Biological mistakes: what they are and what they mean for the experimental biologist

Oderberg, D. S. ORCID: https://orcid.org/0000-0001-9585-0515, Hill, J., Austin, C., Bojak, I. ORCID: https://orcid.org/0000-0003-1765-3502, Cinotti, F. ORCID: https://orcid.org/0000-0003-2921-0901 and Gibbins, J. M. ORCID: https://orcid.org/0000-0002-0372-5352 (2026) Biological mistakes: what they are and what they mean for the experimental biologist. British Journal for the Philosophy of Science. ISSN 1464-3537 doi: 10.1086/724444 Available at https://centaur.reading.ac.uk/109845/

Biological Mistakes: What They Are and What They Mean for the Experimental Biologist

David S. Oderberg, Jonathan Hill, Christopher Austin, Ingo Bojak, François Cinotti, and Jonathan M. Gibbins

Organisms and other biological entities are mistake-prone: they get things wrong. The entities of pure physics, such as atoms and inorganic molecules, do not make mistakes: they do what they do according to physical law, with no room for error except on the part of the physicist or their theory. We set out a novel framework for understanding biology and its demarcation from physics—that of mistake-making. We distinguish biological mistakes from mere failures. We then propose a rigorous definition of mistakes that although invoking the concept of function, is compatible with various views about what functions are. The definition of mistake-making is agential, since mistakes do not just happen—at least in the sense analysed here—but are made. This requires, then, a notion of biological agency that we set out as a definition of the minimal biological agent. The article then considers a series of objections to the theory presented here, along with our replies. Two key features of our theory of mistakes are, first, that it is a supplement to, not a replacement for, existing general frameworks within which biology is understood and practised. Second, it is designed to be experimentally productive. Hence we end with a series of case studies where mistake theory can be shown to be useful in the potential generation of research questions and novel hypotheses of interest to the working biologist.

1. Introduction

If one thing is evident in biology, it is that organisms make mistakes. How do we know? The simple reason is that we are organisms and we make mistakes. We forget to turn off the lights at night; we add up the bill incorrectly at the restaurant; we play the wrong card at bridge. The adage ‘to err is human’ is certainly correct, but are we human organisms special when it comes to mistake-making?

Accepted January 27, 2023; electronically published March 17, 2026.

The British Journal for the Philosophy of Science, volume 77, number 1, March 2026.
 https://doi.org/10.1086/724444

We are special inasmuch as we have properties not shared by other organisms[1]: rationality; language; a distinctive kind of awareness, especially of why we do what we do; free will; responsibility, including the moral kind. All of these play a part in the mistakes to which we humans are prone. It does not, however, follow from the fact that we make mistakes involving one or more of these properties that other organisms do not make mistakes. Nor does it follow from the fact that they make mistakes that they must have the properties we do in virtue of which we make mistakes, such as rationality or freedom. It might be that other organisms make mistakes in virtue of properties humans do not have. More likely, though, is that they make mistakes in virtue of properties we and they share, for example, memory or the power of locating objects in space or time.

Here are some examples of what look like mistakes made by organisms: a frog misses catching an insect with its tongue by a few millimetres; a fish takes the bait; a cat jumps on a shrew, thinking it is a mouse—and promptly spits it out since shrews taste and smell unpleasant to cats. There is no question of size bias since mistakes can be made by the very small: bacteria can be confused by plants mimicking signals used by the bacteria for infection of the plant (Bauer and Mathesius 2004).

