Synthese
https://doi.org/10.1007/s11229-021-03380-5

ORIGINAL RESEARCH

Agential thinking

Walter Veit[1]

Received: 26 April 2021 / Accepted: 20 August 2021

Abstract
In his 2009 monograph, *Darwinian Populations and Natural Selection*, Peter Godfrey-Smith accuses biologists of demonstrating 'Darwinian Paranoia' when they engage in what he dubs 'agential thinking'. But as Daniel Dennett points out, he offers neither an illuminating set of examples nor an extended argument for this assertion, deeming it to be a brilliant propaganda stroke against what is actually a useful way of thinking. Compared to the dangers of teleological thinking in biology, the dangers of agential thinking have unfortunately rarely been discussed. Drawing on recent work by Samir Okasha, I attempt to remedy this omission, through analyzing the nature of agential thinking, and providing a philosophical treatment of the unexamined dangers in this peculiar, yet tempting way of thinking.

Keywords Agential thinking · Intentional stance · Agency · Natural selection

> Maynard Smith's book Evolution and the Theory of Games directed game theorists' attention away from their increasingly elaborate definitions of rationality. After all Insects can hardly be said to think at all, and so rationality cannot be so crucial if game theory somehow manages to predict their behavior under appropriate conditions.
>
> – Ken Binmore, Foreword in Weibull (1995, P. x)

1 Introduction

Humans have long shown a special kind of interest in the behaviour of other animals. Diverse species are kept in zoos and aquariums, drawing visitors from far and wide. While typical documentaries are produced for niche audiences, animal documentaries reach much larger crowds, sometimes even played in cinemas. A striking feature of such documentaries is the colourful storytelling from figures such as David

✉ Walter Veit
wrwveit@gmail.com

[1] University of Sydney, Sydney, Australia

Published online: 27 August 2021

Attenborough, who can reach an almost cult-like level of fame. Their voice-overs turn animal behaviour—such as the refusal of Capuchin monkeys to receive small rewards when others in their group receive much larger ones (see Brosnan & De Waal, 2002)—into rational actions, interesting plots, and stories that can compete with the best dramas, thrillers, romance, and of course comedy movies. To most, these anthropomorphising descriptions—the *intentional language* that portrays the animals as complex human-like agents—appear as nothing more than fantasy or fiction merely used for their educational or entertainment value.

As several philosophers of biology have pointed out, however, evolutionary biologists themselves also often use such intentional language to treat both biological entities and the process of natural selection as agents (see Wilson, 2005; Godfrey-Smith, 2009; Walsh, 2015; Okasha, 2018). Talk of agents is positively rampant within evolutionary biology, as can be seen in evolutionary game theory, signalling theory, behavioural ecology, and lifehistory theory. When pressed, many evolutionary biologists, such as Richard Dawkins (1976, 1986), will inevitably deflect the use of this language as being merely metaphorical.[1] After all, teleological thinking as the received view in evolutionary biology has been banished since the advent of Darwin.

Speaking of purposes in a submission to a biology journal is highly likely to be criticized by reviewers, if not desk rejected out of hand. Daniel Dennett (2011, 2017) has attributed at least part of this anti-teleology attitude—i.e. the unwillingness to talk about purpose and design that is seemingly operating in nature—to a political motivation. This is not an unreasonable suggestion. Biologists, after all, may understandably be quite reluctant to give up any ground in the battle against advocates of creationism and intelligent design. But it is also precisely for this reason that a few evolutionary biologists such as Andy Gardner (2009) positively maintain that "in light of the current incarnation of creationism—which styles itself as 'intelligent design'—evolutionary biologists should not shrink from addressing the central problem of adaptation" and put "an effort to emphasize that Darwinism is the only scientific theory of biological design" (p. 864). Among philosophers the teleology debate is similarly far from settled, despite a now extensive literature on the topic. In fact, one may even consider it as one of the paradigm cases among the set of subjects that doctoral students are warned from engaging in, due to the difficulty of making a substantive contribution.Proponents of teleological thinking in the biological sciences, as Daniel Dennett (2017) points out, must regularly deal with epithets such as "Darwinian paranoia" (Francis, 2004; Godfrey-Smith, 2009) and "conspiracy theorists" (Rosenberg, 2011).[2] Dennett deems such labels to be "brilliant propaganda stroke[s]" against teleological thinking (2011, p. 481), but doubts that they have any real bite to them.

In this paper, I am not concerned with the role of teleology in biology, per se, but a special sort of teleological thinking that seems to have survived pretty much intact within evolutionary biology: i.e. talk of agents and goals. Though R.A.

[1] Which is not to say that metaphors don't play important roles in science (see Veit and Ney 2021).

[2] See Dennett's *From Bacteria to Bach and Back: The Evolution of Minds* (2017, p. 34).

Wilson labelled such intentional language merely a "cognitive metaphor" (2005, p. 75), Samir Okasha's (2018) recent monograph *Agents and Goals in Evolution* has elegantly shown that there is more to it than this.[3] Instead, it is a particular way of thinking about evolution, one dubbed 'agential' by Peter Godfrey-Smith:

> Here we think of evolution in terms of a contest between entities with agendas, goals, and strategies. I see the agential view of evolution as something of a trap. It has real heuristic power in some contexts, but also has a strong tendency to steer us wrongly, especially when thinking about foundational issues. And once we start thinking in terms of little agents with agendas—even in an avowedly metaphorical spirit—it can be hard to stop.
>
> – Peter Godfrey-Smith (2009, p. 5)

This warning against agential thinking is one of Godfrey-Smith's main messages in his 2009 book *Darwinian Populations and Natural Selection*. Though elaborated less than other ideas within it, Dennett's (2011) has gone so far as to call it the book's *punch-line*. While highly praising Godfrey-Smith's contribution as "the best, most thought-provoking book in the philosophy of biology" that he has "read in a long time", Dennett thinks that Godfrey-Smith is mistaken in his "puritanism" intending to keep any and all teleological notions at bay (2011, p. 475). As he points out points out, Godfrey-Smith leaves it to some extent an open question as to why such thinking is problematic, especially since he frequently relies on agential language himself.

> Given Godfrey-Smith's expression of strong disapproval, I was expecting a parade of Bad Examples, shocking or embarrassing instances of agentialists led on a wild goose chase, or blinded to a simpler truth. But I found none in the book [...]
>
> Daniel C. Dennett (2011, p. 483)

Dennett overstates his case when he suggests that Godfrey-Smith has (i) not argued against agential thinking, and (ii) that this critique is central to his arguments. In many ways, his arguments against agential thinking could be removed without affecting most of his conclusions regarding the nature of natural selection. Despite this, Godfrey-Smith urges a rather radical stance on how agential thinking is to be treated within evolutionary biology. Dennett is right to note that justification of the severe disapproval of such thinking throughout the book requires a stronger case, such as providing actual cases where such thinking has led biologists astray. This problem, however, has not so far been further addressed by Godfrey-Smith, and his more recent work on the evolution of consciousness and subjectivity suggests that he might have even changed his position (Godfrey-Smith, 2016a, 2017a, b, 2020; New England Anti-Vivisection Society et al. 2020).[4] There are subtle hints

[3] See Veit (forthcomingb) for a recent essay review.

[4] Indeed, agential language is found throughout his recent monograph *Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness* (2016b). It will hardly be surprising that many see a close connection between the evolution of agency and consciousness (see Ginsburg and Jablonka 2019; Browning and Veit 2021; Veit 2021b, forthcoming a), but this topic is beyond the scope of this paper.

in his 2009 book, however, that reveal a more nuanced position than just a staunch puritanism.

Despite their extended discussion on the topic, neither Okasha nor Godfrey-Smith offer a very precise characterization of what 'agential thinking' is, instead jumping straight into evaluating its role in evolutionary biology. But this makes it hard to evaluate just how widespread it is among evolutionary biologists, why it should be avoided, and how it relates to old debates about teleological notions. This lack of anything like a canonical statement in the literature is lamentable, as it could help bring further clarity to the debate. I suspect that part of the resistance by both Godfrey-Smith and Okasha to provide such a statement, is the fear that quite a lot of different things are being included under 'agential thinking', some of which are about language, others about explanation, others about styles of modelling, and probably more. There is a legitimate fear that any statement of agential thinking that isn't deliberately vague would leave out some of its instances. A very simple and general definition, however, could and should have been provided, if only for the sake of clarity. I propose the following definition: Agential thinking is the modeling of an entity or process as making a choice in order to achieve a goal.

Granted, this is an idealized statement and analysis, but that is precisely why it will be useful as a philosophical analysis of the status and scientific value of agential thinking. One direct implication of this definition is that agential language ought to be more controversial than teleological language since it appears to require something additional: whereas there is now widespread agreement among philosophers of biology that the selected-effects accounts of function[5] can naturalize attributions of purposes to traits beyond mere metaphorical talk, the description of biological entities (or the process of natural selection itself) in terms of an agent making choices is still seen as requiring further justification. Yet, agential thinking is described by both Okasha and Godfrey-Smith as something that is frequently used by evolutionary biologists, contradicting the anti-teleological stance that is typical of the field. This apparent clash between an open embrace of agential language and an aversion to teleological language among evolutionary biologists is puzzling and deserves rigorous philosophical attention.

