

Traits and functions in the evolution of morality

Emanuele Serrelli¹

Pre-submission for the Proceedings of “The Evolution of Morality” workshop

School of Etology, Erice, Italy, June 2012

Edited by Frans de Waal, Stefano Parmigiani, Telmo Pievani

Abstract. This paper is about *evolutionary explanations*. They come in different kinds but mostly need traits and functions. Evolutionary theory requires traits to be inheritable although not in a strong genetic sense: ideas of “inheritance pattern” and “inheritable pattern” are explored. Function is also a necessary concept, but complex and diverse, and it lacks causal power on traits. The debate on the evolution of morality is cautious and already far from naive “just-so story” explanations, but theoretical analysis fleshed into morality-related examples can aid towards the development of critically conscious and up-to-date explanatory hypotheses in this field.

1. Just to start, a basic (and discredited) kind of evolutionary explanation

Evolutionary explanations come in different kinds. One kind Gould and Lewontin (1979) labeled Panglossian paradigm,² has been long criticized and it is not likely

¹ “Riccardo Massa” Department of Human Sciences, University of Milano Bicocca, U6 building, 4th floor, Piazza dell’Ateneo Nuovo 1, 20126 Milano, Italy.

to be scientifically acceptable nowadays. It relies on the identification of *a trait* on the one hand, and of *its function* on the other hand, and it consists in explaining the former with the latter by means of a “just-so story”.³ A trivializing example: why do birds have wings the way they have? Because wings are good to fly, and flying is manifestly good!

So, why do we humans have morality?... Decades of conceptual critique warn us against answering by the explanatory scheme above. In fact, as we can see in this issue, the state-of-the-art of the evolution of morality field is well beyond this point. Yet, I use this to problematize the idea of a trait and the idea of a function. I first concentrate on traits, following a taunt Francis Crick once made to the young Stephen Jay Gould by asking something like: “Why do you evolutionists always try to assess the value of anything even before knowing what it is?” (cf. Gould 1991, p. 138).

2. What is morality? Some evolutionary criteria for trait recognition

We find different definitions of morality embedded or explicated in the various papers in this issue. Richard Joyce raises the problem of the *precisification* of moral sense; Philip Kitcher too cites the terminological issue “ethics vs. morality”, labeling it “just a technical philosophical problem”, which I’m sure he knows it’s not. My sense is that in studying morality either we have scarcity of precisifications, or we tend to hold back from choosing among the existent

² Referring to dr. Pangloss, a character in Voltaire’s *Candide* (17...) engaged in explaining any feature of the world by means of its “usefulness”. Noses are there because they have to support glasses, and earthquakes have a hidden utility for the sake of the world.

³ In reference to Rudyard Kipling.

precisifications, or from declaring such choice, or from designing new ones. This situation might be due to the fact that we are still trying to get a grip, or it might be an explanatory strategy itself. Richard Joyce talks about “indeterminate concepts”, and Patricia Churchland cites Eleanor Rosch to argue for the “radial structure and fuzzy boundaries of concepts”. Yes, but we all know that this doesn’t wash away the scientific need for concepts, and furthermore for operationalizations. If we don’t feel like crystallizing our ideas about morality, the pragmatics of contextualizing results within certain premises will allow us to work.

Evolutionary theory gives us some criteria to check when we identify a trait. *Heritability* is among the most important. There are also *interindividual variation* (as I shall say, among the appropriate units), closely tied to heritability, and the possibility of building a *fitness function* on trait variation. I will briefly touch on these two when I talk about function. Notice that the link between the traits or structures and functions, however complex, cannot be broken in evolution.

The evolutionary relevance of a trait depends on its inheritance pattern. This is a very loose claim: there have been periods in which evolutionary biology was very strict on this criterion, as both a cause and a consequence of focusing on *genes*. Genetic factors are totally inheritable, and with Neodarwinism their patterns of inheritance were accurately modeled by the mathematics of Mendelian population genetics. The Modern Synthesis then emphasized that genes *are all* which is inheritable, thus binding evolutionary biology to them. Would we be working on the evolution of morality within that hardened framework, we would be forced to ask first: what genes do give morality? Although sometimes the “genes for” or “genes of” morality still show up on newspapers and tv programs,

professional scientists, also here, are manifestly working outside this framework. Churchland, for example, insists in affirming that the idea of “genes for a behavior” is meaningless. But, if morality is not a set of inherited genes, the question about heritability remains crucial in order to define what morality is in evolution.