Mistake-making is not, however, limited to organisms, or so it seems. Sub-systems of organisms, and the members and parts of these systems, also look mistake-prone: an antibody is fooled by a pathogen variant that gets around the adaptive immune system's memory response (Wildner 2023); the meningitis bacterium *Neisseria meningitidis* is able to evade the innate immune system by mimicking the appearance of human cells, thereby inducing the complement system (part of the innate immune system) erroneously to refrain from activation against the pathogen (Schneider et al. 2009); competitive inhibitors can trick an enzyme into binding to a molecule different from the correct substrate (Tymoczko et al. 2015, pp. 134–35)[2]; DNA polymerases make mistakes in DNA synthesis (Lee et al. 2016), among the many kinds of errors that are made throughout the system of genetic transcription, translation, and regulation.[3] Again, groups of organisms also can be mistake-prone. Lemmings may not really jump off cliffs to their death (as opposed to being pushed; see Mikkelson 1996), but migrating birds crash into skyscrapers and aeroplanes (a mistake predicated of the whole group as a group as well as of the individual members), and domestic hens insist on trying to hatch golf balls (a mistake predicated of all the individual members of a group but not of the group as a group).[4]

[1] We recognize that there is much disagreement about the contents of this list; but that at least some of these features belong on it should be agreeable to all.

[2] Note that this does not assume that there is a single substrate for every enzyme. Multi-site enzyme binding also occurs and may or may not involve a mistake, depending on context. Competitive inhibition occurs at a single site.

[3] See, for example, (Potapova and Gorbsky 2017).

[4] See <flockjourney.com/golf-balls-dont-hatch-but-they-help-manage-a-broody-hen/> (last accessed 29 January 2026). Farmers use this mistaken behaviour to manage their hens. Another example is the herring gull, which Tinbergen (1969, pp. 145–46) notes is far better at distinguishing its own young than it is at distinguishing its own eggs.

We propose that mistake-making is a universal feature of biology, demarcating it from physics and chemistry and rendering it irreducible to either or both. The simple reason is that whilst biological entities make mistakes, physical (and chemical[5]) entities do not. The only mistakes in physics are those made by physicists. A mistaken physical theory is one proposed by a physicist. Electrons, however, do not make mistakes.[6] There is no such thing as an electron's making a mistake by heading in this direction rather than that or being attracted to this proton rather than that. In biology, by contrast, there are both mistakes made by biologists and mistakes made by biological entities.[7] We submit that the reality of biological mistakes (as opposed to mistakes made by biologists) and the unreality of physical mistakes (as opposed to mistakes made by physicists) should be a lynchpin of the ontological anti-reductionist case; but the reduction question is not our question here.[8]

In addition, we do not argue that every biological entity can make a mistake, at least if 'biological entity' is taken so broadly as to include all components of living beings. For example, electrons in the electron transport chain, part of the respiratory pathway, do not make mistakes: they obey the laws of physics just as electrons do outside a biological system. Electrons outside such a system are purely physical entities, unlike the physical entities that operate within biological systems. Either way, electrons do not make mistakes. By contrast, if someone incorrectly takes a drug (such as tenofovir; see Ramamoorthy et al. 2014) that damages their mitochondria and consequently the electron transport chain that occurs there, they will have made a mistake. Moreover, it is a nice question whether every biological entity can make a mistake even if we exclude the proper objects of physics such as atoms and subatomic particles. Can every cell make a mistake, or every organelle? We do not address this question directly here, but what we have to say later on will help to clarify how to go about answering such a question. That said, our general thesis is that mistake-making is found wherever there are living systems, at least organisms, but also groups and sub-systems. Hence we take mistake-making to be a potential universal framework for biological research, complementary to existing frameworks.

In particular, we take mistake-making to be a way of operationalizing the concept of teleology in biology by generating novel, testable hypotheses of interest to the working biologist. Before giving examples, in section 2 we will outline the distinction between mistakes and what we call 'mere failures'. In section 3 we will give a formal definition of biological mistakes, followed by elaboration and clarification of the definition. In section 4 we give a formal definition of what we call the minimal

[5] 'Chemical' will henceforth usually be omitted unless needing explicitly to be mentioned.

[6] Purely physical artefacts can make mistakes, of course—software, hardware, tools, constructions—but they will only do so by reference to a biological entity, namely, the living thing that constructed, uses, or has a purpose or function for that physical entity. We do not consider artefacts here.

