Philos Stud
DOI 10.1007/s11098-014-0297-9

How (not) to bring psychology and biology together

Mark Fedyk

Abstract Evolutionary psychologists often try to "bring together" biology and psychology by making predictions about what specific psychological mechanisms exist from theories about what patterns of behaviour would have been adaptive in the EEA for humans. This paper shows that one of the deepest methodological generalities in evolutionary biology—that proximate explanations and ultimate explanations stand in a many-to-many relation—entails that this inferential strategy is unsound. Ultimate explanations almost never entail the truth of any particular proximate hypothesis. But of course it does not follow that there are no other ways of "bringing together" biology and psychology. Accordingly, this paper explores one other strategy for doing just that, the pursuit of a very specific kind of consilience. However, I argue that inferences reflecting the pursuit of this kind of consilience with the best available theories in contemporary evolutionary biology indicate that psychologists should have a preference for explanations of adaptive behavior in humans that refer to learning and other similarly malleable psychological mechanisms—and not modules or instincts or any other kind of relatively innate and relatively non-malleable psychological mechanism.

Keywords Evolutionary perspective · Ultimate explanations · Proximate explanations · Consilience · Evolutionary psychology

1 Introduction

Evolutionary psychology (EP) is the attempt to employ the methods of cognitive and social psychology to discover, test, and confirm theories about the nature of the human mind and its role in causing various different patterns of behaviour, and it is

---

M. Fedyk (✉)
Department of Philosophy, Mount Allison University, Sackville, NB E4L 1G9, Canada
e-mail: mfedyk@mta.ca

EP in the sense that it consciously adopts an “evolutionary approach” to searching for and subsequently framing the psychological hypotheses which can potentially be confirmed by such methods.

But EP also routinely operates entirely within the conceptual framework of a specific theory of the mind, a theory which is often thought by its proponents to uniquely fit the adoption of an evolutionary approach to human psychology. As John Cartwright explains in his excellent introductory textbook to EP:

> The whole of EP is based on the premise that behaviour is driven by adaptations. […] Evolutionary psychologists would argue that human behaviour as we observe it today is a product of contemporary environmental influences acting upon ancestrally designed mental hardware. The behaviour that results may not be adaptive in contemporary contexts. We should focus then on elucidating mental mechanisms rather than measuring reproductive behaviour. We should expect to find mind mechanisms that were shaped by the selection pressures acting on our distant ancestors…. The mind is not a blank slate designed to solve general mental problems because there were no general mental problems in the Pleistocene Epoch, only specific ones concerning hunting, mating, travelling, and so on. (Cartwright 2008, pp. 79, 85–86)

This is a sketch of the massive modularity hypothesis (MMH), and Cartwright supplies one of the common arguments used to link the MMH with evolutionary considerations.

However, one of this paper’s primary conclusions is that it is a mistake to hold that such a link exists. There is no inference from the fact that the mind, or its faculties, are adaptations to some version of the MMH. Pursuing an evolutionary approach—and more on what it means to adopt this approach throughout this paper—to human psychology does not require adopting a strongly nativist psychology like the MMH. In fact, quite the opposite seems to be true: an approach to human psychology that is grounded in the most recent theoretical developments in evolutionary biology implies a scientific preference for non-nativist psychological theories and models.

To make the argument for this, I begin by examining the common view that it is possible to make *predictions* about what specific psychological hypotheses are likely to be true using only premises which describe patterns of putatively adaptive behaviour. These patterns of behaviour are of course the “solutions” to the “problems” mentioned by Cartwright. Sects. 2 and 3 demonstrate that this approach—again, grounding psychological predictions in descriptions of patterns of adaptive behaviour—is inconsistent with one of the most secure methodological generalities of evolutionary biology.

This shows that the evidential relationship between psychology and evolutionary biology is not such that theories in evolutionary biology routinely entail particular psychological conclusions. But if the evidential connection between the two fields is not one of prediction or entailment, then what is it? I argue in Sect. 4 that the best analytic concept to use to characterize the relationship is that of “anticipated consilience”. This concept is able to explain why, when EP emerged from sociobiology in the late 1980s, it adopted the MMH. However, if the connection between evolutionary biology and EP is one of anticipated consilience, then, given developments in evolutionary biology since the turn of the millennium, adopting the "evolutionary approach" to human psychology in fact now requires that the MMH be rejected in favour of psychological theories which are compatible with more developmental plasticity than the MMH allows for. The upshot is clear: adopting an "evolutionary approach" to the mind now requires adopting a preference of non-nativist psychological theories.

2 Ultimate explanations, proximate explanations, and the "evolutionary approach"

So, let us begin by looking at one way of pursuing an "evolutionary approach" to human psychology. As indicated above by Cartwright's comments, such an approach can begin with examining the EEA for humans or an ancestral species and trying to find therein insight into what sorts of patterns of behaviours were adaptive. "EEA" stands for *the environment of evolutionary adaptedness*, and it is a technical concept denoting the set of environments in which a species or a near ancestor was subject to the pressures of natural selection. For humans, the EEA is usually taken to be a set of different environments which occurred in the late Pleistocene. Thus, the environment in which humans "solved" the "problem" of hunting by, say, cooperating to bring down game animals, is part of the EEA for humans—and here, the relevant pattern of adaptive behaviour would be cooperative hunting. Similar considerations apply to the environment in which humans "solved" the "problem" of travelling by, say, moving only in mornings and evenings to reduce water loss by perspiration. This environment is also part of the EEA, even if it is separated by both time and geography from the environment in which cooperative hunting was adaptive.

Now, when an evolutionary biologist is able to explain how a particular pattern of behaviour made a significant positive difference to the fitness of some organism(s), this kind of explanation using the concepts of natural selection is called an *ultimate explanation*. Ultimate explanations in biology refer to the historical conditions responsible for causally stabilizing a particular phenotype (or range of phenotypes) in a population (Mayr 1961; West-Eberhard 2003; Alcock 2001). So, ultimate explanations can refer to a kind of natural selection, or migration, or founder effects, or rapid climate changes, or indeed, any other structural feature of the environment in which a species lives that is causally responsible for determining the phenotypes present in that species.[1] An explanation as to why a pattern of behaviour was adaptive in an environment is therefore an *ultimate explanation* of the behaviour in question, but again, ultimate explanations can also refer to other evolutionary forces and other historical but non-evolutionary forces. For this reason, ultimate explanations are not uniquely the provenance of evolutionary biology; they are

---

[1] Accordingly, one way of reading Darwin's *Origin* is to see the point of chapter 1, "Variation under domestication", being that artificial selection is a sufficient ultimate explanation of some kinds of speciation, and then that the argument in subsequent chapters is that natural selection provides an analogous ultimate explanation of species that are, to put it simply, not domesticated.

used also in other historical sciences like astronomy and paleontology (Cleland 2011).

