Australasian Philosophical Review

Volume 6 Issue 4 December 2022

Routledge
Taylor & Francis Group

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rapr20

Are Biological Traits Explained by Their 'Selected Effect' Functions?

Joshua R. Christie, Carl Brusse, Pierrick Bourrat, Peter Takacs & Paul E. Griffiths

To cite this article: Joshua R. Christie, Carl Brusse, Pierrick Bourrat, Peter Takacs & Paul E. Griffiths (2022) Are Biological Traits Explained by Their 'Selected Effect' Functions?, Australasian Philosophical Review, 6:4, 335-359, DOI: 10.1080/24740500.2024.2370630

To link to this article: https://doi.org/10.1080/24740500.2024.2370630

AUSTRALASIAN PHILOSOPHICAL REVIEW
2022, VOL. 6, NO. 4, 335–359
https://doi.org/10.1080/24740500.2024.2370630

Are Biological Traits Explained by Their 'Selected Effect' Functions?

Joshua R. Christie[a], Carl Brusse[a,b], Pierrick Bourrat[a,c], Peter Takacs[a], and Paul E. Griffiths[a]

[a]University of Sydney; [b]Australian National University; [c]Macquarie University

ABSTRACT

The selected effects or 'etiological' theory of Proper function is a naturalistic and realist account of biological teleology. It is used to analyse normativity in philosophy of language, philosophy of mind, philosophy of medicine, and elsewhere. The theory has been developed with a simple and intuitive view of natural selection. Traits are *selected* because of their positive *effects* on the fitness of the organisms that have them. These 'selected effects' are the Proper functions of the traits. Proponents argue that this analysis of biological teleology has the unique advantage that the selected effect function of a trait is also a causal explanation of the trait: the trait exists *because* it performs this function. We show, however, that selected effect functions as currently defined explain the existence of traits only under highly restrictive assumptions about evolutionary dynamics. In many common scenarios in which traits evolve by natural selection, selected effect functions do not explain those traits. This is because definitions of selected effect function extract from any evolutionary scenario only the information that would be explanatorily relevant in the simple evolutionary scenario implicit in those definitions. When applied to more complex scenarios, selected effect functions omit the key information that is explanatorily relevant in those scenarios. The assumptions required for selected effect functions to be explanatory are particularly unlikely to hold in the domain that its proponents care most about—the evolution of representation. A more adequate selected effects theory of Proper functions may be possible, but will require much greater attention to the structure of actual evolutionary explanations.

ARTICLE HISTORY Received 9 June 2021; Accepted 14 October 2021

KEYWORDS Proper function; Millikan, Ruth; frequency-dependent selection; adaptation; teleosemantics

1. Introduction

Teleological language in biology, such as the 'function', 'purpose' or 'goal' of some part or process, can be interpreted as an elliptical reference to natural selection.

> If we ask 'What does a cat have sharp, curved claws for?' and answer simply 'To catch mice with', this does not imply a profession of any mythical teleology, but the plain statement that catching mice is the function whose survival value, by the process of natural selection, has bred cats with this particular form of claw. [Lorenz 1966: 9]

CONTACT Pierrick Bourrat pierrick.bourrat@mq.edu.au

In the Darwinian modern synthesis this idea was known as 'teleonomy' [Pittendrigh 1958], a term chosen to clearly demarcate it from the early twentieth century idea that teleology is an emergent property of living things [Nicholson and Gawne 2015].

The teleonomic approach to biological function has been widely adopted by philosophers as a result of Ruth Millikan's influential *Language, Thought, and Other Biological Categories* [1984] and Karen Neander's influential but unpublished PhD thesis and subsequent papers [Neander 1983, 1991a, 1991b]. These philosophers sought to define the true or 'Proper' functions of biological traits—to explain why it is the Proper function of the nose to warm and filter inhaled air and not its Proper function to support spectacles, however well it performs that function [Voltaire 1759]. Their analysis of teleological language was essentially the same as that popularized in the modern synthesis, 'The gist of this … theory is that the (or a) function of an item … is what it was selected to do' [Neander 2017: 20]. They did not adopt the term 'teleonomy', perhaps because their immediate influences were other philosophers (e.g., Wright [1973]) rather than biologists. Instead, philosophers call this the 'etiological' or 'selected effects' theory of Proper function.

While usage differs, we find it helpful to use 'Proper function' to mean functions that provide some sort of objective criterion of what is normal or pathological, intact or broken, well-functioning or dysfunctional. It is the aspiration of the selected effects theory to identify the Proper functions of traits. We use 'selected effects function' to refer to the actual functions picked out by the selected effects theory, remaining neutral on whether these are in fact Proper functions (or, indeed, whether Proper functions exist).

Advocates of the theory claim that what distinguishes it from other interpretations of teleological language in biology is that only functions so defined—selected effects functions—causally explain why traits exist. The jewel-like colours of the male superb fairy wren exist today because ancestral males with brilliant colours survived and reproduced more effectively than they would have without them. In ancestral wrens, the colours increased fitness by attracting mates, so in today's wrens the selected effect function of these colours is to attract mates. The wren's colours have other effects, such as attracting wildlife photographers, but only the first effect is what colour is for: the colour exists because it has this effect. This special explanatory role is what makes it the 'Proper' function of the trait.

Neander makes clear that this special explanatory role is the basic rationale for the selected effects theory of Proper function:

> That the koala's pouch has the function of protecting its young does seem to explain why koalas have pouches. That the bee's dance is for directing other bees to pollen does seem to explain why bees dance [Neander 1991b: 457]

She continues, '… my view is that function attributions universally and intrinsically justify teleological explanations' [Neander 1991b: 458]. They do so as a result of 'a causally explanatory selection process, during which those items or traits were selected for those effects which are their functions' [Neander 1991b: 467].

Leaping forward twenty-eight years, Justin Garson, the most prominent current advocate of the selected effects theory, still treats the claim that functions causally explain traits as a key argument for that theory:

> consider a puzzling feature of ordinary biological usage, namely, function's explanatory depth: sometimes function statements are causal explanations for traits. When biologists say the

> function of the zebra stripes is to deter biting flies, for example, they're trying to explain why zebras have stripes … If we take explanatory depth seriously, then the traditional selected effects theory, or something in its neighborhood, has no equal. [Garson 2019: 7]

Our concern in this paper is that selected effect functions only explain the existence of the traits that bear them under restrictive assumptions about evolutionary dynamics. When we drop these assumptions, the claim that selected effects functions causally explain traits becomes highly problematic. Selected effects functions almost always provide some information that is explanatorily relevant to the evolution of traits. But in many cases they discard the key explanatory information about the evolution of the trait. They do this because they extract from an evolutionary explanation only information that would be explanatorily relevant if the assumptions about evolutionary dynamics that are implicit in the selected effects theory held. When those assumptions do not hold, explaining the trait by citing its selected effects function is misrepresenting why the trait evolved. The key elements of the actual evolutionary explanation have been deliberately omitted from this 'explanation' and the remaining information corresponds to an evolutionary scenario that did not occur.

We show below that selected effects functions fail to explain traits when the process of natural selection feeds back onto the relative fitness of those traits. Such feedback generates a form of environmental heterogeneity that is endogenous to the selection process. Exogenous environmental heterogeneity can also create evolutionary scenarios in which traits are not explained by their selected effects functions. Both endogenous and exogenous environmental heterogeneity are ubiquitous in evolution. Proper function theorists have not had to confront these problems because they use simple, intuitive examples of natural selection presented via verbal scenarios. These scenarios assume a uniform selective environment. But when that assumption is dropped, then it is not generally true that selected effects functions explain the traits that bear them.

Selected effects functions fail to explain these traits in the straightforward sense that when the actual evolutionary explanation of how such a trait evolved is fed into the definition of selected effect function, the resulting function ascription is not a summary or sketch of the actual evolutionary explanation. By the 'actual evolutionary explanation' we mean the explanation found in the scientific literature. The selected effects function of a trait is derived from facts about the evolution of the trait, and both we and selected effects theorists have no choice but to rely on evolutionary biology to estimate those facts if we want to apply the theory. We give concrete examples below of traits whose evolution is well understood, but when the explanation is fed into selected effects definitions of function, that understanding is discarded and the case treated as if the trait evolved in a very different way.

Our criticism is not tied to any specific philosophical theory of scientific explanation, which is fortunate given the lack of consensus on that topic [Woodward and Ross 2021]. However, we can clarify our criticism using an idea from that literature, namely an 'ideal explanatory text' containing all the information about an event needed to answer any why-question we might pose about that event [Railton 1981]. A scientific explanation extracts some of that information and a good explanation extracts the right information to answer a specific question. We show below that the question the selected effects theory of function asks is always 'what were the circumstances under which the trait increased the fitness of its bearers relative to rivals with alternative traits?' The selected effects theory implicitly assumes that an answer

to this question is also an answer to the question 'why did this trait evolve?' But these two questions are often not equivalent. In many important evolutionary scenarios the two questions extract very different information from the ideal explanatory text. Selected effects function ascriptions almost always convey some information from the ideal explanatory text about the evolution of the trait. In the problematic cases, however, this is not the right information. It is wrong in the straightforward sense that it does not include the key pieces of information that an evolutionary biologist would select to explain why the trait evolved.

