Metascience (2019) 28:365–372
https://doi.org/10.1007/s11016-019-00452-8

SYMPOSIUM

The rationality of mother nature

Samir Okasha: Agents and goals in evolution. Oxford: Oxford University Press, 2018, xiv+254pp, £30.00 HB

Hannah Rubin[1]

Published online: 21 August 2019

Samir Okasha's *Agents and Goals in Evolution* is remarkably well argued and deep for a book that covers so much ground. Okasha clarifies and organizes many formerly disparate ways of using agential thinking in biology, discussing grand ideas with extraordinary clarity and subtly. He is careful to warn against latching on to intuitively appealing arguments and falling prey to the 'temptation to go further' than we are actually licensed to, reminding the reader what is a matter of empirical fact and where there are complicating factors. We are not always given a clean answer about when and why we are justified in using agential thinking, though one suspects it is because there is generally not one forthcoming. Instead, the reader gains an appreciation for how complicated matters really are, along with the means to make some sense out of the morass.

From the start, we are given helpful ways to think about what an agent *is*—an entity that simply does something, that performs intentional actions, that has goal-directed and autonomous behavior, or that behaves 'as if' it is trying to maximize a utility function—and which of these are useful for evolutionary thinking, either literally or metaphorically (or both) (12–15). We are given the unity-of-purpose criterion for treating an organism as an agent: its traits evolved because they contribute to a *single* overall goal. While minor deviations from complete unity-of-purpose are tolerable, without sufficient unity-of-purpose, we cannot treat the organism as if it were an entity with a goal or purpose it is trying to achieve (29–30). This criterion has intuitive appeal, and it becomes clear through repeated application throughout the book just how useful it is, e.g., in talking about group selection versus treating groups of organisms as agents (52–53).

Also in the first chapter, we are given a useful distinction between two types of agential thinking: type 1, viewing the organism as an agent, and type 2, viewing natural selection or 'mother nature' as an agent. These two types of agential thinking are returned to repeatedly and help the reader to contextualize and sort the many

✉ Hannah Rubin
hannahmrubin@gmail.com

[1] Department of Philosophy, University of Notre Dame, South Bend, IN, USA

disparate modes of evolutionary explanation discussed in the book. Okasha artfully brings together rational choice and (philosophy of) biology to make insightful connections. For instance, he justifies the objective rationale of the agential idiom of type 1 by noting a connection between rationality as an adaptation which allows an organism to respond to its environment within its lifetime and the evolutionary process which similarly allows organisms to respond to their environment, albeit between generations (160).

Throughout the book, Okasha argues that type 1 reasoning is 'broadly defensible,' while type 2 reasoning is 'more problematic' (16). However, I think ultimately his defense of this claim is not adequate. So, in the following, I will focus on the cluster of criticisms he offers of type 2 reasoning and attempt to extract the thread of an argument that is otherwise spread throughout different parts the book. My criticism of this argument will center on two points. First, I will argue that the 'natural selection as rational choice' motivation for type 2 reasoning is dismissed too quickly and that the initial description of mother nature's preferences does not match how an evolutionary game theorist would describe the situation. Second, the standards by which we are meant to judge the usefulness of agential reasoning seem to shift when we are considering type 1 and type 2 reasoning; I will argue that it is unclear why this shift occurs.

Mother nature's preferences?

Type 1 agential thinking treats the evolved organism or some other entity as if it were a rational agent 'trying' to achieve a goal, which is generally taken to be something like reproductive success. The focus here is on *products* of evolution—adaptations—and explanations of how adaptive traits conduce to the organism's goal. Type 2 agential thinking, by contrast, treats 'mother nature' as an agent choosing traits according to some goal. In this case, it is the *process* of natural selection which is treated as agent-like and the goal is something like adaptation or optimality (15–16). Okasha argues that type 1 thinking is a legitimate expression of adaptationist thinking, but type 2 thinking, while it can yield insight, is more likely to mislead than help.

Okasha introduces three ways one might try to motivate type 2 reasoning and then dismisses each (16–21, 39–41). I will focus on the first motivation, though it is likely what I say would overlap somewhat with a discussion of Dennett's idea that 'adaptationist theorizing in biology involves reading mother nature's mind' (39), the third motivation Okasha discusses.

The first way to motivate type 2 reasoning, the 'natural selection as rational choice' motivation, is to argue that 'there is a non-trivial parallel between the choices of a rational agent and the "choices" made by natural selection between alternative phenotypes or genotypes' (17). In particular:

> Whenever selection operates on a biological population, the set of alternative phenotypes can be ordered by their fitness in the current environment; if we

> wish, we can treat this as the preference order of a fictitious agent, namely mother nature. (17)

That is, mother nature chooses alternatives according to expected payoff just as a rational player would choose strategies according to their expected payoff. So, one can replace rationality with selection and self-interest with fitness, and connect concepts of equilibrium in rational deliberation, e.g., Nash equilibrium, with concepts of equilibrium in natural selection, e.g., evolutionary stable strategy (ESS).

