Function in the Light of Frequency-Dependent Selection

Samir Okasha

June 24, 2022

Abstract

Christie et al. (2022) claim that the SE theory of function is premised on a simplistic view of evolution. In complex evolutionary scenarios, in particular those involving frequency-dependent selection (FDS), the SE theory fails, they argue, since citing a trait's SE function does not serve to explain why the trait exists. I argue that where FDS leads to a stable equilibrium, at which all individuals' trait values constitute a "best response" to the rest of the population, the SE theory can be straightforwardly applied. However matters are different if the evolutionary dynamics do not take the population to a stable equilibrium.

1 Introduction

According to the "selected effect" (SE) theory, function-attributing statements in biology make implicit reference to evolutionary history. What makes it true that the function of the cactus's spines is to deter herbivores, for example, is the fact that natural selection led cacti to evolve spines because this did deter herbivores. The SE theory thus identifies a trait's function in a modern organism with something that the trait did in the organism's ancestors, in virtue of which the trait was selected over alternative traits.

The philosophical significance of the SE theory is that it promises to naturalize biological teleology by providing objective truth conditions for function attributions in biology. With the SE theory in hand, one could in principle determine the truth-value of any statement of the form “the function of trait T is X”, if one knew the relevant facts about evolutionary history. The SE theory thus demystifies talk of function and purpose in nature, which accounts for much of its appeal.

Proponents of the SE theory point to two further advantages (Garson 2016). Firstly, the theory captures the intuitive distinction between function and side-effect. (Attracting mates is the function of the insect’s mating display, attracting predators a side-effect.) Secondly, the SE theory accounts for the fact that citing something’s function serves to explain why it exists. (“Why do crabs have hard shells? To protect their innards.”) This fact can seem puzzling – how can something’s effect explain why it exists? – but is easily explained once function is taken as referring to selective history.

Christie et al. (2022) offer a powerful and original critique of the SE theory. They argue that the theory only works in simplified cases, since it requires “highly restrictive assumptions about evolutionary dynamics”. Their basic argument is that in order for a trait’s SE to explain the trait’s presence in a current population, assumptions are needed about the process of evolution that led to it. In particular, the selective environment must be assumed uniform and unchanging. Otherwise, to cite a trait’s selected effect, by way of trying to explain why the trait evolved, will be to give an incorrect explanation, they claim.

I agree with Christie et al. that the SE theory is tailored to simple evolutionary scenarios, and that its applicability to more complex scenarios has been insufficiently examined. I also agree that the SE theory requires an implicit assumption about the evolutionary dynamics. However, I do not think that Christie et al. have correctly identified what the required assumption is. I argue that the problem cases they raise for the SE theory can be handled fairly easily; but that there exist other problem cases, not treated in their paper, which pose a more serious problem.

My discussion centers on the issue of frequency-dependent selection (FDS). Negative FDS, as illustrated by the famous Hawk-Dove game of Maynard Smith (1982), is one of the three “complex scenarios” that Christie et al. use to press their objection to the SE theory. However their other two examples – Red Queen style co-evolution and bet-hedging in stochastic environments – in fact involve (implicit) frequency-dependence too. So by examining how the SE theory can handle FDS, we can cut straight to the heart of the issue.

2 Selection, Adaptation, and FDS

The SE theory identifies the function of a trait in a modern organism with something that the trait did in the organism's ancestors, in virtue of which the trait was selected. From this bare-bones formulation, it is clear that the SE theory fits naturally with the following simple picture of Darwinian evolution. In an ancestral population, there was heritable variation with respect to the trait in question; organisms with certain trait-variants enjoyed higher fitness than those with others; so over time, natural selection led the "best" variants to predominate. Thus organisms we see today, that are descended from this ancestral population, have traits that survived the selective filter.

Evolution of this sort will lead to the characteristic "fit" of organism to environment that Darwin marveled at, so long as certain conditions are met. These include: (i) sufficient variation in the ancestral population; (ii) heritability of the trait; (iii) absence of genetic constraints; (iv) selection strong enough to overcome other evolutionary forces; (v) constancy of the selective environment, and thus a fixed function mapping traits to fitness. Where these conditions are met, the evolutionary process is expected to produce organisms that are well adapted to the environment. That is, their evolved traits will be such as to maximize their fitness, relative to the alternative traits that were available for natural selection to choose from.

Where the conditions for Darwinian evolution to lead to adaptation are met, then today's organisms will display traits that they inherited from ancestors, and that have effects in virtue of which the traits were selected. The SE theory then identifies trait functions with these effects. Thus there is an intimate link between the biological concept of adaptation and the philosophical concept of an SE function. This link has been noted by a number of authors, including Bourrat (2021) and Sterelny and Griffiths (1999). Indeed the latter write: "[t]he [SE] functions of a biological trait are those effects for which it is an adaptation" (p.221).

Christie et al. argue, convincingly, that proponents of the SE theory have operated with this simple picture of evolution, which presupposes, *inter alia*, that selection is frequency-independent. In cases of FDS, by contrast, a trait's selective value vis-à-vis alternative traits depends on its relative frequency in the population. This means that the selective environment changes as the population evolves, and thus there is no fixed function mapping traits onto fitness; whether one trait is fitter than another depends on the population composition. Under FDS, evolution typically does not involve a monotonic increase of the optimal trait until fixation, as happens in the simplest models of Darwinian evolution.