Ultimate explanations are to be distinguished from proximate explanations. Proximate explanations refer to causal processes that are both “internal” to an organism and occur “within” the lifespan of an organism. Psychology is by-and-large in the business of offering proximate explanations, but so too are medicine, evolutionary developmental biology, neurology, and molecular biology, amongst other scientific fields. Accordingly, proximate explanations refer to the psychological, physiological, neurophysiological, biochemical, biophysical, etc. processes which occur at some point within the course of an organism’s development and which are responsible for determining some aspect of the organism’s phenotype. So, a proximate explanation for a pattern of adaptive behaviour can refer to, for instance, a hormonal change, if such a change could cause the behaviour.

Let us return briefly to the two problems mentioned by Cartwright above, hunting and traveling. By showing how the fitness of, for example, the cooperative hunters was significantly increased relative to the fitness of the non-cooperative hunters, an evolutionary psychologist thereby supplies an ultimate explanation of cooperative hunting. However, this explanation says nothing whatsoever about what proximate mechanisms are responsible for causing the behaviour. We might have independent reason to think that there is any number of scientifically plausible proximate explanations of the behaviour, but knowing the correct ultimate explanation for cooperative hunting provides no insight whatsoever into which of these various proximate explanations is true. Perhaps a module for cooperation specifically in hunting generates the behaviour, or perhaps people simply imitated the behaviour of a small group of spontaneously cooperative hunters once they observed the benefits of this behaviour. Or, cooperative hunting could have been caused by social learning: some of the best theories in developmental psychology say that social learning is responsible for many patterns of cooperative behaviour, generally speaking (Tomasello 2009). Again, the ultimate explanation for cooperative hunting does not tell us which of these various proximate explanations for cooperative hunting is true, and the determination of which (if any) of these various proximate explanations is true can proceed independently of any related efforts to assess various ultimate explanations of the very same pattern of behaviour. As John Alcock writes, “one can ask purely proximate and purely ultimate questions, each category dealing with different but complementary aspects of a biological phenomenon.” (Alcock 2001, p. 15)

There is a deep explanation for why this is a general methodological truism that applies to the interface of the various proximate and ultimate sciences. For, proximate explanations and ultimate explanations nearly always stand in a *many-to-many* relation. It is usually the case that, for any scientifically plausible proximate explanation of some trait, there will be multiple different scientifically plausible ultimate explanations available for the same trait that are all compatible with the candidate proximate explanation. Likewise, for any scientifically plausible ultimate explanation for some trait, there will be multiple different scientifically plausible proximate explanations available for the same trait, and where each of these is compatible with the relevant ultimate explanation. Because of this, knowing which proximate explanation is most likely to be true for some trait will not ordinarily entail anything about which scientifically plausible ultimate explanation is likely to be true, and vice versa. This is what explains to become an expert in, say, neurophysiology is not ipso facto to become an expert in evolutionary biology. Or, in other words, the many-to-many relationship which holds between proximate and ultimate explanations explains the autonomy from one another of the various different proximate and ultimate sciences.

Let me offer some examples to help make this all clear. David Pfennig and Paul Sherman (Pfennig and Sherman 1995) describe the different proximate mechanisms used by a diverse range of species to implement kin recognition, which is commonly thought to evolve by kin-selection. These proximate explanations include, inter alia, smell (western toad tadpoles), pollen (mountain delphiniums), odor (paper wasps), and order in which eggs have been laid (acorn woodpeckers). Similarly, Mary Jane West-Eberhard writes that "cues that indicate degree of genetic kinship with other group members… are learned by individuals in many kinds of social animals, from isopods to humans." (West-Eberhard 2003, p. 340) So, knowing that kin recognition in some species most likely evolved by kin-selection does not provide any specific insight into what particular proximate mechanisms are used to implement kin recognition; there are many scientifically plausible proximate explanations available.

The same is true for ultimate explanations as well. Take for example the debate over whether or not warning coloration in insects usually evolves by kin or individual selection (Jarvi et al. 1981), or the debate over whether or not there is a specific type of sexual selection which can solve the 'lek paradox' either generally or at least in some particular species (Bradbury 1981; Andersson 1994), or the difficult problem of determining whether group selection for 'ecotypes' offers the best explanation of seemingly adaptive responses in ecosystems formed out of microbial communities (Williams and Lenton 2007). In each of these cases knowledge about what proximate mechanisms are responsible for the various traits and patterns of adaptive behaviour cannot be used to settle the debate over which ultimate explanation is correct. What's more, these various debates about the ultimate explanations for different traits can proceed even in complete absence of any understanding of which proximate explanations are plausible. We simply do not have the resources to determine the molecular and physiological processes that fix the coloration of every insect species; recent estimates of the number of insecta put this number at around 9 million (Mora et al. 2011).

3 Evolutionary psychology

The fact that ultimate and proximate explanations almost always stand in a many-to-many relationship has deep implications for the methodology of EP. To see these, we need first to turn back to John Alcock again, who writes that,

> The new field of EP (Alcock is comparing EP to sociobiology) analyzes proximate mechanisms of human behaviour from an explicitly evolutionary

> perspective, asking questions about why we possess particular psychological attributes and seeking ultimate answers in terms of the contributions these mechanisms might make or have made to the reproductive success of individuals. (Alcock 2001, p. 16)

Alcock describes a research programme focused on determining the correct ultimate explanation for independently identified psychological traits. In other words, Alcock seems to be describing a "present-to-past" or "backward-looking" pattern of inference: ordinary cognitive and social psychology tells us what psychological traits exist, and then we adopt an evolutionary approach when we seek to determine if any of these attributes are adaptations, and we do this by figuring out if there are any well-confirmed ultimate explanations for these traits that refer to some kind of natural selection.

