuous action of natural selection. The adaptive modification of lineages is produced by the recurrent occurrence of the same ultimate causes. At the level of macroevolution, however, when events become too rare to belong to the 'experience' of species, such microevolutionary processes do not necessarily hold. In periods of mass extinctions, for instance, the selective process may become more destructive than productive[10]. In Cretaceous molluscs, the tendency for lineages to increase in body size during background geological time – Cope's classical rule – is not found in the periods of mass extinction[11]. During such events, success is more dependent on the characters allowing survivorship in changing environment than on the characters promoting fitness in normal time. A lineage may possess by chance the characters needed to survive an event it never encountered before. Thus, the macroevolutionary mode can disrupt the microevolutionary mode[12].

Species selection occurs when the selective event favours the survival of one or the other phyletic lineage. If species selection occurs only during great extinctions, it would remain unimportant relative to natural selection[13]. But minor extinctions regularly occur during background geological time. This paves the way for the regular occurrence of species selection, which may then play a major role in evolution. Species selection can 'reverse, dilute, or undo' the adaptations accumulated by natural selection[9]. A character advantageous at a given time frame may become deleterious at another time frame. For example, some lineages have evolved toward asexuality because it is less costly than sexual reproduction[13]. Asexual reproduction is a successful strategy in stable conditions, but the absence of genetic recombination may favour the accumulation of disadvantageous mutations, it impairs the ability of asexual lineages to quickly adapt to changing environmental conditions. Less than 1% of extant species are asexual; they could be those not having yet encountered an event responsible for their disappearance[14].

The organisms we observe today are those whose ancestors went through a succession of events that filtered them at different spacing. The selection induced by an event is likely as intense as the probability of occurrence of the event is low[15]. But the character becomes advantageous only if the selective event having promoted it comes up again. A unique event does not produce adaptation on its own. The same may be said for an event so infrequent that the character conducive to survival is lost through evolution before the next event occurs[15]. There is no single time scale to measure adaptation. Such a perspective is at odds with the uniformitarian view of evolution in which the concept of ultimate cause is rooted. The reversal of the causal arrow between present and past causation presupposes the regular repetition of selective events. Should an event too rare to be 'anticipated' by living beings be counted as an ultimate cause? Thinking in terms of ultimate causation makes it difficult to conceive such mechanisms as random sorting or else selective processes working in opposition to each other at different levels.

The epigenetic dimension of the organism

Proximate causes are those occurring during the lifetime of the individual, they involve the interactions among genotype, phenotype and environment. A basic tenet of the dual scheme is that we must provide both a proximate and an ultimate cause for every phenotypic character to be accounted for. This is a logical consequence of a functionalist line of reasoning, which assumes that natural selection can freely act on organisms. The phenotype is essentially viewed as a set of trade-offs, that is, a benefit realised through a change in one character is linked to a cost paid out through a change in another[16]. Hence the systematic search for ultimate causes. This atomistic stance is largely unrealistic, however, since it implies that characters are held *structurally* independent of each other. Quite to the contrary, organisms are integrated wholes where some characters arise as non-adaptive consequences and side effects of other characters. There are innumerable examples of biological characters that have no direct ultimate causes. In various mammals, for instance, some females may display signs of oestrus while pregnant. Such post-conception oestruses may be an accidental output of the delicate tuning of the steroid hormones that regulate gestation. Nevertheless, several hypotheses have been proposed to ascribe an adaptive function to these infrequent oestruses. They have been

alternatively interpreted as female tactics aiming to conceal ovulation, to attract males or to deceive them about their possible paternity in order to obtain their protection and decrease the risks of infanticide[17]. Gould and Lewontin have denounced this endless hunt for functionalist explanations in their classical 'Spandrels' paper[18].

