Biological Theory
https://doi.org/10.1007/s13752-023-00439-z

THEMATIC SECTION ARTICLE: CONCEPTS OF AGENCY

The Concept of Agent in Biology: Motivations and Meanings

Samir Okasha[1]

Received: 3 November 2022 / Accepted: 8 May 2023

Abstract
Biological agency has received much attention in recent philosophy of biology. But what is the motivation for introducing talk of agency into biology and what is meant by "agent"? Two distinct motivations can be discerned. The first is that thinking of organisms as agents helps to articulate what is distinctive about organisms vis-à-vis other biological entities. The second is that treating organisms as agent-like is a useful heuristic for understanding their evolved behavior. The concept of agent itself may be understood in at least four different ways: minimal agent, intelligent agent, rational agent, and intentional agent. Which understanding is most appropriate depends on which of the two motivations we are concerned with.

Keywords Agency · Concept of agent · Evolution · Organism

Introduction

Biological agency has become something of a hot topic in recent discussions in philosophy of biology and related fields. A number of authors have suggested that certain biological entities, typically organisms, are agent-like in certain respects, or can usefully be treated as agent-like for particular purposes (Clayton and Kaufmann 2006; Barandiaran et al. 2009; Skewes and Hooker 2009; Grafen 2014; Nicholson 2014; Arnellos and Moreno 2015; Moreno and Mossio 2015; Walsh 2015; Okasha 2018; Sultan et al. 2021). This is not an entirely new suggestion; indeed, in some ways it can be seen as a modern incarnation of the Aristotelian notion that living organisms are "causes of themselves." Here I do not wish to dwell on the historical origin of the biological agency concept, interesting though that would be, but rather to examine its meaning and motivation.

One obvious question to ask is what exactly we mean by the term "agent"? In addition to its various vernacular usages, the term is employed in semi-technical senses in fields including philosophy of mind, cognitive science, economics, and artificial intelligence/robotics. Presumably, those scholars who talk of agents and agency in a biological context intend some connection with these more established uses of the term—for their intention, I take it, is not to invest "biological agent" with a wholly new meaning by stipulative fiat. We need to consider, then, which of the various preexisting senses of the term "agent" are best suited to the biological discussion and why.

To address this question, we firstly need to consider what the *point* of talking about agents and agency in a biological context is supposed to be. After all, "agent" is not an entry likely to be found in the index of a textbook in any biological subfield, so a skeptic might reasonably ask why the notion should be thought important for philosophy of biology given that biology itself appears to have little use for it? This is a fair question. By way of answering it, I think that we can discern two distinct motivations for employing the concept of agent in philosophical reflections on biology, corresponding to two different intellectual projects.

Two Motivations

The first project is that of trying to understand what is distinctive about living organisms. The point here is not simply that organisms are quite unlike nonliving entities such as rocks, which no one will dispute, but also that there is something distinctive about organisms compared to biological entities at other hierarchical levels—such as organs, tissues, ecosystems, or species, for example. Organisms seem

✉ Samir Okasha
Samir.Okasha@bristol.ac.uk

[1] Department of Philosophy, University of Bristol, Bristol, UK

Published online: 28 June 2023

to possess a certain autonomy that other biological entities do not—a point emphasized by a number of theorists in the tradition stemming from the work of Varela (1979, 1981). Various concepts have been offered as ways of fleshing out this idea of organismic autonomy, including goal directedness, functional organization, emergence, self-maintenance, and individuality. Agency is another possible candidate for the job. Organisms certainly seem to exhibit some agent-like attributes, such as making choices, learning about the environment, and performing actions, that other biological entities do not. So the idea that what makes organisms distinctive is that they are agents—in some sense of that term—has a certain prima facie appeal.

This first motivation for talk of agency is sometimes accompanied by opposition to gene-centric or reductionistic views of evolution and development. The worry that the gene-centric paradigm either "leaves out the organism" altogether, or wrongly depicts all organismic activity as subservient to gene action, is of course familiar; this was a well-known critique of Richard Dawkins's early views and still resonates today. The organism-as-agent idea is seen by some authors as a potential corrective to gene-centrism, emphasizing as it does the power of organisms to make choices, overcome challenges, modify their environment, and shape their own fate. In a recent paper, Sultan et al. (2021) point to a number of empirical phenomena, including phenotypic plasticity, the evolution of novelties, and niche construction, which, they claim, are easily accommodated by an agential perspective but less easily by rival perspectives on evolution and development. Thus, according to this line of thought, the organism-as-agent notion is not only a defensible metaphysical position but is better suited to understanding aspects of biology.