Now, while "backward-looking" evolutionary analysis of independently identified traits seems to be a viable research programme, it would be a mistake to conclude that it is an important or central inferential pattern in the field of EP. For, pace Alcock, a very different kind of inferential pattern is characteristic of much of the most influential work in EP. This inferential pattern Edouard Machery calls EP's "main heuristic for the discovery of psychology traits", and here is how he describes it:

> Evolutionary psychologists' most important contribution to psychology is perhaps their attempt to offer a method for *discovering* human psychological traits. […] Evolutionary psychologists attempt to identify the adaptive problems regularly encountered by our ancestors. On this basis, they develop hypotheses about which psychological traits might have evolved. Whence the name "*forward-looking*": Evolutionary psychologists hypothesize a past adaptive problem and *predict* either that modern humans should possess a yet unknown psychological trait or that a known psychological trait of modern humans should possess yet unknown properties (Machery 2011, emphasis mine).

More explicitly, the evolutionary psychologist begins by finding a pattern of human behaviour that in the EEA would have been favoured by selection. This is sufficient to show that the pattern of behaviour *could have* been an adaptation; this is of course still a long way from confirming any particular ultimate explanation. Next, the evolutionary psychologist infers that there is a psychological mechanism which is largely innate and non-malleable, and which has the *unique* computational function of producing the relevant pattern of behaviour. Finally, a test for this mechanism is performed. Note that Machery calls this pattern of inference "forward-looking", and, importantly, he describes evolutionary psychologists as making predictions about what psychological (i.e., proximate) hypotheses are true on the basis of ultimate explanations. Of course, this strategy would only be reliable if an ultimate explanation usually entailed some proximate explanation—but because these two kinds of explanation almost always stand in a many-to-many relation, such an entailment does not exist.

However, before examining this problem in more detail, let us first consider an example of the research that Machery calls “forward-looking”, the recent work of evolutionary psychologists Aldo Cimino and Andrew Delton. They present the results of two experiments that were designed to test “whether the mind contains an evolved concept of newcomer—a series of cognitive subroutines designed [specifically] to reason about newcomers and their coalitional impact[.]” (Cimino and Delton 2010, p. 187) Cimino and Delton argue that there is compelling evidence indicating that, in the EEA for humans, intergenerational coalitions were routinely formed and that being a member of an intergenerational coalition was almost always beneficial to an individual’s fitness. Furthermore, they argue that it would have been adaptive for individual members of coalitions to reliably sort newcomers from “veteran” coalition members, and that veterans would generally be more trusting of other veterans than of newcomers. On this basis, Cimino and Delton *predict* that “the mind contains an evolved newcomer concept”, and that therefore:

> the mind should encode tenure length and assign members to categories based on these encodings (e.g. newcomer)… Second, categorizing an individual as a newcomer should elicit specialized motivational responses. Because newcomers are potential free riders, they should be viewed as (1) less trustworthy than veterans, (2) less entitled to coalition benefits, and (3) more worthy of punishment… newcomers should [also] be (4) judged as less competent than veterans and (5) given lowed regard—that is, judged as less likeable. (Cimino and Delton 2010, p. 188)

Cimino and Delton conducted a series of experiments designed to test these claims, and they tentatively conclude that the mind has an evolved newcomer concept (Cimino and Delton 2010, p. 197). Importantly, their testing protocol contained no controls for any of the other scientifically plausible proximate explanations—and I’ll give an explicit example of one in just a moment—for the adaptive patterns of recognition and trust that Cimino and Delton describe.

With that point in mind, it is easy to see the fundamental problem with this pattern of inference. Let us say that a set of proximate explanations PE_1, PE_2, … PE_n is *behaviourally-equivalent* if and only if each member of the set predicts the same pattern of behaviour. Now, since there will almost always be multiple different proximate explanations which are compatible with any given ultimate explanation, this is the same as saying that there will be a set of behaviourally-equivalent proximate explanations for almost any given pattern of potentially adaptive behaviour. Furthermore, if it comes out that there is a *true* ultimate explanation in terms of natural selection for some pattern of behaviour, then this means that the behaviourally-equivalent proximate explanations are *evolutionarily-equivalent*. That is, since each of these proximate explanations predicts the same pattern of behaviour, selection will not distinguish between them, and therefore, so far as the history of natural selection is concerned, *any* of the behaviourally-equivalent proximate hypotheses would have been just as good as any other. As per the argument in Sect. 2, knowing what the right ultimate explanation is for some pattern of adaptive behaviour does not tell you which of the behaviorally- and, if the pattern of behaviour is adaptive, evolutionarily-equivalent proximate explanations for the behaviour is true. It does not even imply that whichever proximate explanation is true must refer to some kind of relatively innate and relatively non-malleable trait, since after all many adaptations are extremely malleable and acquired from developmentally-exogenous sources (Jablonka and Lamb 2006).

Let's return to Cimino and Delton for a second to see how these points interact with their conclusion. The pattern of adaptive behaviour that Cimino and Delton identify is allowing newcomers to cooperate with existing coalitions, but trusting the newcomers less than existing coalition members. The favoured proximate explanation for this behaviour is the putative "evolved" newcomer concept. However, an alternative proximate explanation for this pattern of behaviour is supplied by Michael Tomasello:

> When they first begin to walk and talk and become truly cultural being … human children are already cooperative and helpful[.]… But later in ontogeny, children's relative indiscriminate cooperativeness becomes mediated by such influences as their judgments of likely reciprocity and their concern for how others in the group judgment them…. And they begin to internalize many culturally specific social norms for how we do thing, how one ought to do things if one is to be a member of this group. (Tomasello 2009)

Tomasello calls this hypothesis the "Early Spelke, Later Dweck" explanation of cooperation, and he defends it at some length. But the important point here is that both Cimino and Delton's evolved concept and Tomasello's Early Spelke, Later Dweck hypotheses are behaviourally-equivalent with respect to trusting newcomers to cooperative activities less, but still cooperating with the newcomers. And, if we assume that there is a true ultimate explanation referring to natural selection of the relevant pattern of behaviour, then these proximate explanations are also evolutionarily-equivalent. The result is evolutionary considerations cannot therefore be used to decide between these two hypotheses. And they same can be said for Cimino and Delton's experiment, since they failed to include any controls for proximate explanations like Tomasello's account or indeed any of the other scientifically plausible proximate explanations for cooperation and trust in the EEA (cf. Henrich and Henrich 2007).

