Hanging On To The Edges:
Is it explanation yet?

When I taught my PhD students life-history theory,
I entitled the lecture 'Life-history theory is hell'.
- Richard McElreath, 20th November 2015

One of the most devastating rejoinders you can give an academic is to characterise what they have offered you by way of an explanation as no more than a redescription of the phenomenon at hand. For example, let's say I am interested in the knotty problem of why people of lower socioeconomic position are less likely to successfully quit smoking compared to those of higher socioeconomic position[1]. I could offer you the insights of the theory of planned behaviour, one of the most popular theories in this kind of area[2]. The theory says (very roughly) that people do healthy things when they want to do them, they think they should do them, and they think they can do them. So perhaps the reason people of lower socioeconomic position are less likely to successfully quit smoking is either (a) they don't so much want to; (b) they don't so much think they should; and/or (c) they don't so much think they can. All of these are testable: I could go off and ask a load of people, and come back with some results. Let's say, hypothetically, I find that it's mainly (b). Hurrah, I say, I have now explained the social gradient in smoking cessation—in terms of a social gradient in the belief that it is normatively desirable to cease smoking.

Here's where your wounding rejoinder comes in. All you have done, you say, is to redescribe the social gradient of interest—poorer people are less likely to quit smoking—as a social gradient in the extent to which people believe they should quit smoking. But where does *that* social gradient come from? Perhaps people of lower socioeconomic position don't feel so much normative pressure to quit smoking because fewer of them successfully do so….but wait, isn't that where we started? As so often in social science, we have end up at a place where the thing being invoked to do the explaining (the *explanans*), does not seem entirely independent of the thing we wish to explain (the *explanandum*). And, more pressingly, you want to ask: where the hell did the *explanans* come from anyway? What explainans that? (Sorry.) To account for one pattern, I offered you another, but that other one seems immediately to cry out for a deeper explanation, an explanation that stands entirely free of the phenomena we are studying.

This is about the point where people like me, who advocate evolutionary, *aka* behavioural-ecological, explanations for patterns of human behaviour, pipe up. What we tend to say at this point is something along the lines of: what you other social scientists offer is some kind of *proximate* explanation for the phenomenon at hand: another phenomenon that stands immediately prior to the original one in the chain of causation. That's fine, but it only kicks the can one pace down the road. What we will need sooner or later is to show how the behaviour pattern in question arises from more general principles of surviving and reproducing in different kinds of environments: an *ultimate explanation*. For example, we might point out that people doing dangerous manual jobs or living in hostile environments tend to

---

[1] Kotz, D. & West, R. (2009) Explaining the social gradient in smoking cessation: It's not in the trying, but in the succeeding. *Tobacco Control* 18:43–46.
[2] Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50: 179-211.

die anyway, for other reasons, before the age at which smoking starts to really kill you. Thus, the payoff for foregoing the pleasures of smoking may be less for them than for people living in under other conditions[3]. And we assume that people respond, sooner or later, to payoffs in the currencies of survival and reproduction. This kind of explanation has a few things going for it: it's non-obvious; it uses information not contained in the original observations; and it connects to broader expectations about evolved Darwinian creatures, such as that they should suit their behavioural strategies to the ecological circumstances they experience.

This ultimate explaining is not a bad thing, but we do tend to be rather smug about it. Look at you lot, rearranging your proximate deckchairs on the deck, we seem to imply, whilst we alone are looking beyond the prow, to the causally primal iceberg looming out of the sea. This book is about honest self-assessment, though, and in this spirit, I have to make the following admission about us evolutionary folk. The ultimate explanations we proffer are sometimes not as ultimate as we make them out to be. They also suffer from a lot of the same vagueness and indeterminacy as the more proximate frameworks we like to claim we are going beyond. It's healthy to admit this. And it's also healthy to understand that it science, with the possible exception of theoretical physics of the most fundamental kind, it's always the case that your explanations will themselves require deeper explanations in their turn (yes, even for us evolutionary folk). One person's *explanans* always ends up being someone else's *explanandum*. It's a food web of indefinite size, stretching off in every direction.