Of course, genes do not necessarily lose their importance. On the empirical side, Churchland argues that *oxytocin* and other molecules are essential components of a “neurobiological platform” for morality. And we know that these molecules are synthesized from inherited genes. On the theoretical side, models of *altruism* evolution (e.g., kin selection cited by Norenzayan, or group selection by Wilson & Sober cited by Schloss) are based on population genetics, and they are very important, but they make sense only by assuming a very reliable, Mendelian-like heritability of the behavioral trait; they can therefore offer “how possibly” constraints on how we can treat genes involved in altruistic behaviors.

But heritability has a *larger* scope today, thanks to revisions and extensions of Neodarwinism. Evo-devo and epigenetics studies, for example, are forcing a shift towards talking about inherited processes, networks and developmental pathways. This is a new frame for evolution: the heritability pattern of a dynamic phenotype is a true locus for evolution (e.g. Müller 2010), not captured by the level of genes, and being inherited through a plurality of “channels” (Jablonka & Lamb 2005; Jablonka xxx).⁴ Authors like Churchland (2011) need to talk about phenotypic traits that are not detailed at the molecular level,⁵ e.g. *psychopathy*, and

⁴ Jablonka and Lamb (2005), for example, argue for epigenetic, behavioral, and cultural inheritance channels to interact with the genetic channel (cf. Darcia Narvaez on epigenetics). Odling Smee et al. (2003) contemplate ecological inheritance.

⁵ “The mechanism is unknown”, as Churchland contested to some speakers.

they know that measuring heritability is crucial although difficult.⁶ Now, moving from some identifiable gene or molecule to a behavioral pattern, or even to a disposition to behave, we may have a sensation of **fading**: how can we study heritability and homology of this stuff? But the point is right here: *we are studying evolutionarily relevant traits only if we can assess their reliable reappearance*. Almost always, it will be a statistical measure, but there is actually a solid tradition in population genetics that – as strange as it might seem – doesn't deal analytically with genes (Serrelli 2011, Downes 2010); and we often forget that *fitness too* is nothing but a probabilistic and difficult-to-measure dimension in biology; moreover, current studies on Quantitative Trait Loci are linking it with Mendelian genetics (Hartl & Clark 2007, chp. 8). All this gives us much flexibility in the kinds of traits we can consider in evolution. That is why I propose the most generic definition of a trait as an inheritable pattern, a pattern that is reliably produced generation after generation.

Three short, final remarks on this view of traits.

Firstly, often pattern perception and recording is not the whole story: we work with *models* that integrate patterns. In his writings, de Waal is careful in showing the diversity of models that can explain the observed behaviors.⁷ All psychologists addressing morality, for example, have *models of* morality: Darcia

⁶ "Twin and family studies suggest a heritability of psychopathy in the range of about 70%; childhood conditions such as abuse and neglect may contribute to those who are genetically disposed" (p. 41)

⁷ Van der Weele (2011) emphasizes the cultural and linguistic influences on the conceptualization of morality. Referring to Eisenberg & Strayer (1987) he points out a loss of diversity in the sources: empathy didn't even exist in the English language in Darwin's time. There can be interesting effects of "loss of diversity in the search for the sources of morality".

Narvaez relies on her “triune ethics model”, whereas Melanie Killen holds the “domain theory”, a narrower one as she said. But also in Churchland oxytocin plays *in a model* of “attachment and bonding” including patterns of neural activation, and this in turn is enclosed in a model of morality.