We expect that many selected effects theorists will see this as mere quibbling about technicalities. However complex real evolutionary explanations may be, surely this cannot undermine the basic insight that the appearance of teleology in nature is due to natural selection? But this is not a responsible attitude given the weight that philosophers want to put on the selected effects theory. Philosophers want to identify the Proper functions of traits so that they can, for example, define the truth-conditions of mental and/or linguistic representations [Millikan 1984, 2017; Neander 2017]. They use their theory to distinguish the normal from the pathological in medicine and psychiatry [Wakefield 1992; Neander 1998; Garson 2022]. Here and in other areas of philosophy they try to ground normative claims on facts about Proper function or dysfunction. The humble observation that natural selection reduces the puzzle of apparent design in nature cannot support the ambitious conclusion that multiple domains of normative facts are reducible to facts about natural selection. Instead, selected effects theorists need to demonstrate that this ambitious program can be carried out. Their claim to identify 'Proper' functions rests squarely on the claim that selected effects functions causally explain the existence of the traits that bear them. If this is not true, then the selected effects theory is simply one among many definitions of function and its judgements about what is 'Proper' cannot bear the weight philosophers put on them.

The situation here is similar to the situation in evolutionary theory with respect to the links between natural selection and adaptation. It is widely believed that natural selection consistently leads to adaptation. This is assumed as a working hypothesis in some subfields of evolutionary biology such as behavioural ecology. But being widely believed is no substitute for rigorous demonstration. R.A. Fisher's 1932 'fundamental theorem of natural selection' [Fisher 1999] was long supposed to be such a demonstration. This theorem states that the genetic variance in fitness in a population, at a given time—which measures the strength of natural selection—is equal to the rate of increase in the mean fitness in this population. Mean fitness has been taken as a proxy for adaptation (see Okasha [2018: ch. 3]). However, this interpretation was contentious even before Price [1972] gave the fundamental theorem what is now its standard interpretation. On this interpretation, natural selection has two types of effects: direct and indirect (via the environment). While the direct effects lead to an increase in mean fitness, when the indirect effects are taken into consideration, an overall increase in mean fitness is not guaranteed. So the fundamental theorem does not, in fact, imply that natural selection always produces adaptation. The formal Darwinism project of Alan Grafen [2014] is a more recent attempt to forge a rigorous link between natural selection and adaptation. In this case too, it is unclear that the project delivers the required demonstration (see the special issue on the topic in *Biology and Philosophy* edited by Okasha and Paternotte [2014]). Two philosophers who have examined the question in considerable depth, Jonathan Birch [2016] and Samir Okasha [2018], both conclude that the

hopes for finding a general link between natural selection and adaptation are bleak. Our analysis, although less formal, reveals that the widely accepted link between selected effect function and evolutionary explanation is similarly lacking any firm foundation. It, too, assumes an unwarranted link between episodes of selection and the course of evolution. We hope that selected effects theorists will take this on board and dig into the evolutionary details to see if their program can really be carried out. Until this work is done the selected effects theory runs the risk of merely hand-waving at natural selection to lend an air of respectability to normative intuitions.

In section 2, we compare the most influential statements of the selected effects theory and extract a canonical version. In the process we reveal a surprising degree of vagueness in even the best current statements of the theory. In section 3, the longest section of the paper, we substantiate our criticism by demonstrating the failings of selected effects function in a range of important evolutionary scenarios. In section 4, we identify why the selected effects theory is unable to handle these scenarios and use this analysis to generalize our criticism.

We note that one move available to selected effects theorists is to restrict the scope of the theory to those cases where its assumptions about evolutionary dynamics hold. This would already be a major concession, but in section 5 we also show that those assumptions are least likely to hold in just those domains of greatest interest to philosophers, namely the evolution of representations.

2. Selected Effects Definitions of Proper Function

There are many formulations of the selected effects theory of Proper function. Here are four influential examples:

Definition 1:

> … for an item A to have a function F as a 'Proper function', it is necessary (and close to sufficient) that … A originated as a 'reproduction' (to give one example, as a copy, or a copy of a copy) of some prior item or items that, due in part to possession of the properties reproduced, have actually performed F in the past, and A exists because (causally historically because) of this or these performances. [Millikan 1989: 288]

Definition 2:

> It is a/the Proper function of an item (*X*) of an organism (*O*) to do that which items of *X*'s type did to contribute to the inclusive fitness of *O*'s ancestors, and which caused the genotype, of which *X* is the phenotypic expression, to be selected by natural selection. [Neander 1991a: 74]

Definition 3:

> Where i is a trait of systems of type S, a Proper function of i in S's is F iff a proximal selective explanation of the current non-zero proportion of S's with i must cite F as a component in the fitness conferred by i. ['proximal' means recent episodes of selection] [Griffiths 1993: 415]

Definition 4:

> The function of m is to F iff:
>
> (i) m is a member of family T,
> (ii) members of family T are components of biologically real systems of type S,

(iii) among the properties copied between members of T is Property or Property cluster C, which can do F,
(iv) one reason members of T such as m exist now is the fact that past members of T were successful under selection in the recent past, through positively contributing to the fitness of S, and
(v) members of T were selected because they did F, through having C. [Godfrey-Smith 1994: 359]

Nicholas Shea is another influential selected effects theorist. He defines Proper function for the specific case where a trait has the Proper function of carrying information about a state of affairs, for example, a state of the visual system carrying information about the orientation of a stimulus. Shea's definition (Defn 5) requires that 'an evolutionary explanation of the current existence of the representing system adverts to Rs *[the trait]* having carried information about C *[the state of affairs]*' [Shea 2007: 419].

The differences between these definitions are dwarfed by their similarities. All share the idea that a selected effect function is an effect whose positive effect(s) on fitness plays a role in explaining why a trait evolved by natural selection. Exactly what role the positive effect on fitness must play in an evolutionary explanation is less clear from these definitions. Shea says that the evolutionary explanation 'adverts to' the effect. Neander says that the effect 'caused the genotype … to be selected by natural selection' [Neander 1991a: 174]. Godfrey-Smith says that the trait was 'successful under selection in the recent past, through positively contributing to … fitness' [Godfrey-Smith 1994: 359]. Both Neander and Godfrey-Smith's formulations are naturally interpreted as requiring that the effect caused an increase in the proportion of the trait relative to its alternatives. Griffiths takes a slightly different tack—a selective explanation of the observed proportion of the trait needs to include this effect on fitness if it is to correctly predict the observed proportion. But all the definitions agree that the positive effects of the trait on individual fitness must be used to explain the representation of the trait in the target population.

Selected effects theorists have given surprisingly little attention to how much influence selection must have to count as an explanation of the trait. Griffiths' formulation is both the most precise and the weakest: any positive effect of the trait on fitness whose omission from a model would reduce the ability of that model to predict the course of evolution is a selected effect function. Godfrey-Smith requires that there be positive selection in the recent past, and that this selection is one reason we observe the trait in the target population. Both these formulations suggest that any positive effect of a trait on selection, however weak, is sufficient to confer a function. Population geneticists often think of the 'strength' of a selection pressure in terms of the probability that the trait will become fixed[1] in the population in some period of time. Traits will sometimes become fixed when it was much more likely they would go extinct, a feature of evolution which takes on considerable importance when effective population sizes are small. If the unlikely survival of the trait was *slightly less unlikely* due to some effect, then that effect is a selected effect function of the trait on Griffiths' and perhaps Godfrey-Smith's definitions. But what explains the persistence of the trait in such cases is overwhelmingly luck. Considerations like these have led some philosophers to suggest that traits have Proper functions to some

[1] A trait is fixed when all members of the population have the trait.

degree rather than as an either/or matter [Matthewson 2020]. We will not pursue this suggestion here.

Neander and Millikan's language suggests stronger requirements than Griffiths' or Godfrey-Smith's on what influence selection must have to count as an explanation. Neander says the effect must have 'caused' the selection of the relevant genes. Millikan says the trait 'exists because (causally historically because)' of the effect [1989: 288]. But selection scenarios range from those where it is near inevitable that the trait will be represented in future populations to those where it is possible but unlikely, and we are left to guess where the line is to be drawn. Given that assignments of selected effect function are meant to clarify problematic distinctions in psychology, medicine, and so forth, this is highly unsatisfactory.

In contrast to these neglected areas, there has been extensive discussion of whether the selected effect functions of a trait are the effects that led to its initial spread in the population or whether they are the effects it was recently selected for and, if so, what is meant by 'recent'? [Griffiths 1993; Godfrey-Smith 1994]. But our concerns apply whichever period of evolution is supposed to define function, so we can set this dispute aside.

The most careful definition of Proper function in the philosophical literature is undoubtedly that of Millikan [1984]. Definitions 1–5 all state or imply that the trait which has the Proper function is the same trait, in some sense, as the trait that was selected in ancestral populations. Millikan spells this out in more detail with her definition of a 'reproductively established family' of trait tokens.[2] What Millikan calls a 'family' of traits biologists call a 'homologue'—a set of traits related to one another by descent from a common ancestor.