Can the SE theory be extended to cover FDS? To address this question, let us ask what becomes of the concept of adaptation under FDS. Many evolutionary theorists argue, and I think they are right, that the concept of adaptation *can* apply in cases of FDS so long as it is suitably understood. The required understanding is this: each individual should exhibit a trait (or play a strategy) that maximizes its fitness *conditional on the trait distribution in the rest of the population.* Thus for example, Vincent and Brown (2005) write: "an adaptation is a strategy that maximizes individual fitness...given the circumstances, and these circumstances include the strategies and population sizes of others" (p.197). Another way to express this is to say that to be an adaptation, an individual's strategy must be a "best response" to the rest of the population.

This definition of adaptation may be justified on the grounds that an individual's fitness is always dependent on the environment as well as its own traits; and the environment may legitimately be taken to include the trait-distribution in the rest of the population. So adaptation in the sense of best-response is a straightforward generalization of the standard notion of adaptation that applies when selection is frequency-independent.

Adaptation in the sense of best-response is entailed by two of the standard solution concepts in evolutionary game theory, namely Nash equilibrium and evolutionary stable strategy (ESS). A strategy is a Nash equilibrium if it is a best response to itself; more precisely, if we let $w(y, x)$ denote the fitness of an individual with strategy y in a population where the resident strategy is x, then x^* is a Nash equilibrium strategy if and only if $w(y, x^*) \leq w(x^*, x^*)$ for all $y \in A$, where A is the set of available strategies. A strategy is an ESS if when fixed in the population, it is uninvasible by mutant strategies. ESS is a logically stronger condition than Nash equilibrium, and thus also implies that each individual is maximizing its fitness conditional on the rest of the population.

Where does this leave the SE theory? Given the close link between function and adaptation, this suggests that the SE theory should extend naturally to FDS, at least when the population evolves to a stable equilibrium. This last qualification is important for two reasons: (i) in some models of FDS, no evolutionarily stable equilibrium exists; (ii) when such an equilibrium does exist, the evolutionary dynamics will not necessarily lead there, even if non-selective forces are absent. These points suggest that FDS will not always give rise to trait functions; see section 5. But Christie et al. see a quite different problem with the application of the SE theory to FDS, that applies even in "nice" cases where evolutionary stability *is* attained. They illustrate their point with the Hawk-Dove game, the very example that gave birth to the concept of evolutionary stability.

3 SE functions in the Hawk-Dove game

The Hawk-Dove game describes a pairwise contest over a divisible resource. Organisms are of two types. Hawks are aggressive and ready to fight. Doves are submissive and seek to avoid a fight. Organisms are paired at random from a large population to play the game. If a Hawk is paired with a Dove, the Hawk wins the resource, earning a payoff of V while the Dove gets zero. If two Doves are paired, they split the resource fairly, each earning $V/2$. If two Hawks are paired, they fight; the loser gets zero while the winner gets $V - C$, where $-C$ is the cost of fighting; so each has expected payoff of $(V - C)/2$. Payoffs are summarized in Table 1.

	Hawk	Dove
Hawk	$(V - C)/2$, $(V - C)/2$	V, 0
Dove	(0, V)	($V/2$, $V/2$)

Table 1: **The Hawk-Dove game**

Assuming that payoffs are increments of biological fitness, that reproduction is asexual, and that strategies are hard-wired and perfectly inherited, this defines a simple evolutionary game. Assuming $V < C$, a Hawk does better than a Dove when playing against a Dove, but not when playing against a Hawk. So we have a case of FDS: the expected payoff to each type, and thus its fitness, depends on how common that type is in the population. Each type has a selective advantage when rare, so can spread, but this advantage is lessened as its frequency rises. A polymorphic equilibrium exists in which the frequency of Hawks in the population is $p = V/C$. At this equilibrium, which is evolutionarily stable, expected payoffs to both types are equal, and every individual's trait (i.e. type) constitutes a best response to the rest of the population.

Christie et al. make their argument with reference to a real-life case of the Hawk-Dove game discussed by Kokko et al. (2014), in which two morphs of the northern Australian Goudlian finch persist in wild populations. Red-headed finches, who correspond to Hawks, have elevated production of testosterone which leads them to compete aggressively for nesting cavities. Black-headed finches, who correspond to Doves, have reduced testosterone and are less aggressive, so do less well in competition for nesting cavities, but are able to invest more in parenting so can raise larger broods. Empirically, the equilibrium frequency of red to black is approximately one to three.

Christie et al. focus on the trait *Elevated*, i.e. having elevated testosterone, found in the red-headed finches. What is the function of this trait, they ask? They suggest, plausibly, that the SE theory would assign the function <*defeating rivals in contests over nesting cavities*>, this being the effect of the trait in virtue of which it enjoys a selective advantage (when it does). However – and this is their key point – citing this function is not a good answer to the question "why does *Elevated* exist?" (which they take to mean: "why does *Elevated* have the frequency that it does in the current population?") The correct answer to this question would certainly include the fact that higher testosterone enables red-heads to defeat their black-headed rivals in contests for nesting cavities, but it would include more than this. In particular, the facts that selection is frequency-dependent, that each morph is fitter than the other when rare, and that the two morphs have equal fitness at the equilibrium, are crucial parts of the explanation. Merely citing the SE function of *Elevated* does not explain why one quarter of finches in the population have this trait; and similarly for the alternative trait *Reduced* (i.e. reduced testosterone) found in the black-heads. To explain this, we need to include the fact that *Elevated* reduces an individual's fitness in some population compositions, not just the fact that it increases fitness in others.

