A response to ‘Biological agency: a concept without a research program’

Philip Ball
London, UK
p.ball@btinternet.com

Abstract: In critiquing the value of agency in biological research, DiFrisco and Gawne have tilted at a broad target. They say that "Proponents of the "agency perspective" on biological systems have claimed that agency is not explainable by physiological or developmental mechanisms, or by adaptation via natural selection", and that "proponents of the agency perspective have suggested that embracing agency is an alternative to molecular reductionism." Certainly, these would be bold claims, and DiFrisco and Gawne are correct to say that the threshold of proof for them should be high. However, these statements are simply not true in general. Some proponents of agency have claimed such things; others have not. It is the latter position that I wish to defend: the idea that agency should be regarded as a genuine property of living organisms *even though* it arises from molecular, physiological, developmental and evolutionary mechanisms, much as phase transitions are a real phenomenon even though they arise from (for example) molecular interactions. Agency is a regular property of living organisms, albeit a very general one. It is not inherently anti-reductionistic, nor is it an alternative to explanations of behaviour in adaptive terms. Their critique thus rests on a category error.

Keywords: agency, goal-directedness, mechanism, adaptation, natural selection, emergence

1. Introduction

At the heart of the critique of the notion of agency presented by DiFrisco and Gawne [DiFrisco & Gawne 2024] might be considered biology's central dilemma: do we consider living organisms to be a fundamentally different kind of matter to that which is non-living, or do we elide any such distinctions? To do the former does not entail evoking any kind of *élan vital*; it is simply to recognize that, in living organisms, the laws of physics and chemistry have given rise to an extremely unusual kind of non-equilibrium state of matter that can act in the world and, at least in the form of humans, can know itself. It is possible to study how such a state of matter arises and persists without pretending, perhaps out of some embarrassment at this remarkable state of affairs, that it is not real. Properly applied, agency can be considered a key property, possibly *the* key property, that sets living matter apart [Potter & Mitchell 2022, Mitchell 2023, Ball 2023a, Ball 2023b, Jaeger 2023]. It has the virtue that it seems at least conceivable that it might be operationalized: that we can make progress in identifying the specific and mechanistic features of an entity that enable it to exhibit genuine agential behaviour [Potter & Mitchell 2022].The notion of agency is rooted in observed behaviours and their causes, avoiding the seemingly somewhat arbitrary – or at least, the unresolved – issues that arise in definitions of life [Zimmer 2021].

DiFrisco and Gawne's critique is therefore rooted in a category error: they consider agency to be a theory advanced as an alternative explanation for biological phenomena to conventional reductionistic approaches, whereas, properly applied, it refers to a phenomenon that itself can and should be understood in mechanistic terms. Imagine if such a critique were applied to, say, the concept of photosynthesis. Is this not, in the end, an arbitrary delineation of some of the molecular processes that take place in certain cells, which in reality connect seamlessly to others? What (to paraphrase DiFrisco and Gawne) is explained by invoking the notion of photosynthesis that cannot be equally explained by considering light absorption by particular molecules and consequent electron transfers that result in water-splitting reactions, and so forth? One can go further, and question the usefulness of identifying any higher-order process, whether in living systems or not: what is explained by suggesting that ice undergoes a melting phase transition, rather than simply noting that at a particular temperature the hydrogen bonds between water molecules are overwhelmed by

thermal fluctuations? In other words, agency is a label given to an emergent phenomenon, and in this respect it is no different to countless other scientific concepts.

2. Where are agents to be found?

DiFrisco & Gawne illustrate their critique with a "phage and bacterium" version of the classic Heider & Simmel animation [Heidel & Simmel 1944], implying that attributions of agency in such cases are an example of the innate propensity of humans to project purpose, intention and indeed psychology onto situations where they do not exist. But in fact their invocation of Heider and Simmel's work does the opposite. Their implication is that, in that work, we falsely attribute agency to mere geometric shapes because we anthropomorphize their movements. But such an attribution is in fact not false at all, except for the location of the agency. Heider and Simmel chose complex trajectories for the shapes specifically to suggest an agential narrative, and we correctly infer that this is so, but are apt (without careful reflection or knowledge of how the animation was created) to mislocate the source. Or rather – for observers do not truly somehow suspect that the objects are living beings – Heider and Simmel's point was that we readily construct diverse narratives for such a circumstance, complete with associations of personality and so forth.