Having defined 'reproductively established family', Millikan defines Proper function as follows:

Definition 6:

> Where *m* is a member of a reproductively established family R and R has the reproductively established or Normal character C, *m* has the function F as a direct[3] Proper function iff:
>
> (1) Certain ancestors of *m* performed F.
> (2) In part because there existed a direct causal connection between having the character C and performance of the function F in the case of these ancestors of *m*, C correlated

[2] Millikan writes: 'First we define the notion "ancestor of a member of a reproductively established family":

(1) Any member of a (first-order) reproductively established family from which a current member *m* was derived by reproduction or by successive reproductions is an ancestor of *m*.
(2) Any temporally earlier member of a (higher-order) reproductively established family which member was produced by an ancestor of the device that produced a present member *m* is an ancestor of *m*.
(3) Any earlier member of a (higher-order) reproductively established family that a present member *m* is similar to in accordance with a Proper function of a producer that produced both is an ancestor of *m*.' [Millikan 1984: 27–8]

Note that 'reproductively established families' are not lineages of organisms (e.g., wrens) or even lineages of characters in the familiar sense of a 'transformation sequence' of characters like the transformation sequence from scales to feathers. Instead, they are lineages of individual characters: one blue head is the 'ancestor' of another blue head and all the homologous blue heads are a 'family'.

[3] Millikan's distinction between 'direct' and 'derived' functions will not concern us here. Very briefly, her idea is that objects which do not have a history of natural selection, such as the stones in the case of a Caddis fly larva, can nevertheless have Proper functions 'derived' from the Proper function of the mechanisms that make the stone case.

positively with F over a certain set of items S which included these ancestors and other things not having C.

(3) One among the legitimate explanations that can be given of the fact that *m* exists makes reference to the fact that C correlated positively with F over S, either directly causing reproduction of m or explaining why R was proliferated and hence why *m* exists. [Millikan 1984: 28]

Restating Millikan's definition using conventional biological terminology will make it easier to apply her theory to realistic biological examples. These terms are *organisms*, *populations*, and *characters*. In comparative biology 'character' refers to what philosophers call a 'determinable Property', and 'character state' to what philosophers call a 'determinate Property'. For example, head colour is a character of which blue and brown are possible character states. Turning blue for the breeding season is a character of which March to October and Never are two possible character states. A table of equivalences between Millikan's terminology and the conventional terms in biology is given as an Appendix.

We will also depart from Millikan's formulation by defining Proper function for the members of a population, not for a specific individual. The constant *m* in Millikan's definition refers to a trait of a specific individual, such as the blue head of the fairy wren that one of us saw while working on this article—let's call him Walter. But the actual evolutionary explanations that biology produces explain facts about populations. So in our definition what is assigned a Proper function is not Walter's blue head in particular, but blue heads in the population of which Walter is a member. Given that we are assigning functions to members of a population, we also state the definition in terms of what explains the proportion of the trait in a population, not why it 'exists'. Presumably a trait 'exists' if the proportion in some population is greater than zero.

Our proposed canonical statement of the selected effects theory of Proper function is:

Definition 7: Consider a population of organisms some of whom have character state C_i. These organisms are descendants of organisms with C_i and their character states are homologous. In this population, C_i has effect F as a Proper function if and only if:

(1) In some ancestral populations there was variation in C.
(2) Having state Ci caused some ancestral individuals to produce effect F with higher probability than individuals with alternative character states.
(3) Performing F caused some ancestral individuals with Ci to have greater reproductive output than they would have had if Ci had been changed to an alternative character state extant in that ancestral population (Cj).
(4) The frequency of Ci in the current population is influenced by selection for Ci in these ancestral populations in virtue of the increase in fitness accruing to individuals with Ci from performing F more frequently than individuals with alternative character states.

From Defn 7, we can clearly see the general structure of the selected effects account:

- Condition 7.1 is necessary if natural selection is to operate.
- Condition 7.2 establishes a causal connection between our focal trait and its effect.
- Condition 7.3 mechanistically connects our focal trait and its effect with a fitness advantage over individuals with alternative character states.

- Condition 7.4 connects our trait, its effect, and its fitness advantage over competing traits with the spread of the trait in the population.

When all four conditions are satisfied, F is the selected effects function of the trait.

3. When Do Selected Effects Functions Explain Traits?

In the sections that follow, we will highlight common evolutionary scenarios in which there is incongruence between the actual evolutionary explanation and the explanation offered by a selected effect function. By the 'actual evolutionary explanation' we mean the explanation found in the current scientific literature. The selected effects function of a trait is meant to be derived from the facts about how the trait evolved, and to constitute an explanation of why the trait evolved. If that explanation is not congruent with the actual evolutionary explanation from which it is derived, something has gone wrong.

In each example, we follow a four-step process: (i) introduce the evolutionary scenario and state the background assumptions; (ii) state the actual evolutionary explanation; (iii) fill out Defn 7, obtaining a statement of selected effect function; and (iv) demonstrate the incongruence between the actual evolutionary explanation (from ii) and the explanation offered by the selected effect function (from iii).

3.1 No Coevolution: Selected Effect Function Successfully Explains Zebra Stripes

We will begin by exploring Garson's example: 'the function of zebra stripes is to deter biting flies' [Garson 2019]. For simplicity, imagine that zebra stripes are determined by the action of a single gene and that offspring exactly resemble their parent (more formally, an allele at a single haploid locus with clonal reproduction—this is a common idealization in population genetics). Let us assume that biting flies have a predilection for Non-striped individuals [Caro, Izzo, *et al.* 2014; Caro, Argueta, *et al.* 2019], and that this preference does not evolve.

We now state the actual evolutionary explanation: in an ancestral population of non-striped individuals, there once arose a striped individual through mutation, who on account of being subject to fewer fly bites, was able to survive and reproduce more efficiently than non-striped individuals, a process that continued unabated until all Non-striped individuals had been replaced by striped individuals.

Now we will define the variables and fill out Defn 7. Let the character C be Pattern, of which we assume there are two variants: the striped trait C_i (denoted Stripes), and the non-striped trait C_j (denoted Non-stripes). The effect, F, of Stripes to reduce the probability of being bitten by flies is denoted Reduce_Bites. We can fill out Defn 7 as follows.

Consider a population of organisms, some of whom have Stripes. These organisms are descendants of organisms with Stripes and their character states are homologous. In this population, Stripes has effect Reduce_Bites as a Proper function if and only if:

- Z1. In some ancestral populations there was variation in Pattern.
- Z2. Having Stripes caused some ancestral Striped individuals to Reduce_Bites with higher probability than Non-striped individuals.

- Z3. Reduce_Bites caused some ancestral Striped individuals to have greater reproductive output than they would have had if Stripes had been changed to Non-stripes in that ancestral population.
- Z4. The frequency of Stripes in the current population is influenced by selection for Stripes in these ancestral populations in virtue of the increase in fitness accruing to Striped individuals from Reducing_Bites more frequently than Non-striped individuals.

In this case, the explanation offered by the selected effect function is congruent with the actual evolutionary explanation because the selective regime is invariant with respect to the fitness advantage of Striped individuals over Non-striped individuals.

3.2 Coevolution: Selected Effect Function Fails to Explain Zebra Stripes

In the example above, we made a number of simplifying assumptions. In particular, we assumed that biting flies could not evolve their preference for non-striped organisms. We now relax this assumption and consider a scenario in which there is a coevolutionary arms race between zebra stripes and the preference trait of biting flies.

Imagine that biting flies can evolve their preferences for Pattern (assume that a single fly either specializes on Stripes or on Non-stripes). We assume that biting flies can evolve rapidly and that the more frequent a Pattern, the stronger the selective force acting on biting flies to specialize upon it. The strength of selection increases with the frequency of a Pattern because the more common a Pattern, the more frequent the encounter rate between flies and that type, which increases the advantage an individual fly would gain by specializing on that Pattern. Let us assume that when Stripes evolved, the majority of mammal species upon which the biting fly fed were Non-striped [Jordan, Lee-Jones, and Weitz 1962].

In this scenario, biting flies evolved to specialize on (common) Non-striped species, and Striped zebras simply happened to possess a phenotype to which the biting fly was not co-adapted: Stripes Reduces_Bites because it is a rare phenotype. The actual evolutionary explanation is that being a rare phenotype Reduces_Bites. Indeed, if Stripes were to rise in frequency and become the dominant Pattern—or if Striped species had originally been the majority type—the biting fly would specialize on Stripes. In this case, Stripes would instead increase bites, highlighting why an evolutionary explanation centred around how Stripes Reduce_Bites would be deeply misleading.

Crucially, the Proper function ascription (Z1–Z4) does not change despite the addition of coevolutionary dynamics to this scenario. There is no way for Defn 7 to incorporate the fact that Stripes Reduces_Bites in some contexts (when Non-Striped species are most common) but increases bites in other contexts (when Striped species are most common). As such, there is now an incongruence between the explanation offered by the selected effect function and the actual evolutionary explanation.