Secondly, scaling up inheritable patterns at an evolutionary time scale is a really delicate passage. However, it is necessary, by definition, if what we are looking for is a trait that is able to enter the evolutionary arena.⁸ At the scale of phylogeny, the issue of heritability becomes the issue of homology: the trait shows up in different taxa. Discerning homologies is a challenge, and requires a careful look at development and phylogeny (Pievani & Serrelli 2011) to get around circularity: homology tends to depend on the model we have of the considered traits (Griffiths & Brigandt 2007). Cooperation in mammals is *not at all* homologous to cooperation in social insects – at least, this is what I get from Churchland’s analysis but also from Pier Francesco Ferrari’s presentation. This means that we should cease from talking about “the evolution of cooperation” as if it were a single phenomenon. We saw yesterday that a homology between mammals *and birds* is more likely, if not apparent.

Lastly, patterns can be looked for in units different than the organism.⁹ When authors like Cristopher Boehm or Philip Kitcher or Ara Norenzayan identify sets of cultural ideas or socially decided norms that influence individuals, the relevant unit is the group or the society. In Philip Kitcher or Sober & Wilson (1998) or perhaps in Ara Norenzayan there is also the idea of a measure of fitness varying

⁸ Thanks to Darcia Narvaez for her useful question about the sufficiency Lewontin’s criteria of heritability.

⁹ And corresponding time scales?

in correlation with patterns across groups. There is actually a very important difference whether we consider the domain of cultural evolution as part of biological evolution or as a social artifact, a tool, a device. But deepening this would be too long. Suffice to say that culture as an artifact allows co-evolution, like when Narvaez talks about undercare in our cultures affecting our phenotypes, or when Boehm talks about genetic consequences of certain social organizations.

3. What's the *function* of morality?

As George Williams famously reaffirmed (1966), any trait has many positive effects, but in Neodarwinism not all effects are functions. Proper functions are only those few effects of trait T that have been “seen” by natural selection, thus being causally relevant to the spread of T .¹⁰ In the following equation (Wright 1931):

$$[1] \Delta q_T = [sq_T(1-q_T)]/[1-s(1-q_T)]$$

q_T is the frequency of trait T in a population, coefficient s is the selection pressure on T , and Δq_T is the change in the frequency of T . Positive selection on T corresponds to q_T increasing. But this depends on a lot of other factors, like population size, set of available alternatives to T , heritability pattern (see above), mutation pressure. Moreover, equation [1] says nothing about the reason why T is retained and spread, “in function of what”, T is there. Paraphrasing Kitcher's book (2011), finding a function for T requires *us* to specify a “problem background” in respect to which T is successful.¹¹ The “problem background” is not

¹⁰ See also Sober (1984) on traits that are “selected for” as opposed to simply “selected”.

¹¹ In analogy with technological progress, Kitcher says that “Attribution of functions is straightforward where there are clear-sighted potential users who can express their desires and identify their problems. On other occasions, we can talk of functions even though cognitive beings

there in [1], it is subsumed in parameter s . That is why scholars try and build evolutionary scenarios. Evolutionary theory gives us at least two ways of describing the “problem background”.

In the first way, trait T appears in a fitness function:

$$[2] w = f(A, B, \dots T)$$

Fitness is function of traits in a particular environment.¹² Wing size and structure varies throughout any population of birds, correlating to fitness of individuals bearing them. A description [2] requires a T to vary throughout the population of relevant units (e.g., individuals, or groups) – remember that interindividual variation was also a criterion for identifying an evolutionary trait; it requires as well a measure of fitness; and it implies an exploration of the covariance of the two, allowing then an understanding of the populational spread of T [1].¹³

The second way can be formalized as:

$$[3] T = f(F, r)$$

Trait T is function¹⁴ of F (the overall functioning of the organism bearing T), and r (the role of T in such functioning). Birds fly, and wings have a role in such a functioning, a role that correlates and explains the existence and shape of wings.

with desires and needs are not involved. Biologists and physicians routinely discuss the functions of organs, bodily systems, cells, and molecules” (p. 219). But that transposition is not straightforward.

¹² This is closer to the intuitive idea that T has a function.

