

Review of Samir Okasha, *Agents and Goals in Evolution*.*

New York, Oxford University Press, 2018, xiv+254, \$40.00, ISBN: 9780198815082

Reviewed by Robert A. Wilson, La Trobe University, Melbourne, Australia.

Samir Okasha's focus in *Agents and Goals in Evolution* is a pervasive way of describing and explaining organismic traits, including behaviours, what he calls, with a nod to Godfrey-Smith (2009), *agential thinking*. This is to think of evolved entities—paradigmatically but not only organisms—as agents with certain kinds of interests or goals that are pursued through strategies. Here explaining why organisms and other evolved entities have the phenotypic traits they do involves drawing on a subset of intentional idioms that ascribe psychological states to organisms, as Okasha says, “usually in an extended or metaphorical sense” (p.230).

In *Genes and the Agents of Life* (2005), I referred to this aspect of agential thinking as a reliance on the *cognitive metaphor*, raising questions both of how the cognitive metaphor functions in the biological sciences more generally and why it appears there at all. Agential thinking in evolutionary contexts also involves the analogical transfer of the framework employing those idioms used primarily to understand economic agents in order to describe and explain adaptive behaviour. This is the framework developed by rational choice theory, a more systematic deployment of the cognitive metaphor in the biological and social sciences.

Okasha's overall discussion is admirably clear, focused, and integrative, despite ranging over literatures from evolutionary biology, game theory, rational choice theory, and the philosophy of economics. He brings much order to what can be a confusing set of issues and debates, his most important innovations being simple-sounding distinctions and claims that provide guidelines for how to think about agential thinking and the metaphorical ascriptions it makes to biological individuals.

To make clear the extended or metaphorical sense that Okasha has in mind, consider an example that he briefly discusses (p.26), cactus spines that have evolved to offer protection against herbivorous predators. Here we conceptualize the cactus as an agent with the goal of surviving across intergenerational time, and the spines as strategies (or the result of strategies) that the cactus adopts to achieve this goal. Agential thinking treats an evolved entity (a cactus) that does not literally have the kind of instrumental psychological profile underwriting means-end reasoning as if it were a rational agent. Such agential thinking is ubiquitous in evolutionary biology. We apply it to mobile and to sessile creatures, to multicellular and to unicellular organisms, and to rational and to mindless critters, and it has been especially effective in elevating the gene's-eye view of evolution popularized by Richard Dawkins.

Consider a putative case of genic agency. Worker ants in the red fire ant species *Solenopsis invicta* decapitate queens within their own species that are homozygous at a particular genetic locus, Gp-9. They do so because those workers possess a genotype at that locus different from that of such queens and they possess a way (likely by odor) of detecting their distinct genotype (Keller and Ross 1998). Here Gp-9 exemplifies what Dawkins (1989: ch.6) called the *green beard effect*, particularly a facultative, harm-inducing form of that effect (Gardner and West 2010, Queller 2011), whereby an allele both produces a trait and the recognition of that trait in others, facilitating populational spread of that trait. A sketch of this phenomenon employing agential thinking at the genic level ascribes to the selfish allele at Gp-9 the goal of eliminating its competitor at that locus. It does so by generating behaviours in worker ants to kill queens of the only surviving homozygote (since the other homozygotic queens are inviable) for that competitor allele at Gp-9. Here the greenbeard trait has a signature of having a particular allele (or perhaps allelic pair) at the Gp-9 locus. I shall return to whether Okasha himself would or should accept this explanation sketch as a legitimate form of agential thinking in evolutionary biology.

Okasha's first general aim is to explore the reliance in evolutionary biology on agential thinking, drawing on his conclusions from this first part of the book (chapters 1-2) to inform his discussion of fitness-maximization in Part II (chapters 3-5) and of rationality in Part III (chapters 6-7). There are two primary conclusions drawn in Part I.