[7] Henceforth we will usually speak only of organisms both for convenience and because these are the focal case. We will speak explicitly of non-organismic entities when necessary.

[8] This is addressed in forthcoming work.

biological agent, sufficient to underwrite attributions of mistake-making to biological entities. In section 5 we state and respond to a series of objections to the theory of biological mistakes we have set out—the ones most likely to occur to a biologist or philosopher of biology mildly sceptical of the very idea of mistake-making by biological entities. Finally, in section 6, we show that novel, testable hypotheses can be generated by our framework, focusing on selected systems of interest. It is our hope that working biologists in particular will find fertile ground for further hypothesis generation, thereby demonstrating the productivity of the theory.

2. Mistakes Versus Mere Failures

All living systems are subject to failure—failure to survive being the most obvious. Failure to reproduce, to stay healthy, to avoid damage, to escape predation, to adapt to a new environment: such is the stuff of extinctions, of blights on populations, and also of various kinds of natural selection and novel adaptation. Living systems are subject to standards of correctness applicable to their kinds. An incorrect immune response leads to disease, just as much as starving to death in the desert is a departure from correct operation or function—taking in nutrition and hydration to stay alive. It is not our purpose here to address the foundational question of standards of correctness in biology.[9] Rather, we point out that if biological mistakes are ubiquitous, we need a minimal concept applicable across systems. We have already pointed out that the concept of mistake must not be restricted to the mistakes we humans make. At the other end, however, we need to distinguish mistakes from what we call mere failures. For example, when the dinosaurs were purportedly wiped out by a meteorite, they did not make a mistake. Colloquially, they did not get anything wrong. It is not just that there was nothing they could have done about it; more importantly, it was not anything they did that caused them to be wiped out. It just happened to them. Again, starving in the absence of food, or being ravaged by disease or killed by radiation, all things being equal,[10] are things that happen to an organism, not something it does. These are examples of what we mean by mere failures that do not rise to the level of mistakes even though, in a broader and less strict sense, one could call all such events 'mistaken' inasmuch as they involve the departure from standards of correct functioning for the system in question.

Mere failure is a broad category encompassing various kinds of what we might call disruption or damage to a system. An organism can be destroyed by brute physical contact (witness the dinosaurs[11]). It can suffer chemical damage or disruption (toxins, deficiencies). It can fail through lack of opportunity (starvation). Or it might fail to adapt to a new environment. Some mere failures (starvation through sheer lack of

[9] See (Oderberg 2026).
[10] One can always fill in details to give such scenarios a different interpretation, but leave that aside.
[11] At least in the popular version of the story; the full story of dinosaur extinction is of course far more complex.

food) involve what Garson (2019, p. 128) has called an 'uncooperative environment'. Others involve a kind of bad luck (perhaps being wiped out by a meteorite falls into that class). Still, a mere failure as we define it is not necessarily a case of bad luck or an uncooperative environment. If an organism has, let us suppose, an immune weakness, it will not be due to bad luck or an uncooperative environment if they get a disease that exploits that weakness. Yet this is still a mere failure rather than a mistake by the organism. As we have already noted, parts and sub-systems can make mistakes, so exploitation by a pathogen of an organism's immune disease might still involve mistakes made at a sub-organismic level—say, by antibodies that are fooled by the pathogen. Mere failures, then, are a broader category than bad luck or an uncooperative environment.

The cases we imagine are often *ceteris paribus* situations—though not always, as we will see—inasmuch as a mistake might not have been made by a system or subsystem in a given case: the proverbial deer caught in the headlights made the mistake of not moving out of the way, yet perhaps it could have, or the lights might not have been so dazzling; the beaver could have found dam-making materials in its new habitat but did not look hard enough. The point is that we should not foreclose empirical investigation. In our spirit of operationalizing teleology, we want to ask, for any relevant scenario: has the organism made a mistake, or has it merely failed? This should be a spur to further empirical research, not an *a priori* pronouncement from the philosopher's armchair.