§

This whole business of explanation is very much tied up with having something called a theory. The function of theories is sometimes said to be able to predict future cases, at least up to relative statistical likelihoods. There is some truth in this, but at least as important a role for theories is to shed light on *why* things happen. For example, say I passed a vast archive of historical social and economic variables, and the results of elections from the same years, through a machine-learning algorithm, to try to find regular relationships. Afterwards, I find I can predict election results in novel cases with 75% accuracy. Would I then have a *theory* of electoral outcome? Not, it seems to me, without a lot more work. I would have to show which variables the machine-learning had given most weight to, and then relate these to some kind of general conception of humans as decision-makers: what they like, what they don't like, when they stick, when they shift. Of course, the aggregate behaviour of the electorate might not read off from thinking about a single representative voter; different sectors of the electorate might respond have different experiences and might respond to them in different ways, and there could be complex social dynamics. My theory might need to take this into account. Nonetheless, to have a *theory*, I would need not just to gain predictive statistical power over election outcomes, but also to gain epistemic power, the ability to state why elections turn out as they do, using some generalizations about voters, their voting, and their interactions.

Theories are devices that come in diverse kinds. The kind of theory lay people imagine scientists having is what I will call the *Newtonian.* The properties of a Newtonian theory, in my sense, are as follows. Only minimal and general properties of the situation are needed as inputs for the theory to do its work. If you are going to fire a cannonball into the air, on earth, then I can tell you that if you fire it off on flat ground at 45° at an initial velocity of 100 metres per second, it will travel about 1,020 metres and reach a maximum height of 255 metres at the mid-point of its flight. It will do this because it will be decelerated in the vertical dimension of its motion at a known constant rate due to earth's gravity. It doesn't matter whether the ball is black or pink or bears the colours of West Allotment Celtic; whether you do it on a Wednesday, under Scandinavian egalitarianism, or in anger at being spurned by your lover.

The theory has no real wiggle room from person to person. If two scholars apply Newtonian mechanics to the same problem, they must both conclude with the same predictions. If they don't, at least one

[3] See *The mill that grinds young people old.*

of them has simply made a mistake. It should be relatively straightforward to look at the working and see where this has occurred. There is no: 'she's a Newtonian, but she brings a more modern sensibility and a command of the African evidence to the picture, and so she concludes the ball will fly 1,023 metres rather than 1,020'. The theory is a somewhat stable historical object. The Newtonian mechanics of today is just the same as the Newtonian mechanics of 100 years ago, and makes exactly the same predictions (and there is only possible prediction for a given set of inputs). We learn more about *the world* over time, for example that the theory doesn't do such a good job for things that are very small or moving very fast, but the theory itself is a fairly definite and stably identifiable entity.

Not all things that get called theories have the Newtonian properties. Take, for example, 'social practice theory'[4]. As I understand it, this theory presents an alternative both to rational-actor models, which see people as free decision-makers with inherent preferences that they seek to satisfy, and acculturation models, which say that people do what their culture or society conditions them to do. The basic premise of social practice theory is that people will do what seems best to them, but are not asocial, cultureless, pastless, bodiless, and omniscient decision-making demons as implied by some microeconomic models. Instead, what seems best to them is limited by the habits, rules, norms, and understandings that they have absorbed through their daily lives in their social environments, that they also play an unwitting role in perpetuating. They are agents, but agents situated within a particular local field of social practices, a field that cannot be stood outside, or reinvented from scratch.