Such coevolutionary dynamics between hosts and parasites are well-studied in the context of Red Queen dynamics and can lead to a range of different evolutionary dynamics (Brockhurst, Chapman, *et al.* 2014). For example, when host-parasite coevolutionary feedbacks occur over a short timescale, we expect to see fluctuating selection. Such dynamics might occur if we were to assume that zebras are the only host of biting flies, and that the traits for the biting preference of flies and stripes of zebras evolve at a similar rate. In this case, the frequency of Stripes would oscillate from low to high along

with a time-lagged oscillation of the frequency of the biting fly preference trait [Brockhurst, Chapman, *et al.* 2014]. Under such dynamics, it is clearly problematic to claim that what explains the proportion of Stripes is that it has the effect Reduce_Bites when roughly half the time Stripes actually increase bites. There is, however, no easy way out of this quandary because our selected effect function cannot incorporate information about coevolutionary dynamics.

Our aim in this section is to show that implicit assumptions about evolutionary dynamics are made whenever a Proper function explanation is constructed. We are not trying to show that the historical evolution of zebra stripes *actually* involved coevolution with biting fly preference. It is sufficient that numerous examples of natural selection to which we might hope to apply a selected effect function explanation *do* involve coevolutionary dynamic: for example, parasitic cuckoos [Servedio and Lande 2003], plant–insect mutualisms [Bronstein, Alarcón, and Geber 2006], and, as we will discuss later, the evolution of signalling and communication. An explanation involving coevolutionary dynamics *could* be the correct explanation for any given evolutionary scenario, and whenever this is the case, the explanation offered by the selected effect function will be incongruent with the actual evolutionary explanation. In the next section, we show how another common evolutionary dynamic, negative frequency-dependence, can lead to incongruence.

3.3 Frequency-dependent Selection in the Gouldian Finch

In this section, we show in more detail using a real case how negative frequency-dependence undermines the idea that the selected effect function of a trait is also an evolutionary explanation of that trait. We start with the simple, textbook evolutionary scenario known as the 'Hawk–Dove game'.

In this game-theoretic scenario, two strategy types (Hawkish versus Dovish) can persist in a population despite the fact that the payoff to 'Hawks' always exceeds the payoff to 'Doves' when the two interact. Even though Hawkish individuals always prevail over Dovish individuals, the costs to Hawks of their interactions with other Hawks are high. Hawks can, in effect, be considered 'their own worst enemy'. As the proportion of Hawks in the population rises and Hawk–Hawk interactions become more likely, the average payoff to Hawks decreases and the average payoff to Doves increases. The equilibrium ratio of Hawks to Doves is such that the average fitness of the two types is equal.

A natural example of this evolutionary game has been observed in the spectacular Gouldian finch of northern Australia [Kokko, Griffith, and Pryke 2014]. Simplifying somewhat, the species has two heritable character states or morphs: red-headed and black-headed. The two morphs mate assortatively, their mating preference sustained by the fact that hybrid offspring have reduced fitness. Red-heads have elevated production of testosterone and corticosteroids which causes them to compete aggressively and exclude black-heads from nesting cavities. Red-headed finches can accordingly be depicted as exhibiting the Hawkish strategy in a suitable game-theoretic model. As the proportion of red-heads in a population increases, so too will the cost of competition due to the increased frequency of encounters with other red-headed finches. This reduces the resources (time, energy) available to red-headed birds for parenting. Black-headed birds, whose Dovish strategy invests more resources in parenting and less in conquest, subsequently raise more successful broods. In wild populations, the

two types persist alongside one another at around three black to one red and with occasional gene flow between them. A selected effect function explanation for the existence of either heritable character state should preserve the core of the actual evolutionary explanation, which is that each exists at a level that represents equilibrium in frequency-dependent competition with the other.

What are the functions of these behavioural traits in Gouldian finches? We can apply Defn 7 to determine the selected effect function of the Hawkish phenotype of red-headed finches:

> Consider a population of finches, some of whom have elevated production of testosterone and corticosteroids ('Elevated'). These organisms are descendants of finches with Elevated and their

Figure 1: Gouldian finches, © suwatsir / Adobe Stock

Table 1: Payoff matrix for the Hawk–Dove game (V = value of contested resource, C = cost of conflict)

	Hawk	Dove
Hawk	$\left(\frac{V-C}{2}, \frac{V-C}{2}\right)$	$(V, 0)$
Dove	$(0, V)$	$\left(\frac{V}{2}, \frac{V}{2}\right)$

character states are homologous. In this population, Elevated has defeating rivals for nesting cavities as a Proper function because:

- H1. In some ancestral populations there was variation in production of testosterone and corticosteroids (Elevated versus Reduced).
- H2. Elevated caused some ancestral Elevated individuals to defeat rivals for nesting cavities with higher probability than individuals with Reduced.
- H3. Defeating rivals for nesting cavities was causally responsible for some ancestral Elevated individuals having greater reproductive output than they would have had if Elevated had been changed to Reduced.
- H4. The frequency of Elevated in the current population is influenced by selection for Elevated in these ancestral populations in virtue of the increase in fitness accruing to Elevated individuals from defeating rivals for nesting cavities more frequently than Reduced individuals.

Defn 7 could also be applied, *mutatis mutandis*, to determine the selected effect function of the Reduced phenotype of black-headed finches, whose function will be to conserve resources for investment in the young. This definition assigns a function in both cases, but the claim that the function of each trait explains that trait is highly problematic.

The fact that Elevated has the selected effect function <defeating rivals for nesting cavities> is supposed to answer the question, 'what explains the frequency of Elevated in the current population?' The actual evolutionary explanation is that selection is frequency-dependent and takes the population to an equilibrium frequency where the fitness of the two types is equal. The selected effect function explanation discards this information and includes only information about why Elevated was sometimes fitter than Reduced. This misrepresents how Elevated evolved, attributing it to positive fitness differences in some patches of the environment, rather than to negative frequency-dependence. As well as misrepresenting the actual explanation, the selected effect function simply fails to answer the question. The fact that Elevated is sometimes fitter than Reduced does not explain why the frequency of Elevated is 25 per cent or even why Elevated coexists with Reduced at any frequency, rather than being fixed.

A possible response by the selected effects theorists is to change the question. They can ask 'why does Elevated exist' (why is its frequency greater than zero)? But the selected effects function is also a poor explanation of why Elevated exists. The models used to construct the actual evolutionary explanation of the Gouldian finch polymorphism reveal a 'population fitness valley, [where] hawks are competent enough to take over nest sites from doves and to breed relatively successfully, but are still not successful enough to maintain the productivity of the population in the absence of the healthy population growth that doves provide' [Kokko, Griffith, and Pryke 2014: 4]. In these scenarios the hawkish red-heads really are their own worst enemy, and in the absence of dovish black-heads there is a collapse in the red-head population. The actual evolutionary explanation of why red-heads exist is essentially about the interaction between the two morphs.

It was the desire to fit cases of negative frequency-dependent selection into the theory of selected effect functions that led Griffiths [1993] to offer Defn 3 (see above). Griffiths fitted these cases by weakening the sense in which selected effect functions 'explain' traits. He required only that the selected effect function ascription convey some information that needs to be included in a complete model of the

evolution of the trait. But that information isolated from other key facts of the case may be a very poor explanation, as we have just seen.

It is also worth noting that in some evolutionary scenarios, Elevated simply does not have a selected effects function. Take the scenario where a population of red-headed individuals is invaded by a small number of black-headed individuals. Perhaps red-heads found a new population in unoccupied territory where they initially avoided the costs of conflict through sheer abundance of nest sites. As black-headed birds spread, the proportion of red-headed birds will decline until it reaches equilibrium. Throughout this process red-heads have lower (relative) reproductive fitness than black-heads. Defn 7 will not apply in this case because clause H.3 is unfulfilled. It was never the case in the actual evolution of the observed equilibrium frequencies that Elevated (red-headedness) had higher reproductive output than Reduced (black-headedness). So Elevated does not have a selected effect function if the population reached equilibrium by this trajectory.

This is not a 'counter-example'—this specific scenario is merely a dramatic version of what will happen in any scenario where a population overshoots the equilibrium frequency, by drift, founder effects, selective predation, or whatever. The point is not that Defn 7 fails to 'get it right' in one or more possible cases, but that Defn 7 builds in a fundamentally impoverished picture of how natural selection explains outcomes.

In negative frequency-dependent selection the selective environment changes as a result of selection: selection is an endogenous source of environmental heterogeneity. In the following section, we consider an extension of selected effects theory that has been formulated specifically to deal with heterogeneous environments: Millikan's concept of 'Normal' conditions. We begin by introducing 'Normal' conditions and then show why this idea does not allow selected effect function theorists to deal with selection in heterogeneous environments.