The first is the significance of a simple-sounding distinction in what agential thinking applies to: whether to the *product* of natural selection, the evolved entities that it produces, or to the *process* of natural selection itself. Okasha calls the first of these "type 1" applications and the second "type 2", arguing that while the former are well-grounded and justified, the latter often over-reach and lead to confusions (cf. Martens 2019: 11-13). Okasha's second chief conclusion in Part I is that what he calls "unity of purpose" is a precondition for the successful and appropriate application of type 1 agential thinking. Just as there must be psychological unity for folk psychological and other intentional ascriptions to function explanatorily, so too must there be a unified biological perspective from which to apply agential thinking successfully. That unified purpose is the agential goal of enhancing fitness, typically understood as organismic fitness. Because organisms, rather than genes or groups, readily and generally exemplify unity of purpose, they are the primary examples of biological agents.

The innocent-sounding implication of these conclusions that end up having downstream oomph for Okasha is that type 1 and type 2 applications of agential thinking can come apart, despite adaptation and selection usually being regarded as closely related concepts. In Part II this separation implies that we can't presume that natural selection typically produces adaptations, for the models in which this presumption holds make one or more restrictive assumptions that are not always satisfied. Here Okasha discusses fitness maximization models in evolutionary theory, such as Wright's adaptive landscape models and Fisher's fundamental theorem of natural selection (chapter 3), and Grafen's more recent formal Darwinism project

(chapter 4), the topic of a recent special issue of *Biology and Philosophy* that Okasha co-edited (volume 29, 2014). Okasha's general conclusion is that there is no theoretical justification for the general claim that natural selection maximizes fitness, and so "adaptationism in biology must ultimately be justified on empirical rather than theoretical grounds" (p.114).

In Part III we see the piecemeal, empirical defence of type 1 agential thinking offered in Parts I and II manifest in Okasha's measured view of the relationship between adaptation and rationality. Although there is *prima facie* plausibility to the claim that rationality itself is an adaptation and so subject to type 1 agential thinking, Okasha takes seriously the evidence for the adaptive value of irrationality as reason for caution, the problem being more pronounced with the specifications in terms of rational choice theory or Bayesian decision theory. The norms contained in such specifications appear to be often violated in human behaviour, leading to a long history of questioning whether they accurately characterize human reasoning at all.

In Chapter 7, Okasha provides an informative, analytically-sharp overview of six phenomena discussed in the literature on rationality and in the philosophy of economics, each constituting what he calls a *parting of ways* between adaptation and rationality: cooperation, fairness, trust, intransitivity, risky choice, and inter-temporal choice. Some of these, such as cooperation in a class of prisoner's dilemma situations, suggest that in their choice of whether to cooperate or defect organisms attach significance to factors other than that of utility. Others, such as intransitive choices, suggest that utility maximization can't even get purchase as the goal of agential thinking. Okasha offers responses to all of these parting-of-ways arguments. He does so not so much in the spirit of bringing resolution to debates about each of these phenomena but more by way of providing a good sense of how to continue the dialectic.

This measured feel to much of *Agents and Goals in Evolution* makes it a useful survey of literature that can be technical in places (e.g., Grafen's work) and a guide for those in search of a deeper understanding of evolutionary biology's appropriation of models of rationality at

home in economics. Yet this quality also contributes to leaving larger issues in play either unresolved or resolved in ways that some may find unsatisfactory.

For example, one philosophical question that frames the discussion is “why agential thinking is so widespread” (p.4), casting this as the issue of whether agential thinking is *objective* in reflecting biological facts on the ground, or is instead a symptom of a human tendency to anthropomorphize, projecting human psychological characteristics onto things that do not have them. Okasha summarizes his view in concluding as being “a bit of both” (p.232): there are cases where agential thinking is underwritten by the phenotypic traits that biological agents possess and in virtue of satisfaction of the unity-of-purpose condition—the “yes” side to the objectivity question—and others in which this is not true—the “no” side.

Note that even in the cases where Okasha thinks agential thinking is objective, the relevant phenotypic traits are typically not themselves psychological. So Okasha’s position is not what Carrie Figdor (2018) has recently called *Literalism* about psychological predication, the view that nonhuman agents (e.g., whole critters, neurons, other cellular systems) literally have the psychological states ascribed to them. Rather, the objectivist strand to Okasha’s position implies that the cognitive metaphor is grounded in non-cognitive, biological facts.