The point of the Heider-Simmel experiment was not, then, to show "how easily attributions of intentionality are elicited in the human visual processing system", but rather, to explore the nature and diversity of narrative construction in a system *explicitly and consciously contrived* to elicit such responses. Heider and Simmel's description of the animation they devised explicitly acknowledges the storytelling component of the choices they made in constructing it. One might reasonably argue that the experiment shows how good humans are at detecting agential behaviour even when it is given highly abstract form.

Suppose now one were to observe those same movements in simple inorganic "active particles" [Bechinger et al. 2016]. That would be remarkable, and would demand explanation. It is hard to imagine how such an explanation could be furnished by simple laws of attraction and repulsion between the particles; we would surely find, in such a case, that the particles have complex internal degrees of freedom, feedback

mechanisms between sensing and propulsion, and so forth. It would be not just reasonable but surely correct to say that, barring some bizarre coincidence in which arbitrary and independent particle trajectories happened to coincide so as to give the impression of following or chasing, there is goal-directedness here. (Dynamics like those of the Heider/Simmel shapes have, I would posit, *never* been observed in a situation that lacks goal-directedness – even if the goals are themselves not immediately obvious.) We could then reasonably wonder where that goal-directedness comes from. We would not need to impute any conscious intent to the entities executing those goals – the particles could in fact be nanobots programmed for such behaviour, say.

The same applies for the phase and bacterium: this is no less a pursuit with a goal than is the fox and the hare. The difference is that one involves rather complex cognition, while the other involves much simpler systems-scale feedbacks between sensing and motility. While acknowledging the philosophical arguments about the notion of function in biology [Mossio et al. 2009, Roux 2014], there is nothing remarkable about saying that the function of the phage is to detect and consume bacteria.

But why then invoke agency in an account of such processes? Why is it not sufficient to say that such pursuit is adaptive behaviour: that foxes have adapted to pursue hares, and hares to flee from foxes? Yet these are not alternative explanations. The latter statement is obviously true. It is also predictively barren. It will tell us that a fox will chase hares, and hares will run from the fox. It says nothing about their trajectories, their strategies, the inherent probabilities of capture and escape, and so forth. All of those things require more consideration of developmental and physiological details, but also of historical contingencies. How well fed are the animals? What experiences have affected their degree of alertness, fear, determination? What determines whether a hare runs this way or that, when it turns, whether it can and will exploit the terrain in some way, and so forth. In the end, an explanation for how a particular chase plays out demands an extraordinary level of detail, and indeed more understanding of, say, neurological processing than we currently possess. Adaptation by natural selection is silent about all of that, because it is simply not an explanation of the phenomenon but the precondition for it.

Agency is not, then, an alternative to an "explanation by natural selection"; it does different work. Agency might be considered the capacity that (in this instance) a hare displays in its escape behaviour. Of course one can attempt to take that behaviour apart: to consider, for example, the cognitive processes that lead to the hare suddenly switching direction in an evasive maneuver. Presumably any given instance of such switching might involve neural integration of great deal of information: Have I switched direction recently? How close is my pursuer? What is the terrain like here? How tired am I? By recognizing the pursuit of hare by fox, and of bacterium by phage, both as agential phenomena, we can start to pose questions that do not dissolve into the ephemeral minutiae of a specific situation. What capacities are involved in such decisions? How do they depend on sensory modalities? How are environmental affordances evaluated and processed? And perhaps in particular: how is the goal-directedness created and sustained throughout this behaviour? None of this is in any sense "anti-reductionistic".

3. Where do goals come from?

Goal-directedness lies at the root of this discussion. DiFrisco and Gawne do not challenge the idea that goal-directedness exists in biology – and after all, this is surely not seriously in question. Mayr [Mayr 2004] attests that "purposive behaviour that is clearly goal-directed is widespread among animals, particularly among mammals and birds", while Monod [Monod 1972] goes further: "One of the fundamental characteristics common to all living beings without exception... [is that they are] objects endowed with a purpose or project."