I'll call theories like this 'recipe for a recipe' theories. Their properties are mirror-image of the Newtonian ones. Let's say we want to understand how people will respond to a social change, such as petrol being made 10% more expensive due to concerns about carbon emissions and climate change. Social practice theory does not make a simple general prediction, like 'car use will fall by 17%'. Instead, it says, we would need to know a lot of things about the context. How is car use embedded in people's daily practices; what social rules are there; what are normative pressures on them; what practical knowledge do they have of alternative modes of transportation; and so forth. The theory does not give us a prescriptive recipe for cooking up a prediction, as Newtonian mechanics did. Instead, it points us toward a flexible but not completely open-ended list of ingredients we ought to seek more information on in order to begin studying the problem, and hence making an appropriate recipe to then cook up a prediction (or more likely, retrodiction). The theory does not uniquely pre-specify what the relative proportions of these ingredients will be for the present case, nor how they will interact. It follows that two scholars can both employ social practice theory competently and without error, and yet come up with very different expectations, not just for two slightly different cases, but even for the very same case. There can be different emphases within the broad envelope of the theory, and the theory itself will drift over time, with different elements becoming more or less central.

The corollary of the comforting elasticity of 'recipe for a recipe' theories is that it is quite hard to say that they are wrong. If the cannonball in our previous example doesn't fly in a parabola and go 1020 metres plus or minus a few centimetres, then Newtonian mechanics gets chucked out as a theoretical framework for projectiles on earth. Discipline for social practice theory is more nebulous: almost any pattern of findings, *ex post*, can be parsed in a way that is compatible with the theory. The theory itself can update in the light of new evidence and new priorities; or it can fall from use in favour of some other recipe for making recipes that, like a new musical style, seems more interesting to the current generation. It might at best be shown to be more or less useful; there is almost no observation I can think of that would inflict it a critical blow.

[4] See Schatzki, T. (1996). *Social Practices* (Cambridge University Press, 2009); or Reckwitz, A. (2002). Toward a theory of social practices: A development in culturalist theorizing. *European Journal of Social Theory*, 5, 243–263. These two texts have rather different things in mind. This reinforces the point I am making, since I have seen them both cited as descriptions of what social practice theory, or the theory of social practices, contends.

There's a third type of theory I would like to mention, and that's the *inductive*. An example is the 'purse versus wallet' theory[5]. This theory says that increasing household income through giving it to mothers has greater positive effects on childrens' outcomes than via giving it to fathers, because of different ways the two genders spend their money. Unlike a 'recipe for a recipe' theory, it's pretty clear what this theory predicts for a given case within its domain (maybe not the size of the effect, but certainly its direction). Unlike Newtonian mechanics though, the main grounds for this theory seem to be largely, 'we have looked at some previous cases, and that's how it often worked out before', rather than any more general principles. (A quibble: Newton had undoubtedly looked at some previous projectiles and seen that they flew in parabolas prior to coming up with his theory; and perhaps you could found the 'wallet versus purse' theory on some more general first principles to do with the two sexes and evolutionary fitness. Nonetheless, the two cases do feel rather different.) So now we have defined three species of theory: Newtonian; recipe for a recipe; and inductive. It's time to ask: when we construct evolutionary theories of human behaviour, which species of theory are we constructing?

§

When we make the evolutionary gambit, we sincerely feel Newtonian. It feels as if by beginning our Introduction 'Evolutionary theory predicts….', we have connected our claims to the might of biology, and specifically to the considerable epistemic and formal power of Darwinian algorithms. But in reality, lots of 'evolutionary' theories about human behaviour are far from Newtonian, and don't have such solid explanatory or formal support as you might think. They are really just recipes for recipes, or patterns discovered by induction, and a good theoretical biologist would still want to ask 'yes, but under what conditions would *that* evolve?'. In this regard, we are not so different from any other kind of social scientist. I have a case study that illustrates this very clearly: the recent enthusiasm for 'life history theory' as an explanation for diverse human behavioural phenomena from risk-taking to obesity, schizophrenia to savings. If you want to see the kind of research I am talking about, just type 'life-history theory' or 'life-history strategies' into your literature search engine of choice and follow up the recent human-focused references[6].