It is easy to find examples of such "forward-looking" inferences in EP; Machery is right that this pattern of inference is absolutely central to the methodology of the field. So let us consider just one more example of a "forward-looking" inference, and the critique of it that the many-to-many relationship between proximate and ultimate explanations allows us to generate. The example is from a famous paper by John Daly and Margo Wilson, the very first words of which state a "forward-looking" inference:

> Child-rearing is a costly, prolonged undertaking. A parental psychology shaped by natural selection is therefore unlikely to be discriminate. Rather, we should expect parental feeling to vary as a function of the prospective fitness value of the child in question to the parent… When people are called upon to fill parental roles towards unrelated children, we may anticipate an elevated risk of lapses of parental solicitude (Daly and Wilson 1985, p. 197).

This prediction about parental feeling was tested by analyzing rates of child abuse in the Hamilton-Wentworth area in Southern Ontario, where Daly and Wilson found significantly higher rates of abuse for children living with at least one step-parent than children living with both biological parents. Daly and Wilson conclude, "As predicted from Darwinian considerations, step-parents themselves evidently constitute a risk factor for child abuse." But there are other proximate explanations that are behaviourally-equivalent to Daly and Wilson's view of parental feeling and that are consistent with the Hamilton-Wentworth data. Parents may become more easily annoyed with step-children than biological children, as they will almost certainly have spent more time bonding with their own biological children than with any step-children who come into their care. Step-children may be less tempered when interacting with a new care-giver who is replacing the role of a previous caregiver with whom the child had already formed a bond. The socioeconomic impact of a divorce and remarriage may also simply make parenting more difficult than it otherwise would be. These and other alternative proximate hypotheses are discussed further in (Cartwright 2008, p. 220).

Let us see where we have come. EP is committed to a methodology whereby a pattern of putatively adaptive behaviour is identified, and on this basis, it is predicted that there exists a module-like mechanism that has the unique psychological function of causing the relevant pattern of adaptive behaviour. This is, in effect, to deduce a proximate explanation from an ultimate explanation. However, in evolutionary biology proximate and ultimate explanations stand in a many-to-many relationship, and so it will routinely be possible to identify, for some pattern of adaptive behaviour, a set of scientifically plausible behaviourally-equivalent proximate explanations. But evolutionary psychologists routinely fail to control for the members of this set, and this fatally impairs EP's ability to confirm its predicted proximate hypotheses.

Or, at least that seems to be the lesson so far. In the next section, however, I consider one strategy that may rationalize a scientific preference for either a single, or at least a small number, of the behaviourally-equivalent proximate hypotheses that are compatible with some ultimate explanation. That is, I consider a strategy which may rationalize EP's methodological tendency of only seriously considering largely innate and largely non-malleable proximate explanations.

However, before moving on to examining this issue, there is an important clarification to the argument above that needs to be made. It needs to be emphasized that the argument in this section is neutral with respect to the nativism/non-nativism controversy. As it happens, evolutionary psychologists have a tendency to posit only those proximate explanations that refer to relatively innate and non-malleable psychological mechanisms, but this tendency is not the target of the argument above. Rather, the argument focuses on the practice of seriously examining only a *single* scientifically plausible proximate explanation for a pattern of adaptive behaviour, when nothing less than the methodology of evolutionary biology suggests that there will routinely be multiple scientifically plausible and behaviourally-equivalent proximate explanations available. This means that the argument above would still apply in the possible world in which evolutionary psychologists routinely only considered single proximate explanations that refer just to, say, psychological states produced as the output of domain-general associative learning mechanisms. The difference would be that, in this alternative scenario, the proximate explanations that get ignored would include hypotheses which refer to developmentally endogenous modules, and which are nevertheless behaviourally-equivalent to the proximate explanations which refer to the outputs of associative learning.[2]

4 Anticipated consilience

So, it is tempting to conclude that the argument in the previous section demonstrates that "forward-looking" EP is simply methodologically unworkable. Once it is understood that there will ordinarily be many different behaviourally-equivalent proximate explanations that are compatible with an ultimate explanation of some pattern of behaviour for humans in the EEA, it is hard to see what possible course of experimentation or analysis could successfully rule out all but one of these proximate hypotheses, for this is what would be required to confirm such a hypothesis. Indeed, since for many putatively adaptive patterns of behaviour the list of plausible proximate hypotheses will include fairly context-specific causes (i.e., local cultural norms) about which we cannot collect much meaningful data, it is hard to see how any scientific progress could be made determining which proximate explanation is (or at least was) true.

However, there is a strategy which seems to be able to short-circuit this problem, at least partially. Here's the idea. When scientists are engaged in interdisciplinary research like EP, there must a number of 'auxiliary disciplines' from which they draw both conceptual resources and empirical results. If one of the auxiliary disciplines happens to be an extremely mature, extremely successful scientific

---

[2] There is also an important caveat to the claim that, generally, true ultimate explanations do not entail the truth of any particular proximate explanations. For, if some particular pattern of adaptive behaviour did occur in the history of some organism, then it does follow, in a trivial sense, that the organism either has, or at least had, some psychological (or some other kind of proximate) mechanism that has (or had) the function of being able to cause the relevant pattern of behaviour. But it does not follow from this that there is a psychological (or proximate) mechanism the *only* function of which is to produce the relevant behaviour, since many different psychological mechanisms can satisfy such a functional description. For, again, it is a truism that, if an organism is able to produce a pattern of behaviour B, then some part of it has the function of *being able to produce B*. Call whatever has this function trait T. Now, note the language used to talk about T does not indicate whether or not T has any other functions. So T could of course be a psychological module, in which case its *only* significant function may be to produce behaviour B and it is also true that the possession of T is relatively innate and non-malleable. But T could also be information that was acquired from one-off peer learning, in which case its only function may be to produce B, and yet in this case the possession of T is extremely context dependent. What's more, T could also be a domain-general psychological faculty, or even a system of faculties. In this case the *total* functional description of T would include the ability to cause B along with a host of other causal functions; indeed, this list might be infinitely long for a sufficiently complex system. These three different (types of) psychological mechanisms—a module, contextually-acquired information, and domain-general faculties or systems—can all be said to have the function *being able to produce B*. For this reason, it would be a mistake to infer that any specific proximate conclusions follow from the trivial fact that, if behaviour B was produced by an organism, some part of the organism has the functional ability to cause B.

discipline, then scientists working in an interdisciplinary research programme can use the mature auxiliary discipline to coordinate the direction of their own research. One of the more specific ways of doing this is to rely upon what can be called the principle of anticipated consilience, which says that it is rational to have a prima facie preference for those novel theories commended by previous scientific research which are most likely to be subsequently integrated in explanatorily- or inductively-fruitful ways with the relevant mature discipline as it expands. The principle will be reliable simply because the novel theories which are most likely to be subsequently integrated into the mature scientific discipline as it expands are just those novel theories which are most likely to be true.