The issue then is about the origin of goal-directedness. DiFrisco and Gawne say it "is a product of natural selection rather than the inherent agency of organisms". Again, this is a category error. We should say that it is natural selection – or rather, all the mechanisms of the evolutionary process – that awards organisms inherent agency and goal-directedness, just as it awards them inherent morphology and developmental potential. "The capacity for goal-directed behavior" is not "explained" by Darwinian natural selection acting in populations of individuals; it is delivered by that process. Could one truly advance an "explanation" for vision, say, based only on "Darwinian natural selection acting in populations of individuals", with no mention of optics or

neuroscience? Just as the universality of the evolutionary process in biology does not somehow render it unnecessary to ask any further questions about developmental biology, so it does not somehow imply that there is nothing more to be said about agency.

DiFrisco and Gawne go on to say that "An organism's "goals"... can be understood as outcomes that enhance Hamiltonian inclusive fitness, such as finding a mate or immune defense against infection. Being in the state of seeking these outcomes—"goal-directedness"—is due to positive selection in the past, and is thus equivalent to adaptation." Let us now replace "goal-directedness" with another, more specialized attribute, such as appetite: "An organism's appetite can be understood as an outcome that enhances Hamiltonian inclusive fitness, such as the need to sustain metabolism. Being in the state of hunger is due to positive selection in the past, and is thus equivalent to adaptation." This statement says nothing other than "appetite and hunger are adaptive". (It certainly does not say, as DiFrisco and Gawne would seem to imply, that "hunger is a metaphor for natural selection".) What then are the mechanisms and attributes that create a sense of hunger? How is that state represented in the brain? What strategies does the organism pursue in response to it? Does the fact that hunger is an adaptive response mean that these questions are moot, because then hunger has already been "explained" by evolution?

More specifically, DiFrisco and Gawne say that "In order for goal-attributions to explain anything, goals would need to be linked to some empirically detectable feature of the system other than the actual outcomes of its behavior. Otherwise, these explanations would be circular and uninformative. It is not clear that this can be done without reference to natural selection." This is a good point: by what criteria can we impute goals, especially if are not realised?

Yet seems uncontroversial to say that the goal of the fox is to catch the hare, regardless of whether it does so. It is also uncontroversial, and empirically detectable, to say that such behaviour is governed by decision-making circuits in the fox's brain. Do similar considerations apply also to the phage? It is not obvious why not; failure to "catch" the bacterium does not mean that, in that instance, the phage was executing independent motions that, by a strange coincidence, happened to track it. And the

processes governing a phage's trajectory are in principle detectable and quantifiable. Its agency has mechanisms; it does not arise from some kind of immaterial *telos*.

How one attributes goals can surely be subtle, but the notion is routinely invoked even at the molecular level – for example, in mechanisms for error correction in DNA replication, protein translation [Banerjee et al. 2017], chromosomal separation in mitosis [Ha et al. 2024], as well as embryology [Hoijman et al. 2021] and neurogenesis [Rajan & Denburg 1996], for errors can only exist in the presence of goals defined by normativity. In all these cases the existence of error-correcting mechanisms can of course be attributed to adaptation. And that is precisely the point: evolution introduces goals.

Thus the idea that "goal-directedness, at least in organisms without complex nervous systems and advanced cognition, is nothing other than adaptation" is itself an untestable statement: not exactly wrong, but rather, hard to assign any real meaning. We simply do not say that a process, characteristic or behaviour that has arisen through evolutionary adaptation simply *is* adaptation, nor that recognizing it as adaptive somehow explains how it works.

Moreover, the mere existence of goals does not tell us how or if they will be pursued. We might with good reason suppose that adaptation has predisposed foxes to chase hares. This does not mean that, on seeing a hare, a hungry fox will chase it as if a neural switch has been flipped. A fox might decide the pursuit is not worth the effort (perhaps the hunger is not bad enough, or the fox considers the case unlikely to succeed). Or the fox might be more preoccupied with other issues, such as tending its young. Or the pursuit might expose the fox to too much danger. Of course, all these responses have adaptive origins too! Whatever the fox does, we can attribute it to some adaptive tendency. So adaptation as such predicts nothing.