There are here at least three open questions we can ask: (i) what is the nature of agential thinking?, (ii) why should agential thinking be avoided in evolutionary biology?, and (iii) how should we distinguish good from bad agential reasoning? In this paper, I attempt to at least partially answer these questions, illuminating the dangers of agential thinking within evolutionary biology. As Dennett (2011) notes, in order to argue that it should be avoided, one has "to demonstrate that these are harmful ways of thinking, and that has not been done [yet]" (p. 484). Agential thinking, compared to teleological thinking, has unfortunately been getting the short end of the stick within the philosophy of biology—perhaps because it was seen as a mere subset of illicit teleological notions in biology. But this assessment of the relationship between purpose and agency within evolutionary biology may well turn out to be wrong, instead each requiring their own distinctive discussions. The

[5] See Millikan (1984); Neander (1983).

omission of such a discussion shall be remedied here, providing an account that will strengthen the bite of supposed epithets such as *conspiracy theorizing* and *Darwinian Paranoia*.

1.1 Article outline

This article is organized as follows. Section 2 discusses teleology, agential thinking, and why we shouldn't rely on conceptual analysis to address the legitimacy of agential language in (evolutionary) biology. Section 3 investigates the peculiar psychological pull and usage of agential thinking, arguing that it is especially hard to escape its temptation and that this should make us much more wary of applying it too hastily. Section 4 explores a variety of cases in which agential thinking can lead us in the wrong direction, providing support for Godfrey-Smith's (2009) assertion that "when we think foundationally about evolution, the agential perspective can be seen to steer us wrongly" (p. 10). Finally, Sect. 5 covers some remaining objections and concludes the discussion.

2 From teleological to agential thinking

In order to make sense of and strengthen Godfrey-Smith's warning against agential thinking, it is best to take a minor step backwards and consider his original motivation. In his 2009 book, Godfrey-Smith sets out on a crusade against *typological thinking* in biology. Drawing on Ernst Mayr's (1994) distinction, he argues that this old conceptual way of treating and understanding the living world is of a tenacious sort, and needs to be replaced by population thinking—here introducing his core idea of a *Darwinian Population*. In order to contrast typological thinking with population thinking in evolutionary biology, he determines at least three (problematic) elements of the former. Only one of these is relevant for my analysis of agential thinking and stands in a particularly interesting relationship wich such: i.e. the "teleological outlook on biological activity" (p. 12).

2.1 Teleology, design, and Darwin

As alluded to in the introduction, the subject of teleology has been one of the most discussed areas in the philosophy of biology (see Allen & Neal, 2020). There are two reasons why it is useful here to have a brief discussion of teleology. Firstly, agential thinking is often seen as a special form of teleological thinking, so whatever consensus is about to emerge on the status of teleology within biology is undoubtedly going to impact our understanding of agency. Secondly, naturalist philosophers have long been concerned with the question of whether talk of purpose, goals, intentions, and functions can be naturalized or ought to be eliminated from the scientific vocabulary. Before Darwin, of course, such talk was not only common but seen as irreplaceable. Those who argue that Darwin banished any *telos* from the living world often turn as an analogy to Newton's banishment of teleology from the

physical realm. Before Newton, things were taken to happen for a reason—not just *a* reason, but a *raison d'être*, i.e. a purpose and not merely a causal process explanation.[6] Indeed, it was only natural to describe the entire physical realm as one full of purposes.

Nowadays, however, it would be exceedingly strange to describe physical processes in teleological terms. Sterelny and Griffiths (1999), in their introduction to philosophy of biology, highlight this difference between biology and physics by comparing the heart to the sun; an often discussed example we can usefully reiterate here. Unlike the function of the heart, which is to pump blood, no such distinction between 'purposeful' and accidental effects (such as the beating sound of a heart) can be drawn in the case of non-living entities. The sun has many effects: for instance, it warms all the planets, and thus enables the continued existence of life on our planet. But as they point out, the sun is not *meant* to do anything and not *for* anything either. There is no purpose, normativity, or goal-directedness. Without a creator or designer, who put the sun it its place to ensure our survival, no purpose is to be had. The latter part is of particular importance here. For the sun to have a purpose, not only would it have to be placed there by God, but it would require an intention—an intended effect that explains its existence. This is why natural selection is treated within selected-effects accounts of function as a way out of this dilemma: it offers us 'design without a designer'.

This is nicely illustrated in much of Richard Dawkins' work, particularly his 1986 book *The Blind Watchmaker*. Natural selection itself is treated as a sort of agent, explaining (or explaining away) the 'purposiveness' in nature. As I shall argue, natural selection is also the key to understanding, or rather justifying, agential language. What then distinguishes biology from physics—or chemistry for that matter—is natural selection, i.e. roughly the blind ecological filtration of individuals within a population due to the variation between them. The result are adaptations: traits with a particular function that they have been selected *for*. It is not clear, however, what this would entail for the status of teleology within biology. Many philosophers (and biologists) have joined this debate, arguably fought since Darwin (1859) published his *On the Origin of Species*.

In fact, let us delve quickly into the various verdicts that have been drawn, starting with Karl Marx:

> Despite all deficiencies, not only is a death blow dealt here for the first time to "teleology" in the natural sciences but their rational meaning is empirically explained.
>
> – Karl Marx (1861)

Marx offers an interesting interpretation of Darwin's contribution that Dennett (2017) thinks echoes through the entire debate. It is a certain ambiguity found in the writings of many biologists who routinely refer to functions in the biological world. Traits, such as the wings of a bird, are *for* something, behaviour is

[6] The English language is unfortunately ambiguous in its use of the term 'reason' between these contexts.

explained by referring to the their *rationale*, and natural selection itself is treated as a process of *design*. How to make sense of this if Darwin has banished teleology from the natural sciences? Dennett thinks that the equivocation in Marx's analysis can be divided into two options:

> We should banish all teleological formulations from the natural sciences
>
> [or]
>
> now that we can "empirically explain" the "rational meaning" of natural phenomena without ancient ideology (of entelechies, Intelligent Creators and the like), we can replace old-fashioned teleology with new, post-Darwinian teleology.
>
> Daniel C. Dennett (2017, p. 34)

The answer to this question is not as straightforward as the banishment of teleology from the physical sciences suggests. Godfrey-Smith describes the result of Darwinism as follows:

> Sometimes Darwinism is seen as demolishing the last elements of a teleological outlook, but at other times Darwinism is seen as constructively domesticating these ideas, showing that they have a limited but real application to biological processes.
>
> Peter Godfrey-Smith (2009, p. 12)

Neither Dennett nor Godfrey-Smith sketch the entire continuum of possible positions however. Here it is useful to compare their options directly.

Let us start on the positive side: if Darwinism can save teleology, to which extent does it do so? Here, Godfrey-Smith only allows for a *limited domestication* of teleology. This certainly sketches the dominant view among those who think teleology can survive at least to some extent. However, there is far more radical, niche view that has been popularized and repeatedly argued for by Dennett himself since at least 1995:

> The biosphere is utterly saturated with design, with purpose, with reasons. What I call the "design stance" predicts and explain features throughout the living world using the same assumptions that work so well when reverse-engineering artifacts made by (somewhat) intelligent human designers.
>
> Daniel C. Dennett (2017, p. 37)

In this picture, teleology is not banished, but its rationale transformed with waves that reach far beyond the usual subject matter of biology.[7] We move from a sort of 'magical' or 'occult' teleology to a post-Darwinian teleology. For this design stance to work, natural selection must provide at least a rough replacement for the prowess of God and provide us with a naturalist form of biological normativity (Veit, 2021a).

The alternative is the complete 'banishment' of teleology. Here, GodfreySmith might be read in a stronger way: Darwinism not only banishes teleology from the natural sciences but should do so from our entire vocabulary. I do not think,

[7] See Dennett (1995).

however, that Godfrey-Smith has such a reading in mind. His 2009 book after all is purely concerned with evolutionary biology. Another polar opposite to Dennett in this regard is found in Alex Rosenberg's critique of Darwinian conspiracy theorizing.

> Ever since Newton, physics has ruled out purposes in the physical realm. If the physical facts fix all the facts, however, then in doing so, it rules out purposes altogether, in biology, in human affairs, and in human thought processes too.
> Alex Rosenberg (2013, p. 19)

According to Rosenberg, Darwin not only banished teleology from biology, but from all of the social sciences and humanities. If Rosenberg is right, then Darwinism also not only banishes agential thinking from biology, but even from the humanities; or at least calls their scientific status into question should it remain. His view would lead us to eliminative materialism—i.e. the denial of intentional states such as beliefs and desires—not only for other animals but also for humanity, indeed making their existence a conceptual impossibility.