Now, the principle of anticipated consilience is obviously not a bad principle to use in any interdisciplinary research programme that has a mature science like contemporary evolutionary biology as one of its auxiliary disciplines. For, no reasonable person should deny that evolutionary biology and its various subfields together are some of the most successful scientific theories of all time. Amongst its many other epistemological virtues, this theory has immense explanatory power, extremely inductively fruitful conceptual resources, and impressive amounts of internal coherence and consistency. It is extraordinarily unlikely that we will be able to develop a general biological theory that could overturn the Darwinian consensus. It will therefore be rational to prefer amongst the novel theories commended to the relevant interdisciplinary research programme only those novel theories that have a reasonably strong chance of being integrated—even though they are not yet—in explanatorily or inductively significant ways with theory in evolutionary biology as it continues to expand.

Thus, the suggested strategy is to rely upon considerations of anticipated consilience to narrow down the range of proximate explanations that must be investigated while pursuing research in EP. For example, considerations of anticipated consilience may rationalize preferring just those proximate explanations which refer primarily to innate modules *if and only if* in the relevant auxiliary disciplines to EP—evolutionary biology most importantly, but also cognitive science and other branches of psychology—there exists a significant tendency to confirm only module-like proximate explanations.

To see how this would work, consider, for example, two fairly general categories of proximate explanations for adaptive behaviours in humans, nativist psychological hypotheses which posit some kind of module and non-nativist psychological hypotheses which posit some kind of learning routine. Considerations of anticipated consilience would only tell us to prefer the nativist hypotheses and reject the non-nativists hypotheses if, for instance, the evidence from the best available evolutionary biology says that adaptive behaviours almost always are produced primarily by modules. Conversely, if the lesson from evolutionary biology is that learned adaptive behaviours are ecologically ubiquitous, then the principle of anticipated consilience tells us to prefer the non-nativist hypotheses. And finally, if the best available theories in the relevant auxiliary disciplines provide no indication one way or another as to whether learned or modularized adaptive behaviours are significantly more common, then the principle of anticipated consilience cannot be used as a reason either to prefer or to reject either members of either category of hypotheses.

5 Modules, instincts, cognitive science, and the history of evolutionary psychology

What I now want to argue is that it is plausible that, for approximately the first decade of research in EP following its emergence out of sociobiology in the 1980s, considerations of anticipated consilience would have likely rationalized a preference for proximate explanations which refer to modules and similar types of proximate mechanisms. The specific evidence for this is an interesting intersection of two trends: first, the popularity of modularized explanations in cognitive science stemming from the publication of Fodor's (1983) and the influence of Chomskyan linguistics, and second, the relative paucity of efforts in evolutionary biology at the time to assess proximate hypotheses.[3] My argument relies upon Paul Griffiths (2008) historical account of the emergence of EP from sociobiology.

This history begins in the 1940s with Konrad Lorenz's theory of instincts. Lorenz was committed to the view that instincts were discrete neurological mechanisms, which he treated as analogous to a complicated hydraulic pumping system. The idea was that impulses to perform specific patterns of behaviour are continuously generated by the nervous system, but these impulses do not automatically flow into causing behaviour because they are, instead, stored up in some kind of neural reservoir. These stored-up impulses are only are able to cause behaviour when the environment supplies stimuli which triggers a releasing mechanism, or when the neural pressure becomes too great and the instinctive behaviour spontaneously emerges "in a vacuum".

Griffiths writes:

> The theoretical framework of ethology evolved rapidly in the 1950s and 60 s. Three important developments were: (1) The abandonment of Lorenz's identification of instinctive behavior with behavior which is innate as opposed to acquired; (2) The abandonment of the hydraulic model; (3) The integration of ethology with evolutionary ecology, resulting in an increased focus on documenting the adaptive value of behavior. […] By the mid 1980s one would have been hard pressed to find a young student of animal behavior who regarded their work as a contribution to ethology, as opposed to behavioural ecology or sociobiology. (Griffiths 2008)

One of the central methodologies of this revised approach to studying animal behavior was the reliance upon population genetics, a tendency which is underscored in E.O. Wilson's famous dumbbell prediction of the trajectory of research in sociobiology and related fields. According to Wilson's model,

---

[3] That is, at least until the emergence of evolutionary developmental biology. Importantly, the field's leading journal, *Evolution and Development*, was founded in 1999.

neurobiology and sociobiology together would eventually swallow all other areas of research on animal behavior. Importantly, little attention was paid to proximate questions in these new approaches to the study of animal behaviour that developed between the 1960s and the 1980s. Thus:

> Studies of [proximate] causation and ontogeny could, in principle, have contributed to behavioral ecology in a very direct way …. but in almost all cases those studies were not advanced enough to provide this information…. Hence, instead of developmental biology making a positive contribution to behavioral ecology, it appeared only in the negative role of 'developmental constraints'." (ibid.)

So by the time EP emerged in reaction to human sociobiology in the 1980s, Lorenz's old hydraulic model of instincts really was the last positive model in biology of the proximate causes of adaptive behaviour.

Furthermore, Griffiths writes that, "EP uses the same population genetic and evolutionary game theory models as sociobiology, and there is often little difference in the actual explanations which the two schools offer for human behavior." (ibid.) Griffiths here is referring to ultimate explanations only, since sociobiologists and behavioural ecologists have traditionally been reluctant to make proximate claims. But as for proximate explanations in these fields, Griffiths echoes the argument given in Sect. 3 above:

> Behavioral ecology does not predict anything about the specific neural structures that underlie behavior. Models in behavioral ecology predict which behaviors would have been selected in the ancestral environment [i.e., the EEA], but they cannot distinguish between different mechanisms that produce the same behavioral output. (ibid.)