Note that Christie et al.'s complaint with the SE theory is not that it leads to an incorrect assignment of function. They have no objection per se to the claim that <*defeating rivals in contests over nesting cavities*> is the function of *Elevated*. Rather, their point is that one of the main advantages of the SE theory is meant to be that, unlike other theories of function, it respects the truism that citing a trait's function serves to "explain why it exists". Christie et al. allow that this is true in simple cases of Darwinian evolution, where selection leads to fixation of the optimal trait. (Thus the function of gills is to enable fish to breath underwater, and this explains why all fish have gills.) But in cases of FDS, where the selective environment is not constant, this does not hold true. If someone asks why one quarter of finches have the *Elevated* trait, or why a stable polymorphism of any sort exists, telling them that the trait's function is <*defeating rivals in contests over nesting cavities*> is not a good answer; it omits crucial information.

Another way to see Christie et al.'s complaint is this. Imagine that, contrary to fact, red-headed finches drove out the black-heads thanks to their enhanced ability to compete for nesting cavities, so *Elevated* was fixed in the population. This would occur if the black-heads enjoyed no compensatory benefit in the form of increased brood size. Hence this would be a case of simple Darwinian evolution in which one trait evolves to fixation while the other is eliminated. Now the SE theory would assign the same function to *Elevated* as it does in the actual world – helping to defeate rivals in contests over nesting cavities. But the evolutionary explanation for why *Elevated* is found in the population is now quite different. So the actual world and the imagined world constitute distinct evolutionary scenarios, and yet the trait in question has the same SE function in each. So citing the trait's SE function does not tell us which of these scenarios obtains.

How should the SE theorist respond to this objection? Christie et al. are surely correct that in cases of FDS, citing a trait's SE function is not equivalent to giving a full explanation for why the trait "exists" (whether this means why it originally evolved, why it is maintained in the population today, or why it is has the actual frequency that it does).[1] But it remains true that a trait's SE function is an indispensable part of the explanation. The presence of the *Elevated* trait in the finch population could never be understood without knowing that it has the function of helping defeat rivals for nesting cavities. Moreover, even in cases of simple Darwinian evolution, where the selective environment is constant, citing a trait's SE function hardly amounts to giving a *full* evolutionary explanation for its presence today. The full

---

[1] Christie et al. claim that their critique applies whether the SE theory is formulated in terms of a trait's initial spread or its selective maintenance today. For the most part this is true, though one of their objections to the SE theory, discussed at the end of this section, only works against the former version.

evolutionary explanation of the presence of any trait in a population will have to include information about the ancestral population, the trait's genetic basis, the alternative traits available to selection, and the fitness landscape; the SE function of the trait is only ever part of the story, even if it may often be the most explanatorily salient part. There appears no difference in kind, in this respect, between cases of simple Darwinian evolution and cases of FDS.

Now Christie et al. do anticipate this response. But they argue that if the SE theorist opts for the weaker claim that a trait's SE function is explanatorily relevant to its existence, rather than explains its existence in toto, this sacrifices the signature advantage of the SE theory over rival theories of function. However, I think this is too quick. The main philosophical attraction of the SE theory, as I see it, is that it provides a fully naturalistic basis for claims about what organisms' traits are for, without appeal to final causes, intelligent designers, or what is "good" for the organism. And this aim is achieved so long as it possible, in principle, to read off a trait's SE function from the complete evolutionary history of the relevant population or species. *It is not necessary that we can do the reverse*, that is, read off the evolutionary history from the SE function. But Christie et al.'s objection essentially amounts to rueing the fact that the latter cannot be done.

To see this point, consider again the *Elevated* trait in both the actual world and the counterfactual scenario sketched above. The evolutionary history and dynamics are different in the two cases, yet the trait's SE function is the same in both. But this does not alter the fact that, from the complete evolutionary history of the relevant population – be it the actual history that led to stable polymorphism or the counterfactual history that led to fixation of *Elevated* – it is possible to *deduce* the function of *Elevated*, given the SE theory. (I assume that the complete evolutionary history includes information about the ecological causes of fitness differences). Therefore, the SE theory permits us to answer the sceptic who insists that function-talk in biology cannot be made properly objective, or is mere anthropomorphism. In short, the relation between complete evolutionary history and SE function can be many-one rather than one-one, compatibly with the aim of naturalizing teleology being achieved.

Christie et al. also make a secondary argument about FDS, that harks back to an argument of Griffiths (1993). They claim that in some evolutionary scenarios, the SE theory will be unable to assign *Elevated* a function at all. They consider a population initially monomorphic for red-heads that is invaded by a small number of black-heads, who increase in frequency until stable polymorphism is attained. In this case, *Elevated* declines in frequency from 100% to the equilibrium frequency. Since a trait's SE function is meant to be an effect in virtue of it was selected for, *Elevated* lacks an SE function, they argue, for is it continually selected against. I find this unconvincing. The assignment of an SE function to *Elevated* in this example is straightforward, if the SE theorist allows that a trait's being selected for can include selective maintenance as well as initial spread (as recommended by Godfrey-Smith (1994) and others). For at equilibrium, what maintains the *Elevated* trait in the population is precisely that it serves to defeat rivals in contests for nesting cavities, irrespective of whether the trait ever increased in frequency for this reason. If *Elevated* did not do this today, selection would quickly eliminate it from the population.