¹³ Notice that equation [2] portrays the situation at a certain moment: different fitness values are obtained by checking out individuals with varying versions of T , not by transforming T through time.

¹⁴ Talking about function playing and imitating a mathematical language, where “is function of” means “depends systematically on”, some curious word games emerge: a trait does not have a function, rather, it *is* a function of something else.

Role is an important parameter because, for example, feathers or tails have a roles too, and of course differing roles yield different traits. Perhaps this is the sense in which Frans de Waal yesterday says that “the function of sex is social regulation”, or when Pier Francesco Ferrari talks about the “functional properties of groups of neurons”. But if [3] is considered the explanation for *T* – de Waal and Ferrari, I guess, do not – then it is the Panglossian explanation, and it is fatally undermined by the ubiquitous observation that functioning and roles are not causal upon traits and their structure.

An example from Churchland’s book. She observes that “the palette of neurochemicals affecting neurons and muscles is substantially the same across vertebrates and invertebrates [...]” (p. 45). Oxytocin is a very ancient and simple peptide, found in all vertebrates. In mammals, “Oxytocin ... is at the hub of the intricate network of mammalian adaptations for caring for others” (p. 14). The striking thing – Churchland continues – is that “modest modifications in existing neural structures ... can lead to new outcomes”.¹⁵ Among mammals, then, There is a large “...range of social patterns [...], but underlying them are probably different arrangements of receptors for oxytocin and other hormones and neurochemicals” (p. 32). For Churchland, this is crucial to argue that behavioural and evaluative processes grounded in neurochemical mechanisms – she assumes that morality consists in social values¹⁶ – are not confined to *Homo sapiens*. For my goal, their

¹⁵ DISTRESS AT SOCIAL SEPARATION as a modification of a more ancient place preference? (Jaak Panksepp, textbook 1998).

¹⁶ “That nonhuman mammals have social values is obvious; they care for juveniles, and sometimes mates, kin, and affiliates; they cooperate, they may punish, and they reconcile after conflict” (p. 24). Churchland’s “Rule use” approach opposes rule-based accounts of morality. “...take a close, and

presence in the most different mammals implies *T*'s large insensitivity to *F*. What we normally call “the function of a trait” cannot cause the trait itself. The existence and structure of oxytocin *simply cannot* be due to its function in caring for offspring and in wider forms of sociability: “...the evolution of the mammalian brain adapted oxytocin to new jobs” (p. 14).

If equation [3] seems to be misled, perhaps another, dynamic expression might better reflect some current evolutionary approaches:

$$[4] \Delta T = f(\Delta F, \Delta r)$$

That is, in evolution, a trait *may change* – oxytocin, for example, does not – if there's a change in the overall functioning of organisms (including ecological conditions) and/or in the role the trait plays therein. In sum, in Churchland's words, “Biological evolution does not achieve adaptations by designing a whole

skeptical, look at the commonplace look that rules, and their conscious, rational application, are definitive of morality” (p. 166). Stresses the EVALUATION aspect: “...my aim is to explain ...how humans are able to evaluate a law as a bad law or good law or a fair law, /without/ appealing to a yet deeper law - something that they actually do, and do regularly” (p. 166). “Moral theories that leave room for exceptions to rules have tended to seem incomplete” (p. 168), so many moral philosophers (Aristotle, Kant, Mill & Bentham etc) search for exceptionless rules. On the contrary, “The ability to appreciate when a circumstance is a fair exception, or which rule to follow when rules conflict, embodies some of the most refined aspects of social understanding. Going through life, we all acquire a lot of subtle, and often inarticulate, knowledge through our experience – stories, examples, and observation” (p. 167). In this light, Churchland criticizes moral psychology experiments: “A strategy for probing how people make moral decisions consists in contriving moral dilemmas designed to pit *kill-one-save-many* against *kill-no-one-and-let-many-die*. Subjects read scenarios and then rate the moral propriety of the alternatives. Not surprisingly, response vary... The detail-stripping is intended to eliminate confounds, but it introduces a new flaw: the scenario is so artificial that the brain's normal reliance on morally relevant facts to guide problem-solving is undermined ... everyone uses case-based reasoning, but given our individual histories and temperaments, we may draw on different cases to guide judgement concerning the present case” (p. 183). Churchland sees rationality too as not restricted to humans (Ibidem).

new mechanism from scratch, but modifies what is already in place, little bit by little bit. Social emotions, values, and behavior are not the result of a wholly new engineering plan, but rather an adaptation of existing arrangements and mechanisms that are intimately linked with the self-preserving circuitry for fighting, freezing, and flight, on the one hand, and for rest and digest, on the other” (p. 46).