The second motivation for invoking the concept of agency in a biological context is rather different. A number of authors have argued that it may be helpful to treat certain biological entities—paradigmatically individual organisms though in some cases other entities such as genes and groups—*as if* they were agents (in some sense of that term) for the purposes of understanding how evolution has shaped their phenotype, especially their behavior (cf. Okasha 2018, Chap. 1). By contrast with the first motivation, the suggestion here is not that the organisms in question really do have the attributes of agency—though they might—but rather that it may be heuristically valuable to treat them as such for the purposes of scientific understanding.[1]

This idea comes in various versions. Dennett (1983, 2013, 2017) makes a powerful case for the utility of taking the "intentional stance" towards organisms as a way of understanding the "rationale" of their evolved behavior. This means treating the organism as if it were a rational agent in pursuit of a goal, even if the organism's own powers of rational deliberation are limited or nonexistent. In part, Dennett is motivated by the observation that adaptive explanations can often be naturally couched in the intentional idiom. For example, we might explain a female rat's infanticidal behavior by saying that it knows that a malformed offspring will not survive and doesn't want to waste resources on it. When expressed this way, the adaptive explanation characterizes the rat's behavior as a rational response to the circumstance it finds itself in, akin to what the rat would do if it were a conscious agent able to engage in complex means-end reasoning. Other authors who defend the practice of "psychologizing" adaptive explanations in this way include Dawkins (1976) and Okasha (2018).

In a different though related vein, Grafen (2008, 2014) defends what he calls an "individual as maximizing agent analogy" as a way of formalizing and justifying the Darwinian idea that evolution will lead to well-adapted organisms. Grafen describes an evolved organism as an "agent" trying to solve an "optimization program" by wielding a phenotypic "instrument," that is, by choosing a phenotype from some "feasible set" so as to maximize the value of some "objective function." The objective function is intended as a measure of biological fitness, but its precise definition is left open. Grafen then argues that under many conditions, natural selection will lead organisms to solve the optimization program, so long as the objective function is suitably chosen. One interesting result is that where organisms interact socially, the objective function must be defined as inclusive fitness sensu Hamilton (1964) (see Grafen 2006). This (partly) supports the notion, widespread among Hamilton's contemporary followers, that by treating an organism as "trying" to maximize inclusive rather than personal fitness, we can thereby rationalize its social actions (Gardner 2009). Note that the suggestion here is not that organisms consciously aim to maximize inclusive fitness in their social interactions, but rather that they behave as if they do.

These two motivations for treating organisms as akin to agents are importantly different, though in principle they could complement each other. Let us call the first the "organisms-are-agents" thesis (OAT) and the second the "organism-as-agent" heuristic (OAH). Therefore, OAT says that organisms really do exhibit some or even all of the attributes of agency, while OAH says that it can be heuristically useful to treat organisms as if they were agents for certain intellectual purposes. Another difference is that OAH is explicitly evolutionary, in that the intellectual purpose in

[1] This heuristic should be sharply distinguished from heuristically treating the evolutionary process *itself* as an agent (which Sober (1998) calls the "heuristic of personification"). The distinction between these two heuristics corresponds to the distinction between "agential thinking" of type 1 and type 2 in Okasha (2018).

question is that of giving adaptive explanations; while OAT derives from what organisms do in the here and now, and so would in principle remain valid even if the organisms had been manufactured in a laboratory.

Concepts of Agency

Let us now return to the question of what we actually mean by the term "agent." In previous work, I have argued that we can identify four distinct concepts of agency that are employed in different contexts, scientific and everyday (Okasha 2018, Chap. 1). These are: the minimal concept, the intelligent agent concept from AI, the rational agent concept from economics, and the intentional agent concept from philosophy. Let us look briefly at each, before examining which is best suited for use in a biological context, bearing in mind the two different motivations we have identified—OAT and OAH—for introducing talk of agency into biology in the first place.