To summarize: in cases where FDS leads to stable equilibrium, as in the polymorphism of the Hawk-Dove game, adaptation *sensu* best-response does obtain, so evolved traits intuitively should have functions; and the SE theory correctly identifies those functions. Christie et. al.'s objection shows only that in such cases, a trait's SE function is not the full explanation for the trait's existence; but the requirement of full explanation is arguably unattainable even in simple cases, and is unnecessary for the aim of naturalizing teleology.

However there is a different objection, not discussed by Christie et al., that could be raised against the SE theory's handling of the Hawk-Dove game, and of FDS more generally. The objection stems from the use of mixed strategies.

4 Mixed strategies and SE functions

In the simple version of the Hawk-Dove game above, Hawk and Dove are hard-wired behaviours each encoded by a different genotype; this is the case that Christie et al. focus on. But as is well-known, there is an alternative version of the game in which what is hard-wired is an individual's strategy, defined as its probability of playing Hawk on a given encounter, denoted p. Thus there are two possible pure strategies – $p = 0$ and $p = 1$ – and a continuum of mixed strategies for natural selection to choose between. In this version, a stable equilibrium exists in which each individual uses the strategy $p^* = V/C$;

this strategy is the unique ESS of the game.

Since at the mixed-strategy equilibrium, each individual's evolved strategy is a best-response to the rest of the population, it follows that individuals are well-adapted to their (social) environment, so by our previous reasoning trait functions should exist, and we should expect that the SE theory applies. However, given how the SE theory is usually formulated in the literature, there is in fact a problem here.

To see the problem, note that the SE theory has it that "traits" are the bearers of functions: it attempts to say what it is for a given trait to have a particular function. But what is the relevant trait, when mixed strategies are allowed? Is it the action of playing Hawk, or the strategy of playing Hawk with a certain probability? (Similarly for Dove). Intuitively the former makes more sense, since playing Hawk, i.e. being aggressive in contests over a resource, is the sort of thing of which we can ask "why does the animal do that?" Moreover, playing Hawk has a definite effect – securing the entire resource without conflict when one's opponent plays Dove – in virtue of which playing Hawk is (sometimes) advantageous. By contrast, playing Hawk with probability p is a much less natural function-bearer. For although we can still ask: "why does the individual play Hawk with probability p (rather than q)?", we cannot easily point to an *effect* of playing Hawk with probability p. It is actions that have effects, earn payoffs and have fitness consequences; probability mixtures of actions do these things only derivatively, if at all.

Suppose, therefore, that we take actions to be the putative function-bearers in the mixed-strategy game. A problem then looms. For the SE theory, to recall, identifies the function of a trait in a modern organism with an effect of that trait in the organism's ancestors, which presupposes that the organism did actually inherit the trait from its ancestors.[2] But if the trait is an action that forms part of a mixed strategy, this presupposition may fail. For it is the strategy that is inherited, not the action. Conceivably, an organism who plays Hawk in a resource contest may have had ancestors who only ever played Dove. So if the function of playing Hawk is meant to be an effect that playing Hawk had in the organism's ancestors, the SE theory implies that playing Hawk, in this organism, has no function. This is an unwanted result.

---

[2] This presupposition is explicit in many formulations of the SE theory, for example Millikan (1984), who talks about "reproductive families of traits". It is also explicit in Christie et al.'s statement of the SE theory.

Is this a deep problem for the SE theory, or merely a technicality? On the one hand it highlights a serious point, which is that inheritance occurs at the level of the genotype (strategy), but we recognize function and adaptation at the level of the phenotype (trait, action). Typical formulations of the SE theory assume a simple relation between genotype and function-bearing trait, and thus talk directly about the inheritance of traits from ancestors; but this is an oversimplification, especially where the traits in question are behaviours. On the other hand, this example is more troubling for the letter of the SE theory than its spirit. Even if our Hawk-playing individual never had a direct ancestor that played Hawk, there must have been *some* Hawk-playing members of the ancestral population from which the individual comes, in whom the Hawk-playing trait had its fitness-enhancing effect. So it seems that a relatively minor modification of the SE theory can accommodate this sort of case. The required modification says that a trait's function, in a modern individual, is the effect in virtue of which it was selected in the ancestral population from which the individual derives, rather than just in its direct ancestors.

It is worth noting that a modification along these lines is necessary anyway, if the SE theory is to capture the phenomenon of adaptive phenotypic plasticity. Where a single genotype evolves to have different phenotypic effects in different environmental conditions, it is entirely possible that a given individual will express a phenotypic trait that is adaptive, and that intuitively has a function, but that none of its direct ancestors expressed. For example, in certain species of the freshwater crustacean *Daphnia*, juveniles exposed to chemical cues from predators develop neck-teeth at the back of the head (Christjani et al. 2016). Neck-teeth serve to reduce mortality from predation, so clearly have a function. But it need not be true that the presence of neck-teeth, in an individual crustacean today, has an effect in virtue of which the individual's ancestors gained a selective advantage. Conceivably, none of the individual's ancestors developed neck-teeth, for they may not have been exposed to the chemical cue during the embryonic period. If so, then the SE theory, in its official formulation, will not assign the trait a function. But the required modification is fairly obvious. For the neck-teeth trait does have an effect – reducing mortality from predation – in virtue of which *some* individuals bearing the genotype in question, in the ancestral population from which our contemporary individual derived, gained an advantage; even if they were not its direct ancestors.