Yet, as we have seen, "forward-looking" EP is just the effort to ground a particular proximate conclusion in an ultimate explanation. So while EP borrows from behavioural ecology its techniques for assaying ultimate explanations, its concern with deducing proximate conclusions from ultimate explanations represents a significant departure from the methods of behavioural ecology.

Let us take stock. When in the late 1980s and early 1990s evolutionary psychologists began to do what behavioural ecologists and sociobiologists were reluctant to do, namely infer substantive proximate conclusions from ultimate hypotheses, the biology of adaptive animal behaviour was largely silent as to what types of proximate explanations should be preferred. That said, the similarity between Lorenz's hydraulic model of instincts and the notion of a module as was being developed in cognitive science at the time is remarkably striking. Instead of some kind of neural pressure, a module implements a program that when given some appropriate environmental stimuli as input produces as output a specific pattern of behaviour (or motivation for that pattern of behaviour). Yet both the module's program and the instinct's neural impulses are innate, and both mechanisms operate as switches that link features present in the flow of environmental stimuli received by the senses with discrete neural circuits that, in turn, can cause specific patterns of behaviour. The main difference is that Lorenz's instincts are analog switches (specifically, fluid switches), whereas a module is of course a very complicated digital switch. Thus, we have the following situation in the late 1980s: researchers in cognitive science were articulating an approach to the proximate causes of human behaviour which resembled rather strikingly the last positive proximate hypotheses considered in the biology of adaptive animal behaviour. Accordingly, if an evolutionary psychologist had to make an educated guess as to the type of proximate hypotheses that were most likely to eventually be subsumed by future developments in two of the auxiliary fields of EP, evolutionary biology and cognitive science, considerations of anticipated consilience would have most likely rationalized a preference for just those proximate hypotheses which posited modules.

6 Plasticity and biology without modules

However, the direction of research in evolutionary theory has become increasingly complicated since the 1990s. There is of course evolutionary developmental biology, and evolutionary anthropology (Richerson and Boyd 2006). However, what I want to argue is that, in the last decade or so, a new understanding of the biological importance of *developmental plasticity* has implications for EP. Whereas previously considerations of anticipated consilience with evolutionary biology and cognitive science may have provided support for those proximate hypotheses which posited modules, I argue in this section that these very same considerations now support significantly non-nativist proximate hypotheses. The argument, put simply, is that traits which have high degrees of plasticity will be more evolutionarily robust than highly canalized innately specified non-malleable traits like mental modules. The upshot is that a mind comprised mostly of modules is not plastic in this specific sense, and is therefore ultimately unlikely to be favoured by natural selection. But a mind equipped with powerful, domain general learning routines does have the relevant plasticity.

But let us go through the argument in more detail. Developmental plasticity is a technical concept which refers to the ability of an organism to modify its development in response to environmental conditions. (Moczek et al. 2011) Mary Jane West-Eberhard defines it as “the ability of an organism to react to an environmental input with a change in form, state, movement, or rate of activity.” (West-Eberhard 2003, p. 34) Thus, developmental plasticity does not denote any particular chemical, molecular, neurological or physiological function, but instead stands for the general ability, realized at various different levels of physiological organization, to express some members of a set of potentially adaptive traits as a function of environmental conditions.

Importantly, centrality of developmental plasticity in evolution requires that we adopt a more sophisticated understanding of how an organism’s genotype is linked with its phenotype. Genes must be conceived of as providing for both unexpressed and expressed traits, where traits of the former kind can remain unexpressed for generations upon generations at a time. This approach contrasts with the more traditional neo-Darwinian picture in which traits must be expressed in order to be preserved by selection. The traditional view therefore assigns to the environment the function of a "screen", whereby through a largely passive and random process certain genetic mutations are blocked from being reproduced in subsequent generations, while others are allowed through. Environmental conditions therefore determine the nature and course of selection, and development serves only as the process whereby a genotype is translated into a phenotype. However, the developmental plasticity approach represents the environment as being able to reach into organism and release previously "stored" genetic variation. In this way, development is conceived of as *highly* environment dependent. Moczek et al. (2011) write, "Evolution by genetic accommodation therefore expands beyond a traditional neo-Darwinian model by recognizing that the interplay between environment and development provides a mechanism for both the accumulation and the rapid release of genetic variation in the face of novel environmental challenges."

Bones provide an interesting and concrete example of developmental plasticity. Wolff's Law states that bones will remodel themselves in response to the loads they are required to bear, becoming stronger or weaker in proportion to their typical loads. Bone remodeling is an example of developmental plasticity, and it is to be contrasted with the view that the density or strength of a bone is innately fixed and largely constant throughout mature ontogeny. West-Eberhard again:

> Given our inability to perceive change in growing bones, it is easy to underestimate the extreme condition sensitivity of bones. (Wimberger 1991) A similar malleability characterizes the development of wood in trees, which responds to wind and position to produce species-specific variants so consistent in particular environments while absent in others that they could easily be mistaken for genetically specified, constitutive traits.

Bones, unlike mental modules, change in response to environmental conditions. Bones are plastic in a way that modules are not; but more on this subject in just a moment.

What has emerged in the last decade is a near consensus view amongst evolutionary biologists that developmental plasticity occurs in all areas of the biological world and that developmental plasticity is fundamental to evolution. The simplest argument for this latter claim is that the ability to replicate, which is the mark of organic life, requires molecules that are able to respond dynamically to the environments in which they occur (Schlicting and Smith 2002). However, there is also an emerging consensus that learning is a crucial example of developmental plasticity. West-Eberhard one last time:

> Learned traits have greater evolutionary potential than do mutational ones, since many individuals of a population may simultaneously and suddenly learn the same things in the same circumstances, or due to mimicry of other individuals. At the same time, learning can match trait to conditions, thereby bringing a learning-linked trait complex under consistent selection in the environment where it is likely to be adaptive…. [For these and other reasons,] Learning has a special quality that sets it apart from other mechanisms of plasticity and makes it unusually important for adaptive evolution….