This position might not be as absurd as it initially seems: after all, prior to Newton and Darwin many people believed that teleology was simply an ineliminable feature of all the 'domains' of reality. Noticeably, even Kant fell victim to this essentialist thinking:

> [W]e can boldly say that it would be absurd for humans even to make such an attempt or to hope that there may yet arise a Newton who could make comprehensible even the generation of a blade of grass according to natural laws that no intention has ordered; rather, we must absolutely deny this insight to human being.
>
> – Immanuel Kant (2000, p. 5:401)

Yet, this is just what Darwin (1859) did, and Rosenberg argues that we should similarly not reject the eliminative materialist position out of hand. Natural selection is a blind, foresightless, and definitely intentionless process, without any need for the guidance of a designer. As Dennett (2017) observes, philosophers are often too quick to postulate the impossibility of a certain task, particularly when they are not engaged in, or familiar with, the relevant empirical research. The reason for Kant's assertion is, of course, the allure of agential thinking, something Dennett has called the *intentional stance*.[8] Pre-Darwin, however, it was almost impossible to look at the apparent design found in nature, and not be convinced of the necessity of an intentional designer.

There are at least two options then: either teleology is entirely naturalized or banished. *Figure 1* elegantly illustrates the continuum in this debate. Most thinkers fall somewhere on this spectrum. What connects the 'extremes'—here Rosenberg and Dennett—is their agreement that all of the historical folk concepts are utterly dissolved in what Dennett (1995) has dubbed "Darwin's universal acid". Thinkers

[8] The allure of this 'stance' will be target of Sect. 3.

sitting between their radical views may hold the conviction that Darwin only affects some more limited domain of intentionality, rationality, goals, purposes, and agency, but both Dennett and Rosenberg argue that this is a mistake. Whereas one argues that the concept is too tied up with the manifest image and needs to be eliminated, the other argues that we merely need to revise the concept of teleology by drawing on the resources of science.[9] We may reasonably ask how much disagreement there really is between these thinkers, though their conclusions could not be more different. Is this a mere difference between seeing the glass as half-full vs half-empty? Rosenberg (2013) suggests not.[10] He asks us to consider what the difference is between treating "Darwin's achievement as expunging purpose from nature versus treating it as making purpose safe for causality?" (p. 34). I think that this debate comes down to a difference between the philosophical analysis and explication of concepts. This should become much clearer once we consider agential thinking, to which we shall turn now.

2.2 Darwinian conspiracies and paranoia

Agential thinking has played a surprisingly large role in biology. Not only are organisms the paradigm case for being regularly treated as agents: since Dawkins', (1976) book *The Selfish Gene*, the genes-eye view of evolution has become a popular view, that treats genes as agents with their own goals and motives. Only metaphorically, of course, so Dawkins insists retrospectively regretting the title of his book, preferring the alternative "the immortal gene". In the introduction to its 30th year anniversary edition, he writes: "[m]any critics, especially vociferous ones learned in philosophy as I have discovered, prefer to read a book by title only" (2006, p. vii).

Dawkins, however, does not think that the 'gene as agent' metaphor is a bad one, but rather that people misunderstood the implications of the title, wrongly inferring that genes have actual desires and that humans are inherently selfish rather than altruistic. In fact, he seems to think that evolution *must* be thought of in agential terms: "[g]iven that the Darwinian message is going to be pithily encapsulated as The Selfish *Something*, that something turns out to be the gene" (p. viii). Sometimes Dawkins expresses this quite radically, treating the gene as the only true level of selection, with the organism being a mere vehicle for (the purposes of) the gene. As Okasha (2006) and Godfrey-Smith (2009) argued, this sort of essentialist thinking is a mistake. Godfrey-Smith goes farther than Okasha, however, criticizing the *genes-eye view* itself as a flawed mode of agential thinking, rather than a mere lack of recognition for the other levels of selection. As Godfrey-Smith's warning of agential thinking mayb e considered an overreaching, it is useful here to look at how Godfrey-Smith describes the agential perspective[11]:

[9] See Veit (2018) for an earlier discussion of mine of these different responses to the nature of 'purpose' by Dennett and Rosenberg.

[10] Dennett (2016) regards a rigid anti-teleological stance as a conceptual mistake.

[11] It is worth noting that when Godfrey-Smith discusses agential thinking, he unfortunately restricts himself to a criticism of the *selfish gene* point of view—a view he has several problems with that go beyond those related to agential thinking.

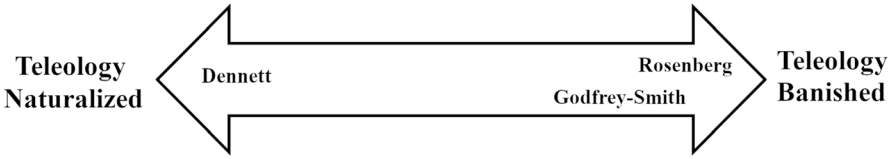

Fig. 1 The Darwinian continuum from the naturalization to the banishment of teleology

> Evolution is treated as a contest between entities that have purposes, strategies, and agendas. This sort of description can be applied to many biological entities; organisms, for example, might be said to "battle to increase the representation of their genes in future generations." But it is often now applied to genes and other replicators themselves. Replicators act to further their own replication. The agential perspective on evolution has always been an uneasy mix of the metaphorical and the literal.
>
> Peter Godfrey-Smith (2009, p. 10)

While Dennett (2011) agrees there is such a mixture, he does not recognize this as problematic:

> [The mix of metaphorical and the literal is] a feature, not a bug; it is just another instance of Darwinian anti-essentialism: drawing a "principled" dividing line between *genuine* belief-talk or agent-talk and mere *as if* belief-talk and agent-talk is the sort of task Jerry Fodor insists on, pounding his fist on the table, not a methodological maxim any Darwinian should have any truck with.
>
> Daniel C. Dennett (2011, p. 481)

Dennett's reference to Fodor links us to the second puritanist next to GodfreySmith in the debate. Alex Rosenberg (2011) argued that evolution has shaped us into what he calls *conspiracy theorists*, making us liable to see plots everywhere and be dissatisfied with purely mechanistic explanations—even if that's all there is. Indeed, as I will argue in Sect. 3, there is considerable empirical support from psychology and neuroscience for the peculiar power this thinking seems to have over our minds, something we should carefully consider before making any sweeping conclusions. After all, the agential mode of thinking is employed within biology—for example, even when there is scientific consensus that bees and other eusocial insects do not possess a sufficiently sophisticated cognitive architecture to attribute them with mental states, they are still often described using intentional language.[12] As the quote from Kin Binmore in the epigraph illustrates, rational choice models and game theory were readily applicable to problems in biology.[13] Perhaps surprisingly,

[12] Naturally, much the same can be said for genes and groups, both of which are sometimes treated as agents in their own right.

[13] Simple agent-based models for evolutionary dyamics of say social behaviour can be intepreted both culturally and genetically (Veit 2019c).

the use of agential thinking in the form of rational choice models in economics has received more criticism than their counterpart in biology.[14]

Aristotle, in trying to find an essence to what it means to be human, defined humans as *the rational animal*, distinguishing us from other animals supposedly lacking this trait. The initial knee-jerk reaction of many might be to simply assert that agential thinking is perfectly legitimate in the case of humans, but a case of illegitimate anthropomorphism when used on other animals. But as much work in experimental economics, behavioural economics, and psychology more generally shows, we ourselves are far off from the rational agent model employed in economics, or even folk psychology for that matter. We tend to rationalize not only the behaviour of others—as expressed in expressions like 'what were her reasons?', 'what are his beliefs and desires?', and 'why did they do that?'—but also of ourselves, when we engage in the post-hoc attribution of beliefs and desires to ourselves in order to explain past behaviour (see Cushman, 2020; Veit et al. 2019). The success condition for such explanations however, whether in psychology or economics, is as we shall see, precisely the same as in biology.

An important observation is the fact economists not only use rational actor models to explain all kinds of human choices, but even the decisions of organizational entities such as households, firms, and nations, i.e. entities to which most people would not attribute beliefs and desires. The conditions for decision theory to work here (i.e. to treat these as agents) are the same as those in the case of agential thinking in biology. Roughly, conflict between its components needs to be minimized in order to treat the larger whole as an 'agent', whether this is the transition from a group of single-cell organisms towards a multicellular organism or the transition from a group of diverse human agents towards a sort of group agency with shared goals and interests, e.g. a political movement. Indeed, Okasha (2018) explicitly draws on previous work on agency in the philosophy of psychology to defend an analogous position concerning the use of agential talk in biology:

> agency [...] requires a unity-of-purpose both at a time, in order that we may eliminate conflict among our motives and do one thing rather than another, and over time, because many of the things we do form part of longer-term projects and make sense only in the light of these projects and plans.
>
> – Kennett and Matthews (2003, p. 307)

Accordingly, Okasha (2018) argues that agential thinking applied to an organism, at least implicitly, presupposes a "unity-of-purpose", which for him entails the *adaptationist* assumption that the organism's traits all contribute to a single goal (i.e. fitness-maximization). Even though this definition is vague, it allows us to draw at least a rough distinction between justified and unjustified instances of agential thinking, and Okasha argues that just such a distinction can enable us to keep harmful uses of agential thinking at bay.

[14] See for example: Thaler (2015), Sugden (2015), Kahneman and Tversky (1979), and Loewenstein (1999).