These equilibrium concepts characterize a presumed end point of evolution; yet, as Okasha notes, when we look at how the evolutionary dynamics might play out, we get some results that seem to undermine the link between rationality and evolution. Evolution may not lead to an ESS if one exists, and may sometimes even cycle indefinitely, never settling on an equilibrium. He also mentions branching points, where payoffs are actually minimized. I will return to this case in the next section. While it is true that in a variety of cases a population might not reach an ESS, the implications for type 2 reasoning are far from straightforward.

One thing to note on this front is that, while an ESS may not always be reached, depending on modeling assumptions or intricacies of evolution in any real population, there are many results connecting concepts from rational choice to the dynamics of evolution. For instance, we can look at a couple results regarding the replicator dynamics, which is often used in evolutionary game theory as a simplistic representation of natural selection.[1] A Nash equilibrium is always an equilibrium, i.e., rest point, in the replicator dynamics, and if a population state is a weakly stable, i.e., Lyapunov stable, equilibrium in the replicator dynamics then it is a Nash equilibrium (Weibull 1997, Ch. 3). This does not mean that selection will always lead to a particular Nash equilibrium or ESS, but there is more to the connection between rational choice and the evolutionary process than what we get from characterizing presumed end points of evolution in terms of rationality inspired equilibrium concepts like ESS. One might also consider the process of rational deliberation and associated stability concepts (e.g., Skyrms 1990) in addition to equilibrium concepts from classical game theory. This is to say, it is not clear whether the results Okasha discusses 'partly undermine the rational choice/natural selection parallel' (18) so much as they warn against over-reliance on ESS analysis or other equilibrium concepts.

Further, Okasha argues that 'the complexities of evolutionary dynamics in game theoretical scenarios suggest that personifying natural selection is not always particularly apt, and may mislead' (18). It seems that his argument against the 'natural selection as rational choice' motivation for type 2 reasoning rests in large part on the contention that when selection is frequency dependent this means mother nature's

[1] The replicator dynamics describes the change in x_i, the frequency of trait i, with the following equation:

$$\dot{x} = x_i[w_i - \bar{w}]$$

where w_i is the fitness of trait i and $\bar{w}$ is the average fitness in the population, which both depend on the current population composition.

Table 1 Hawk–Dove game

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

preferences are changing, while we would expect a rational decision maker to have stable preferences (18). He says:

> It is as if mother nature continually chooses the phenotypic alternative she most prefers, only to find that her tastes have changed a moment later, so she needs to choose again; and this process can continue indefinitely. Viewed at a single point in time, mother nature's choices may resemble those of a rational agent; but viewed over time, she looks more like a fickle child. (18–19)

But why should we view this re-evaluation of phenotypes over time as mother nature's tastes changing?

We can maybe start to find an answer by looking at how Okasha makes agential talk precise, using rational choice theory, in Section 1.7. We can view an agent as having a complete and transitive preference order R over some finite set of options X, which they are faced with. These preferences give rise to a utility function u which assigns some real value to each of the options, and we can say an agent's goal is to maximize this utility function. Further, in the simple case where an organism's fitness depends only on its own phenotype, X can be thought of as the set of possible phenotypes (35). Under this picture, where preferences are over phenotypes, mother nature's preferences would look like they are changing because a phenotype with high fitness at one point in time need not have high fitness once the population composition changes.

Of course, Okasha notes that this is only the simple case and he complicates the picture in later chapters, but it is still perplexing that in dismissing the 'natural selection as rational choice' motivation for type 2 thinking, he equates mother nature's changing evaluation of phenotypes over time with her preferences changing over time. To see this, consider how the rational choice formalism would be applied to a biological case where an organism's fitness depends not only on its own phenotype, but on the phenotypes of other organisms as well. In this case, we would standardly say that X represents a set out outcomes, which can be thought of as resources, territory size, access to mates, or whatever it is that is relevant to the situation at hand. The fitness function then maps these outcomes to some measure of fitness. To determine which phenotype is preferred we have to consider its expected fitness, which takes into account how much fitness is attached to each of these outcomes and how likely it is that those outcomes occur, given the current population composition.