> Learning, then, is a fitness-enhancing mechanism par excellence, a mechanism formed by natural selection in such a way that it mimics natural selection as a multiplier of adaptive traits. (West-Eberhard 2003)

Developmental plasticity is fundamental to evolution, and learning is an "unusually important" kind of developmental plasticity.

The point should now be clear: if you had to guess at what kind of proximate mechanism was responsible for a pattern of adaptive behaviour in humans in the EEA, considerations of anticipated consilience with evolutionary biology would *now* recommend going in for some kind of learning mechanism, not a module. The reason this follows is that the mind, according to the MMH, is like a bone which is unable to remodel itself in response to increases or decreases in its load. The MMH says that there are developmentally-endogenous computational modules which come pre-programmed with algorithms for those patterns of behaviour that "solved" adaptive problems in the EEA. The mind is thought to be nothing more than, well, a massive collection of these modules. John Cosmides and Leda Tooby call this idea the "monomorphic mind thesis" (Tooby and Cosmides 1992; see also Kurzban 2011). Importantly, since each of these modules is largely hardwired with specific programs which must cause specific behavioural patterns in response to specific environmental conditions, learning, according to the MMH, cannot be a frequent cause of adaptive patterns of behaviour.

What's more, the evolutionary process hypothesized to be responsible for selecting the modules by evolutionary psychologists is the very same neo-Darwinian view of evolution that is challenged by the growing understanding of developmental plasticity in evolutionary biology. Tellingly, John Cosmides and Leda Tooby write that, "selection, interacting with sexual recombination, enforces a powerful tendency towards unity in the genetic architecture underlying complex functional design at the population level and usually the species level as well." (Cosmides and Tooby 1990, p. 393) So the MMH goes hand-in-hand with a particular conception of natural selection that no longer seems plausible in light of the most recent developments in evolutionary biology.

We have, therefore, an interesting dialectical situation. The MMH may have once seemed to be favoured by considerations of anticipated consilience running from cognitive science and evolutionary biology to EP, at least in the 1980s and early 1990s. However, in the intervening years, evolutionary biology has evolved, and the new understanding of developmental plasticity is ipso facto a renewed interest in the proximate mechanisms responsible for adaptive behaviour. This means that, given a choice between proximate explanations which reflect a commitment to the MMH and proximate explanation which, instead, reflect an approach to the mind which privileges learning (Gopnik and Wellman 2012; Xu and Kushnir 2012; Tomasello 2000; and for the minds of non-human animals, see Syal and Finlay 2011; Mineka and Cook 1988), the latter is most plausible in light of evolutionary biology.

This demonstrates that EP is not, well, EP. It is committed—as per Machery, Cartwright, and others—to a methodology which stands in tension with the best available evolutionary biology. This is not just because the use of "forward-looking" inferences to make psychological predictions is incompatible with one of the deeper methodological truths in evolutionary biology, that proximate and ultimate explanations almost always stand in a many-to-many relation. We can also see that, to put the point simply, the MMH is implausible in light of an understanding of developmental plasticity.

That said, this conclusion also points to a potential direction that "forward-looking" EP can take, should its proponents care to align the discipline with evolutionary biology. When faced with a range of behaviourally-equivalent proximate explanations for some pattern of behaviour in the EEA, considerations of anticipated consilience with contemporary evolutionary biology can be used to "screen off" as unlikely-to-be-true those hypotheses which posit modules or some other kind of relatively innate, relatively non-malleable proximate mechanism. In some cases, this approach may reduce the number of candidate proximate hypotheses sufficiently to allow for their rigorous scientific testing.

7 Conclusion

Let's recap. Proximate explanations and ultimate explanations stand in a many-to-many relationship, and this means that it is not possible to make predictions about what proximate mechanisms are true on the basis of an analysis of what patterns of behaviour would have likely been adaptive in the EEA for humans. This presents a deep problem for EP, as this field is committed by-and-large to a methodology whereby "forward-looking" predictions about the existence of particular modules are derived from ultimate explanations. We saw above that there will routinely be several behaviourally-equivalent scientifically plausible proximate hypotheses which are compatible with almost any given ultimate explanation, and one of the problems this raises for "forward-looking" EP is that it will be hard to conduct a proper scientific evaluation of the full set of proximate explanations.

But perhaps there is a way to decrease the number of behaviourally-equivalent proximate hypotheses which must be tested. Above I argued that considerations of anticipated consilience may be able to do just that. Specifically, I argued that considerations of anticipated consilience with cognitive science and evolutionary biology may have once rationalized a preference for strongly nativist proximate hypotheses. However, I have also argued that the growing understanding of the evolutionary importance of developmental plasticity has changed this, so that considerations of anticipated consilience now most likely rationalize a preference for strongly non-nativist proximate hypotheses.

If this argument is right, then it is yet another example of the deep way in which scientific success is beholden to the accuracy of the background scientific theories which all scientific practice relies upon. Many philosophers of science have called attention to the way in which background conditions contribute to scientific success (Boyd 1985, 1990, 2001; Psillos 1995; Cartwright 1983; Giere 1983, 1990; Fine 1984), but we have in the preceding section an interesting illustration of a way in which background conditions can also be a source of scientific error. Since EP became an autonomous scientific discipline in the 1990s, it has developed its own internal theories and methodologies. Evolutionary psychologists can now by-and-large publish in their own specialist journals without the direct involvement of its various auxiliary fields. However, if the conception of evolution found within EP does not agree with the conception of evolution as it exists in the best available contemporary evolutionary biology, it is surely the evolutionary psychologist's conception of evolution which needs to give.

So, after all this, let us return to one of the questions we began with: What does it mean to adopt an "evolutionary approach" to human psychology? This could be the view that attempts to characterize human nature should begin with efforts to understand what patterns of behaviour would have been adaptive in the EEA for humans. However, as we've seen, it is very hard to extract any specific psychological conclusions from even very well-confirmed ultimate explanations. So, an alternative way of pursuing an "evolutionary" study of human nature would be to make use of considerations of anticipated consilience with evolutionary biology. This would mean that a general rule for pursuing evolutionary approaches to human psychology—holding everything else equal, of course—is: set the initial plausibility of any given novel proximate hypothesis proportional to the plausibility of significantly similar (kinds of) proximate hypotheses as can be found in any of the biological sciences which have as their subject adaptive patterns of behaviour.