The question is then no longer 'is this *really* an agent?' but rather 'can we usefully describe this biological phenomenon in agential terms?' I am thus rejecting the idea that the properties of biological entities must conform to what philosophers have often described as 'conceptual analysis', i.e. the investigation of the social, linguistic, or perhaps even 'true' meaning of a term. Instead, I am interested in concepts in the sense of explication (Carnap, 1950) or naturalist conceptual engineering (Veit & Browning, 2020). This approach is pragmatic and revisionist in nature and asks whether we can construct the concept of agency in such a way as to play a useful role in biology.

This brief interlude enables us to draw some important distinctions and sketch an alternative continuum of the debate in Fig. 2. Putting the positions in the debate on a line allows us to emphasize a crucial difference between Rosenberg's critique of agential thinking as "conspiracy theorizing" and Godfrey-Smith's "Darwinian paranoia". Rosenberg's argument is fundamentally a metaphysical one grounded in conceptual analysis. Because he thinks that there really are no agents in nature, he asserts that agential explanations *must* always be inadequate. Godfrey-Smith's worry is different. Even if there are agents, it is not the case that all evolutionary processes involve agents, and certainly not in any essentialist manner. Agential thinking, however—as we shall see in the following section—has a strong psychological force that inevitably invites such essentialist thinking. It can, hence, lead us astray when thinking about evolution.

What matters then, is whether agential talk in biology can play important roles. Godfrey-Smith and Okasha, despite different conclusions in regards to the legitimacy of agential talk can therefore be placed roughly in the middle between Dennettian revisionist explication and traditional conceptual analysis of the sort Rosenberg engages in. Here, I am not interested in the more traditional philosophical project of trying to understand what agential talk 'truly' consists in, but the methodological project of trying to determine the benefits and costs of agential thinking in (evolutionary) biology. Unfortunately, I suspect that the psychological allure of this mode of thinking has stopped us from seriously investigating the ways it can lead us astray. Let us therefore now turn its peculiar psychological pull and how agential thinking is used in practice.

3 The psychological pull and usage of agential thinking

In the previous section, I have argued that the question of agential language in biology requires explication, rather than conceptual analysis. Nevertheless, if Rosenberg is right, and agential thinking is an unavoidable feature of our evolved human psychology, we should strongly resist any initially plausible story of the apparent usefulness of agential thinking, as it may also be an artifact of this credulous tendency. Dennett, for instance, seems to prematurely buy into the usefulness of agential thinking through his commitment to adaptationism. Okasha (2019) describes the difference between himself and Dennett as one of emphasis, himself favouring a neutral position on the status of adaptation in nature. Because of Dennett's anti-essentialist

Fig. 2 The Darwinian continuum from explication to the conceptual analysis of agential language

position and his stance that design is ubiquitous in nature, he is quick to become a strong defender of agential thinking:

> Peter Godfrey-Smith may call this "Darwinian paranoia" and Alex Rosenberg may call it a "conspiracy theory." These are just the latest overreactions to the recognition that the intentional stance is a strategic tool of undeniable power, not a describer of unvarnished facts. Sometimes, they should realize, varnish is just what is called for, an indispensable Good Trick.
>
> – Daniel C. Dennett (2013, p. 62)

Okasha (2018) similarly views agential thinking as a type of adaptationist thinking with the implicit rationale "to understand evolved traits in terms of their contribution to fitness" (p. 2). I think this is a mistake. It confuses the purported justification of agential thinking with the internal structure of its logic. Agential thinking in biology goes farther back than the discipline itself, and more importantly, predates Darwin's theory of natural selection.

Though talk of agents and goals in evolution is rarely intended to be literal, Okasha (2018) argues that the use of agential language within biology should satisfy higher criteria for application, such as the presence of behavioural plasticity. This is somewhat surprising, as Okasha himself notes that agential thinking can come in a variety of modes. Out of these, he draws a general distinction between two types of agential thinking in evolutionary biology: one concerning the *process* of natural selection the product, the other concerning its *product* itself.

Let us start with a kind of *agential thinking* Okasha (2018) suggests we should discard entirely: *Type 2* thinking, that treats natural selection as an agent itself—mother nature. While Dennett (2017) frequently engages in this way of thinking, Okasha (2018, 2019) insists that it should be avoided, because it reinforces the common misconception of natural selection as a directed process, leading to ever-higher fitness, and perfectly adapted organisms. The agential metaphor, hence, invites a way of thinking about evolution that gets things wrong. The problem is decidedly not that natural selection is not actually an agent with beliefs and desires, but that invites an erroneous way of thinking about evolution.

Type 1, Okasha argues, is the more familiar form of agential thinking that treats genes, organisms, or groups as 'rational agents' in their own right.

This way of thinking is familiar from economics, where we can treat firms or households as agents even though any 'conceptual analysis' of the idea would fail to classify them as such. What matters is that multicellular organisms, i.e. groups of single cells, can only be usefully treated as an agent if they have a "unity of

purpose", which is just the justification used when applying the rational-agent model in economics to group entities. Agential models need not presuppose intentional states, but perhaps a sort of 'goal-directedness' is needed, one that justifies talk of what an entitity is 'trying to achieve', i.e. agential thinking.

In a recent review of Okasha (2018), Gardner (2019) offers a slightly misguided criticism. He suggests that there is nothing odd about agential thinking, even when there is consensus that, say, ants do not have beliefs (at least in any sophisticated human sense). Agential thinking here is merely a form of modelling, idealizing the world and thereby emphasising some important features of the world, such as the population dynamics of different strategies studied in evolutionary game theory. Insisting that all models are technically false will not do, however, as Okasha (2019) points out: the mere fact that many evolutionary biologists frequently engage in agential thinking does not attest to its usefulness, if there are compelling reasons to believe that it is an almost addictive way of thinking, applied even when there is no clear benefit what the agential perspective has to offer. Gardner makes an important assertion when he argues that agential thinking allows us to understand evolutionary phenomena that would have otherwise remained mysterious. Dawkins' idea of a Selfish Gene and Maynard Smith's introduction of evolutionary game theory to biology are perhaps two wonderful cases for the benefits the agential perspective has to offer. However, this does not provide the agential mode of thinking with a get-out-of-jail-free card, if it as I shall argue in Sect. 4—also obscures the understanding of biology, and in particular evolution.

As Godfrey-Smith points out, "all talk of benefits and agendas comes with a peculiar psychological power" (2009, p. 10). Unsurprisingly Dennett (2011) responds that this 'peculiar power' is the "power of the intentional stance (Dennett 1971, 1987)" to treat entities as agents with beliefs and desires (p. 481). Dennett goes on to argue that the intentional stance "enables us to think strategically about all manner of phenomena, from our fellow human beings and animals, to computers and even to evolutionary processes" (p. 481). The easy attribution of plots, goals, and intentions to objects in the biological world are not merely creative inventions by commentators.

Rather, they highlight a particular fact about human psychology, one that that has long been emphasized by Dennett. Indeed, one may be tempted to treat agential thinking as merely one instance of the intentional stance, i.e. as applied to (evolutionary) biology. My minimal definition certainly lends itself to that, though I suspect that agential choice alone need not always imply a firm intentional stance with the attribution of beliefs and desires. Rather, the intentional stance can be understood as the psychological pull underlying agential thinking in evolutionary biology.

Here it is best to quote Dennett (1987) directly from *The Intentional Stance*:

> Here is how it works: first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs.

> A little practical reasoning from the chosen set of beliefs and desires will in most instances yield a decision about what the agent ought to do; that is what you predict the agent will do.
>
> – Daniel C. Dennett (1987, p. 17)

What matters for the intentional stance are fundamentally pragmatic considerations of the sort: is it useful to think of x as an agent? Nevertheless, Dennett thinks this is all there is to it. Using the intentional stance we are able to discover 'real patterns' or stable regularities in nature that are perfectly able to ground talk of agency. There is a problem, however, when this agential mode of thinking highlights particular, perhaps unimportant, patterns, and clouds others with more scientific relevance. As Davies (2010) points out, microbiologists—contrary to popular belief—happily anthropomorphise their targets of analysis, a move that can have harmful consequences for our understanding of the biological world, as I will illustrate in Sect. 4. Indeed, I think that it is all too easy to attribute (minimal) intentionality. Dennett's anti-essentialism simply embraces this. There is no *essential* difference between the preferences of, say, a thermostat, and an adult human being. The difference is merely one of complexity and sophistication. Both, according to Dennett can be usefully treated as agents.

The neuroscientist Michael Graziano (2016) offers perhaps the best support for Dennett's intentional stance as a psychologically tempting phenomenon. Unlike many others, Graziano is better able to get ideas across in talks than in papers or books. Indeed, there might be is an important gap between the idea of the intentional stance in the abstract and in practice. In order to demonstrate this, Graziano only requires one little helper: a small orangutan puppet named Kevin. If one attends a talk on consciousness by Graziano—a experienced ventriloquist—it is almost certain that Kevin will play an important role. Graziano (2016) argues that in some sense consciousness can be considered an illusion (though he would deny this), a useful trick that our brains evolved to play in order to navigate us through social situations. By engaging in brief conversations with Kevin, Graziano emphasizes the impossibility not to feel a sense of Kevin's awareness, consciousness, and intentionality. Of course, we all know (with perhaps the exception of very young children) that Kevin is not really conscious, but one cannot stop our brain from instinctively applying the intentional stance.