This leaves open the question of just *why* biologists (and others) rely on the cognitive metaphor, i.e., why they make use of agential thinking at all. I have suggested the view that agential thinking functions positively to *crystallize agency*, bringing a focus on the causal agency of biological individuals by assimilating them to the paradigm of human agency (Wilson 2005, ch.4). In his critical treatment of the application of agential thinking to genes, Peter Godfrey-Smith (2009: 144-145) draws on Richard Francis’s (2004) idea of *Darwinian paranoia* to identify a negative effect of a more general psychological tendency that agential thinking rests on. Future engagement with either of these suggestions, as well as with Figdor’s

defence of Literalism, would enrich and perhaps sharpen Okasha's response to the overarching question about the pervasiveness of agential thinking.

For Okasha, agential thinking about evolutionary individuals serves three explanatory purposes (pp.22-28). Given the kind of realism he defends, it is thus grounded in one or more of three kinds of biological facts: the goal-directedness of behaviours and processes, the behavioural flexibility of biological individuals in response to their environment, and the exhibition of adaptations, "traits that confer a fitness advantage and have evolved for that reason" (p.25). When Okasha turns in chapter 2 to consider—surprisingly briefly—genes and groups, he points out that genes display neither goal-directedness nor behavioural flexibility. So the basis for legitimately employing agential thinking to genes rests solely on the exhibition of adaptations at the genic level. Consider now the earlier example of *Solenopsis invicta*, and whether Okasha should consider this a legitimate case of agential thinking or not. Here general features of his discussion pull in opposing directions.

On the one hand, Okasha generally follows proponents of the gene's-eye view in looking to cases of outlaw genes as paradigms for when we need to "downshift" to the genic level in order to understand natural selection and adaptation. These are cases, such as meiotic drive, that undermine the unity-of-purpose necessary at the organismic level for agential thinking to operate. Greenbeard genes, such as the posited "decapitator allele" Gp-9 in heterozygotic workers, are typically assimilated to such cases. Okasha supportively mentions a possible greenbeard gene (p.33), referencing David Haig's kinship theory of genomic imprinting, and appeals to a similar case—that of stingless bees in the *Melipona* genus—to illustrate how appeals to groups as agents may fail to meet the unity-of-purpose condition (p.58).

On the other hand, do we need to posit *genic* rather than organismic agency to understand this phenomenon? Suppose that the relevant worker ants both have heterozygosity

at Gp-9 and detect its absence in homozygotic juvenile queens, leading them to attack and kill those queens. This is a trait of whole organisms, and it's not clear why we can't employ agential thinking simply at the organismic level. A consideration that Okasha introduces that make him more circumspect about group agency is the availability of agential thinking at both the individual and genic levels. Yet extensions of the agential metaphor either to groups or to genes should rest on the necessity (rather than possibility) of frameshifting up or down.

Okasha issues caution about both genic and group agency, but treats the two differently. When it comes to groups, Okasha points to how unity of purpose can be undermined by the selfish interests of group members, leading him to claim that only in superorganism-like cases can groups be properly viewed as agents. Yet considerations of unity of purpose seem missing when it comes to attributions of agential thinking to genes, unless that unity is achieved simply because genes are held to have just one purpose: to create more copies of themselves, or (more accurately) to "outcompete its alleles" (p.45). Such thinning out of the concept of agential thinking seems problematic, however, in this context.

What of the unity-of-purpose condition? Okasha's discussion here (pp.28-34), given the rest of the book, sometimes comes across as strangely *a prioristic*. He says that an agent satisfies this condition when "its different traits have evolved because of their contributions to a *single* overall goal: enhancing the organism's fitness. Where this unity does not obtain, the organism cannot be regarded as agent-like, and treating it as such will impede, not facilitate, understanding of its features in adaptationist terms" (p.29). The first sentence suggests that if *any* of an organism's traits evolved for any other reason, unity of purpose is undermined, making it unlikely that any organisms satisfy the condition, given phenomena like meiotic drive and genomic imprinting. The modal strength of the second sentence makes it false, understood descriptively, and so is perhaps best seen as issuing conditional, normative advice.