To conclude: the mixed-strategy version of the Hawk-Dove game poses a prima facie problem for the SE theory, distinct from the problems that Christie et al. identify. However, the problem is not fatal; rather, it highlights a further simplification inherent in canonical formulations of the SE theory that has not been widely appreciated.[3] Relaxing the simplification poses no difficulty of principle, and is necessary anyway if the SE theory is to license function attribution in cases of adaptive phenotypic plasticity, as it should do.

5 Non-Equilibrium Dynamics

To this point we have argued that the SE theory generalizes easily to FDS. Our argument was that that the concepts of function and adaptation go hand-in-hand; and that at evolutionary stability we do observe adaptation in the sense of best-response, so SE functions can be identified and do play an explanatory role. We rejected as overly strong Christie et al.'s demand of a one-one correspondence between SE function and an answer to the question "why does the trait exist?". We identified a prima facie challenge posed by mixed strategies, but found that the SE theory, suitably re-formulated, could survive it.

Crucially, however, this whole line of argument presupposes that FDS will lead to evolutionary stability in the first place. It is well-known that this is not always true. In some evolutionary scenarios, no population state is evolutionarily stable; while in others, a stable equilibrium does exist but the evolutionary dynamics does not lead there. In such cases, then even if natural selection is the sole evolutionary force affecting the population's composition, it does not lead to phenotypic adaptation, and trait functions (intuitively) do not exist.

A scenario of the first sort is the Rock–Paper–Scissors game from evolutionary game theory. Pairs of individuals are drawn at random from a large population to play the game. There are three possible actions (and thus pure strategies). As in the children's game, Rock beats Scissors, Scissors beats Paper, and Paper beats Rock. An illustrative payoff matrix is given in Table 3.[4] As usual, reproduction is asexual, strategies breed true and payoffs are incremental fitness

---

[3] That SE theorists have not said much about phenotypic plasticity is noted by Griffiths and Matthewson (2018) p. 314.

[4] Note that this is a special case, since it incorporates a linearity assumption. The generic case is given by replacing each 1 with a variable α, where $0 < \alpha < 2$.

gains.

	Rock	Paper	Scissors
Rock	1	0	2
Paper	2	1	0
Scissors	0	2	1

Table 2: **The Rock–Paper–Scissors Game**

Rock–Paper–Scissors has a unique mixed-strategy Nash equilibrium in which individuals randomize equally between the three actions, i.e. choose Rock, Paper or Scissors with probability 1/3 each. However this equilibrium is not an ESS, since it is not stable against mutation (Hofbauer and Sigmund 1998). To see this, suppose that starting from the mixed-strategy Nash equilibrium, a mutant appears that always plays Rock. This mutant enjoys equal payoff as the residents, so can increase in frequency. This will then destablize the equilibrium, as once the mutant has risen in frequency, others do better if they play Paper rather than Scissors. Thus a mutant that always plays Paper will be able to invade. This then creates conditions in which playing Scissors is advantageous, and so-on. Thus an evolutionary cycle will occur in which the population moves indefinitely around a closed trajectory in the strategy space centred on the point (1/3 Rock, 1/3 Paper, 1/3 Scissors), but never settling down to any point (Figure 1).[5]

Note that in this example, natural selection does not lead to adaptation. At every point on the closed trajectory around which the population cycles, most individuals employ a strategy that is not a best response, so is not adaptive. The only point at which all individuals play a best response is where each strategy has a frequency of 1/3. But the evolutionary dynamics do not lead to this point; and if (perchance) the population were to find itself there, it would not stay there for long.

Such a scenario is genuinely troubling for the SE theory. To see why, consider the real-life case of Rock–Paper-Scissors reported by Sinervo and Lively (1996). In certain Californian populations of the side-blotched lizard *Uta Stansburiana*, there exists a polymorphism in

---

[5] Technically, the cycle may either be a limit cycle, i.e. attracting nearby cycles into its orbit, or a neutral oscillation, depending on the payoff matrix; see Nowak (2006) for explanation.