The minimal concept of agency is simply that of *doing* something, or behaving. As Dretske (1988) pointed out, there is an intuitive distinction between something that an entity does and something that happens to it. To borrow Dretske's example, consider the difference between a captive rat moving its paw towards a lever to get food and a scientist manipulating the rat's paw. The former is something that the rat does—a rat behavior—while the latter is something that is done to the rat. But what exactly is the difference? Dretske argues, plausibly, that it depends on whether the proximate cause of the movement is "internal" or "external" to the rat. Though there is clearly more to be said about what exactly the internal versus external contrast amounts to, Dretske's distinction is a real one and serves to define a minimal notion of agency.[2,3]

In this minimal sense, biological entities of many sorts, organisms and non-organisms alike, will count as agents. Cells divide, hearts beat, mitochondria make ATP, bacteria swim, and insect colonies swarm. In all these cases, the proximate cause of what occurs is internal to the entity—though external factors may be necessary background conditions. By contrast, an example of a biological entity that is not an agent in the minimal sense is a species. Although we talk about a species "going" extinct or "producing" a daughter species, the active voice here is misleading. Extinction is something that happens to a species when all its members die, and speciation something that happens to it when it is split into two lineages that then diverge; these are not species' behaviors. Thus, in the minimal sense, many but not all biological entities count as agents; and many non-biological entities will count as agents too.[4]

At the opposite end of the continuum from the minimal concept lies the intentional concept, which is how agency is usually understood in the philosophy of mind and action. In this sense, to be an agent is not just to behave but to *act*, where this means that the behavior is suitably caused by the agent's psychological states—beliefs, desires, and intentions. This notion of agency is intimately linked to the ability to engage in practical reasoning, or to work out what course of action one should take, and thus to means-end rationality. Whether any nonhuman organisms count as intentional agents, and if so which, is a much-debated question in the field of comparative cognition. Some authors believe that intentional agency is exclusive to *Homo sapiens*, others that it is found widely among vertebrates.[5] But whatever one's view on this, it is clear that most biological organisms are not intentional agents. A bacterium that swims towards an oxygen gradient does not do so because it believes that the oxygen lies upstream and wants to get to it; the correct explanation of its movement is nonpsychological. I take it that this would be agreed to by all parties, wherever exactly they wish to place the line between the intentional and the non-intentional.

The other two concepts of agency are more demanding than the minimal concept but less demanding than the intentional concept. In AI, an intelligent agent is defined as any entity that senses its environment and changes its behavior in response (Russell and Norvig 1995). Examples include simple control systems such as thermostats, software agents, and robots. The key attribute of agency in this sense is flexibility: an intelligent agent does not always do the same thing. The simplest intelligent agent is a "reflex agent" whose behavior depends only on its current percept; it thus implements a set of stimulus-response conditionals. More sophisticated intelligent agents can learn from experience; have a goal which they are trying to achieve; and in some cases can engage in search and forward planning in order to achieve their goal. The behavior of such agents is not just flexible but goal-directed (though this does not mean that the agent must have a mental representation of the goal).

Virtually all organisms, from microbes to animals, will qualify as intelligent agents in this sense, for they all exhibit adaptive responses to environmental stimuli. Think, for

[2] Dretske himself does not talk about agency; his aim is rather to elucidate the nature of behavior.

[3] What I am calling the minimal concept of agency is less restrictive than what some other authors have referred to as minimal agency, in particular Moreno and Mossio (2015) who include a requirement of goal-directedness in addition to internal causation (pp. 92–98); and Burge (2009) who seeks to exclude prokaryotes from counting as minimal agents.

[4] Dretske notes that the behavior of many artifacts and man-made machines has an internal proximate cause.

[5] See Carruthers (2006) and Andrews and Monsó (2021) for discussions of this issue.

example, of a bird migrating in response to a fall in ambient temperature, an insect following a pheromone trail, or a plant growing towards light. These behaviors are clearly both flexible (within certain bounds) and goal-directed. Sub-organismic entities such as cells and organelles will also count as intelligent agents, as too will some organismic subsystems, for instance, the vertebrate immune system. So although this concept of agency is more demanding than the minimal concept, requiring as it does flexible/goal-directed behavior rather than behavior period, it still casts quite a wide net.