That Okasha does not hold the view, which I have suggested above, that the unity of purpose condition precludes any level of disunity is made clear in his surrounding discussion. This puts more pressure on the condition being necessary for agential thinking and on its use in motivating shifts between levels at which such thinking can be applied. Intragenomic conflict, such as in the case of cytoplasmic male sterility in flowering plants (p.30), illustrates how an organismic failure to satisfy unity of purpose motivates genic agential thinking. Okasha points out that while we cannot think of male sterility as a goal of the plant, we can think of it as a goal of the mitochondrial genes that cause it, since it functions to promote their presence in future generations.

Yet why not think of the plant in richer agential terms, appealing not only to preferences and goals but a wider range of psychological states? Surely we can say that the plant *allows* its mitochondrial genes to cause male sterility, or *tolerates* parts of its bodily machinery to operate non-optimally, or *has decided on the whole* to put up with male sterility. And to undermine the shift from the organismic level we could appeal to the point that unity of purpose doesn't require complete unity. Much the same could be done with the greenbeard effect in the case of *Solenopsis invicta*. The point is that forced moves in the ascription or abandonment of agential thinking are harder to ensure than Okasha thinks, particularly once we both deepen agential thinking and recognize a softer reading of the unity-of-purpose condition.

The putative problems here may be brought out by turning to Okasha's endorsement of the claim that "unity-of-purpose is fundamental to human agency" (p.29), citing work at intersection of the philosophy of action and ethics. Our common understanding of human agency tolerates much disunity, this being part of the messiness of our full-blown psychological lives, with that understanding breaking down only in extreme cases. The scope and complexity of the messiness becomes more apparent once we thicken up our psychological descriptions beyond appeals simply to preferences and goals. Okasha's empirically-driven discussion in

Part III, particularly his discussion of parting-of-ways arguments against the adaptiveness of rationality, support this messier view of human psychology and rationality. And for this reason some of the claims he makes about human agency early in the book in support of the unity of purpose condition can seem *a prioristic*. Yet if unity of purpose is rarely an all-or-nothing matter, instead being something that both human agents (in psychology) and biological individuals (in evolutionary biology) possess to higher and lower degrees, then it is less clear just how the unity-of-purpose condition can support, limit, and shift applications of agential thinking.

That said, those looking for a clear overview of agential thinking in the evolutionary sciences will learn much from *Agents and Goals in Evolution*, and it will be important reading for philosophers of the biological and social sciences more generally.

* Thanks to Samir Okasha for feedback on a draft of this review.

References

- Dawkins, Richard, 1989 [1976], *The Selfish Gene*. New York: Oxford University Press.
- Figdor, Carrie, 2018, *Pieces of Mind: The Proper Domain of Psychological Predicates*. New York: Oxford University Press.
- Francis, Richard C., 2004, *Why Men Won't Ask for Directions: The Seductions of Sociobiology*. Princeton: Princeton University Press.
- Gardner, Andy, and Stuart A. West, 2010, "Greenbeards", *Evolution* 64(1): 25-38.
- Godfrey-Smith, Peter, 2009, *Darwinian Populations and Natural Selection*. New York: Oxford University Press.
- Keller, Laurent and Kenneth G. Ross, 1998, "Selfish Genes: A Green Beard in the Red Fire Ant", *Nature* 394: 573-575.

Martens, Johannes, 2019, “Inclusive Fitness as a Measure of Biological Utility”, *Philosophy of Science*, 86(1): 1-22.

Queller, David C., 2011, “Expanded Social Fitness and Hamilton’s Rule for Kin, Kith, and Kind”, *Proceedings of the National Academy of Sciences* 108(2): 10792-10799.

Wilson, Robert A., 2005, *Genes and the Agents of Life: The Individual in the Fragile Sciences: Biology*. New York: Cambridge University Press.