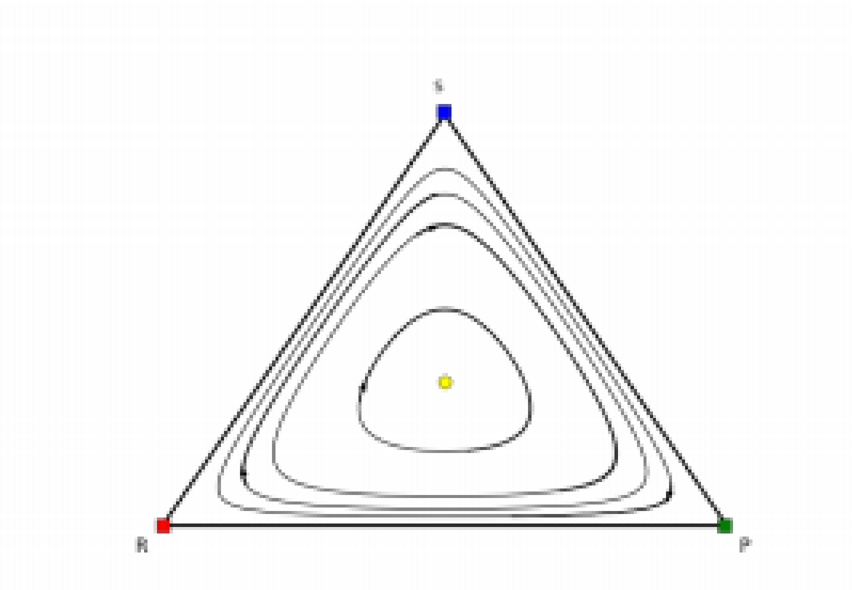

Figure 1: **Evolutionary Cycle in Rock-Paper-Scissors**

throat colour among males, associated with different strategies for territory defence and thus access to females. Orange-throated males are aggressive and defend large territories. Blue-throated males are less aggressive and defend smaller territories. Yellow-throated “sneaker” males, who look like females, do not defend their own territories but rather try to mate on others. An individual lizard’s fitness, as measured by its number of surviving offspring, depends not just on its own strategy but also on its neighbours. Empirically, it turns out that the sneaker strategy of the yellow males defeats the ultra-aggressive strategy of the orange males, which in turn is defeated by the less aggressive strategy of the blue males; however the blue strategy is defeated by the orange. Thus the payoff intransitivity characteristic of Rock–Paper–Scissors obtains. Consistently with this, Sinervo and Lively found that the observed frequency of the three male morphs oscillated periodically, with no morph able to reach fixation.

In this example, do the three male strategies have functions? Intuitively I think not. At any point in time, most of the lizards in the population employ a strategy that is not adaptive, so is selected against, and is thus declining in frequency. Functional attribution is therefore out of place. If a researcher, on discovering the polymorphism in male strategies, asks about the function of each strategy, their question has a false presupposition. Of course, we can give an evolutionary explanation for why each strategy has the frequency that it does at any point in time, and for why its frequency oscillates periodically. But not all evolutionary explanations are functional, even if natural selection is the only force at work.

The SE theory, however, *does* license functional attributions in this case. Take the orange males' ultra-aggressive strategy. In those parts of the cycle where the orange strategy increases in frequency, there will be some effect in virtue of which the strategy has a selective advantage – such as securing access to a territory containing many females. The fact that the orange strategy is often not advantageous doesn't alter the fact that in those historical periods when it was advantageous, it was so in virtue of a specific effect. So the SE theory implies that the ultra-agressive strategy in a contemporary male does have a function, and similarly for the other two strategies.

To me this seems wrong, given that at any time, most of the lizards use a strategy that does not enhance their fitness. Might an SE theorist simply dismiss this intuition? To do so would be to sever still further the connection between a trait's SE function and the answer to "why does the trait exist?" For it need not be true that the effect of the orange strategy in virtue of which it was selectively advantageous, when it was, even partly explains its presence (i.e. current frequency) in the population. Although this claim will plausibly be true at those points in the cycle when orange is increasing – given that the population's composition at time $t + 1$ is explained by its composition at time t and whatever led it to change over the interval – at just as many other points it will be untrue. At a point in the cycle where orange is decreasing, the effect of the orange strategy that explains its current frequency is an effect in virtue of which it is selected against, not selected for.

An SE theorist might try a different tack by imposing a "current utility" requirement, that is, requiring that a trait be currently fitness-enhancing (as well as or instead of in the past), in order to have an SE function. Such a move has sometimes been made to avoid having to attribute functions to vestigial traits, or to traits that are mismatched to the current environment but were advantageous in past environments (such as humans' sugar craving). But even if this move is well-motivated in those cases, the Rock-Paper-Scissors case is hardly analogous. And while this move would block some of the unwanted functional attributions in the lizard example, it would have the odd consequence that SE functions keep popping into and out of existence, as the population cycles round its trajectory.

The crux of the matter, I think, is that the SE theory is premised on the assumption that natural selection will lead to phenotypic adaptation. But where FDS leads to an evolutionary cycle, as in Rock-Paper-Scissors, the assumption fails. The SE theory then either over-attributes functions or – if modified in the direction of "current utility" – implies that functions are highly transitory properties.

The second type of case is where an evolutionarily stable equilibrium does exist but the population does not evolve to it. This case has been much studied in the framework of adaptive dynamics, which deals with FDS on continuous phenotypic traits (Diekmann 2004, Doebeli 2011, Metz 2012). Space does not permit discussion of this framework here. But the key point is that despite its name, adaptive dynamics shows that a population *may* evolve to an equilibrium at which individuals' trait-values maximize their fitness conditional on the rest of the population, but not necessarily. There are other dynamical possibilities too. For example, the population may be continually driven to an (evolutionarily unstable) point at which individual fitness is actually *minimized*; mutants will then be able to invade, but this causes the fitness landscape to change in such a way that the population is driven back to the fitness minimum![6] Again, this means that SE theory will attribute functions where intuitively there are none.

How often this phenomenon – failure of convergence to a stable equilibrium – arises in real biological populations is unknown. But like the evolutionary cycle example, it is an important theoretical possibility and highlights a real limitation of the SE theory of function. Since the SE theory equates a trait's having a function with the trait's having evolved by natural selection, the theory presupposes that the evolutionary dynamics are not of this sort. In short, the SE theory extends easily to cases of FDS where stable equilibrium results, but not to non-equilibrium cases.