Rather, it is adaptive instincts coupled to specifics of immediate circumstance, including the fox's current mental and physiological state (its internal settings, as it were), that ultimately determine behaviour. Could we, given enough information, predict the outcome? Perhaps, although it is far from clear that current understanding of neurobiology or techniques of data gathering would be up to the task. The argument made here is that a scheme for knowing how to integrate all these factors, including the

setting and prioritization of immediate goals and the resolution of conflicts between them, is precisely the kind of objective a true theory of agency might set itself.

It also explains why we might reasonably regard agency in cognitive terms [Levin & Dennett 2020]. DiFrisco and Gawne's charge that "biological agency is in fact cryptically cognitive—i.e., a psychological style of explanation applied to organisms not normally regarded as having a mind" is a warning worth heeding: to imply that all organisms make choices on the basis of desires is evidently to step beyond the bounds of anything that can currently be justified empirically. But this is not necessarily what is implied by suggestions that organismal behaviour, even at the level of single cells, be considered "cognitive". Rather, that term may be used to refer to particular informational competencies; as Dennett and Levin put it [Levin & Dennett 2020],

> Agents, in this carefully limited perspective, need not be conscious, need not understand, need not have minds, but they do need to be structured to exploit physical regularities that enable them to use information (following the laws of computation) to perform tasks, beginning with the fundamental task of self-preservation, which involves not just providing themselves with the energy needed to wield their tools, but the ability to adjust to their local environments in ways that advance their prospects.

As they explain, "the point is not to anthropomorphise morphogenesis – the point is to naturalise cognition." DiFrisco and Gawne's insistence that only humans, with a uniquely sophisticated psychology, can be regarded as "cognitive" not only imputes some unknown evolutionary transition that awarded us (and us alone) this capacity but also risks making it a kind of "cognitism", reminiscent of vitalism, that itself cannot be understood mechanistically in terms common to many if not all other organisms.

4. Demystifying agency

DiFrisco and Gawne says that "Downward causation and context-dependence are "mechanistic" in the sense relevant to experimental biology. They are not mysterious processes that require adopting the teleological form of investigation provided by an agency perspective." I agree – and recognize that some discussions of biological agency at least appear to suggest otherwise. This is not an argument against agency, but simply a statement that the downward causation and context-dependence often

evident in agential behaviour demand a rigorous theoretical understanding that imputes nothing mysterious to them.

DiFrisco and Gawne claim that "agency is a psychological concept with origins in heuristic ascriptions of intentionality." This is not an argument against a wider conception of agency (of which human-like intentionality becomes a special and particularly complex case), but merely an arbitrary prohibition against it: the argument is closed by definitional fiat, where no logic demands it. Their criticism of a gradualist view of agency is advanced on those grounds: we should not call it agency if it does not involve conscious deliberation. But why not? Compare the phenomenon of multicellularity. The definition of multicellularity seems relatively unambiguous: while accepting that there are transitional organisms that may exhibit temporary multicellularity, it clearly does not help to suggest that the earliest prokaryotes were multicellular but in groups of 1. Likewise, DiFrisco and Gawne object to extending the notion of agency to cases where no conscious deliberation can plausibly be invoked. But there are perfectly meaningful definitions of agency that do not need to impute conscious deliberation at all: a substantial literature does not adhere to so restrictive a definition [Ball 2023b].

In the end, the concept of agency does not invoke some sort of immaterial fairy dust but, on the contrary, avoids the trap of attributing magical properties to evolutionary adaptation: of suggesting that organisms do what they do (whatever it is) only and always because it is adaptive. Why does a hand grow in the human embryo? Because hands are adaptively useful, certainly; but that says nothing about proximal mechanisms. Meanwhile, to admit those proximal mechanisms but not their emergent result is to permit only a particular scale of phenomena the status of being real ones. To suggest that there is no such thing as agency is, in this view, equivalent to saying that there is no such thing as a hand but only arrangements, movements and differentiation of cells.