3.4 Normal Environments

The idea of the Normal (big 'N') is an important part of Millikan's theory of selected effect function: 'I capitalize *Normal* —to distinguish it from *normal* in the sense of *average*' (Millikan [1984: 34], italics in original). To be Normal is to be the way things were on actual, historical occasions on which a trait successfully performed its functions. Normal conditions are 'conditions in which [the *trait*] has historically been when it actually performed F—these conditions being uniform over as large a number of historical cases as possible' [Millikan 1984: 33]. A Normal explanation is one that explains how F is performed under these Normal conditions. A Normal character is a character state that features in a Normal explanation (see Appendix).

Millikan's Normal conditions are not the same thing as the normal environment the trait encountered in its evolution. The normal (small 'n') environment in that sense includes both environments in which the trait was historically successful and environments in which it was historically unsuccessful. It is both normal for rabbits to escape predators and normal for rabbits to be eaten. But for Millikan, only environments in which the rabbit escapes, and moreover escapes in the same way it escaped in the past, are Normal conditions in which the rabbit's anti-predator adaptations perform their Proper functions in accordance with a Normal explanation. Normal conditions are

the benign environments in which the trait can perform its function successfully in just the way it performed it in the evolutionary past.[4]

Millikan says that Normal conditions 'are the conditions to which the device that performs the Proper function is biologically adapted' [Millikan 1984: 34]. This suggests that the problems identified in sections 3.2 and 3.3 can be solved by focusing only on evolution in Normal environments, since these are the ones in which adaptation occurs. But rather than solving the problem, this is merely another way to reveal it. In the scenarios we have described it is not possible to explain the evolution of the trait by concentrating solely on how it evolves in Normal conditions. Whether a trait succeeds in evolutionary competition depends on its relative fitness in the expected environment, a weighted average of environments in which the trait occurs, not merely its relative fitness in environments to which it is particularly well suited.[5] 'Biologically adapted to' in Millikan's theory simply does not refer to evolutionary adaptation, it means something like 'has high fitness in'.

The idea of Normal explanation thus divides the environment into Normal patches, where the character state is advantageous, and abNormal patches, where the character state is not advantageous. The abNormal patches are ignored in the Normal explanation of how the character state affects fitness and hence in the selected effects function explanation of why we observe that character state in the target population (see Defn 6). This means that these selected effects explanations will often be poor explanations, since actual evolutionary models of selection in patchy environments must consider how the trait performs in both advantageous (Normal) and disadvantageous (abNormal) patches. We will now explain this using an example of bet-hedging in heterogeneous environments.

3.5 Heterogeneous Environments and Bet-hedging

Evolution in patchy environments, whether spatial or temporal, does not involve an evolutionary response to a single kind of patch independently of what is happening in other patches [Brandon 1990]. Organisms evolve to deal with the heterogeneity inherent in the environment, responding to a probability distribution over possible environments with which they may have to cope. Evolutionary responses to such patchy environments often take the form of traits that embody a 'bet-hedging' strategy. A bet-hedging strategy is one that minimizes variance in reproductive output

[4] Millikan needs her concept of Normal for two reasons. The first is to distinguish a trait performing its function 'Normally' (Proper functioning) from the trait performing the same function merely adventitiously. When a male superb fairy wren has an extra-pair copulation because the female wren remembers that he was the first local male to turn blue in spring, his blue head performs its Proper function in accordance with a Normal explanation. But suppose a male wren has an extra-pair copulation because the owner of an aviary picks the bluest wrens in an effort to breed males who are constitutively blue. According to Millikan, the blue trait does not have its effect as a Proper function in this case because it does not produce the effect in accordance with a Normal explanation.

The second role of Millikan's Normal is to explain how a trait can evolve when it almost always fails to perform its Proper function. Most sperm are unable to perform their Proper function of fertilization because the Normal condition of the presence of an egg is not met. Most sperm are in an abNormal environment [Millikan 1984: 34].

Reliance on this notion of Normality rather than the notion of a normal environment as that in which the trait evolved creates problems when selected effects theorists try to understand diseases as failure to perform Proper functions [Matthewson and Griffiths 2017; Bourrat and Griffiths forthcoming].

[5] This is the simplest case—sometimes a more complex statistic than the expectation is needed [Sober 2001].

[Olofsson, Ripa, and Jonzn 2009]. This aim can be achieved in a couple of ways. An organism can adopt a 'conservative' bet-hedging approach in which it follows the same low-risk strategy for all environments (e.g., laying a clutch of same-sized eggs). Even if this strategy is suboptimal for any given environment, it can minimize variance by avoiding particularly poor reproductive performance in unfavourable environments. On the other hand, an organism might adopt a 'diversified' bet-hedging approach in which it utilizes a wider range of strategies (e.g., laying eggs of different sizes). Despite this approach being risky for any specific strategy, it can lower reproductive variance because at least one strategy is likely to be well-matched for any given environment [Olofsson, Ripa, and Jonzn 2009].

When two strategies have the same arithmetic (additive) mean reproductive outputs, the effect of minimizing the variance in reproductive output for one of those two strategies is to increase its geometric (multiplicative) mean reproductive output. One implication of this is that, if environmental conditions vary enough to cause temporal fluctuation in a trait's effects on reproductive output, an invading mutant trait with a lower arithmetic mean reproductive output may spread on account of its geometric mean reproductive output being sufficiently larger than the resident's [Gillespie 1974; Beatty and Finsen 1989]. To illustrate, consider a simple example of conservative bet-hedging. Imagine two environments—one good and the other bad—that occur with equal probability. Suppose that an organism that does not hedge its bets has a relative reproductive output of 1 in the good environment and a relative reproductive output of 0.5 in the bad environment. That organism's arithmetic mean reproductive output is 1 * 0.5 + 0.5 * 0.5 = 0.75 and its geometric mean reproductive output is $\sqrt{1*0.5} = 0.707$. Consider now an organism employing conservative bet-hedging that has a reproductive output of 0.71 in both good and bad environments. Its arithmetic and geometric fitnesses are both 0.71. Since natural selection is a multiplicative or compounding process, the bet-hedging organism has a higher expected number of offspring (a good proxy of fitness), despite having a lower arithmetic expectation.

A canonical biological example of bet-hedging is seed dormancy in annual plants. For many annual plants, the environmental conditions for seedling establishment and reproduction can vary widely from year to year. Although a strategy in which all seeds germinate the following year is optimal when conditions are predictably good, such a strategy risks catastrophic failure if the following year is a drought. To get around this, some annual plants evolve a diversified bet-hedging strategy in which they produce seeds that germinate in different years (a so-called 'seed bank' [Childs, Metcalf, and Rees 2010]). We will now use an example of bet-hedging in plants to demonstrate how Normal explanations can be incongruent with actual evolutionary explanations.

Imagine that each year the environment is either good or drought. All seeds that germinate in a good year survive, but all seeds that germinate in a drought year die. Assume that trait C_i causes half the seeds to germinate in the following year and half the year after, and a trait C_j causes all seeds to germinate the following year. Furthermore, assume that the effect F of C_i is to decrease the proportion of seeds that germinate in drought, which it does more efficiently than C_j.

The diversified bet-hedging works by limiting losses in drought years so that ultimately more seeds can germinate in good years. If all seeds die once a drought hits, then there are no seeds left to germinate in the next good year. Crucially, to understand whether the bet-hedging trait will spread, we must account for drought *and* good

environments as well as their respective likelihoods (the 'small n' normal environment). Millikan's concept of Normal conditions, however, isolates one of these environments and attempts to explain a trait's existence solely based on its performance in this Normal environment.

In this example, we believe that Millikan would apply the idea of Normal explanation to individual seeds. For example, we might say that the Normal environment for a seed is the good one, and that the Normal explanation for how the seed performs its function of germination is that it germinates in the good environment. But this omits the key piece of explanatory information concerning the bet-hedging strategy, namely that the reason some seeds produced by a plant with the bet-hedging trait can germinate in a good environment is *because they did not germinate during drought years*. We could try reversing the explanation and assigning drought as the Normal environment. In this case, the function of the seed would be to not germinate in the drought environment, but this omits explanatory information about germinating in good years. Whichever choice we make, the Normal explanation only tells part of the story. Ultimately, to make sense of how traits evolve in heterogeneous environments, we must invoke both the Normal and abNormal environments (i.e., we must consider the 'small n' normal environment). Since a selected effects function only contains information about events in the (big-N) Normal environment, it will not be a good explanation of how the trait evolved.

In the next section we will draw together the threads from these three examples and identify the same underlying problem with selected effects explanations in each case.

4. When and Why Selected Effect Functions Do Not Explain

The special virtue of the selected effects theory of Proper functions is supposed to be that selected effects functions explain the traits that bear them. Male fairy wrens have blue heads because those heads attract mates. The selected effects theory is built upon the following picture of evolution: An organism with character C_i has selected effect function F because performing F gave it a reproductive advantage over an organism with character C_j. Repeated rounds of interaction between C_i and C_j individuals, typically resulting in a reproductive advantage for C_i, explain the existence of C_i. We have shown that this simple, intuitive picture often does not hold.