This tendency is often very useful. It enables us not only to enjoy animated movies, which would be impossible without attributing agency, but is also applied to ourselves. As Sellars (1963; 2007) has repeatedly argued, we do not come to understand ourselves and others as mere vehicles of evolutionary change, but rather as rational agents subject to the operation of norms. We have a 'theory' of mind that makes us liable to attribute agency across the board. Indeed, as Rosenberg (2011) points out, this shapes us into *conspiracy theorists*, enabling us to see plots everywhere in nature, and makes us addicted to stories. Rosenberg argues that it was precisely this more or less sophisticated theory of mind that enabled us to climb from the brink of extinction in the Savannah to the most 'influential' species on our planet.

This is emphasized even further by the Moving Shapes Paradigm developed by Heider and Simmel (1944). In their original experiment they showed adults an animated clip in which a circle and two triangles move in and out of a larger square, with a sort of door. This is already a use of our ability to see purposes everywhere, even a mere two-dimensional and incredible simple animation. Adults and children attribute intentions and goals (i.e. agency) to these shapes. These surprising results from experimental psychology suggest that we should be careful when attributing agency, when there is a strong psychological pull even to attribute mental life to mere two-dimensional forms.

Now we can make sense of Godfrey-Smith's (2009) assertion that agential thinking "engages a particular set of concepts and habits: our cognitive tools for navigating the social world" (p. 10). Indeed, let us return to Gardner (2019) who—while maintaining that we can, in principle, do all of evolutionary biology in terms of agent-free population genetics—is convinced that agential thinking facilitates scientific progress. This echoes the sentiment of David Haig (2012), another evolutionary biologist, who is similarly convinced of the importance of agential thinking—that it is a "way for us to be smart" (Godfrey-Smith, 2009, p. 10). Drawing on the evolutionary psychologists Cosmides and Tooby (1992), Godfrey-Smith argues, similarly to Rosenberg, that our minds are particularly powerful in dealing with social situations containing human agents. However, this need not be a benefit in scientific investigations. As Godfrey-Smith notes: "[w]hen we think about agents and agendas, we think differently and more acutely than we do about abstract logical and causal relations" (2009, p. 10). Evolutionary theorists may thus tempted to rely on their evolved cognitive machinery to analyse the more 'strategic' aspects of evolution, as for instance Maynard Smith and Price (1973).

We should thus now be able to make sense of Godfrey-Smith's argument against 'Darwinian Paranoia':

> And once taken up or switched on, these are psychological tools that are hard to put down; they have a compelling, almost addictive, narrative appeal, and tend to send us down specific paths. This claim, again, applies to all agential views of evolution, not just those invoking replicators. But the introduction of small and hidden agents has a particular power. It can lead to an acute form of what Richard Francis (2004) has—dramatically but accurately—called "Darwinian paranoia." Darwinian paranoia is the tendency to think of all evolutionary outcomes in terms of reasons, plots, and strategies. An agenda is a powerful explainer. Once introduced to the possibility of understanding a phenomenon in terms of a grand rationale, we become reluctant to settle for less. One agenda might be exchanged for another, but this becomes the kind of understanding we are after. The application of an agenda to the empirical facts might be indirect and constrained, but an agenda "makes sense" of things for us in a way that no mere catalog of efficient causes can.
>
> – Peter Godfrey-Smith (2009, p. 10)

The problem with these assertions is that Godfrey-Smith only provides sparse evidence that this way of thinking actually misleads biologists. I nevertheless hope to have here given a much firmer foundation for the strong opposition Godfrey-Smith

holds against agential thinking, with its psychological pull being confused with actual explanatory power. By contrast, Robert Wilson's (2005) minimal conception of agency as something that is merely physically bounded and play some causal role, stands on much shakier ground, compared to my goal-centric definition given above. Indeed, it serves to illustrate how easy it is to think that the agential mode of thinking provides something distinctly valuable. As Godfrey-Smith (2009) points out, Wilson could count any causal 'individual' as an agent, be it a "carbon atom or a brick" (p. 11). But why should we? The question we have to ask here is the Dennettian one: *cui bono*; who benefits? But the target is different: how does it benefit scientists to engage in this attractive, albeit misleading sort of talk, if it benefits them at all? It is now that I will turn to the dangers of agential thinking, a discussion that must focus on actual scientific practice. As Okasha (2018) notes, the "ultimate justification [for agential thinking] is empirical rather than theoretical" (p. 230). Let us therefore now move to discussion of a set of examples that should satisfy at least some of the initial 'thirst' Dennett exhibits for when he demands a "parade of Bad Examples, shocking or embarrassing instances of agentialists led on a wild goose chase, or blinded to a simpler truth" (2011, p. 483).

4 The dangers of agential thinking

Despite the insistence of many biologists that teleology has no role in biology, agential thinking plays a surprisingly large role in biological theory. This might suggest that agential thinking is unproblematic, as long as it is merely taken to be metaphorical. This, however, is a mistake. In this section, I attempt to draw attention to an open question that I raised in the conclusion of a paper on the role of cheaters in the evolution of multicellularity: what is the actual impact of anthropomorphic language on research?[15]While there has been some engagement with the role of agential thinking regarding higher vertebrates (see Kennedy, 1992), I am here more interested with more basic evolutionary units. No one can deny that agential thinking has often played a useful role in evolutionary biology, however, one should also consider the dangers of such thinking. That a metaphor *can* be useful is trivial. To merely show examples in which it is useful is not the same as legitimizing its widespread usage. Even if agential thinking can play a positive role in evolutionary biology, this may be utterly outweighed by the potential negative effects such thinking invites. Hitherto, however, debates on agential thinking have mostly focused on ontological concerns, an indicative complaint being something like "but corvids are not *really* plotting". Such ontological concerns are to some degree separate from the usefulness such thinking might otherwise provide. Usefulness alone, however, is not enough when the negative effects of such thinking outweigh the positive ones. It is a serious omission in the literature that this issue has not yet received sufficient attention. Accordingly, I argue that there are at least two distinct problems with agential

[15] See Veit (2019a).

language that have a negative effect on evolutionary biology: (i) the restriction of scientific creativity, and (ii) the revival of essentialist thinking.

4.1 Restriction of scientific creativity

The first problem I shall address concerns the restriction of scientific creativity and generation of alternative ways of thinking. This problem may seem odd at first sight, given that agential thinking is commonly defended in virtue of its success in promoting creativity and the generation of new and fruitful hypotheses. By treating a biological system or natural selection as an agent, we can use a particular way of thinking that is familiar from engineering, i.e. reverse engineering.[16] If a biological system seems to pursue one or more goals, we can ask how it is able to pursue these goals. Furthermore, it allows for a forward-looking perspective of evolution. Given an organism with apparent 'goals', we may ask what phenotypic changes would allow for an improvement in the pursuit of these goals and thus potentially predict future change.

These points are well taken, however, there is no necessary inconsistency here with the idea that agential thinking can *prevent* the generation of new hypotheses that are at odds with the agential descriptions, particularly when agency and goals are misdescribed. A conceptual tool might be useful for some purposes, but utterly mislead us when applied to an inappropriate context. This is not simply the claim that metaphors are misleading when used outside of their correct context.[17] Rather, my methodological point is that the psychological pull of agential thinking, and its usefulness in some cases, have led to an epistemic culture within some domains of evolutionary biology in which ideas will be expressed in agential form. This will then actually restrict, rather than aid, our imaginative capacities.

If a challenge is made to this method of modeling of evolution in terms of agents making choices, its proponents (such as Gardner or Dennett, as I have discussed in the previous sections), will misread the very objective of Okasha's, (2018) monograph into one about the ontological status of agency, or even a flat-out endorsement. It is not surprising that Dennett (2019) treats Okasha as a guide and proof for the usefulness and almost inevitability of agential thinking, when Okasha's motivation was only to tease out its methodological limits. And it is this part of his work that left both adaptationists unfazed. Unfortunately, here I will have to partially side with both Dennett and Gardner. While philosophical analysis has its virtues, scientists have rarely been convinced by methodological arguments from the armchair. To show that a particular way of modelling is flawed and should be used much more carefully must involve actual cases in science. We need to see where agential thinking has led us horribly astray and proved itself to be more akin to blinkers, rather than a useful lens through which to see the world. After all, it is within science itself that a particular mode of thinking must stand trial. Since such a parade of empirical

[16] A point vigorously defended by Dennett (2017).

[17] I thank an anonymous reviewer for raising this point.

examples has neither been provided by Godfrey-Smith nor Okasha, I will try to remedy this here.