In other words, it is precisely via "complex multiscale feedback mechanisms evolve[d] through natural selection" (and operating through principles explicable by, say dynamical systems theory) that agency arises. We could leave the result of those evolved mechanisms as a miscellany of so many bespoke outcomes for each and every organism, or we could do what science generally does: to identify common aspects and

seek for generic explanations in a theory of agency. One benefit of doing the latter is that it might help to elucidate what is required of artificial systems, for example in robotics or AI, that they be able to demonstrate genuine agency (as opposed to executing instructions in automatic fashion).

DiFrisco and Gawne raise important questions about the extent to which agents need to be regarded as single, holistic entities or, on the contrary, as systems of component parts with specific and identifiable causal powers. These considerations touch on the widely recognized problem of defining biological individuality [Clarke 2010]. However, selection happens at the level of the whole organism, and a component – a brain circuit, say – that prompts a particular action might be considered in some sense to be acting "for" the whole. Typically, an action demands the coordinated involvement of many subsystems: we might be able to pinpoint the impulse for insect navigation to specific regions of the insect brain, but they depend on other regions and organs for the requisite sensory data, and still others to actuate that impulse. DiFrisco and Gawne are of course correct to say that such coordinated actions can be studied mechanistically, but this does not undermine a program to consider the whole organism from an agential perspective as arriving at behavioural choices that it then puts into practice. On the contrary, that is the whole basis of the study of organismal decision-making [Hills 2019, Brembs 2020].

5. Discussion

DiFrisco and Gawne have presented a detailed and thoughtful critique of the suggestion that agency is a property of many and perhaps all living organisms. They make some valid points about the over-reach of some discussions of agency in, for example, development and evolution, which adds to the cogent arguments already put forward by Potter and Mitchell [Potter & Mitchell 2025].

But they have failed to address or perhaps to grasp the role allotted to agency in careful treatments of the topic. Their central challenge is to ask: what questions do the introduction of the concept of agency answer that could not be answered without it? This is, however, to misunderstand the nature of the concept. One could say precisely the same about the notion of life. From a materialist position that regards living organisms as being composed of interacting molecules and nothing more, "life" has no

explanatory purpose in a reductionist program; one simply seeks to understand those molecules and those interactions in a specific set of entities. At the molecular level, there is nothing that obviously distinguishes the chemistry of living organisms from, say, organic chemistry in a laboratory. Do we conclude, then, that the notion of "life" alludes to nothing real – that it is a mere "as if" concept, a "stance" comparable to Dennett's "intentional stance" [Dennett 1987]? One is reminded of Szent-Györgyi's statement [Szent-Györgyi 1972]:

> My own scientific career was a descent from higher to lower dimension, led by a desire to understand life. I went from animals to cells, from cells to bacteria, from bacteria to molecules… On my way, life ran out between my fingers.

We can say just the same for agency. To argue that biological property X is not a real property because it can be studied in reductionistic terms and explained mechanistically is to say that biology has not yet caught up with nineteenth-century physics, let alone that of the twenty-first. The reasons why this is nonetheless said are interesting, complex, and to some extent sociological. They are also connected to the long and problematic debate in biology about teleology [Mayr 2004, Dresow & Love 2023], and the confusions that arise from the fact that a process without apparent purpose or agency – evolution – gives rise to entities that possess both. DiFrisco and Gawne's critique performs the valuable function of cautioning against installing agency as a kind of omnipotent force that, like adaptation, can be summoned as a kind of "explanatory force" for any biological observation. But it should thus serve to distinguish useful and tractable approaches to agency from ones that do indeed attribute it undue explanatory power.

When agency is denied as a genuine characteristic of organisms, it has a habit of creeping back in in cryptic form. DiFrisco and Gawne cite the obviously metaphorical agency that is attributed to genes within a "selfish gene" view of evolution [Dawkins 1976]. This view denies organisms all agency by making them passive vehicles controlled by their genes – genes, moreover, that have their own agendas (primarily, to replicate). Yet here this agential aspect of genes is not "merely" metaphorical at all. It is not simply a colourful manner of speaking about how populations of genes change during the evolutionary process, but becomes a putative explanatory framework for the observed agential behaviour of organisms in which such behaviour is "explained away"

in terms that become deniably metaphorical *even though the metaphor itself is required to do causal work*. This is a classic example of how metaphor tends to be deployed in science in the naive belief that it is just a "manner of speaking" rather than an invitation to adopt a particular point of view.