When human behavioural scientists invoke 'life history theory' as an explanatory framework, there are a number of related things they might actually be up to. I will call three prominent ones 'enterprise 1', 'enterprise 2', and 'enterprise 3'. Not all work on 'life-history theory' falls into any of these enterprises. Moreover, I have no objection to any of them—indeed, have contributed to some—but the role being played by the term 'theory' within them does bear some examination.

Enterprise 1: often, what researchers invoking life-history theory are doing is asking whether the behaviour under study covaries with a number of other traits, particularly those to do with the timing of reproduction (for example, age at first menarche, age at first sexual intercourse, or age at first childbearing). The idea here is that human psychological and reproductive traits, rather than each varying independently, covary along a principal axis, the 'fast-slow continuum'[7]. At the fast end, we have early maturation and childbearing, along with which allegedly go a high rate of future discounting, proneness to violence and coercion, impatience, obesity, certain moral and social attitudes, and even certain psychiatric disorders. At the slow end, we have late childbearing, high parental investment, and all the opposite behavioural and psychological traits. Sometimes the mere

---

[5] Discussed by Cooper, K. and K. Stewart (2013). Does money affect childrens' outcomes? A systematic review. *The Joseph Rowntree Foundation* (www.jrf.org.uk).
[6] Many of the critical points I make here about the application 'life history theory' to human behavioural variation echo those made by Purzycki, B. G. et al. (2018). Material security, life history, and moralistic religions: A cross-cultural examination. *PloS ONE* 13: e0193856.
[7] For a review and critique of this enterprise, see: Copping, L. T., Campbell, A., & Muncer, S. (2014). Psychometrics and life history strategy: the structure and validity of the High K Strategy Scale. *Evolutionary Psychology* 12: 200–222.

demonstration that different traits covary along a principal axis is presented as if this were an explanation of these traits, and also a confirmation of the utility of life-history theory. There are, however, lots of reasons things might be correlated with one another, and demonstrating a correlation is very different from explaining it.

Enterprise 2: Sometimes scholars invoking life-history theory are doing more than just establishing covariation between traits. They are also trying to demonstrate that those traits relate to the ecology in which people live. In particular, they may be testing whether 'fast' behaviours are differentially likely to occur in places where life prospects are poor or uncertain[8]. This relates to an intuitively appealing argument that if the environment is harsh (for example, extrinsic mortality is high), then you need to get on with life and at least get some reproduction done while you can, whereas if the environment is benign you can take longer and invest more in temporally distant outcomes. Note that enterprise 2 is in principle independent of enterprise 1: it could be that traits covary along a principal axis, but for some other, completely different kind of reason than the one argued in enterprise 2.

Enterprise 3: If 'fast' behaviours occur particularly in 'harsh' environments, there are a number of ways this could come about, from the slow march of genetic selection at one extreme to rational real-time decision-making at the other extreme. Enterprise 3 concerns the particular claim that experiences in the first few years of childhood are particularly important in setting how 'fast' or 'slow' a person turns out later on[9]. The idea here is that people are born not knowing what their adult environment is like, but that things like the stability of the family, how their parents behave, and so forth, serve as cues that over evolutionary time have carried useful information about their adult worlds. Thus, natural selection has favoured mechanisms that effectively say 'if this crazy stuff is going on in childhood, I need to get ready for a world where I am going to need to be a fast adult'.

There's actually a lot of evidence compatible with the idea that childhood adversity affects reproductive behaviour and many other adult outcomes besides: with that I have no quibble. I just want to point out that enterprise 3 is not deducible from the other two enterprises. You could believe that there is a fast-slow continuum, but that it is not related to environmental harshness; that there is a fast-slow continuum related to environmental harshness but childhood experiences do not serve as cues to speed you up or slow you down; or that there is a fast-slow continuum and childhood experiences move you along it, but for reasons that have nothing to do with those experiences being evolutionarily valid cues to prevailing environmental harshness. So which of these various enterprises is the core claim of 'life-history theory'; and, more importantly, which of the various enterprises has its explanatory basis in evolutionary theory?