The first example comes from recent work by Paul B. Rainey, an experimental biologist looking at the transition to multicellularity through the role of 'cheats' rather than the traditional emphasis on co-operation (Rainey & Kerr, 2010; Veit, 2019a). Using the *Pseudomonas fluorescens* wrinkly spreader system, Hammerschmidt et al. (2014) initially propagated 140 beakers with a *P. fluorescens* population. While the ancestral population of these singlecell eukaryotes (SM = smooth) floats individually within the broth, mutations quickly occur within them leading to a new phenotype (WS = wrinkly spreader) of cell–cell glue production.[18] Under usual conditions these WS cells normally have a lower fitness than their SM counterparts. Due to the adhesive, daughter cells are unable to detach themselves from their parents, suffering the costs of glue-production and a life in close proximity. However, though non-buoyant, these groups of WS cells are able to attach themselves at wall of the beaker, taking over the interface between broth and air. This allows them to reap the benefits of access to oxygen by contributing to this public good, taking over the entire surface. Due to their high mutation rate, however, 'cheats' arise that, while benefiting from their access to oxygen within the mat, do not provide the stabilizing cell–cell glue necessary to keep the integrity of the mat. These cheats quickly multiply and lead to the fall (i.e. doom) of the mat. While groups of such previously solitarily living cells quickly arise in the course of evolution, these 'group entities' go extinct just as quickly.[19] At this point, Rainey and Kerr (2010) have suggested that these cheats, while causing the 'doom' of the mat, could also serve as their 'savior' by playing the role of propagules able to detach themselves from the inevitable fate of the group.

In their experiment, Hammerschmidt et al. (2014) were able to show that—perhaps counterintuitively—a cheat-embracing life-cycle achieved a higher fitness for the 'group' than a cheat-purging regime. These surprising results have not gone unnoticed. In a recent article on the study of social behavior in microbiology, Corina Tarnita (2017) offered an entire section on "The dangers of anthropomorphizing: when does cooperation cease to be a useful concept?" (p. 21). Drawing on Rainey's work, she argues that value-laden terms such as cooperation and cheating should be omitted and replaced with neutral ones.[20] I have also raised this problem in my philosophical analysis of Rainey's experiments: "by questioning the very notion of cheats being always bad simply in virtue of their evaluative name […] gametheoretic analyses, excluding such roles for cheats, must necessarily cloud the generation of new hypotheses" (2019a, p. 17). If anthropomorphic agential language is merely metaphorical and heuristic, one may wonder whether this danger is a serious one. In hindsight, however, Rainey's hypothesis may be

[18] A variety of alternative mutations allows for this phenotype.

[19] But as Hammerschmidt et al. (2014) argue, it is not meaningful to speak here of a Darwinian individual in its own right, unless it has some way of reproduction. See also Okasha (2006) and Godfrey-Smith (2009).

[20] See also Davies (2004).

considered as an almost obvious solution—so simple and trivial, in fact, that one might wonder why no one came up with this explanation before. I argued that Rainey's research project calls for a reconceptualization of cheats, but there is an alternative conclusion—that is to insist that we were merely mistaken in treating these proto-propagules as cheats to begin with. If it was a mistake, the question arises as to where the mistake resides – is it treating them as cheats when they are in fact not, or in the very act of thinking in agential terms? As this distinction reveals, it is far from easy to judge a priori whether an application of the intentional stance is going to be useful or harmful.

A second similar example comes from the work of Tarnita (2017), who suggests a positive role for cheats in the bacterium *Myxococcus xanthus*. She argues that a positive role for cheats can be recognized once evolutionary innovations are introduced into the equation (see also Fiegna et al., 2006). Innovations, however, are rarely if ever considered in evolutionary models of cooperation that mostly focus on two strategies, i.e. cooperation and cheating. This, while perhaps appropriate in behavioural ethology, does not neatly transfer to experimental evolution studies in microbiology. Tarnita argues that mutations, and hence innovations, play a much bigger role when natural selection is operating on a faster time scale.

M. xanthus is a species of predatory myxobacteria that live in a variety of population structures. Usually these bacteria can be found living together in a loose swarm in the form of a biofilm. What makes this species particularly interesting, however, is their complex multi-cellular developmental life-cycle when exposed to a lack of nutrition. Under conditions of starvation, these bacteria form fruiting bodies, of which only some will survive as spores. A costly process such as this clearly allows for the possibility of cheats. But as we have seen, this need not be a problem for the evolution of multicellular organisms.

> One such cheater fails to produce viable spores in monoculture, which makes it an obligate social cheater whose survival during starvation is dependent upon chimeric fruiting body development with a social host. In lab experiments, this obligate cheater, which led to the downfall of cooperation, eventually mutated into a novel social type; moreover, it did so via mutations that generated novel genetic interactions rather than by a simple reversal of its defects. Thus, 'a temporary state of obligate cheating served as an evolutionary stepping-stone to a novel state of autonomous social dominance' (Fiegna et al., 2006).
>
> – Corina E. Tarnita (2017, p. 22)

Tarnita uses these two examples, stemming from experimental evolution studies on *P. fluorescens* and *M. xanthus*, to advocate a replacement of terms such as cheat and cooperator by "[m]ore neutral terms such as producers and non-producers […] as they allow for multiple alternative hypotheses to be considered" (p. 22). Tarnita makes a stronger claim here than my suggestion that this sort of anthropomorphic language makes the generation of alternative hypotheses more difficult. She seems to argue that anthropomorphic language leads to the complete abandonment of hypotheses that conflict with the value-laden nature of terms such as cheats.

This stronger claim is not without appeal. Rather than being tested, hypotheses—such as Rainey's cheats as proto-germline hypothesis—might initially be discarded

because of the mental association of cheating with something bad and something to be avoided, i.e. a problem that requires solving. That there is the potential for agential language to lead science astray is well accepted. However, biologists may underestimate the extent of these dangers, judging them to be outweighed by the benefits of agential thinking. It is, as Tarnita points out, precisely because we understand the causal mechanisms of social behaviour better in the case of eusocial insects than the evolution of multicellularity that the ascription of agency seems less problematic in the former. But in that case the work was not done by agential thinking alone; rather agential thinking *of a particular sort* becomes justified.

Does agential thinking block the generation of hypothesis entirely? No, but agential thinking may make initially counterintuitive results seem far less attractive, and even absurd. Unlike other modes of thinking, it has a unique psychological gravity that makes it hard to think in non-agential terms once one has entered into its sphere, positively stifling scientific creativity and progress. These examples will help to understand why Godfrey-Smith warns of agential thinking as a return to older essentialist ways of carving up the world.

4.2 The revival of essentialism

The only example of agential thinking that Godfrey-Smith discusses in detail concerns the genes-as-replicators view. Dennett (2011) does him an injustice by not considering his arguments in detail. Indeed, I will argue that despite Dennett's insistence that agential thinking can be thought of in nonessentialist terms, this way of thinking is particularly vulnerable to falling back into essentialism. The second danger of agential thinking within biology we shall explore is thus the *revival of essentialism*. The discussions about natural selection, selfish genes, and multi-level selection significantly shaped evolutionary biology over the last forty years (see Okasha, 2006 for an overview). Godfrey-Smith sees agential thinking as an unfortunate side-effect of this process, however:

> This process was accompanied by successive shifts in use of the language of agency and benefit. Such language often has a significant communicative role. When a student is told that the gene is the ultimate unit of evolutionary self-interest, for example, he or she is supposed to hear that as gesturing towards one family of evolutionary mechanisms—which can be more precisely described in other terms—and away from another. In the case of descriptions from a genic point of view, however, these formulations developed an unusual power and role. They became more than a shorthand, being used not just to summarize complicated ideas but to shape foundational descriptions of evolution.
>
> Peter Godfrey-Smith (2009, p. 143)

One might read into this passage that the agential mode can lead us away from alternative hypotheses, but it seems more likely that Godfrey-Smith intends this as a critique of the gene as the sole 'benefector' of evolution. Indeed, he argues that the replicator approach has been deliberately constructed to fit an agential understanding

of evolution. Dawkins (1976), in *The Selfish Gene*, frequently talks of goals, strategies, and interests. Godfrey-Smith (2009) does not consider the Dennettian approach of treating this talk as genuine, an approach that—though interesting—does not seem to capture the way agentialists like Dawkins, Haig, and Gardner defend their accounts as mere heuristics. To some extent, Dennett's approach rests on the success of agential thinking, but as I have discussed, this needs to be demonstrated itself.

Godfrey-Smith sees Dawkin's Selfish Gene concept as something distinctively born out of an agential view on natural selection:

> What is true is that if we want to describe evolution using an agential framework, where a process extended over long periods is described in terms of the pursuit of goals by some entity, then that entity must persist, at least in the form of copies. Otherwise it cannot realize or fail to realize its goals. So within the framework that collapses populational processes to the activities of agents, Dawkins is expressing a real constraint. But that is an argument against the agential framework—an argument that it cannot be applied to all cases. It is not an argument that evolution cannot occur unless long-term persisting entities are present. So it is not just overtly teleological features that the agential framework imputes to evolution—features which we can quickly say are only being metaphorically attributed. Subtle structural features are being imposed as well. The agential framework treats the evolutionary process in terms of the persistence of special entities through superficial change, rather than in terms of the successive creation of new entities, with similarities and other dissimilarities to earlier entities.
>
> Peter Godfrey-Smith (2009, p. 5)

Indeed, Godfrey-Smith's (2009) book Darwinian Populations and Natural Selection is one extended treatment of this mistaken idea. Natural selection does not require agents to receive benefits—there are many proto, sort-of, marginal cases of evolution—perhaps even the majority—that do not neatly fit into the paradigm conception of natural selection as operating on neatly delineated individuals. It is precisely those other cases where an agential view is problematic. An important further instance of this is the role of the immune system (Pradeu, 2011) that perhaps altogether undermines the idea of eukaryotes as individuals; which is not to deny that we can treat individuality as a useful model. Gene-thinking is not a replacement but rather an instance of agential thinking, as the following quote by Patten (2019) illustrates: "[t]he self looks more like a democracy comprising various political factions when genetic conflicts are recognized" (p. 89). So it is precisely essentialism we need to be concerned about.