A second assumption of the dual scheme is that proximate and ultimate causes are independent since they belong to different time scales. This again lies on a neglect of the integrated nature of organisms and their development. Mayr[1] merely states that the organism arises from 'the decoding of the programmed information contained in the DNA code of the fertilized zygote' (p. 1502). Epigenetic processes are absent from his perspective. As noted by Ho and Saunders[19], 'Although it is often said that the genotype interacts with the environment during development, this statement reflects a certain degree of sloppy thinking. It is rather the epigenetic system which interacts with the environment and ultimately generates those variations on which selection can act. The epigenetic system belongs, strictly speaking, to the phenotype rather than the genotype.' (p. 579). The field of evolutionary developmental biology is actively filling the gap between evolution and development. The study of the genetic bases of major developmental patterns has provided for a great stability of body plans in distant taxa. Cascades of gene action and developmental controls fix the structure of organisms and canalise the evolutionary pathways open to them. Further constraints act at other levels of the phenotype. The space of possibilities open to organisms is limited, some adaptive solutions may be out of reach because of material reasons or the need for internal consistency. The complexity of the organism itself makes changes difficult to achieve because the options chosen during phylogenetic history have become entrenched into the development and structure of the phenotype[20–22].

Constraints may be defined as limits on the variability of the phenotype, caused by the structure and dynamics of the epigenetic system[23]. If evolutionary changes are directed both by external selection and internal constraints this is not a minor issue, 'it does mean that the variation presented to selection is not random'[21]. Thinking in terms of proximate and ultimate causation hardly grasps this reality. Characters are more or less nested in the structure of the organism. Tightly bound characters may induce phylogenetic inertia and stasis at the level of phyletic lineages. The speed and magnitude of evolutionary changes depend on a balance between the degree of entrenchment of characters and the intensity and frequency of selective events. The more deeply embedded characters may be those that persist through long evolutionary time scales in spite of major selective events, which might be a conservative strategy[15].

The study of the social behaviour of macaques provides an instance of different degrees of character entrenchment. The societies of the different macaque species represent variations of the same basic social organisation, they belong to a single family of forms[24]. A phylogenetic analysis showed that the core of the social organisation and the levels of dominance asymmetry between group members could remain constant during several million years. By contrast, the seasonality of reproduction was not correlated with phylogeny[25]. Macaque species have experienced a switch between temperate and tropical climates more than once during recent geological time, each climate being related to a different mode of reproduction, seasonal or aseasonal. The point is that the reproductive output of males results from an interaction between seasonality and dominance asymmetry. If matings occur year-round, there is usually no more than one receptive female at a time in a group and the top-ranking male has priority of access to her. When breeding is seasonal, several males may mate with receptive females, since no single male is able to monopolise all the females. This entails the following outcome: in non-seasonal species with limited asymmetry between individuals, social hierarchy has more influence on the reproductive success of males than in species in which asymmetry is marked but where females' fertility is synchronous. At the evolutionary scale, shifting between seasonal and aseasonal reproduction may represent an easy modular change regarding physiological constraints, whereas modifying the level of dominance asymmetry – and by the way the whole system of social relationships and the individual characters on which they rest – may be difficult to achieve. We would not be able to understand the differential reproductive success of males without considering both constraints and selection.

Conclusion

As previous critics stated, the proximate/ultimate scheme is not all-inclusive. It carries along a pan-selectionist research program, which is blind to processes like genetic drift or environmental inheritance, and to any process other than natural selection[3,8,26]. I have further argued that a uniformitarian perspective cannot account for the different scales of the evolutionary time and for the epigenetic constraints that direct the changes open to living beings. Evolution is not only the selection of random variation, it also includes random sorting and directed variation.

In fact a number of biologists, including the evolutionists, rarely use the concepts of ultimate and proximate causes. They do not need them. By contrast these concepts are of common use among those studying behaviour and ecology, who are primarily concerned by the action of external factors. The statement 'We can progress towards understanding the evolution of adaptations without understanding how relevant structures develop'[27] reflected mainstream thinking[28,29] for a while. Today many of us are keen on integrating mechanisms, development, adaptation

and historical contingency in a unified theory of evolution. To this aim, however, trying to reconcile proximate and ultimate causation is counter-productive, for this very formulation perpetuates their separation. For those who really want to integrate mechanistic and adaptive functions, please, give it just one more go: drop the dichotomy!