Further, it is not as though mistakes and mere failures cannot co-exist. On the contrary, they are likely always found together, but again this is a matter for further investigation. For example, an organism struck by disease may not have made any mistake—compare drinking polluted water to being contaminated by radioactive rain—but perhaps one of its antibodies did make a mistake, such as being tricked by a pathogen. What is a mistake at one level (speaking loosely of 'levels') might cause a mere failure at another: eating poisoned food might cause a cardiac arrest. At the same level, being hit by a car due to texting on one's phone is a failure caused by a mistake. Mistakes can also be caused by mere failures: walking dizzily off a cliff after being hit on the head by a rock is a mistake caused by a mere failure at the same level; immune deception caused by exposure to a disease would be a mistake caused by a failure at different levels. Sometimes the pairs of mistake and failure are attributed to the same entity (for example, one and the same organism), in other cases to different entities (for example, organism and antibody). In short, there is likely a network of causation between mistakes and mere failures, at the same level, and predicated of the same or different entities, as well as at different levels.[12]

[12] In addition, we take omissions to be potential mistakes as well, so when we talk about agency we include omissions, not merely 'positive' actions. A failure to tie one's shoelaces before leaving the house is a mistake by omission, as is a deer's failure to cross the road in a timely way. Again, it is likely that all mistakes are a combination of omission and commission. That said, the question of how to identify and individuate omissions has been long debated, and raises metaphysical questions that cannot be addressed here.

A further distinction is between mistakes that are avoidable and those that are unavoidable: neither is a mere failure. For example, it is known that domestic hens will try incessantly to hatch golf balls and other roundish, vaguely egg-looking objects. Indeed, farmers exploit this behaviour to manage broody hens, so in this sense the behaviour is a fortuitous mistake for farmers, but a mistake nevertheless by hens. It is simply incorrect for a hen to try to hatch a golf ball. But the mistake, as far as we know, is unavoidable.[13] We do not know whether any farmer has ever tried to condition their hens not to sit on golf balls, but we suspect it would be a waste of time. Such behaviour is an unavoidable mistake because it is not in the nature of the domestic hen to be able to make fine enough discriminations to sit only on eggs. If it is unavoidable, why not then call it a mere failure? The twofold reply is that, first, it is agential behaviour: the hen tries to do something it cannot do. It is not a failure that results from something happening to the hen, but from something the hen does (sit on the wrong kind of thing) and tries, impossibly, to do (hatch that thing). Second, the kind of unavoidability involved—due to its kind membership, not due to exogenous factors—means the hen has not failed to be a hen. If it caught a disease that made it topple over rather than hatch eggs, it would have failed to live up to the standards of the normal hen. But when a normal—healthy, well-functioning—hen sits on a golf ball, there is no further standard to which it has failed to live up. The same goes for a mouse or fish that, given its nature, is unavoidably attracted to a certain kind of bait. In short, behaviour is no less mistaken for being unavoidable due to the very nature of the entity involved.

Other mistakes might be avoidable. For all we know, it is possible to train cats—or maybe some breeds of cat—not to jump on shrews. Perhaps a squirrel that miscalculates the distance it has to jump to get to the feeder tray can be trained to calculate more accurately. Equally if not more interesting are sub-systemic cases, where much research needs to be done. A mere failure of the immune system would be something like damage, say by radiation or drugs, that leads to a failed antibody response. An unavoidable mistake, however, would be lungs inhaling toxic fumes in an oxygen-depleted environment: it is simply in the nature of the organism to breathe, especially when starved of oxygen, even if the air is noxious. But can gut microbiota be trained to resist pathogens that would otherwise overwhelm them? Or do they unavoidably act in such a way as to make for a hospitable environment for such invaders?[14]

Although mere failures raise all sorts of interesting questions, for instance concerning reductionism and the demarcation between biology and physics, it is mistakes in

[13] Neander (2017, p. 116) briefly mentions the example of a toad trying to catch a cardboard cut-out 'worm', reporting that neuro-ethologists describe the response as 'appropriate'. What she says seems consistent with the thought that the response, while not strictly a malfunction, might still be a mistake. Yet the context is not clear, and the vague term 'inappropriate' could be understood in multiple ways. On our theory, this looks to be a case of mistake without malfunction—about which we discuss further below.