> Overtly agential description of evolution is part of a larger family, or a graded series, of metaphorically loaded usages. At the extreme end we have talk of strategies and cabals. These shade into less tendentious talk of welfare and goals, and those shade into talk of costs and benefits understood directly in terms of components of fitness—chance of survival, number of matings, and so on. It can be unclear where metaphor ends and literal usage begins. Talk of this kind can also have several different intended roles. It may be seen as a

> metaphorical expression of a deep truth (as in Dawkins, 1976), or merely as a practical tool for thinking about some complex matters in a simple way (Haig 1997).
>
> - Peter Godfrey-Smith (2009, p. 142)

I think it is here that Godfrey-Smith points to a slippery slope towards essentialism. As Sect. 3 illustrated, there is something particularly addictive about agential thinking, drawing on an old but ingrained way of thinking.

If it was merely a practical tool like other models this would be unproblematic, but that is not the case, and hence deserves thorough philosophical treatment. Many evolutionary biologists appear to skip over the important difficulties in agential thinking, but as the foregoing analysis should demonstrate there are no easy answers, and many biologists would be well-advised to take these issues more seriously.

Sometimes biologists might shrug this off by saying that it is merely a perspective, one of multiple angles we might take: the gene, the organism, or perhaps the group. But if there are various alternative ways to attribute agency, of which only one is right, this already suggests that agential thinking—while helpful for the generation of new hypotheses—can also often have a negative effect by misplacing the 'goals' of units subject to natural selection. Even the simplest living systems have a sort of directedness, as often emphasized by Godfrey-Smith, but what are they directed towards? The answer is simple: goals. Goals of course, very thinly, are simply to be understood as something an organism needs to strive towards in order to increase its fitness. This, however, can be put in entirely neutral and mechanistic terms, without any need for agential descriptions and the risks that accompany such usage.

5 Concluding discussion

As the foregoing section illustrated, *agential thinking* is intertwined with a complex array of philosophical and empirical problems. If my presentation here succeeded, I should have convinced the reader that agential thinking in biology, contrary to Dennett, is neither indispensable nor always a good trick. Instead, we need to recognize that agential thinking is a double-edged sword with a lot of anthropomorphic baggage. It can help us to come up with new hypotheses and alternative explanations, but these are liable to be completely off-track and may cloud the right explanations in a shroud of non-agential fog. Once we have begun to think in agential terms about a particular area of evolutionary biology it can be hard to stop, creating an agential gravity that hinders scientific creativity like blinkers on a horse. A useful lens is then turned into a dogma of necessity. Where agential thinking works, it is because we have correctly determined the units of selection and the adaptive pressures. This is often the case and seems to be what Okasha (2018) has in mind when he argues that agential thinking of type 1 is frequently useful. But in this scenario, it is unclear what the agential perspective adds beyond perhaps neatly summarizing what we know. As it seems we are—as Godfrey-Smith (2009) seems to argue—perfectly able to replace agential by neutral talk. Perhaps agential thinking can serve as

a useful bridge prior to an analysis of the biological mechanisms and processes, but as Rainey's work on 'cheats' illustrates, the opposite might also be true. Agential thinking can serve as a troll on a bridge, making it impossible to cross towards a deeper understanding of the biological phenomena.

This seems to be closer to Godfrey-Smith's understanding of agential thinking as something that is inherently bound up with old philosophical ways of thinking, a way of thinking that is liable to mislead us when applied to evolution. These folk concepts, similar to our folk physics, cannot just be discarded. Even a Nobel Prize-winning physicist might be unable to escape their intuitive folk psychological notions of how physics works. Much of science goes against 'common sense', so we should be open to challenging received wisdom of what are good, legitimate, and more importantly *necessary* ways of understanding particular phenomena. Before Newton, many thought that there was no way to conceive of the world without purposes, and to some extent these thinkers were right. If one survives the climb to the summit of Mt. Everest one might feel hard-pressed not to be overcome by a sense of *pathos*—that there must be some sense to all this beauty.[21] Even the die-hard physicalist will not deny that it seems like there is more, but there is nothing explanatory in this 'seeming', almost as if there was a creator, a designer above the clouds. Similarly, indeed, with much more force, natural selection has shaped us into agency-detectors that makes it almost impossible for us not to interpret the behaviour of other animals in rational terms. This approach will often work in the case of animal agents, allowing us to discover something Dennett (2017) has repeatedly called "free-floating rationales", but once we turn to the fundamental workings of nature and life, this way of thinking is liable to get things wrong. As Veit et al. (2019), argue, there is a 'Rationale of Rationalization', but this rationale has evolved for a specific purpose, one that above all helps to predict the behaviour of others and discover some adaptative rationales, not to discover how natural selection operates in the abstract.[22] As Tarnita (2017) points out, it is precisely because we have gained such a detailed understanding of eusocial insects that it does not seem harmful to apply agential language to them. From slave-making ants to dancing bees there is no shortage of agential descriptions. It is unclear, however, what role agential language played here beyond the forming of some hypotheses. The context of discovery and the context of justification, are indeed sometimes quite distinct. Godfrey-Smith is right then, as much work on the major transitions and the units of selection shows the agential perspective can easily lead us astray. Work by neuroscientists and psychologists should make us wary of these dangers and try especially hard to resist the temptation of agential thinking. It is not just an alternative modelling strategy that we should simply embrace for the sake of a diversity of different models as some scientific pluralists may argue.[23] Yet, I have previously expressed doubts that

[21] Assuming one did not reach the summit in a snow-storm and is hence unable to see anything beyond a one meter radius.

[22] In a forthcoming paper, I argue that we can understand natural selection itself as an ecologically scaffolded process (see Veit forthcomingc).

[23] See for instance Veit (2019b, c).

traditional game-theoretic terms such as cooperator and cheat can be, as Tarnita seems to suggest, "cast-off over night" (Veit, 2019a, p. 17). The foregoing analysis shows why it is so hard to banish agential thinking from evolutionary biology. But there are further, more sociological, reasons that could also be considered troubling. Indeed, articles that are published with attractive titles, including agential terms such as cheaters, slave-making, choice, queens and kinds might be easier to publish and reach a larger audience. More careful articles, with neutral headlines, seem to have less of an impact factor, as measured by citations. In some sense, agential language allows even uneducated readers to skip over the biological details, quickly grasping some important information. This is of particular importance in science communication. In light of this, the best we can do is to develop our theories further and understand these biological systems, such that agential language is no longer misleading, but it is unlikely that we will ever be able to banish it entirely.

Acknowledgements I would like to thank Samir Okasha, Heather Browning, Andy Gardner, two anonymous reviewers, and finally the Theory and Method in Biosciences group at the University of Sydney for their feedback on my manuscript.

Funding This research was supported under Australian Research Council's Discovery Projects funding scheme (Grant No. FL170100160).

References

Allen, C., & Neal, J. (2020). Teleological notions in biology. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Spring 2020 ed.). Metaphysics Research Lab, Stanford University.

Brosnan, S. F., & De Waal, F. B. (2002). A proximate perspective on reciprocal altruism. *Human Nature, 13*(1), 129–152.

Browning, H., & Veit, W. (2021). Evolutionary biology meets consciousness: essay review of Simona Ginsburg and Eva Jablonka's *The Evolution of the Sensitive Soul*. *Biology and Philosophy, 36*(5), 1–11. https://doi.org/10.1007/s10539-021-09781-7

Carnap, R. (1950). *Logical foundations of probability University of Chicago Press*. University of Chicago Press.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. Volume 163, pp. 163–228

Cushman, F. (2020). Rationalization is rational. *Behavioral and Brain Sciences 43*.

Darwin, C. (1859). *On the origin of species by means of natural selection*. John Murray.

Davies, M. (2004). Epistemic entitlement, warrant transmission and easy knowledge. *Proceedings of the Aristotelian Society, 78*, 213–245.

Davies, J. (2010). Anthropomorphism in science. *EMBO Reports, 11*(10), 721–721.

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

Dawkins, R., et al. (1986). *The blind watchmaker: Why the evidence of evolution reveals a universe without design*. WW Norton & Company.

Dawkins, R. (2006). *The Selfish Gene: 30th Anniversary Edition–with a new Introduction by the Author* (3 ed.). New York: Oxford University Press.

Dennett, D. C. (1987). *The intentional stance*. MIT press.

Dennett, D. C. (1995). *Darwin's dangerous idea: Evolution and the meanings of life*. Simon and Schuster.