§

There is a body of evolutionary biological theory called 'life-history theory'. In fact, it is not any single theory, but a body of mathematical methods for making theories, theories about how natural selection would shape patterns of growth, reproduction and ageing under different ecological circumstances[10]. These methods have been applied to many different scenarios, and the general conclusion seems to be: all kinds of different things can evolve, depending on the details of the ecology and demography. And that seems to be borne out in nature: we see everything from salmon that go out in a single blaze of reproductive glory, to puffins that do a little bit of reproduction year after year for ages. There certainly are life-history models that show that if the rate of extrinsic mortality is high,

---

[8] See for example: Nettle, D. (2010). Dying young and living fast: Dying young and living fast: variation in life history across English neighborhoods. *Behavioral Ecology* 21: 387-95.

[9] See for example: Brumbach, B.H. et al. (2009). Effects of harsh and unpredictable environments in adolescence on development of life history strategies: A longitudinal test of an evolutionary model. *Human Nature* 20: 25-51.

[10] Stearns, S. C. (1992). *The Evolution of Life Histories* (New York: Oxford University Press) is a classic text, and possibly more widely cited than read.

one should expect early reproduction to evolve, even at the expense of growth or self-repair[11]. This prediction depends on a lot of things, though: small tweaks in assumptions about, for example, what limits population growth, and when in the life cycle mortality acts, can lead to the prediction that higher extrinsic mortality will delay reproduction, or that extrinsic have no effect[12].

Because of this, it is not really correct to ever say 'life-history theory predicts x…'. Really what one ought to say is 'this particular life-history model, using this particular set of assumptions, predicts x…..'. Then as well as testing prediction x, you would also want to establish that the assumptions were appropriate for the system you were working on. Now you might argue in the following way: in practice, we know that animals facing higher mortality regimes often evolve earlier reproduction. We know this not just from correlational evidence, but even from experimental evolution[13]. So the best class of theoretical models is probably the class that correctly recovers this phenomenon. I have some sympathy with this argument, but note that theory and evidence have changed places. Rather than life-history theory predicting *a priori* that this phenomenon will occur, we see that phenomenon often does occur, and then use that discovery to fix the theory. So it is not so much a case of 'life-history theory predicts that environment harshness will lead to the evolution of earlier reproduction…' as 'in practice, environmental harshness often leads to the evolution of earlier reproduction, and this motivates us to search for selective reasons why that might be true'. If you say 'life-history theory predicts that environment harshness will lead to the evolution of earlier reproduction….', as we often do in enterprise 2, then you are using the word 'theory' in an inductive, not a Newtonian sense.

What about enterprise 1, the idea that multiple different behaviours covary along a 'fast-slow' principal axis. Authors in enterprise 1 are unanimous in expressing the idea that such covariation is a basic prediction of life-history theory. Kimberely Mathot and Willem Frankenhuis recently conducted a systematic review of relevant models and concluded, perhaps surprisingly, that 'there is, at present, little formal theory' relating to the reasons why a single fast-slow principal axis would evolve[14]. If there is 'little formal theory' on the question, one has to ask, why do so many people believe the existence of such an axis to be a basic prediction of life-history theory?

The origins of the idea of the fast-slow continuum are in fact empirical more than theoretical. If you get empirical data on different species, for variables like age at first reproduction, litter size, duration of gestation, duration of lactation, etc., and stick them into a big correlation matrix, then *empirically* you discover that there is a principal axis, with late, slow and long species at one end, and early, fast and short species at the other[15]. So we can only really say that the axis is predicted by theory if by theory, we mean induction. What is in fact an empirical regularity has somehow morphed into being widely considered a theory. But importantly, in the original empirical analyses, the unit was the species, not the individual, and all the traits entered into the analysis were reproductive ones. The idea that you would get a single axis when comparing different individuals of the same species, and in particular that non-reproductive behavioural traits would also fall along this same axis as reproductive

---

[11] For example, Cichoń, M. (1997). Evolution of longevity through optimal resource allocation. *Proceedings of the Royal Society: B* 264: 1383–1388.