[14] Training in some cases seems to be possible according to the latest research, but there is much we do not know; see (Stacy et al. 2021).

the strict sense—both unavoidable and avoidable—that can particularly stimulate novel ideas and hypotheses for empirical investigation. Mistakes raise questions about agency in biology and where it is found. They encourage thinking about mistake-prone operations and correlative notions. Whereas mere failure raises interesting questions about, for instance, prevention and repair, mistakes (both unavoidable and avoidable) raise further important questions to do with the agential aspect of incorrectness, such as learning, conditioning, correction, avoidance, memory, control, and regulation.

3. How to Define Biological Mistakes

We need to provide a rigorous definition of biological mistakes. It will be seen that the various parts, as we elaborate them, defuse several objections to the very concept of a biological mistake that may arise in the mind of a sceptical reader. We first present the definition and then provide explanatory notes.

Def$_{\mathbf{m}}$: Action A is a mistake, M, made by system S = df A threatens (i) the correct function F of S in environment E (i/e, t, p) or (ii) the correct function F_p of $S_p < S$ in E or (iii) the correct function F_w of $S_w > S$ in E.

Def$_{<,>}$: $X <,> Y$ = df (i) X is a proper part of Y or (ii) X is a member of Y, (i) X has Y as a proper part or (ii) X has Y as a member.

Def$_{\mathbf{t}}$: A threatens the correct function F of S (mutatis mutandis for F_p, S_p, F_w, S_w) = df (i) A actually departs from F or (ii) A would depart from F if A were to remain unmitigated.

Def$_{\mathbf{d}}$: A departs from F = df A causes a state of affairs S* inconsistent with F.

Def$_{\mathbf{g}}$: A is mitigated with respect to F = df the causal link is broken between A and a state of affairs S* inconsistent with F.

3.1. Environment

Def$_{\mathrm{M}}$ refers to the environment E because, on our account, behaviour is mistaken only in relation to the environment in which S finds itself. The environment is the spatiotemporal location in which S can act or be acted on, hence the reference to time and place (and, by implication, whatever is in that time and place). Further, E can be external (e)—the environment within which S is located; or internal (i)—the environment located within S. An example of the latter would be S's ingesting a toxin that damaged its internal organs.

3.2. Parts and members

Def$_{\mathrm{M}}$ is broadly formulated, via < and >, such that a mistake can be made not only by S relative to itself, but also relative to a part or a member of itself (subscript p)

or relative to some whole (subscript *w*) of which it is a member or a part. So mistakes can be made by members of species, parts of members of species, and species themselves.

3.3. Threat

One of the hallmarks of post-Darwinian biology is that all systems operate with risk. There is risk to my function every time I walk down the street, but this does not entail that walking down the street is always a mistake. Every organism and every species faces threats to its function, including the inability to reproduce (hybrids and natural sterility aside) and, ultimately, death. Merely being alive is risky but living is not a mistake. Our definition must respect this fact. Why not, then, stipulate that threat to function must be 'significant'? Significance cannot be quantified, even in a species-specific way. There can, we submit, be no magic threshold quantity that turns a non-mistake into a mistake or vice versa. Perhaps this might be the case for some specific systems in specific environments where there is very little room for manoeuvre, but this would have to be the exception, not the rule. Moreover, it is something to be tested for, not pronounced upon in advance of evidence. We should, therefore, interpret the term 'threat' broadly and qualitatively, albeit subject to the clauses in the definition: *A* threatens correct function *F* just in case *A* actually departs from *F* or would do so if *A* were not mitigated, with the definition of mitigation in place.