1. Mayr, E., Cause and effect in biology. *Science*, 1961, **134**, 1501–1506.
2. Baker, J. R., The evolution of breeding seasons. In *Evolution: Essays on Aspects of Functional Biology* (ed. de Beer, G. R.). Clarendon, Oxford, 1938, pp. 161–177.
3. Dewsbury, D. A., On the problems studied in ethology, comparative psychology, and animal behaviour. *Ethology*, 1992, **92**, 89–107.
4. Huxley, J. S., Bird-watching and biological science. Some observations on the study of courtship in birds. *Auk*, 1916, **33**, 142–161.
5. D'Arcy Thompson, W., *On Growth and Form,* Cambridge University Press, Cambridge, 1917.
6. Francis, R. C., Causes, proximate and ultimate. *Biol. Phil.*, 1990, **5**, 401–415.
7. Armstrong, D. P., Levels of cause and effect as organizing principles for research in animal behaviour. *Can. J. Zool.*, 1991, **69**, 823–829.
8. Ariew, A., Ernst Mayr's 'ultimate/proximate' distinction reconsidered and reconstructed. *Biol. Phil.*, 2003, **18**, 553–565.
9. Gould, S. J., The paradox of the first tier: an agenda for paleobiology. *Paleobiology*, 1985, **11**, 2–12.
10. Raup, D. M., *Extinction: Bad Genes or Bad Luck.* Norton, New York, 1991.
11. Jablonski, D., Body-size evolution in Cretaceous molluscs and the status of Cope's rule. *Nature*, 1997, **385**, 250–252.
12. Jablonski, D., Background and mass extinctions: the alternation of macroevolutionary regimes. *Science*, 1986, **231**, 129–132.
13. Ridley, M., *Evolution*, Blackwell, Boston, 1996, 2nd edn.
14. Gouyon, P. H. and Gliddon, C. J., The units of selection. *Trends Ecol. Evol.*, 1989, **4**, 204–208.
15. Simons, A. M., The continuity of microevolution and macroevolution. *J. Evol. Biol.*, 2002, **15**, 688–701.
16. Stearns, S. C., *The Evolution of Life Histories*, Oxford University Press, Oxford, 1992.
17. Thierry, B., Adaptation and self-organization in primate societies. *Diogenes*, 1997, **180**, 39–71.
18. Gould, S. J. and Lewontin, R., The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proc. R. Soc. London*, 1979, **B205**, 581–598.
19. Ho, M. W. and Saunders, P. T., Beyond neo–Darwinism – An epigenetic approach to evolution. *J. Theor. Biol.*, 1979, **78**, 573–591.
20. Kauffman, S. A., *The Origins of Order*, Oxford University Press, Oxford, 1993.
21. Raff, R. A., *The Shape of Life*, University of Chicago Press, Chicago, 1996.
22. Gould, S. J., *The Structure of Evolutionary Theory*, Belknap Press, Cambridge, MA, 2002.
23. Maynard Smith, J. *et al.*, Developmental constraints and evolution. *Q. Rev. Biol.*, 1985, **6**, 265–287.
24. Thierry, B., Social epigenesis. In *Macaque Societies: A Model for the Study of Social Organization* (eds Thierry, B., Singh, M. and Kaumanns, W.), Cambridge University Press, Cambridge, 2004, pp. 267–294.
25. Thierry, B., Iwaniuk, A. N. and Pellis, S. M., The influence of phylogeny on the social behaviour of macaques (Primates: *Cercopithecidae*, genus *Macaca*). *Ethology*, 2000, 106, 713–728.
26. Jamieson, I. G., The functional approach to behavior: is it useful? *Am. Nat.*, 1986, **127,** 195–208.
27. Maynard Smith, J., *Evolution and the Theory of Games*, Cambridge University Press, Cambridge, 1982.
28. Wilson, E. O., *Sociobiology: The New Synthesis*, Belknap Press, Cambridge, MA, 1975.
29. Krebs, J. R. and Davies, N. B., *Behavioural Ecology*, Blackwell, Oxford, 1978.