A different notion of agency is found in the rational actor model in economics (Kreps 1988). In this field, a rational agent is defined as one whose choices maximize their utility. This is not intended as a psychological description but rather as a behavioral characterization, i.e., the maximization is strictly in an "as if" sense. (In effect, this is a de-psychologized version of the notion of agency as intentional action.) In the simplest case of choosing between certain options, so long as an agent's binary choices meet simple consistency conditions, such as transitivity, then the agent behaves as if they have a real-valued utility function on the options and always prefers the option with the highest utility. More complicated cases work in essentially the same way: the agent's choices or preferences are assumed to meet consistency conditions, which then imply the existence of a utility function with the agent behaving as if trying to maximize it. Thus agency in this sense means rational pursuit of a goal, which boils down to consistency of choice.

The rational actor model was designed to describe human choice behavior, but given its psychological neutrality it can apply to nonhumans too. Any organism that can choose between alternatives, or make decisions, is potentially a rational agent. Exactly what ``choice'' amounts to is not entirely obvious, but it is clear that many organisms with nervous systems, even simple ones, are capable of making within-lifetime choices:[6] butterflies choose what plant to oviposit on, birds choose who to mate with, and primates choose who to groom. Indeed, researchers have studied whether the choice behavior of rats, pigeons, and even insects satisfies the axioms of the rational actor model, such as transitivity of choice (Kagel et al. 1995; McFarland 2016). Agency in this sense is found quite widely in biology.

With these distinctions in mind, let us ask which of the concepts of agency (if any) is best suited for articulating (or understanding) the organism-as-agent thesis (OAT) and the organism-as-agent heuristic (OAH). In the case of OAT, I suggest that the intelligent agent concept—possibly strengthened somehow, for example to exclude artifacts and other non-biological objects—comes closest to capturing the intended meaning. The minimal agent concept seems clearly too weak; the intentional agent concept too demanding; and the economic agent concept not quite to the point, since choice-making is not the primary focus in most defenses of the OAT. By contrast, the attributes of behavioral flexibility and goal-orientation *do* capture an important part of what proponents of the OAT are getting at when they describe organisms as agent-like. However, since these attributes are also exhibited by sub-organismic entities, as noted above, they are on their own insufficient to capture what is distinctive about organisms vis-à-vis other biological entities. Given that part of the point of the OAT is precisely to capture organismic distinctiveness, it seems that none of our four agent concepts is up to the task without further supplementation. The underlying reason, I suggest, is that none of the four concepts captures the idea of autonomy, which intuitively is what distinguishes organisms from other flexible goal-directed systems in nature. This conclusion fits with the position of Moreno and Mossio (2015), who argue that agency is *one* central component of organismic autonomy rather than the whole story. But it conflicts with the view of Burge (2009), who seeks a notion of agency that applies to organisms but not to their subsystems. There may be such a notion, but it cannot be straightforwardly equated with any of the four concepts of agency identified above.

Turning to the OAH, I suggest that either the intentional or the economic agent concept is most germane. For the Dennettian version of the OAH, which involves treating an organism as agent-like in order to better understand its evolved behavior, the intentional agent concept seems to be implicated. The idea is precisely to exploit the analogy between rationalizing an organism's behavior in terms of its adaptive significance and rationalizing a person's behavior in terms of their beliefs and desires. Thus for this version of the OAH to work, the relevant notion of agent is intentional agent. By contrast, in the Grafen-style version of the OAH, the economic agent concept is closer to the mark. Grafen's notion that a well-adapted organism "tries" to maximize its objective function by wielding an "instrument" is conceptually close to the as-if utility maximization of the rational actor model, particularly if the organismic phenotype in question is behavioral (as Grafen is well aware). Indeed, in a sense what Grafen is doing in treating organisms as "maximizing agents" is precisely analogizing them to the utility-maximizing agents of economic theory, with utility replaced by a suitable measure of Darwinian fitness.

[6] Proponents of "basal cognition" would extend this claim to organisms that lack nervous systems, e.g., Lyon et al. (2021).