3.4. Discussion

This leaves the all-important term 'function' for some elaboration. We are not here, it should be emphasized, seeking to reopen or advance the long-standing debate about biological function.[15] Our general framework presupposes no more than a broadly normative concept of function, whereby an organism (or group, part, member, or trait) functions properly or normally when it acts effectively in its environment, which means at a minimum surviving, being healthy and integral, and reproducing.[16] Whether a selected effects theory gives us the right criterion for identifying such functions (see Millikan 1984; Neander 1991), or a causal role theory (Cummins 1975), or some other, cannot be assessed here. Instead, we confine ourselves to remarks directly relevant to the project of a theory of mistakes.

We distinguish between two senses of function, what might be called the 'genitive' and the 'agential'. The genitive has either the form 'the function of *x* is (to) *F*' (for example, the function of the heart is circulation of blood) or '*F* is a function of *x*' (growth is a function of plants). The form of the agential is '*x* functions as a *Y*' (the

[15] For an excellent overview, see (Garson 2016).

[16] Which environment? Any it finds itself in, or the environment for which it was naturally selected, or any environment to which it can adapt? We do not offer an answer here. For interesting discussion, see (Brandon 2014).

bird's tail functions as a rudder),[17] '*x* functions to *Y*' (teeth function to grind down food), '*x* functions by *Y*-ing' (the elbow functions by bending), and expressions similar in meaning. There are obvious conceptual connections between these two senses, but they are not the same target of investigation. Okasha (2018, p. 29), for instance, wrongly restricts the term 'function' to talk of traits, by contrast with talk of agency. This fails to make explicit that traits themselves can be a form of functioning (for example, pumping blood) and that organisms function, agentially, according to their traits (walking as a manifestation of bipedality), in furtherance of traits (eating food so as to maintain a healthy size), and so on.

Mistake-making is essentially concerned with both the agential and genitive senses of 'function', albeit in different ways. Agential function is fundamentally about how an organism operates given its nature as the kind of thing it is; for instance: 'platelets function by aggregating via the changed shape of their GPIIb/IIIa receptors' or 'bats function by flapping their wings'. It is often clarified by the genitive sense—how *S* operates when carrying out one or more of its functions; for example: how platelets operate given that it is a function of platelets to stop blood loss, or how a bull behaves given that it has horns. If the operation or behaviour is mistake-prone, the kinds of mistakes that can occur will be identified as mistakes, in large part, in terms of the genitive function subserved by that operation or behaviour. Further, we can ask what implications a mistake made by *S* has for one or more of its functions—remembering all the while that our definition encompasses wholes, parts, and members across systems and sub-systems, so that a mistake made by *S* might have an implication for the function of a larger system to which it belongs, or of a system belonging to it, and so on. For example, when carrying out its locomotive function (genitive), a fish swims towards (agential) and tries to swallow (agential) a plastic bait, which leads to its failure to survive (genitive), and perhaps, if its conspecifics all do likewise, the extinction of the species of which the individuals are members.

4. The Minimal Biological Agent

Given that agency plays a central role in mistake-making—mistakes do not just happen, they are made—and given that we consider mistake-making to be found across all of biology, we are bound to offer a suitable definition of the relevant kind of agency. What could it be? Here, we propose that we need a concept of the minimal biological agent (MBA). Minimality is needed because of the proposed universality, parallel to our claim that we should not treat the distinctively human characteristics of mistake-making as essential features of mistake-making in general.

[17] The sense of 'functions as' here is not that highlighted by Millikan (1989), for whom the expression denotes non-proper functions such as 'the rock functions as a paperweight'. The agential sense used in our discussion, as in 'the bird's tail functions as a rudder', or 'the heart functions as a pump', presupposes that the function being discussed is proper, as per our broadly normative conception.