[12] Baldini, R. (2015). Harsh environments and 'fast' human life histories: What does the theory say? *BiorXiv,* https://doi.org/10.1101/014647. This paper is an enigmatic subcultural classic. Hard to understand, and only ever published as a preprint, if correct and general, it changes a lot. It's like some seminal set of early Dylan songs that exist only on 8-track somewhere, but circulates amongst aficionados.

[13] Reznick, D.A. et al. (1990). Experimentally induced life-history evolution in a natural population. *Nature* 346: 357-9.

[14] Mathot, K. J., & Frankenhuis, W. E. (2018). Models of pace-of-life syndromes (POLS): a systematic review. *Behavioral Ecology and Sociobiology* 72.

[15] Promislow, D.E.L. and P.H. Harvey (1990). Living fast and dying young: A comparative analysis of life-history variation among mammals. *Journal of Zoology* 220: 417-37.

ones, is certainly out there in the literature of non-human biology, but not really supported by current evidence[16].

It follows that when we do enterprise 1 in humans, we are not testing a prediction stemming directly from formal evolutionary theory in some Newtonian manner. We are taking an empirical pattern of co-variation seen across species, and arguably perhaps within some non-human species, and then looking for something vaguely analogous in human behavioural variability. I don't think there is anything wrong with doing this, but one has to wonder in what sense we are 'using life-history theory' or 'testing the predictions of life-history theory'. And indeed, one has to wonder exactly what the theoretical entity is that is subject to potential falsification here. Let's say we do an empirical study of a whole set of psychological traits within a human population and find that they don't really vary along a single axis. What exactly would be the endeavour whose credibility is undermined? Evolutionary life-history theory? Its application to humans? The generalization of between-species variation in reproductive traits to within-species variation in different traits? I don't know.

§

So far I seem to have argued that 'life-history theory' as it gets used in application to human behavioural traits is really a kind of extension of an inductive regularity, rather than a Newtonian theory. Actually, it's not even always that: it's sometimes a recipe for a recipe. My exemplar here is interesting experiments showing that people from different childhood backgrounds seem to respond very differently, in terms of their behavioural intentions for the future, to imagined scenarios evoking a world of harshness and scarcity[17]. So far so good, but these experiments are explicitly framed within 'life-history theory'. Thus the implication is that life-history theory either predicted these different responses *a priori*, or at least provides some major explanatory insight into them.

Life-history theory here is clearly being interpreted in terms of enterprise 3: there are fast and slow ways of behaving, and your childhood affects where you are on the continuum. Fine. But the experiments have two independent variables: childhood experience, and the imagined scenarios (either harsh world, or control). All enterprise 3 says here is 'somehow childhood experience will turn out to matter'. Equally compatible with the general contention of enterprise 3 would be: *only* childhood experience, and not the content of the current scenario, affects behavioural intentions; childhood experience and the current scenario both matter, and their effects are additive; or they matter synergistically in some kind of way (adverse childhood experience deadens you to cues of current harshness, or childhood experience sensitizes you to cues of current harshness). In short, 'life-history theory', as the phrase is being used, would be compatible with any conceivable pattern of results other than the one in which childhood experience does not matter at all. Thus, I have to ask: what sense of 'theory' is it, when 'life-history theory' as applied in this instance is compatible with almost all of the possible empirical outcomes?