Whether agency demands any modification of the Modern Synthesis [Ball 2023a, Ball 2023b, Walsh 2015] is an open question, but not one that bears on the issue of whether agency is an attribute of living organisms at all. And certainly it need not be regarded as some kind of alternative to mechanistic explanations of biological phenomena. It demands only that we take seriously what we observe.

References

Ball P. 2023a. *How Life Works*, Chicago: University of Chicago Press.

Ball P. 2023b. Organisms as agents of evolution. John Templeton Foundation. https://www.templeton.org/news/organisms-as-agents-of-evolution-new-research-review

Banerjee K, Kolomeisky AB, Igoshin OA. 2017. Elucidating interplay of speed and accuracy in biological error correction. *PNAS* **114**, 5183-5188.

Bechinger C, Di Leonardo R, Löwen H, Reichhardt C, Volpe G, Volpe G. 2016. Active particles in complex and crowded environments. *Rev. Mod. Phys. 88*, 045006.

Brembs B. 2020. Towards a scientific concept of free will as a biological trait: spontaneous actions and decision-making in vertebrates. *Proc. R. Soc. B 278*, 930-939.

Clarke E. 2010. The problem of biological individuality. *Biol. Theory 5*, 312-325.

Dawkins R. 1976. *The Selfish Gene*. Oxford: Oxford University Press.

Dennett DC. 1987. *The Intentional Stance*. Cambridge: MIT Press.

DiFrisco J, Gawne R. 2024. Biological agency: a concept without a research program. *J. Evol. Biol.* Voae153. Doi 10.1093/jeb/voae153.

Dresow M, Love AC. 2023. Teleonomy: Revisiting a proposed conceptual replacement for teleology. *Biol. Theory* https://doi.org/10.1007/s13752-022-00424-y

Ha G, Dieterle P, Shen H, Needleman DJ. 2024. *PNAS 121*, e2323009121.

Heider F, Simmel M. 1944. An experimental study of apparent behavior. *Am. J. Psychol. 57*, 243-259.

Hills, TT. 2019. Neurocognitive free will. *Proc. R. Soc. B 286*, 20190510.

Hoijman E *et al*. 2021. Cooperative epithelial phagocytosis enables error correction in the early embryo. *Nature 590*, 618-623.

Jaeger J. 2023. The fourth perspective: evolution and organismal agency. In Mossio M (ed), *Organization in Biology*, 159-186. Springer.

Levin M, Dennett DC. 2020. Cognition all the way down. *Aeon* https://aeon.co/essays/how-to-understand-cells-tissues-and-organisms-as-agents-with-agendas.

Mayr E. 2004. *What Makes Biology Unique?* Cambridge: Cambridge University Press.

Mitchell K. 2023. *Free Agents*. Princeton: Princeton University Press.

Monod J. 1972. *Chance and Necessity; An Essay on the Natural Philosophy of Modern Biology*. Vintage, London.

Mossio M, Saborido C, Moreno A. 2009. An organizational account of biological functions. *Brit. J. Phil. Sci. 60*, 813–841.

Potter HD, Mitchell K. 2022. Naturalising agent causation. *Entropy 24*, 472.

Potter HD, Mitchell KJ. 2025. A critique of the agential stance in development and evolution. In Fábregas-Tejeda A, Baedke J, Prieto GI, Radick G. *The Riddle of Organismal Agency: New Historical and Philosophical Reflections*. Milton Park: Routledge.

Rajan I, Denburg JL. 1996. Error correction during guidance of pioneer axons in the leg of the cockroach embryo. *Roux's Arch. Dev. Biol. 205*, 476-485.

Roux E. 2014. The concept of function in modern physiology. *J. Physiol. 592*, 2245-2249.

Szent-Györgyi A. 1972. *The Living State*. New York: Academic Press.

Walsh DM. 2015. *Organisms, Agency, and Evolution*. Cambridge: Cambridge University Press.

Zimmer C. 2021. *Life's Edge*. New York: Dutton.