It will be seen immediately that the concept of agency on which we focus is not conditioned by historical notions or the exigencies of evolution by natural selection. Much of the discussion of agency in contemporary philosophy of biology (not that there is a lot of it in the first place), by contrast, concentrates on specific questions such as whether natural selection itself is agential in character and whether organisms act, in a literal sense, so as to maximize their own or inclusive fitness according to one or other theory of what maximal fitness in evolution amounts to.[18] Whatever the answers to such questions, and while agreeing that historical considerations—how a trait came into existence, how an organism came to be naturally in one environment rather than another—can offer good evidence of agency here and now, our concern is precisely with the current status of agency in a given system. Without wishing to contradict whatever is true on other conceptions of biological agency, then, we propose a more abstract and metaphysical conception suitable for grounding a theory of biological mistakes.

First, we have to organize the idea of the MBA around the intuitive and, to our minds, inescapable distinction between what an entity does and what happens to it—action and passion, to use traditional philosophical terminology.[19] An organism has powers to act (active powers) as well as powers, sometimes called liabilities or susceptibilities, to be acted upon (passive powers).[20] It would, of course, be circular to define the action/passion distinction in terms of the active–passive powers distinction. Moreover, it is likely that the distinction between acting and being acted upon is so fundamental to ontology that it resists definition in other terms, much like causation itself, or identity, or existence. That said, a definition of some phenomenon, *P*, can also function as a criterion of what counts as a case of *P*, on condition that the scope of the definition is at least extensionally equivalent to the scope of *P*. In this spirit, and to this end, we offer the following definition of the MBA:

(1) x is a MBA_{tr} (on y) in a given interaction if and only if
  (1.1) x persists and y ceases to exist; or
  (1.2) x persists and y comes into existence; or
  (1.3) x does not change intrinsically but y changes intrinsically.
(2) x is a MBA_{tr} on itself if and only if
  (2.1) there is an interaction $I(y,z: y,z < x)$; and
  (2.2) y is a MBA_{tr} on z or vice versa.
(3) x is a MBA_{intr} (with respect to y) on a given occasion if and only if y is the goal for which x is a MBA_{tr} on itself.

[18] See (Okasha 2018) for an excellent survey and discussion, and also (Walsh 2015).

[19] Okasha (2018, p. 12) begins with the same intuition but does not subject it to metaphysical scrutiny, and it does not figure heavily in the rest of the book. See also (Dretske 1999). Note that 'passion' here does not mean 'emotion' or 'feeling'.

[20] The active–passive powers distinction is a staple of Aristotelian metaphysics but is also famously aired—with frustratingly little analysis—by Locke (1975, II.xxi.2, p. 234ff.).

Again, the definition needs some unpacking. First, the concept of an interaction is assumed, and by this we mean a causal interaction; but we can do this without presupposing that the interaction is symmetrical or asymmetrical, and whether, if asymmetrical, it marks a genuine case of action or passion. Moreover, we assume that one or more elements of the definiens obtain due (explanatorily) to the interaction and not some other, simultaneous interaction that may or may not involve some third entity. We do not offer here an account of the explanation involved or of the individuation of interactions.

Second, whereas the first two clauses concern transitive agency,[21] be it by an agent on another entity or by an agent on itself, the third concerns intransitive agency. The distinction is needed because an entity can be a genuine agent—doing something rather than suffering something done to it[22]—with respect to an external entity (that is, not the agent itself or any of its parts) without doing anything to that entity. Think of an organism's identifying some other entity, watching out for it, measuring the distance of the entity from the organism, avoiding it, and so on. The organism is a genuine agent rather than patient, and it is so with respect to some other thing, but without doing anything to that other thing.

Third, clause 1 takes its cue from a recent article by Kuykendall (2024), in which he takes a broadly Aristotelian view of the action–passion distinction.[23] Kuykendall (2024, p. 443), however, defines (informally) an agent–patient interaction as one in which 'one of the two (or more) interacting entities undergoes a change in its kind membership, structure, causal powers, or intrinsic properties as a result of the interaction, while the other does not'. There is much to be said both for and against his account, for which there is no room here. We simply note that a change in kind membership amounts to an essential change for Kuykendall, and this would be included under sub-clause 1.1 of our definition. A change in structure or causal powers would be a change in intrinsic properties, which is covered by sub-clause 1.3. Kuykendall does not include production of a new entity,[24] whereas we do under 1.2. Perhaps more importantly, though, we define a thing's acting on itself (clause 2) and a thing's acting intransitively (clause 3), neither of which is discussed by Kuykendall.