I think the answer is that 'life-history theory' is being used as a recipe for a recipe. It denotes the general expectation that behaviours can be thought of as concerning doing things soon and fast, or later and slowly, and that one's childhood experience will make some kind of difference to one's propensities along this continuum. Exactly what kind of difference, and how childhood experience will combine with other situational factors, is a matter for further determination. The theory does not say. This is fine, I suppose, but we need to take away two things. First, this kind of theorising has only the vaguest of connections to the formal body of life-history models constructed by evolutionary

[16] See: Royauté, R., Berdal, M. A., Garrison, C. R., & Dochtermann, N. A. (2018). Paceless life? A meta-analysis of the pace-of-life syndrome hypothesis. *Behavioral Ecology and Sociobiology*, 72.
[17] Griskevicius, V., et al. (2011). The influence of mortality and socioeconomic status on risk and delayed rewards: a life history theory approach. *Journal of Personality and Social Psychology* 100: 1015–1026; Griskevicius, V. et al. (2011). Environmental contingency in life history strategies: the influence of mortality and socioeconomic status on reproductive timing. *Journal of Personality and Social Psychology* 100, 241–254.

biologists. And second, 'life-history theory' as used here is no more Newtonian than any other social-science theory. In fact, it looks almost exactly like the theory of life-course epidemiology, which I have written about elsewhere[18]. Life-course epidemiology basically says, all the things that happen to you over the course of life, including in particular childhood, are going to affect patterns of health and disease. How the different influences combine (additively, non-additively, etc.) is subject to further determination; indeed, the theory itself will be narrowed down in this regard according to what we find out empirically.

§

What lessons do I take away? First, narrowly, to people like me who want to apply evolutionary ideas to human behaviour. We shouldn't ever be so sloppy to say 'Evolutionary theory predicts x….', or 'life-history theory predicts x…' Evolutionary theory can predict a lot of things. Instead, if what we mean is 'the evolutionary model by Bloggs (2018) makes the prediction, that, under assumptions 1, 2 and 3, x will happen', then we should say exactly this. If what we mean is 'Empirical findings from non-human animals suggest that x often happens', then we should say that and cite the evidence. And if what we mean is 'I cooked up some ideas which I want to endow with the gravitas and authority of the most famous and respected meta-theory in the life sciences because it makes them sound better', then, well, that's just bad.

Another take-home is that we should actually do more life-history modelling, indeed more mathematical modelling in general, to try to provide stronger theoretical underpinnings to the observations we make. Mathematical models are only heuristic devices, and they don't solve all your scientific (or even theoretical) problems. They are very useful though, as instantiations of what your theory really is: formalising makes it into a stable, reproducible entity that can easily be queried. Models are useful in forward-engineering from starting assumptions to predictions, because verbal arguments are notoriously ambiguous, and informal intuitions about what would follow from what are often just wrong. Having a mathematical model is a way of showing rigorously that if I make this particular set of assumptions, then the prediction to which I am led is exactly the following. Rather than: once I got the data (or someone else gathered some data) I realised that quite possibly the predictions of my theory were not necessarily the ones I originally thought they were. And models are useful in reverse-engineering, from phenomena back to explanation. Say aggression is correlated with the risk of predation. Maybe there is some adaptive reason these two things get coupled. Now, under what ecological and demographic assumptions could such a coupling emerge? Then you can start to ask whether those assumptions seem plausible for ancestral humans (or whatever system you are studying).

I should note, however, that even if we made such models, even if those models made predictions, even if we tested those predictions, and those predictions were supported, it still would not be (complete) explanation yet. The classic models of life-history theory, indeed of behavioural ecology in general, are mostly only approximations. That's because they have no explicit population genetics, or only something very simplified. Yet what would actually evolve, presumably, would be alternate forms of genes in populations, populations where many genes were interacting and traits had a complex genetic basis. So even when we are done with making optimality models which to our satisfaction explain why men are more aggressive than women, or people in harsh environments reproduce young, a whole other group of people, theoretical evolutionary geneticists, would see that as just as a heuristic starting point, a sketch of a proposal still to be properly explained. And they would need their own theories and models to do their bit. Explanation is never done: it's just passed along the row. It reminds me of a conversation I once heard between two physicists. One said that he had been able to prove mathematically that some effect should occur under some set of circumstances. 'I mean

[18] Nettle, D. and M. Bateson (2017). Childhood and adult socioeconomic position interact to predict health in mid life in a cohort of British women. *PeerJ* 5:e3528.

prove to the satisfaction of a physicist, of course', he added. Proof to the satisfaction of a mathematician was a completely different issue.