Acknowledgments Thanks to the following for their helpful comments, criticisms, questions, and in one particular case, for several very good ideas too: Amy Allcock, Richard Boyd, Barbara Koslowski, Jane Dryden, Kate Cober, Robbie Moser, Roopen Majithia, Tamar Kushnir, and several anonymous referees.

References

Alcock, J. (2001). *The triumph of sociobiology*. Oxford: Oxford University Press.

Andersson, M. (1994). *Sexual selection*. Princeton: Princeton University Press.

Boyd, R. (1985). Observations, explanatory power, and simplicity. In P. Achinstein & O. Hannaway (Eds.), *Observation, experiment, and hypothesis in modern physical science* (pp. 349–378). Cambridge: MIT Press.

Boyd, R. (1990). Realism, approximate truth and philosophical method. In W. Savage (Ed.), *Scientific theories, Minnesota studies in the philosophy of science* (Vol. 14, pp. 355–391). Minneapolis: University of Minnesota Press.

Boyd, R. (2001). Reference, (in)commensurability, and meaning: some (perhaps) unanticipated complexities. In P. Hoyningen-Huene & H. Sankey (Eds.), *Incommensurability and related matters* (pp. 1–63). Dordrecht: Kluwer.

Bradbury, J. W. (1981). The evolution of leks. In R. D. Alexander & D. W. Tinkle (Eds.), *Natural selection and social behavior* (pp. 138–169). New York: Chiron Press.

Cartwright, N. (1983). *How the laws of physics lie*. Cambridge: Cambridge University Press.

Cartwright, R. (2008). *Evolution and human behavior: Darwinian perspectives on human nature* (2nd ed.). Cambridge: MIT Press.

Cimino, A., & Delton, A. W. (2010). On the perception of newcomers: Toward an evolved psychology of intergenerational coalitions. *Human Nature, 21*(2), 186–202.

Cleland, C. E. (2011). Prediction and explanation in historical natural science. *British Journal of Philosophy of Science, 62*(3), 551–582.

Cosmides, L., & Tooby, J. (1990). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology, 11*, 375–424.

Daly, M., & Wilson, M. (1985). Child abuse and other risks of not living with both parents. *Ethology and Sociobiology, 6*(4), 197–210.

Fine, A. (1984). The natural ontological attitude. In J. Leplin (Ed.), *Scientific realism* (pp. 83–107). Berkeley: University of California Press.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge: MIT Press.

Giere, R. (1983). Testing theoretical hypotheses. In J. Earman (Ed.), *Testing scientific theories, Minnesota studies in the philosophy of science* (Vol. 10, pp. 269–298). Minneapolis: University of Minnesota Press.

Giere, R. (1990). Evolutionary models of science. In N. Rescher (Ed.), *Evolution, cognition, and realism* (pp. 21–32). Lanham: University Press of America.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin, 138*(6), 1085–1108.

Griffiths, P. E. (2008). Ethology, sociobiology, evolutionary psychology. In S. Sarkar & A. Plutyinski (Eds.), *Blackwell's companion to philosophy of biology* (pp. 393–414). Oxford: Oxford University Press.

Henrich, N., & Henrich, J. (2007). *Why humans cooperate*. Oxford: Oxford University Press.

Jablonka, E., & Lamb, M. J. (2006). *Evolution in four dimensions*. Cambridge: MIT Press.

Jarvi, T., Sillén-Tullberg, B., & Wiklund, C. (1981). Individual versus kin selection for aposematic coloration: a reply to Harvey and Paxton. *Oikos, 37*(3), 393–395.

Kurzban, R. (2011). *Why everyone (else) is a hypocrite: evolution and the modular mind*. Princeton: Princeton University Press.

Machery, E. (2011). Discovery and confirmation in evolutionary psychology. In J. Prinz (Ed.), *The Oxford handbook of philosophy of psychology*. Oxford: Oxford University Press.

Mayr, E. (1961). Cause and effect in biology. *Science, 134*(3489), 1501–1506.

Mineka, S., & Cook, M. (1988). Social learning and the acquisition of snake fear in monkeys. In T. R. Zentall & B. G. Galef Jr (Eds.), *Social learning: psychological and biological perspectives* (pp. 51–73). Hilladale, NJ: Lawrence Erlbaum.

Moczek, A. P., Sultan, S., Foster, S., Ledon-Rettig, C., Dworkin, I., Nijhout, H. F., et al. (2011). The role of developmental plasticity in evolutionary innovation. *Proceedings of the Royal Society Biological Sciences, 278*, 2705–2713.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., & Worm, B. (2011). How many species are there on earth and in the ocean? *PLoS Biology, 9*(8), e1001127.

Pfennig, D. W., & Sherman, P. W. (1995). Kin recognition. *Scientific American, 272*(6), 98–103.

Psillos, S. (1995). Is structural realism the best of both worlds? *Dialectica, 49*, 15–64.

Richerson, P. J., & Boyd, R. (2006). *Not by genes alone: How culture transformed human evolution*. Chicago: University of Chicago Press.

Schlicting, C. D., & Smith, H. (2002). Phenotypic plasticity: linking molecular mechanisms with evolutionary outcomes. *Evolutionary Ecology, 16*, 189–211.

Syal, S., & Finlay, B. L. (2011). Thinking outside the cortex: Social motivation in the evolution and development of language. *Developmental Science, 14*(2), 417–430.

Tomasello, M. (2000). *The cultural origins of human cognition*. Cambridge: Harvard University Press.

Tomasello, M. (2009). *Why we cooperate*. Cambridge: MIT Press.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford: Oxford University Press.

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

Williams, H. T. P., & Lenton, T. M. (2007). Artificial selection of simulated microbial ecosystems. *Proceedings of the National Academy of Sciences, 104*, 8918–8923.

Wimberger, P. H. (1991). Plasticity of jaw and skull morphology in the neotropical cichlids *Geophagus brasiliensis* and *G. steindachneri*. *Evolution, 45*, 1545–1563.

Xu, F., & Kushnir, T. (Eds.) (2012). *Rational constructivism in cognitive development*. Academic Press—Elsevier, Waltham.