6 Signalling and Teleosemantics

Christie et al. offer an interesting critique of teleosemantics, the project of using the SE theory to develop an account of represen-

---

[6] I discuss this phenomenon of "fitness minimization" in Okasha (2018), ch.3

tational content. They focus on the "signalling system" version of teleosemantics that uses Lewis-style sender-receiver games as a model (Stegmann 2009, Shea et al. 2018, Artiga 2020). Christie et al. note that "the evolution of signalling systems is highly frequency-dependent", since "sender and receiver behaviours each constitute the selective environment of the other, and dynamically co-evolve". This statement is certainly true. Since Christie et al. take themselves to have shown that the SE theory cannot handle FDS, they conclude that the project of trying to reduce signal content to SE function cannot succeed.

However this pessimistic conclusion is premature if our foregoing arguments are correct. Suppose that the signalling system evolves to a stable equilbrium at which perfect information transmission occurs. At such a point, senders' and receivers' strategies are best responses to each other, and hence maximize their fitness conditional on the rest of the population. In this case the SE theory will apply nicely, so the way is clear for the teleosemanticist to base an analysis of signal content on SE function.

To illustrate, consider the very simplest Lewis signalling game. Pairs of organisms, one in the sender and one in the receiver role, are chosen at random from a large population to play the game. The sender observes the state of nature (n_1 or n_2) then sends one of two possible signals (s_1 or s_2). The receiver attends to the signal then chooses one of two possible actions (a_1 or a_2). Action a_1 is appropriate to state of nature n_1, while a_2 is appropriate to n_2. Both sender and receiver get a payoff of 1 if the receiver chooses the action appropriate to the state, zero otherwise (so their interests are exactly aligned). A sender's strategy specifies which signal to send as a function of the state of nature; while a receiver's strategy specifies which action to take as a function of the signal received.

A *signalling equilibrium* occurs when senders use the strategy "s_1 if n_1, s_2 if n_2" and receivers use the strategy "a_1 if s_1, a_2 if s_2". This means that the receiver always chooses the appropriate action; so both sender and receiver payoff is maximized. Moreover, this signalling equilibrium is evolutionarily stable (for it is a *strict* Nash equilibrium), so if the population reaches it, mutant strategies will be unable to invade. SE functions are then easily assigned, to both strategies and the state-contingent actions that they specify. The function of the sender's equilibrium strategy ("s_1 if n_1, s_2 if n_2") is to induce the receiver to perform an appropriate action whatever the state of nature;

while the function of sending signal s_1, on those occasions when it is sent, is to induce the receiver to perform action a_1 when the state of nature is n_1. The teleosemanticist can then appeal to these functions to develop their account of the signal's representational content.

But will the signalling equilibrium actually evolve? The answer depends on model assumptions, but it turns out that in many cases, the answer is yes – which ties in with the empirical fact that evolved signalling systems do exist in nature. Huttegger et al. (2010) find that in the special case of the above model where the two states of nature are equiprobable, the signalling equilibrium always evolves, but that things are less clearcut when there are three or more states, signals and acts. In that case, the replicator dynamics sometimes leads to the signalling equilibrium, but sometimes to a "partial pooling" equilibrium at which information transmission is imperfect (since the sender uses the same signal in more than one state of nature). And in more complex signalling games, such as those where senders and receivers have opposing interests, non-equilibrium dynamics may arise (Huttegger et al. 2014).

The teleosemanticist project works fine if the signalling equilbrium evolves, but what about the other possibilities (pure pooling, partial pooling, and non-equilibrium dynamics)? These cases are not necessarily a problem for teleosemantics, simply because in most such cases signals will (intuitively) lack representational content anyway. The project of reducing content to SE function requires that wherever a signal has a content, it has an SE function; but this does not require that the signalling equilibrium need always evolve. So although non-equilibrium dynamics do pose a problem for the SE theory, as we have seen, this problem does not infect the application to teleosemantics.

7 Conclusion

The SE theory enjoys the status of orthodoxy in philosophy of biology, and Christie et al. deserve credit for challenging it. They are right to demand that the SE theorist explain how they can handle complex evolutionary scenarios involving frequency-dependence. I agree with Christie et al. that the SE theory requires an implicit assumption about the evolutionary dynamics, but not about what the assumption is. As I see it, the required assumption is simply that the evolutionary dynamics do not prevent selection giving rise to adaptation. And so long as we understood adaptation in the sense of best-response, which is a necessary condition of evolutionary stability, this means that the SE theory is potentially applicable wherever FDS leads to a stable equilibrium.

Christie et al. regard FDS as inherently problematic, even in "nice" cases were stable equilibrium results, since they regard the SE theorist as committed to the equivalence of "citing a trait's SE function" and "explaining why the trait exists". However, even in cases of simple Darwinian evolution, a trait's SE function is only a part of the explanation of why it exists. So the equivalence in question is not something that the SE theorist should be saddled with, and is not needed for the SE theory to achieve its fundamental aim of naturalizing teleology. For that aim, a many-one relation between evolutionary history and SE function is perfectly fine.