In sections 3.2 and 3.3, we described a hypothetical case with biting flies and an actual case with Gouldian finches where the interactions between C_i and C_j individuals change the selective environment. The selective environment includes not only biotic and abiotic factors external to the population but also the relative frequency or density of conspecific competitors. The process of selection is an endogenous source of heterogeneity in the selective environment. As selection changes the selective environment this feeds back and alters the payoffs for C_i and C_j, such that C_i's reproductive advantage may vanish or even reverse. The fate of the population depends on the dynamics that emerge from this feedback process. However, the definition of selected effects function means that a function ascription cannot convey this crucial information. It can only convey information about the subset of environments in which C_i is fitter than C_j.

In section 3.4, we examined a case of exogenous environmental heterogeneity: seed-banking in plants. The essential feature of an evolutionary explanation of this kind is that it considers both what happens in patches where a trait is fitter than its competitor

and in patches where it is less fit than in its competitor. Once again, the definition of selected effects function means that a function ascription cannot convey this crucial information. It can only convey information about the subset of environments in which C_i is fitter than its competitors. We saw that this limitation is also built into Millikan's concept of a Normal explanation.

The heart of the problem is that the selected effects theory assumes that the most explanatorily relevant information about the evolution of a trait is information about occasions when that focal trait was fitter than its rivals. This assumption is correct only when selection occurs in a uniform environment. In the common and important evolutionary scenarios we have described above the assumption is false. When there is a heterogeneous selective environment, occasions where the focal trait is less fit than its rivals are equally relevant. Whether the focal trait evolves depends on both. When environmental heterogeneity is produced endogenously, by the selection process itself, then the most explanatorily relevant information is the dynamics produced by the feedback between selection and relative fitnesses. That is what explains the fate of the population, and whether the focal trait evolves.

In essence, the selected effects definition of function takes the actual evolutionary explanation of a trait and selects the information that would explain the evolution of that trait if it evolved in an environment composed predominantly of patches where that focal trait is fitter than its rivals. When this assumption is correct, the selected effects function explains the trait. When it is false the selected effect function is either a poor explanation or the definition implies that the trait has no selected effect function, despite the fact that it evolved by natural selection. Equally importantly, in all the scenarios we have outlined selected effect functions are misleading explanations. They select just the information that would be relevant if the trait had evolved in a way that is not how it actually evolved. Presenting the function as an explanation suggests that this is how the trait evolved, which is misleading.

We suspect that many selected effects theorists will respond to our arguments by admitting that the theory has trouble with *complex* cases but insisting that it remains a good, simple starting point for thinking about biological function. But a simple model is only truly simple if it makes the right simplifications [Weisberg 2007]. Explanations in evolutionary biology typically take the form of verbal or mathematical models. All such models are of course abstractions or re-presentations of the phenomena studied. These abstractions are useful because they isolate elements that influence important features of the modelled system. In mathematical modelling, a well-designed model abstracts away extraneous aspects while retaining those that meaningfully influence the dynamics of the process under investigation. If a modeller removes important components, then their inferences will go astray. Consider a real-life example. In the early 1990s, models were used to support the hypothesis that uniparental (maternal) inheritance of mitochondria drove the evolution of mating types (primitive sex cells) [Hurst and Hamilton 1992; Hutson and Law 1993]. Twenty years later, another paper came to the opposite conclusion [Hadjivasiliou, Lane, *et al.* 2013]. What happened? The early models made a simplifying assumption that inadvertently removed negative frequency-dependent dynamics from the models, which altered the evolutionary dynamics of the traits and led to the mistaken inference [Hadjivasiliou, Lane, *et al.* 2013]. Considerations such as these are precisely why evolutionary biology is built upon quantitative principles. Models serve as a check on intuition, and they safeguard against minor conceptual errors that might lead our

inferences astray. The shortcomings exhibited by current definitions of selected effect function are just the kind that make for unreliable, potentially misleading generalizations from the simple case to the general case. The examples presented in section 5 will make this dramatically clear.

We should not overstate our case against selected effect function explanations. We do not deny that selected effect functions can explain traits when none of the evolutionary phenomena highlighted above are present. Nor do we hold that trait-environment feedback *always* poses a problem for selected effect function explanations. For example, negative frequency-dependent interactions might be weak, such that the fitness of trait C_i diminishes as it spreads but nevertheless always remains higher than the fitness of alternative C_j until C_i has displaced C_j. Likewise, positive frequency-dependent dynamics show how trait-environment feedback can cause C_i's fitness to increase as its frequency relative to C_j increases. These cases (and possibly others) are similar enough to the simple, intuitive evolutionary scenarios implicit in the definition of selected effect function as to not cause too many problems. But, as we have shown, numerous examples of evolutionary dynamics do not correspond even approximately to this simple model.

In the following section, we hope to motivate selected effects theorists to take our criticisms seriously by showing that they apply with particular force to the case many of them care about most, the evolution of representation.

5. Signalling and Selected Effects Function

The inadequacy of the selected effects theory is particularly marked in the domain of greatest interest of many selected effects theorists: the ascription of functions to *representations* used in signalling and communication. This was Millikan's primary aim in developing her theory of Proper function, and explaining representational states is also the central aim of Neander [2017] and of Shea [2007]. However, the evolution of representational states is the frequency-dependent process *par excellence*, because the best naturalistic models of how communication evolves are inescapably coevolutionary, in ways similar to the examples discussed in section 3 (see also Maynard-Smith and Harper [2003]; Scott-Phillips [2008]; Skyrms [2010]).

Signals succeed in representing states of the world because senders and receivers have adapted to each other's conditional behaviours. In the natural world, such information-bearing signalling systems evolve for a variety of competitive or cooperative reasons. They evolve between different species, such as in the stotting displays of Thompson's gazelles to their predators, and within the same species, such as with sexual signalling or alarm calls [Zahavi 1975; Grafen 1990; Skyrms 2010]; and even within the same organism, as in gene regulation [Calcott, Griffiths, and Pocheville 2018]. Following Lewis [1969], the same game-theoretic framework is applied to the evolution of language and other human conventions, modelling alignments of interest (and other details) in a variety of ways [Huttegger 2007]. Human representational communication likely evolved as early hominins became more reliant on conspecifics and divisions of labour, with language both driving and exploiting the evolution of more complex cognition, social structures, and associated incentives [Planer and Sterelny 2021]. Different mechanisms are involved in the evolution of signals in these different contexts, but they all fit the sender-receiver framework and all can be modelled using evolutionary game theory [Skyrms 2010; Huttegger and Zollman 2013; Brusse 2020].

But if these explanations are correct, then they cannot be re-expressed as function ascriptions via the selected effects theory. As in the case of the Gouldian finches, the evolution of signalling systems is highly frequency-dependent. The fitness or success of sender and receiver behaviours in any given context depends on the prevalence of complementary behaviours which would complete the informational circuit. There is no advantage in sending a veridical signal if no one is responding favourably to it, and no advantage in being poised to respond to signals that are not forthcoming or are poorly correlated with reality. Sender and receiver behaviours each constitute the selective environment of the other, and dynamically coevolve. There is an extensive body of simulation work that demonstrates this. The likelihood of signalling systems evolving is highly sensitive to the initial mix of behaviours, especially as various modelling idealizations are relaxed [Hofbauer and Huttegger 2008; Pawlowitsch 2008; Huttegger and Zollman 2013; Brusse and Bruner 2017]. And even in relatively simple signalling games there are also common, evolutionarily significant conditions under which the frequencies of sender and receiver behaviours permanently fluctuate, due to being sometimes advantageous and sometimes disadvantageous in relation to each other [Zollman, Bergstrom, and Huttegger 2013; Kane and Zollman 2015; Bruner, Brusse, and Kalkman 2017]. This is akin to the coevolutionary dynamic we see between hosts and parasites (as in the zebra stripes example), and can be expected in realistic cases when the interests of senders and receivers are not in perfect alignment.

The problems here for the selected effects theory run deeper though, because the rare special cases where it works best are also least desirable as standard bearers for a theory of communication and representation. There are indeed real-world examples where either signal or response behaviours 'hold still' for the other to adapt to, and we can identify a selected effects function which aligns with its evolutionary explanation. But these are all clear-cut cases of parasitic exploitation or manipulation. A famous example is the *Ophrys speculum* orchid's mimicry of the morphology and pheromones of a receptive female *Campsoscolia ciliate* wasp, manipulating the highly conserved mating strategy of the male wasps. In the human context, we could likewise attribute clear explanatory functions to lies told to the hopelessly gullible. But these are all unnatural, exceptional cases, in the sense that some factor requiring further explanation (mating strategies, maladaptive stupidity) acts to restrain receivers from adapting to the deception. And they are also decidedly awkward to serve as paradigm cases of communication; awkward at least for proponents of the selected effects theory.

6. Where Now for Selected Effects Theories of Proper Function?

The aspiration of selected effects theories is to identify the Proper functions of traits. Proper functions are the 'real' or 'true' functions of those traits and so can be used to assess whether traits are normal or pathological, intact or broken, well-functioning or dysfunctional. The claim that selected effects functions are Proper functions rests squarely on the claim that selected effects functions, alone among the many competing definitions of biological function, causally explain the existence of the traits that bear them. Some will dispute whether this special explanatory role gives these functions the normative force their proponents claim, but without the special explanatory role their claim to have that normative force is much diminished.