4. Conclusion

To summarize, there are several *kinds* of evolutionary explanations. The scientific acceptability of each kind varies throughout decades as the community of evolutionists matures. New kinds are invented, and overall there is a need for diversity and integration of explanations.

In most evolutionary explanations, structures and functions are required. However, they aren't going to come for free. Choices have to be made, ambiguities can be tolerated, but I think that at least the aspects I presented (and there would be many more) have to be taken into careful consideration in order to build meaningful hypotheses on the evolutionary origin of morality.

We have to deal with inheritable patterns of morality and their variation. We need models, we need to overcome circularity, and we must specify the relevant units of evolution. It is not impossible that our heritability analyses end up by de-composing the trait we were interested in, perhaps even rendering it meaningless. Despite our interest in morality, its evolution can turn out to be the evolution of something else, unless we are open to re-define it. There are also possible overlaps in what we call trait or structure and what we call function –

what's the function of oxytocin? Is it relative to a neurochemical mechanism? But isn't this mechanism an inheritable trait itself? These overlaps are due to the polisemy and complexity of the involved terms.

As for functions, they are full of conceptual problems that are object of a rich debate in philosophy of science (e.g. Cummins 1975, Wright 1973, Millikan 1984, 1989, Griffiths 1993, Godfrey-Smith 1994, cf. Casebeer 2003).¹⁷ On the one hand, it is clear that functions can't cause traits. On the other hand, functions are often the sights through which we look for traits. Moreover, function is a crucial element of Darwinian explanations. Functions and effects have to be carefully included in the explanation of morality,¹⁸ also taking into account that non-selected effects and creative re-use are part of the deep nature of evolution. Richard Joyce cited spandrels and exaptations, and Frans de Waal responded defending an argument like "no need of special explanations for pervasive traits".

¹⁷ A trait plays a role in a system, but such role is rarely the explanation of its structure. Calling this role "a function" would be ok for the representative author Cummins (1975): "Relative to the "doorstop system", the hammer *has the function* of holding the door open. However, we would not find this a satisfying explanation of for why the hammer came to have the structure it did" (cit. in Casebeer 2003, p. 51). Calling it a function is not satisfactory for Neodarwinism (Williams), nor for authors representative of the etiological meaning of function (Wright 1973, Millikan 1984, 1989). It can explain some modifications it recently had. According to Godfrey-Smith's (1994) "modern history view" of functions, the temporal reach of this explanation is limited: "Episodes of selection become increasingly irrelevant to an assignment of functions at some time, further away we get. The modern history view does, we must recognize, involves substantial biological commitments. Perhaps traits are, as a matter of biological fact, retained largely through various kinds of inertia [...] There is no avoiding risks of this sort" (Ivi). About the evolution of morality, Casebeer (2003) writes: "...we don't want to gravitate to either extreme. If we gravitate to the distal, super-historically laden conception of function, then the only content we can squeeze out of function is that the ultimate function is to reproduce [...]. On the other hand, if we move instead to the proximate, "instantaneous" analysis of a Cummins function, all historical content is lost" (p. 52).

¹⁸ Is morality a function or an effect? Functionalist view makes us blind to "pure effects".

They are both right, because of how evolution works, and because of our growing understanding of that.