Let us consider a simple example where the fitness function is on outcomes not phenotypes, which Okasha also mentions when discussing frequency-dependent selection: the Hawk–Dove game (108). This game represents conflicts over resources and is summarized in Table 1. A 'hawk' refers to an aggressive animal that fights for a resource, and 'dove' refers to an animal that is unwilling to fight. A

hawk then always gets the contested resource when encountering a dove, but will split the resource and cost of entering into a conflict, e.g., injuries incurred, with another hawk. A dove will always surrender the resource to a hawk, but will split the resource peacefully with another dove. The resource and injuries are assigned some numerical value, V and C, respectively, representing their effects on fitness. Let us assume, as is often done, that $V < C$ so that hawk is not a dominant strategy.

We might say that mother nature prefers organisms with more resources and fewer injuries. How to achieve these goals, though, will depend on how many hawks and doves there are currently in a population. If there are too many hawks, she will prefer to have more doves because they avoid the all too common costly conflict. If there are too many doves, she will prefer to have more hawks because they usually get more of the resource and the chance of conflict is not too high. That is, her goals or preferences are not changing, but the current best means to achieve them are. And, at least in this case, at the population level the predicted outcome of the evolutionary process will be a polymorphic state consisting of a mixture of hawks and doves corresponding to the mixed strategy Nash equilibrium of the game.[2] This is an example where it seems that type 2 reasoning works well and, importantly, it does not look as though mother nature's tastes are changing over time. Mother nature orders strategies according to their fitness in the population and acts to increase her preferred strategies until there is the right balance in the population. That is, mother nature stops once an equilibrium is reached and so acts as if she has reached the goal she has been striving for.

A criterion for the usefulness of agential thinking?

Later on in the book, Section 4.3, we are given a more nuanced way to talk about agential thinking which includes the possibility of frequency dependence. Okasha notes that there is an ambiguity in how people talk about optimization: does it mean optimizing a function from individual trait value to fitness or does it 'include "best response," that is, having a trait that maximizes individual fitness conditional on the trait distribution in the rest of the population'? (108) He prefers to include the notion of best response; it seems like we should say that an ESS is optimal or adaptive even if it does not yield the highest possible average fitness in the population. At an ESS, every trait is a best response to itself and so each individual is maximizing their fitness given the population they find themselves in (81–82, 108).

But even allowing for this, he argues that type 2 reasoning can mislead us. In particular, he discusses branching points, as predicted by adaptive dynamics. This is an extremely interesting example, to which I cannot do justice here, but for details see pages 109–114 (for more details, see Kisdi 2001; Doebeli 2011). The basic idea is that if there are two possible habitats organisms can find themselves

[2] Assuming only selection is at work (e.g., using the replicator dynamics with standard assumptions about population structure, etc.) and that the possible phenotypes correspond to pure strategies of the game.

in and there is a trade-off between being well adapted to one habitat versus being well adapted to the other, then selection can lead toward a sort of generalist phenotype that is not particularly well adapted to either. Under certain assumptions about the trade-offs involved, this generalist phenotype will not be a best response to a population of all generalists; in fact, surprisingly, in a population of all generalists, having the generalist phenotype can actually minimize an organism's fitness.

Rather than diving into the technical details, I will only note a couple of important points. First, while evolution leads to a population composed of individuals whose phenotype is not optimal, no such optimal phenotype exists; we have found ourselves in a situation where there is no phenotype to aim for that maximizes the fitness of its bearers conditional on the population composition. This, I think, makes it complicated to try to draw conclusions regarding whether we can think of selection as aiming for optimal strategies or as choosing strategies like a rational agent would.

Second, in this case, Okasha sets discussions of the empirical significance of branching points to the side in order to make a conceptual point. It seems as if this discussion is meant to provide a counterexample to the generality of type 2 reasoning. In Chapter 1, as discussed above, Okasha argued that type 2 reasoning 'would really only apply well if Darwinian evolution had an inherent directional tendency, such as tending toward maximum fitness' (113) and here adaptive dynamics shows us there is no such directional tendency, even if we replace 'maximum fitness' with a notion of optimality that includes best response. In Chapter 1 and in many places throughout the book, type 2 reasoning is presented as something that is in principle limited to only cases where the environment is constant, which will be in danger of misleading us whenever there is frequency dependence. Yet, when he returns to talk about frequency dependence, he provides a counterexample without discussion of how likely it is that the counterexample will be empirically relevant. This seems to be in tension with the fact that type 1 reasoning is considered to be fruitful, but just how fruitful it is will depend on empirical facts.