References

[1] Artiga, M. (2020) “Models, Information and Meaning”. *Studies in History and Philosophy of Science Part C* 82, 101284. https://doi.org/10.1093/bjps/axx031

[2] Bourrat, P. (2021) “Function, Persistence, and Selection: generalizing the selected-effect account of function adequately *Studies in History and Philosophy of Science Part A*, 90, 61–67. https://doi.org/10.1016/j.shpsa.2021.09.007

[3] Christie, J. R., Brusse, C., Bourrat, P., Takacs, P. and Griffiths, P. E. (2022) “Are Biological Traits Explained by their Selected Effect Function?”, *Australasian Philosophical Reviews* (forthcoming)

[4] Christjani, M., Fink, P. and von Elert, E. (2016) ”Phenotypic Plasticity in Three *Daphnia* Genotypes in Response to Predator Kairomone: evidence for an involvement of chitin deacetylases”. *Journal of Experimental Biology* 219 (11): 1697–1704. doi: https://doi.org/10.1242/jeb.133504

[5] Cooper, W. S. and Kaplan, R. H. (1982) “Adaptive “Coin-Flipping”: a decision-theoretic examination of natural selection for random individual variation”, *Journal of Theoretical Biology* 94, 1, 135–51.

[6] Diekmann, O. (2004) “A Beginner’s Guide to Adaptive Dynamics”. In Rudnicki, R. (ed.) *Mathematical Modelling of Population Dynamics* p. 47–86, Polish Academy of Sciences, Warsaw.

[7] Doebeli, M. (2011) *Adaptive Diversification.* Princeton: Princeton University Press.

[8] Garson, J. (2019) *What Biological Functions Are And Why They Matter.* Cambridge: Cambridge University Press.

[9] Godfrey-Smith, P. (1994) “A Modern History Theory of Functions”. *Noûs* 28 (3): 344–62. https://doi.org/10.2307/2216063.

[10] Grafen A. (1999) “Formal Darwinism, the Individual-as-Maximizing-Agent Analogy and Bet-Hedging”. *Proceedings of the Royal Society B: Biological Sciences* 266(1421), 799. https://doi.org/10.1098/rspb.1999.0708

[11] Griffiths, P. E. (1993) “Functional Analysis and Proper Functions”. *British Journal for Philosophy of Science* 44 (3): 409–22. https://doi.org/10/ddkn3x.

[12] Griffiths, P. E. and Matthewson, J. (2018) "Evolution, Dysfunction, and Disease: A Reappraisal". *British Journal for the Philosophy of Science* 69: 301–27. https://doi.org/10/ddkn3x.

[13] Hofbaeur, J. and Sigmund, K. (1998) *Evolutionary Games and Population Dynamics*. Cambridge: Cambridge University Press.

[14] Huttegger, S. M., Skyrms, B., Smead, R. and Zollman, K. J. S. (2010) "Evolutionary Dynamics of Lewis Signaling Games", *Synthese* 172(1), 177–191. http://www.jstor.org/stable/40496031

[15] Huttegger, S. M., Skyrms B., Tarrès P. and Wagner, E. (2014) "Some Dynamics of Signaling Games". *Proceedings of the National Academy of Sciences USA* 111 Suppl 3 :10873-80. doi: 10.1073/pnas.1400838111.

[16] Kokko, H., Griffith, S. C. and Pryke, S. R. (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.

[17] McNamara, J. M. (1995) "Implicit Frequency Dependence and Kin Selection in Fluctuating Environments." *Evolutionary Ecology* 9, 185–203. https://doi.org/10.1007/BF01237756

[18] Maynard Smith, J. (1982) *Evolution and The Theory of Games*. Cambridge: Cambridge University Press.

[19] Metz, J. A. J. (2012). "Adaptive Dynamics". In Hastings, A. and Gross, L. J. (eds.) *Encyclopedia of Theoretical Ecology*, p. 7–17. Oakland, CA: California University Press.

[20] Millikan, R. (1984) *Language, Thought and Other Biological Categories*. Cambridge MA: MIT Press.

[21] Neander, K. (1983) *Abnormal Psychobiology*. PhD thesis, La Trobe University.

[22] Neander, K. (2017) *A Mark of the Mental*. Cambridge MA: MIT Press.

[23] Nowak, M. (2006) *Evolutionary Dynamics*. Cambridge MA: Harvard University Press.

[24] Okasha, S. (2018) *Agents and Goals in Evolution*. Oxford: Oxford University Press.

[25] Shea, N. (2018) *Representation in Cognitive Science*. Oxford: Oxford University Press.

[26] Shea, N., Godfrey-Smith, P. and Cao, R. (2018) "Content in Simple Signalling Systems". *British Journal for the Philosophy of Science*, 69, 1009–1035. https://doi.org/10.1086/605820

[27] Sinervo, B. and Lively, C. (1996) "The Rock–Paper–Scissors Game and the Evolution of Alternative Male Strategies". *Nature* 380, 240–243. https://doi.org/10.1038/380240a0

[28] Skyrms, B. (2010) *Signals: Evolution, Learning and Information*. Oxford: Oxford University Press.

[29] Starrfelt J and Kokko H. (2012) Bet-Hedging–a triple trade-off between means, variances and correlations. *Biological Reviews of the Cambridge Philosophical Society* 87(3):742-55. doi: 10.1111/j.1469-185X.2012.00225.x

[30] Stegmann, U. E. (2009) A Consumer-Based Teleosemantics for Animal Signals. *Philosophy of Science*, 76(5), 864-875. https://doi.org/10.1086/605820

[31] Sterelny, K. and Griffiths, P. (1999) *Sex and Death*. Chicago: University of Chicago Press.

[32] Vincent, T. L. and Brown, J. S. (2005) *Evolutionary Game Theory, Natural Selection and Darwinian Dynamics*. Cambridge: Cambridge University Press.