Dennett, D. C. (2011). Homunculi rule: Reflections on Darwinian populations and natural selection by Peter Godfrey Smith. *Biology & Philosophy, 26*(4), 475–488.

Dennett, D. C. (2019). Clever evolution. *Metascience, 28*(3), 355–358.

Dennett, D. C. (2013). The evolution of reasons. In B. Bashour & H. D. Muller (Eds.), *Contemporary philosophical naturalism and its implications* (pp. 47–62). Routledge.

Dennett, D. C. (2016). Darwin and the overdue demise of essentialism. In D. Livingstone Smith (Ed.), *How biology shapes philosophy: new foundations for naturalism*, pp. 9–22.

Dennett, D. C. (2017). *From bacteria to Bach and back: The evolution of minds*. WW Norton & Company.
Dennett, D. C. (1971). Intentional systems. *The Journal of Philosophy*, pp. 87–106.
Fiegna, F., Yuen-Tsu, N. Y., Kadam, S. V., & Velicer, G. J. (2006). Evolution of an obligate social cheater to a superior cooperator. *Nature, 441*(7091), 310–314.
Francis, R. C. (2004). *Why men won't ask for directions*. Princeton University Press.
Gardner, A. (2009). Adaptation as organism design. *Biology Letters, 5*(6), 861–864.
Gardner, A. (2019). The agent concept is a scientific tool. *Metascience, 28*(3), 359–363.
Ginsburg, S., & Jablonka, E. (2019). *The evolution of the sensitive soul: Learning and the origins of consciousness*. Cambridge: MIT Press.
Godfrey-Smith, P. (2009). *Darwinian populations and natural selection*. Oxford University Press.
Godfrey-Smith, P. (2016a). Mind, matter, and metabolism. *The Journal of Philosophy, 113*(10), 481–506.
Godfrey-Smith, P. (2016b). *Other minds: The octopus, the sea, and the deep origins of consciousness*. Farrar.
Godfrey-Smith, P. (2017b). The subject as cause and effect of evolution. *Interface Focus, 7*(5), 20170022.
Godfrey-Smith, P. (2017a). The evolution of consciousness in phylogenetic context. In K. Andrews & J. Beck (Eds.), *The Routledge handbook of animals minds* (pp. 216–226). Routledge.
Godfrey-Smith, P. (2020). *Metazoa: Animal minds and the Birth of consciousness*. Harper Collins.
Graziano, M. S. (2016). Consciousness engineered. *Journal of Consciousness Studies, 23*(11–12), 98–115.
Haig, D. (2012). The strategic gene. *Biology & Philosophy, 27*(4), 461–479.
Hammerschmidt, K., Rose, C. J., Kerr, B., & Rainey, P. B. (2014). Life cycles, fitness decoupling and the evolution of multicellularity. *Nature, 515*(7525), 75–79.
Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology, 57*(2), 243–259.
Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, pp. 263–291.
Kant, I. (2000). *Critique of the Power of Judgment*. Cambridge: Cambridge University Press. (The Cambridge Edition of the Works of Immanuel Kant) (P. Guyer, Ed.; E. Matthews, Trans.).
Kennedy, J. S. (1992). *The new anthropomorphism*. Cambridge University Press.
Kennett, J., & Matthews, S. (2003). The unity and disunity of agency. *Philosophy, Psychiatry, & Psychology, 10*(4), 305–312.
Loewenstein, G. (1999). Experimental economics from the vantage-point of behavioural economics. *The Economic Journal, 109*(453), F25–F34.
Marx, K. (1861). Letter to F. Lassalle, 16 January 1861. *Marx Engels Archive*. https://www.marxists.org/archive/marx/works/1861/letters/61_01_16-abs.htm.
Maynard Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Nature, 246*(5427), 15–18.
Mayr, E. (1994). Typological versus population thinking. *Conceptual issues in evolutionary biology*, pp. 157–160.
Millikan, R. G. (1984). *Language, thought, and other biological categories*. MIT Press.
Neander, K. (1983). *Abnormal psychobiology*. Ph. D. thesis, Unpublished doctoral dissertation, Bundoora, Australia: La Trobe University.
New England Anti-Vivisection Society, American Anti-Vivisection Society, The Physicians Committee for Responsible Medicine, The Humane Society of the United States, Humane Society Legislative Fund, J. Jacquet, B. Franks, J. Pungor, J. Mather, P. Godfrey-Smith, L. Marino, G. Barord, C. Safina, H. Browning, and W. Veit (2020). Petition to Include Cephalopods as "Animals" Deserving of Humane Treatment under the Public Health Service Policy on Humane Care and Use of Laboratory Animals. *Harvard Law School Animal Law & Policy Clinic*, pp. 1–30. https://doi.org/10.13140/RG.2.2.27522.30401.
Okasha, S. (2006). *Evolution and the levels of selection*. Oxford University Press.
Okasha, S. (2018). *Agents and goals in evolution*. Oxford University Press.
Okasha, S. (2019, August). Reply to Dennett, Gardner and Rubin. *Metascience 28*(3), 373–382.
Patten, M. M. (2019). The X chromosome favors males under sexually antagonistic selection. *Evolution, 73*(1), 84–91.
Pradeu, T. (2011). *The limits of the self: Immunology and biological identity*. Oxford University Press.
Rainey, P. B., & Kerr, B. (2010). Cheats as first propagules: A new hypothesis for the evolution of individuality during the transition from single cells to multicellularity. *BioEssays, 32*(10), 872–880.

Rosenberg, A. (2013). Disenchanted Naturalism. In B. Bashour & H. D. Muller (Eds.), *Contemporary Philosophical Naturalism and its Implications* (pp. 17–36). Routledge.
Rosenberg, A. (2011). *The atheist's guide to reality: Enjoying life without illusions*. New York: WW Norton & Company.
Sellars, W. (1963). Philosophy and the scientific image of man. In *Science, perception and reality*, pp. 1–40. Routledge & Kegan Paul London.
Sellars, W. (2007). In the Space of Reasons: Selected Essays of Wilfrid Sellars. Harvard University Press.
Sterelny, K. and P. E. Griffiths (1999). *Sex and death: An introduction to philosophy of biology*. University of Chicago Press.
Sugden, R. (2015). Looking for a psychology for the inner rational agent. *Social Theory and Practice, 41*(4), 579–598.
Tarnita, C. E. (2017). The ecology and evolution of social behavior in microbes. *Journal of Experimental Biology, 220*(1), 18–24.
Thaler, R. H. (2015). *Misbehaving: The making of behavioral economics*. WW Norton New York.
Veit, W. (2018). Existential nihilism: The only really serious philosophical problem. *Journal of Camus Studies*. https://doi.org/10.13140/RG.2.2.26965.24804
Veit, W. (2019a). Evolution of multicellularity: Cheating done right. *Biology & Philosophy, 34*(3), 34. https://doi.org/10.1007/s10539-019-9688-9
Veit, W. (2019b). Model pluralism. *Philosophy of the Social Sciences, 50*(2), 91–114.
Veit, W. (2019c). Modeling morality. In Angel Nepomuceno-Fernández, L. Magnani, F. J. Salguero-Lamillar, C. Barés-Gémez, and M. Fontaine (Eds.), *Model-based reasoning in science and technology*, pp. 83–102. Springer. https://doi.org/10.1007/978-3-030-32722-4_6.
Veit, W. (2021a). Biological normativity: A new hope for naturalism? *Medicine, Health Care and Philosophy*. https://doi.org/10.1007/s11019-020-09993-w
Veit, W. (2021b). *Health, agency, and the evolution of consciousness*. Ph.D. thesis, University of Sydney. Manuscript in preparation.
Veit, W. (2021c). Review of Nancy Cartwright's nature, the artful modeler: Lectures on laws, science, how nature arranges the world and how we can arrange it better. *Philosophy of Science, 88*(2), 366–369. https://doi.org/10.1086/711505
Veit, W., Dewhurst, J., Do-lega, K., Jones, M., Stanley, S., Frankish, K., & Dennett, D. C. (2019). The rationale of rationalization. *Behavioral and Brain Sciences, 43*, 53. https://doi.org/10.31234/osf.io/b5xkt
Veit, W., & Ney, M. (2021). Metaphors in arts and science. *European Journal for Philosophy of Science*. https://doi.org/10.1007/s13194-021-00351-y
Veit, W., & Browning, H. (2020). Two kinds of conceptual engineering. Preprint. http://philsci-archive.pitt.edu/17452/
Veit, W. (forthcomingc). Scaffolding Natural Selection. *Biological Theory*.
Veit, W. (forthcomingb). Samir Okasha's Philosophy. *Lato Sensu: revue de la Société de philosophie des sciences*.
Veit, W. (forthcominga). Consciousness, Complexity, and Evolution. *Behavioral and Brain Sciences*.
Walsh, D. M. (2015). *Organisms, agency, and evolution*. Cambridge University Press.
Weibull, J. W. (1995). *Evolutionary Game Theory*. MIT Press.
Wilson, R. A. (2005). *Genes and the agents of life: The individual in the fragile sciences biology*. Cambridge University Press.