This leads to the question: what is the relevant difference between type 1 and type 2 reasoning that makes type 1's usefulness a matter of empirical fact, while type 2 ought to be generally true in order to not be regarded with suspicion or considered 'inauspicious'? Type 1 reasoning is meant to be about a product of a process, whereas type 2 reasoning is about the process itself—so perhaps one could argue that type 1 reasoning is about the predictions of a theory and type 2 reasoning is about 'a general theoretical principle' (231) or how the theory functions, or something along similar lines. But it is not clear that we have to assume that type 2 reasoning captures a general fact about evolutionary theory in order for it to be useful.

To see why, let us return to a philosophical distinction that Okasha discusses in Chapter 1, between two ways of trying to explain a phenomenon:

> The first involves subsumption under law: showing that the phenomenon had to happen as a matter of nomological necessity. The second involves rationalization: showing that the phenomenon makes sense, or is appropriate to a particular end. (40)

He points out that adaptationist reasoning is typically of the second kind, a post hoc attempt to understand why organisms have certain traits rather than an attempt to show that their having these traits follows from certain laws. He argues though, that in adaptationist reasoning, appeals to mother nature are under-motivated because it is not mother nature that is the agent with goals and intentions, or who receives the benefits of the trait (40). In this discussion, he seems to allow for the possibility that type 2 reasoning can be used in these rationalizations, though, he argues type 1 better captures adaptationist reasoning, but in other places he seems to treat type 2 reasoning more like an attempt at subsumption under law, e.g., in taking a counterexample to show its inauspiciousness, as described above.

I think we can view type 2 reasoning as a way to show that a phenomenon makes sense. That is, we see a population with certain traits and want to explain why that is the case. We do so by explaining why an agent would be led by rational deliberation to the same equilibrium. Why might one think this is meaningfully different from observing an organism with a trait, then explaining why it has that trait by showing that it is the trait a rational agent would choose to have? One difference might be that type 2 reasoning is about the process, and we like to think of the process as law-like in some way. However, we might just observe that sometimes this, possibly law-like, process coincides with the process of rational deliberation, and when it does, post hoc rationalization by means of viewing mother nature as a rational agent provides us some deeper understanding. Perhaps one might claim the difference is that, in order to tell when our rationalization is empirically valid or not, we would need some access to history to tell us if selection did indeed occur as if it were a rational agent choosing between options. However, it seems like a similar drawback would apply to type 1 reasoning: in order to tell whether the trait we are explaining is indeed an adaption we need to know whether it was *selected for*, whether it had a causal role in the selection process (Sober 1984). For that we would similarly need access to the history of the population. It is not obvious that type 1 reasoning is useful as a rationalization, whereas type 2 is not.

Additionally, there are plenty of cases where it seems that type 2 reasoning works better or is more natural than type 1. Okasha discusses the evolution of preferences that cannot be considered rational from an expected utility point of view, but arise from the different ways selection takes into account aggregate versus idiosyncratic risk (218–220). For another such case, we can return to the previous discussion of the Hawk–Dove game. The equilibrium in this game is a mixture of hawks and doves, and it is not entirely clear how to explain the evolutionary products in terms of type 1 reasoning—for any particular organism, how can we explain why it is hawkish or dovish in terms of that trait contributing to its reproductive success? Both traits have equal fitness at the equilibrium. Rather, if we observe this population, we might instead use type 2 reasoning and explain the selection process leading to it in terms of a fictitious agent, mother nature, preferring hawks or doves depending on the population composition, as described in the previous section. My impression is that, in cases like these, type 2 reasoning makes sense of phenomenon, functioning as a post hoc rationalization, and appeals to an entity like mother nature are not under-motivated in comparison with type 1 reasoning.

The points discussed here do not speak in favor of type 2 over type 1 reasoning in any sort of general way. However, I do think they show that type 2 reasoning, and in particular the 'natural selection as rational choice' motivation for it, was dismissed too quickly. There are a number of interesting questions regarding this type of agential thinking which were left unaddressed in Okasha's book, but which the framework and tools provided in the book could shed light on.

Acknowledgements Thanks to Simon Huttegger and Mike Schneider for helpful feedback, and to Corey Dethier, Emanuele Ratti, Lynn Joy, Justin Bruner, Brian Skyrms, and Elliott Chen for joining me in discussions on the book.

References

Doebeli, M. 2011. *Adaptive diversification*. Princeton, NJ: Princeton University Press.

Kisdi, É. 2001. Long-term adaptive diversity in Levene-type models. *Evolutionary Ecology Research* 3 (6): 721–727.

Skyrms, B. 1990. *The dynamics of rational deliberation*. Cambridge: Harvard University Press.

Sober, E. 1984. *The nature of selection: Evolutionary theory in philosophical focus*. Chicago: University of Chicago Press.

Weibull, J.W. 1997. *Evolutionary game theory*. Washington: MIT Press.