We have shown that in a wide range of realistic and important cases of natural selection, selected effect functions as currently defined do not explain the traits that bear them. In most cases, if a trait has a selected effect function, then that function conveys some explanatorily relevant information about the evolution of the trait. But in the cases we have described, that information is merely a fragment and very far from an explanation of the trait. One possible response for selected function theorists is to weaken the claim that selected effects functions explain traits to the claim that selected effects functions have some explanatory relevance to why traits exist. This, however, would undermine the claim that selected effects functions have a unique explanatory role, since rival definitions of function such as the 'organizational' theory also meet this new, weaker criterion [Mossio, Saborido, and Moreno 2009; Garson 2016]. The 'organizational function' of a trait is the role that trait plays in the ability of a system to maintain itself in a living state—this surely has some explanatory relevance to why the trait exists! Moreover, even the most marginal role in natural selection would confer a selected effects function on this new and weaker criterion. Selected effects theorists like to contrast the Proper function of the heart in circulating the blood with the mere effect of the heart in making a noise. But stethoscopes have been around for a long time and have saved many lives, and under the new interpretation this would give them some explanatory relevance and thus a selected effect function. Perhaps, *contra* Voltaire, noses have the proper function of supporting spectacles after all!

Alternatively, selected effects theorists may wish to maintain the strong link between selected effect function and evolutionary explanation by restricting the application of the theory to just those traits that evolved according to the simple, intuitive picture of evolution implicit in the theory. This more restrictive approach is no less problematic, however. Frequency-dependent selection, environmental heterogeneity, and more generally the coevolution of traits and their selective environments—which may include other organisms of the same or different species—are extremely widespread [Yoshimura and Clark 1993; Asmussen, Cartwright, and Spencer 2004; Brockhurst, Chapman, *et al.* 2014].[6] Moreover, we have shown that key domains of interest to selected effects function theorists, such as representational states in communication, violate the assumptions about evolution implicit in the theory. To limit the domain of selected effects function to traits where those assumptions hold would render the account of very limited interest to philosophy.

Selected effects function theorists have long recognized that where the theory is applicable depends on the nature of the selection process: 'Differences between biological and artifact functions arise because of the different nature of the selection process involved in the two cases' [Neander 1991a]. We have taken this a step further by showing that the theory as currently formulated applies to only some simple forms of natural selection. If the theory is to apply to a wider range of cases it will need to be reformulated to take account of the nature of selection in those cases. Confronting the complexities of real evolutionary explanations may allow for the development of a richer and sophisticated theory that reforges a strong link between function and evolutionary explanation. Reformulating the theory quantitatively may be required

[6] It will also be important to assess the viability of the selected effects theory of function in light of niche construction theory [Odling-Smee, Laland, and Feldman 2003] but here we appeal only to issues that arise within the most conventional, neo-Darwinian picture of evolution.

[Matthewson 2020]. But until this work is done, the idea that normative intuitions can be naturalized by appealing to evolution, influential in so many areas of philosophy, lacks a firm foundation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Asmussen, Marjorie A., Reed A. Cartwright, and Hamish G. Spencer 2004. Frequency-dependent Selection with Dominance: A Window onto the Behavior of the Mean Fitness, *Genetics* 167/1: 499–512.

Beatty, John, and Susan Finsen 1989. Rethinking the Propensity Interpretation: A Peek Inside Pandora's Box, in *What the Philosophy of Biology Is: Essays Dedicated to David Hull*, ed. Michael Ruse, Nijhoff International Philosophy Series, Dordrecht: Springer Netherlands: 17–30, https://doi.org/10.1007/978-94-009-1169-7_2

Birch, Jonathan 2016. Natural Selection and the Maximization of Fitness, *Biological Reviews* 91/3: 712–27, https://doi.org/10.1111/brv.12190

Bourrat, Pierrick, and Paul E. Griffiths (forthcoming). The Idea of Mismatch in Evolutionary Medicine, *The British Journal for the Philosophy of Science*, https://doi.org/10.1086/716543

Brandon, Robert N. 1990. *Adaptation and Environment*, Princeton, NJ: Princeton University Press.

Brockhurst, Michael A., Tracey Chapman, Kayla C. King, Judith E. Mank, Steve Paterson, and Gregory D.D. Hurst 2014. Running with the Red Queen: The Role of Biotic Conflicts in Evolution, *Proceedings of the Royal Society B: Biological Sciences* 281/1797: 20141382, https://doi.org/10.1098/rspb.2014.1382

Bronstein, Judith L., Ruben Alarcón, and Monica Geber 2006. The Evolution of Plant–Insect Mutualisms, *New Phytologist* 172/3: 412–28, https://doi.org/10.1111/j.1469-8137.2006.01864.x

Bruner, Justin P., Carl Brusse, and David Kalkman 2017. Cost, Expenditure and Vulnerability, *Biology and Philosophy* 32/3: 357–75, https://doi.org/10.1007/s10539-017-9563-5

Brusse, Carl 2020. Signaling Theories of Religion: Models and Explanation, *Religion, Brain & Behavior* 10/3: 272–91, https://doi.org/10.1080/2153599X.2019.1678514

Brusse, Carl, and Justin Bruner 2017. Responsiveness and Robustness in the David Lewis Signaling Game, *Philosophy of Science* 84/5: 1068–79, https://doi.org/10.1086/694156

Calcott, Brett, Paul E. Griffiths, and Arnaud Pocheville 2018. Signals That Make a Difference, *The British Journal for the Philosophy of Science* 71/1, https://doi.org/10.1093/bjps/axx022

Caro, Tim, Yvette Argueta, Emmanuelle Sophie Briolat, Joren Bruggink, Maurice Kasprowsky, Jai Lake, Matthew J. Mitchell, Sarah Richardson, and Martin How 2019. Benefits of Zebra Stripes: Behaviour of Tabanid Flies around Zebras and Horses, *PLOS ONE* 14/2: e0210831, https://doi.org/10.1371/journal.pone.0210831

Caro, Tim, Amanda Izzo, Robert C. Reiner, Hannah Walker, and Theodore Stankowich 2014. The Function of Zebra Stripes, *Nature Communications* 5/1: 1–10, https://doi.org/10.1038/ncomms4535

Childs, Dylan Z., C.J.E. Metcalf, and Mark Rees 2010. Evolutionary Bet-hedging in the Real World: Empirical Evidence and Challenges Revealed by Plants, *Proceedings of the Royal Society B: Biological Sciences* 277/1697: 3055–64, https://doi.org/10.1098/rspb.2010.0707

Fisher, Ronald Aylmer 1999. *The Genetical Theory of Natural Selection: A Complete Variorum Edition*, ed. J.H. Bennett, Oxford: Oxford University Press.

Garson, Justin 2016. *A Critical Overview of Biological Functions*, Dordrecht: Springer.

Garson, Justin 2019. *What Biological Functions Are and Why They Matter*, Cambridge University Press.

Garson, Justin 2022. *Madness: A Philosophical Exploration*, New York & Oxford: Oxford University Press.

Gillespie, J.H. 1974. Natural Selection for Within-Generation Variance in Offspring Number, *Genetics* 76/3: 601–6.