References

- Boehm C (2000), Conflict and the evolution of social control. *Journal of Consciousness Studies* 7(1-2), "Evolutionary Origins of Morality. Cross-Disciplinary Perspectives", pp. 79-101.
- Churchland PS (2011). *Braintrust: What Neuroscience Tells Us About Morality*. Princeton: Princeton University Press.
- Cummins R (1975). Functional analysis. *Journal of Philosophy* 72:741-765.
- Downes SM (2010). Heritability. *The Stanford Encyclopedia of Philosophy (Fall 2010 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2010/entries/heredity/>.
- Eisenberg N, Strayer J (1987) Empathy and its development. Cambridge UP, Cambridge.
- Flack JC, de Waal FBM (2000), 'Any animal whatever': Darwinian building blocks of morality in monkeys and apes. *Journal of Consciousness Studies* 7(1-2), "Evolutionary Origins of Morality. Cross-Disciplinary Perspectives", pp. 1-29.
- Godfrey-Smith P (1994). A modern history theory of functions. Reprinted in: Allen C, Bekoff M, Lauder G, eds. *Nature's Purposes - Analyses of Function and Design in Biology*. Cambridge, MA: MIT Press, 1998, pp. 453-477.
- Gould SJ (1991), "Capezzoli maschili e glande clitorideo", in trad. it. *Bravo Brontosauo*, Milano, Feltrinelli Universale Economica, 2002, pp. 125-139.

Gould SJ, Lewontin RC (1979), The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme, *Proc. R. Soc. Lond. B*, 205, 581-598. Trad. it. di M. Ferraguti, *I pennacchi di San Marco e il paradigma di Pangloss: critica al programma adattazionista*. Piccola biblioteca *on line*, Einaudi, Torino, 2001.

<http://www.einaudi.it/einaudi/ita/pdf/gould-lewontin.pdf>

Griffiths PE (1993). Functional analysis and proper functions. *British Journal of the Philosophy of Science* 44:409-422.

Griffiths PE, Brigandt I, eds. (2007). "The Importance of Homology for Biology and Philosophy", *Biology and Philosophy* 22.

Hartl DL, Clark AG (2007). *Principles of Population Genetics*, Fourth ed. Sunderland, Mass.: Sinauer Associates.

Kitcher P (2011). *The Ethical Project*, Harvard University Press, Cambridge, Mass. & London, UK.

Jablonka E, Lamb MJ (2005). *Evolution in Four Dimensions*. Cambridge (MA): The MIT Press, revised in 2010.

Millikan R (1984). *Language, Thought and Other Biological Categories*. Cambridge, Mass: MIT Press.

Millikan R (1989). In defence of proper functions. *Philosophy of Science* 56:288-302.

Müller G.B., 2010. Epigenetic innovation. In Pigliucci M, Müller GB, eds. *Evolution The Extended Synthesis*. Cambridge-London: MIT Press, pp. 307-332.

Odling-Smee J, Laland K, Feldman MW (2003). *Niche Construction*. Princeton: Princeton University Press.

- McPartland JM, Matias I, Di Marzo V, Glass M (2006). Evolutionary origins of the endocannabinoid system, *Gene* 370:64-74, 10.1016/j.gene.2005.11.004.
- Pievani T, Serrelli E (2011). Exaptation in human evolution: how to test adaptive vs exaptive evolutionary hypotheses. *Journal of Anthropological Sciences* 89, pp. 9-23. [DOI 10.4436/jass.89015]
- Serrelli E (2011). The structure of population genetics. Poster at the meeting of the International Society for History, Philosophy, and Social Studies of Biology (ISHPSSB), Salt Lake City (Utah, USA) July 10th-16th 2011. [<http://hdl.handle.net/10281/28245>]
- Sober E (1984). *The Nature of Selection. Evolutionary Theory in Philosophical Focus*. University of Chicago Press, Chicago.
- Sober E, Wilson DS (1998). *Unto Others: the Evolution and Psychology of Unselfish Behavior*. Harvard University Press, Cambridge (MA).
- Williams GC (1966). *Adaptation and Natural Selection*. Princeton: Princeton University Press.
- Wright L (1973). Functions. *Philosophical Review* 82:139-168.
- Wright S (1931). Evolution in mendelian populations. *Genetics*, 16 (97): 97-159.