Conclusion

The notion that organisms are, or are akin to, or can usefully be treated as, agents, occupies center stage in a number of recent discussions in philosophy of biology and theoretical biology. The aim of this short article has been to clarify the motivation and meaning of this idea. Clearly there is more to be said; but I hope that the distinctions we have drawn between two motivations for introducing talk of agents into biology, and between four different concepts of agency, will prove useful in future discussions.

Acknowledgments Thanks to Lenny Moss, Matteo Mossio, and Richard Watson for their comments on an earlier draft of this article.

Funding This article is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement number 101018533).

Declarations

Competing Interests The author has no competing interests.

References

Andrews K, Monsó S (2021) Animal cognition. In: Zalta EN (ed) The Stanford encyclopedia of philosophy (spring 2021 edition). https://plato.stanford.edu/archives/spr2021/entries/cognition-animal/

Arnellos A, Moreno A (2015) Multicellular agency: an organizational view. Biol Philos 30(3):333–357

Barandiaran XE, Di Paolo E, Rohde M (2009) Defining agency: individuality, normativity, asymmetry, and spatio-temporality in action. Adapt Behav 17(5):367–386

Burge T (2009) Primitive agency and natural norms. Philos and Phenomenol Res 79(2):251–278

Carruthers P (2006) The architecture of the mind. Oxford University Press, Oxford

Clayton P, Kaufmann S (2006) On agency, emergence and organization. Biol Philos 21(4):501–521

Dawkins R (1976) The selfish gene. Oxford University Press, Oxford

Dennett DC (1983) Intentional systems in cognitive ethology. Behav Brain Sci 6:343–355

Dennett DC (2013) The evolution of reasons. In: (2014) Bashour B, Muller HD (eds) Contemporary philosophical naturalism and its implications. Routledge, New York/Abingdon, pp 47–62

Dennett DC (2017) From bacteria to Bach. Norton, London

Dretske F (1988) Explaining behavior: reasons in a world of causes. MIT Press, Cambridge

Gardner A (2009) Adaptation as organism design. Biol Lett 5:861–864

Grafen A (2006) Optimization of inclusive fitness. J Theor Biol 238:541–563

Grafen A (2008) The simplest formal argument for fitness optimization. J Genet 87:421–433

Grafen A (2014) The formal Darwinism project in outline. Biol Philos 29:155–174

Hamilton WD (1964) The genetical evolution of social behaviour. J Theor Biol 7:1–52

Kagel J, Battalio R, Green L (1995) Economic choice theory: an experimental analysis of animal behaviour. Cambridge University Press, Cambridge

Kreps DM (1988) Notes on the theory of choice. Westview Press, Boulder

Lyon P, Keijzer F, Arendt D, Levin M (2021) Reframing cognition: getting down to biological basics. Philoso Trans Royal Soc B 376:2019075020190750. https://doi.org/10.1098/rstb.2019.0750

McFarland D (2016) The biological bases of economic behaviour. Palgrave Macmillan, Basingstoke

Moreno A, Mossio M (2015) Biological autonomy: a philosophical and theoretical enquiry. Springer, Dordrecht

Nicholson DJ (2014) The return of the organism as a fundamental explanatory concept in biology. Philos Compass 9(5):347–359

Okasha S (2018) Agents and goals in evolution. Oxford University Press, Oxford

Russell SJ, Norvig P (1995) Artificial intelligence: a modern approach. Prentice Hall, Englewood Cliffs

Skewes JC, Hooker CA (2009) Bio-agency and the problem of action. Biol Philos 24:283–300. https://doi.org/10.1007/s10539-008-9135-9

Sober E (1998) Three differences between evolution and deliberation. In: Danielson P (ed) Modelling rationality: morality and evolution. Oxford University Press, Oxford, pp 408–422

Sultan SE, Moczek AP, Walsh D (2021) Bridging the explanatory gaps: what can we learn from a biological agency perspective? BioEssays 44(1):2100185. https://doi.org/10.1002/bies.202100185

Varela FJ (1979) Principles of biological autonomy. North Holland, New York

Varela FJ (1981) Autonomy and autopoiesis. In: Roth G, Schwegler H (eds) Self-organizing systems: an interdisciplinary approach. Campus Verlag, Frankfurt, pp 14–24

Walsh DM (2015) Organisms, agency, and evolution. Cambridge University Press, Cambridge