Godfrey-Smith, Peter 1994. A Modern History Theory of Functions, *Noûs* 28/3: 344–62, https://doi.org/10.2307/2216063
Grafen, Alan 1990. Biological Signals as Handicaps, *Journal of Theoretical Biology* 144/4: 517–46, https://doi.org/10.1016/S0022-5193(05)80088-8
Grafen, Alan 2014. The Formal Darwinism Project in Outline. *Biology and Philosophy* 29/2: 155–74, https://doi.org/10.1007/s10539-013-9414-y
Griffiths, Paul E. 1993. Functional Analysis and Proper Functions, *British Journal for Philosophy of Science* 44/3: 409–22, https://doi.org/10.1093/bjps/44.3.409
Hadjivasiliou, Zena, Nick Lane, Robert M. Seymour, and Andrew Pomiankowski 2013. Dynamics of Mitochondrial Inheritance in the Evolution of Binary Mating Types and Two Sexes, *Proceedings of the Royal Society B: Biological Sciences* 280/1769: 20131920, https://doi.org/10.1098/rspb.2013.1920
Hofbauer, Josef, and Simon M. Huttegger 2008. Feasibility of Communication in Binary Signaling Games, *Journal of Theoretical Biology* 254/4: 843–49, https://doi.org/10.1016/j.jtbi.2008.07.010
Hurst, Laurence D., and William D. Hamilton 1992. Cytoplasmic Fusion and the Nature of Sexes, *Proceedings of the Royal Society of London, Series B: Biological Sciences* 247/1320: 189–94, https://doi.org/10.1098/rspb.1992.0027
Hutson, Vivian, and Richard Law 1993. Four Steps to Two Sexes, *Proceedings: Biological Sciences* 253/1336: 43–51.
Huttegger, S.M., and K.J.S. Zollman 2013. Methodology in Biological Game Theory, *The British Journal for the Philosophy of Science* 64/3: 637–58, https://doi.org/10.1093/bjps/axs035
Huttegger, Simon M. 2007. Evolution and the Explanation of Meaning. *Philosophy of Science* 74/1: 1–27.
Jordan, A.M., Frances Lee-Jones, and B. Weitz 1962. The Natural Hosts of Tsetse Flies in Northern Nigeria, *Annals of Tropical Medicine & Parasitology* 56/4: 430–42, https://doi.org/10.1080/00034983.1962.11686144
Kane, Patrick, and Kevin J.S. Zollman 2015. An Evolutionary Comparison of the Handicap Principle and Hybrid Equilibrium Theories of Signaling, *PLOS ONE* 10/9: e0137271, https://doi.org/10.1371/journal.pone.0137271
Kokko, Hanna, Simon C. Griffith, and Sarah R. Pryke 2014. The Hawk–Dove Game in a Sexually Reproducing Species Explains a Colourful Polymorphism of an Endangered Bird, *Proceedings of the Royal Society B: Biological Sciences* 281/1793: 20141794, https://doi.org/10.1098/rspb.2014.1794
Lewis, David K. 1969. *Convention: A Philosophical Study*, Cambridge, MA: Harvard University Press.
Lorenz, Konrad Z. 1966. *On Aggression*, trans. Marjorie Kerr Wilson, New York: Harcourt, Brace and World.
Matthewson, John. 2020. Does Proper Function Come in Degrees? *Biology and Philosophy* 35/4: 39, https://doi.org/10.1007/s10539-020-09758-y
Matthewson, John, and Paul E. Griffiths 2017. Biological Criteria of Disease: Four Ways of Going Wrong, *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine* 42/4: 447–66, https://doi.org/10.1093/jmp/jhx004
Maynard-Smith, John, and David Harper 2003. *Animal Signals*, Oxford Series in Ecology and Evolution, New York: Oxford University Press.
Millikan, Ruth Garrett 1984. *Language, Thought, and Other Biological Categories: New Foundations for Realism*, MIT Press.
Millikan, Ruth Garrett 1989. In Defense of Proper Functions, *Philosophy of Science* 56/2: 288–302.
Millikan, Ruth Garrett 2017. *Beyond Concepts: Unicepts, Language, and Natural Information*, Oxford: Oxford University Press, https://doi.org/10.1093/oso/9780198717195.001.0001
Mossio, Matteo, Cristian Saborido, and Alvaro Moreno 2009. An Organizational Account of Biological Functions, *The British Journal for the Philosophy of Science* 60/4: 813–41, https://doi.org/10.1093/bjps/axp036
Neander, Karen 1983. Abnormal Psychobiology, LaTrobe University.
Neander, Karen 1991a. Functions as Selected Effects: The Conceptual Analyst's Defense, *Philosophy of Science* 58: 168–84.
Neander, Karen 1991b. The Teleological Notion of 'Function', *Australasian Journal of Philosophy* 69/4: 454–68.
Neander, Karen 1998. Mental Illness, Concept of, in *Encyclopedia of Philosophy*, ed. E. Craig, London: Routledge, http://www.rep.routledge.com/article/V021

Neander, Karen 2017. *A Mark of the Mental: In Defense of Informational Teleosemantics*, Life and Mind, Cambridge, MA: MIT Press.
Nicholson, Daniel J., and Richard Gawne 2015. Neither Logical Empiricism nor Vitalism, but Organicism: What the Philosophy of Biology Was, *History and Philosophy of the Life Sciences* 37/4: 345–81, https://doi.org/10.1007/s40656-015-0085-7
Odling-Smee, F. John, Kevin N. Laland, and Marcus W. Feldman 2003. *Niche Construction: The Neglected Process in Evolution*, Monographs in Population Biology, No. 37, Princeton: University Press.
Okasha, Samir 2018. *Agents and Goals in Evolution*, First Edition, Oxford: University Press.
Okasha, Samir, and Cedric Paternotte (eds) 2014. *The Formal Darwinism Project*, Special issue of Biology and Philosophy, Vol. 29, https://doi.org/10.1007/s10539-013-9424-9
Olofsson, Helen, Jrgen Ripa, and Niclas Jonzn 2009. Bet-hedging as an Evolutionary Game: The Trade-off Between Egg Size and Number, *Proceedings of the Royal Society B: Biological Sciences*, August, https://doi.org/10.1098/rspb.2009.0500
Pawlowitsch, Christina 2008. Why Evolution Does Not Always Lead to an Optimal Signaling System, *Games and Economic Behavior* 63/1: 203–26, https://doi.org/10.1016/j.geb.2007.08.009
Pittendrigh, Colin S. 1958. Adaptation, Natural Selection, and Behavior, *Behavior and Evolution* 390: 416.
Planer, Ronald J., and Kim Sterelny 2021. *From Signal to Symbol: The Evolution of Language*, Life and Mind: Philosophical Issues in Biology and Psychology, Cambridge, MA: MIT Press.
Price, George R. 1972. Fisher's 'Fundamental Theorem' Made Clear, *Annals of Human Genetics* 36: 129–40.
Railton, Peter 1981. Probability, Explanation, and Information, *Synthese* 48/2: 233–56, https://doi.org/10.1007/BF01063889
Scott-Phillips, T.C. 2008. Defining Biological Communication, *Journal of Evolutionary Biology* 21/2: 387–95, https://doi.org/10.1111/j.1420-9101.2007.01497.x
Servedio, Maria R., and Russell Lande 2003. Coevolution of an Avian Host and Its Parasitic Cuckoo, *Evolution* 57/5: 1164–75, https://doi.org/10.1111/j.0014-3820.2003.tb00325.x
Shea, Nicholas 2007. Consumers Need Information: Supplementing Teleosemantics with an Input Condition, *Philosophy and Phenomenological Research* 75/2: 404–35, https://doi.org/10.1111/j.1933-1592.2007.00082.x
Skyrms, Brian 2010. *Signals: Evolution, Learning, & Information*, Oxford; New York: Oxford University Press.
Sober, Elliott 2001. The Two Faces of Fitness, in *Thinking About Evolution: Historical, Philosophical, and Political Perspectives (Festschrift for Richard C. Lewontin Vol. 2)*, ed. R. Singh, Costas B. Krimbas, Diane B. Paul, and John Beatty, Cambridge: Cambridge University Press: 309–21.
Voltaire 1759. *Candide, Ou L'Optimisme*, France: Cramer.
Wakefield, Jerome C. 1992. The Concept of Mental Disorder: On the Boundary Between Biological Facts and Social Values, *American Psychologist* 47/3: 373.
Weisberg, Michael 2007. Three Kinds of Idealization. *The Journal of Philosophy* 104/12: 639–59.
Woodward, James, and Lauren Ross 2021. Scientific Explanation, in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Summer 2021, Metaphysics Research Lab, Stanford University, https://plato.stanford.edu/archives/sum2021/entries/scientific-explanation/
Wright, Larry 1973. Functions, in *Conceptual Issues in Evolutionary Biology*, *2nd Edition*, ed. Elliott Sober, Cambridge, MA: MIT Press: 27–47.
Yoshimura, Jin, and Colin W. Clark 1993. *Adaptation in Stochastic Environments*. Berlin: Springer.
Zahavi, Amotz 1975. Mate Selection—A Selection for a Handicap, *Journal of Theoretical Biology* 53/1: 205–14, https://doi.org/10.1016/0022-5193(75)90111-3
Zollman, Kevin J.S., Carl T. Bergstrom, and Simon M. Huttegger 2013. Between Cheap and Costly Signals: The Evolution of Partially Honest Communication, *Proceedings of the Royal Society B: Biological Sciences* 280/1750, https://doi.org/10.1098/rspb.2012.1878

Appendix: Millikan's Technical Terminology and Conventional Biological Terminology

Millikan's term	Our translation	Illustrative examples
First-order reproductively established family	Homologue of genes or genotypes. In molecular biology, orthologs, paralogs and xenologs will all count as 'families' of genes	16S rRNA gene, clonal line of asexual organisms
Higher-order reproductively established family	Homologue of phenotypes, including behavioural phenotypes	16S rRNA, femur, courtship display of the Great Crested Grebe
Device	A mechanism with a heterocatalytic function (it makes something that is not a copy of itself)	Machinery of DNA transcription and translation, stem cell, brains ('making behaviours')
Proper function	An effect for which a character is an adaptation (where 'adaptation' includes exaptations that have been subject to purifying selection in the recent past)	Proper function of 16SrRNA gene is to code for 16SrRNA; Proper functions of femur include strengthening hindlimb, haematopoesis, etc.; Proper function of courtship display is to create pair bond
Normal	How something was historically when the function was successfully performed, as in 'Normal environment', 'Normal character', 'Normal explanation'. See section 3.4 for discussion	Normal environment for a spermatozoon to perform the function of swimming in accordance with a Normal explanation (see Millikan [1984: 34]) includes acidic vaginal fluid, fibrinolysin enzymes, etc.