A similar point can be made about providing a mechanism. People often say to us behavioural ecologists, yes, that's an interesting evolutionary theory, but what's the mechanism by which it would be delivered? I was studying early-life adversity and ageing in starlings. I was pleased with myself because we were measuring oxidative stress, a possible ageing factor, at the cellular level. I explained my plans to some cellular biologists. Interesting, they said. If you did find a difference in oxidative stress, what do you think the mechanism would be? What? Oxidative stress was, for me, the mechanism. Indeed, it was about the most mechanistic I had ever got. For them, my oxidative stress measure was just some crude phenotypic summary. How the oxidation picked up in the assay actually came about was a hole where a mechanism needed to be placed. Will it never end? Probably not. Just as one person's *explanans* is another person's *explanandum*, one person's mechanism is clearly another person's black box.

More broadly, it's clear that when researchers use the term 'theory', they are not referring to a homogenous class of entities. It would perhaps be helpful to use more precise terminology to refer, respectively, to specific hypotheses, inductive generalizations, mathematical models, recipes, recipes for recipes, and so on. Given the indefinitely large food-web in which we all operate, it would also be quite handy if theories, like buses, had an origin and destination clearly displayed on them. The theory of planned behaviour, for example, is a well-formed inductive theory whose destination is behavioural decisions, and whose origin is immediate psychological factors like beliefs. It never claimed to serve earlier stations on the line, and should not therefore be criticized for not doing so. A theory might usefully say: I'll pick you up at known regularities of individual human cognition, and drop you off at cross-cultural regularities in the content of literary stories. If you want to get on any earlier (e.g. where do the known regularities in individual cognition come from?), you will need to take an additional bus.

This picture casts an unflattering light on the idea, sometimes raised with rather messianic zeal, that the human sciences might one day be unified under a single grand unified theory. That idea is like saying that every bus stop in the city should be served by the same bus: hardly a recipe for getting from A to B any time soon. Surely the unification we should be looking for is not that a single bus goes everywhere, but rather that, using a network of buses that is reasonably well integrated, it is eventually possible to get from any starting point to any destination, using several types of theory along the way. The vision is the one beautifully expressed by Melvin Konner in the preface to *The Tangled Wing*[19]*:*

> A good textbook of human behavioral biology, which we will not have for another fifty years, will look not like Euclid's geometry—a magnificent edifice of proven propositions deriving from a set of simple assumptions—but more like a textbook of physiology or geology, each solution grounded in a separate body of facts and approached with a quiverful of different theories, with all the solutions connected in a great complex web.

And by the way, in closing, this for me is where we can shed a bit more light on the special explanatory status of evolution for the life sciences, and therefore for their subset the social sciences. We all know Dobzhansky's famous dictum 'Nothing in biology makes sense except in the light of evolution'. Continuing with my affordable public transport analogy, we could read this dictum in two ways. The strongest reading is the requirement that every bus, whatever its destination, must have evolution as its origin point. Even I, an enthusiast, can see that this is much too strong a requirement to be sensible. The second reading gives evolution a weaker, but still rather special, status. If you rode each bus back to its point of origin, and there picked up another bus and rode that one back to its origin, and so on

[19] Konner, M. (2003). *The Tangled Wing: Biological Constraints on the Human Spirit* (2nd ed., New York: Holt), p. xv. I am grateful to Karthik Panchanathan for introducing me to this quote.

and on, then wherever you started out, evolution would sooner or later, in one way or another, be the place you ended up. You might visit a lot of different and exotic locations along the way, but you can't really avoid getting back to evolution at some point, because we are embodied creatures who arose through a historical process that also produced the other organisms with which we share the earth. 'What we are supplying', as Ludwig Wittgenstein put it, amount in the end to 'remarks on the natural history of human beings'[20].

[20] Wittgenstein, L. (1953). *Philosophical Investigations* (Oxford: Blackwell), section 415.