That said, his basic insight is correct in our view—that one thing's acting on another (which does not preclude their acting on each other at the same time) must involve a change in the patient, whether substantial or (intrinsically) accidental. One of the examples he cites is that of an enzyme's acting on a substrate. The enzyme destroys the substrate by contorting and ultimately breaking its bonds in the service of speeding up a chemical reaction within the system. Indeed, the agential language

[21] We use the terms 'transitive' and 'intransitive' in the technical sense here without respecting the precise grammatical or logical meaning.

[22] Again, in a technical metaphysical sense having nothing to do necessarily with pain or harm.

[23] Kuykendall is responding to the criticisms of the active–passive distinction found in the work of Martin (1993), Heil (2012), and Mumford and Anjum (2018), among other writers—particularly powers theorists.

[24] Except in passing; see (Kuykendall 2024, p. 450).

employed in the standard description of enzyme behaviour is remarkable (Tymoczko et al. 2015, pt. 1, sec. 3). If production and destruction literally occur in biology, then the active–passive distinction must be real and not a mere matter of perspective or interest. Further, if changes are not miracles, then biological entities are literally changed by each other: what does the changing is active and what is changed is passive—in that particular interaction.

As for a thing's acting on itself (clause 2), we define this in terms of a transitive interaction between the parts or members (clause 1). We expect that whenever a biological agent acts on itself, at least two of its parts are in a clause 1 agent–patient relationship. Perhaps this is too much to expect, but it is difficult to know what else could explain an entity changing itself. Note that this thought is not a reductionist one, as though the entity was just the sum of its parts. It remains that an organism has its own active powers, as an organized, integral whole, to change itself in various ways; our point is simply that this should manifest itself in the form of organismic parts changing each other.

As for clause 3, the idea is that an agent does not need a patient, at least in biology. Here is where we need to deal with what might seem like a reductio of our definition but is in fact a modus ponens we are happy to embrace. A critic might complain that inorganic agents also exist, assuming, plausibly, they satisfy clause 1 or 2, and maybe even clause 3. But there are no inorganic agents, so the definition of the MBA is faulty. Our response is that there are inorganic agents, they do satisfy clause 1, probably clause 2, though improbably clause 3. (This is an investigation for another occasion.) Kuykendall's (2024) example, a very good one in our view, is the dissolution of salt by water. Water is the agent, and salt is the patient[25]: H_2O molecules break the ionic bonds between sodium and chloride ions in the NaCl ionic compound. The salt does not destroy the water, but the water does destroy the salt. Dissolution, in general, is about as good an example of inorganic agency as one can find.

What, then, makes our definition specifically biological? The answer is simple: the definition of the MBA is, tautologically, of the agent as found in biology. Hence the MBA is a special case of the more general concept of the minimal agent. So when reading and applying the definition, one should presuppose that we are talking about biological entities—entities with features not found in the purely physical, such as normativity and correlative mistake-proneness. The definition of the MBA does not demarcate the biological from the non-biological without presupposing those uniquely biological features, of which mistake-proneness is the subject of the present discussion. It is mistake-making that, in our view, is an important way of demarcating the biological from the physical.

Why not, though, build some of the key concepts of biological behaviour, such as mistake-proneness, explicitly into the definition—also function, survival, health,

[25] A position also endorsed by Lowe (2013), though we do not agree with his overall brief account of the active–passive power distinction.

goals, purpose, and so on? Perhaps we could make all the presuppositions of the definition explicit in the definition itself, but we do not think it necessary or useful. When considering whether something is an MBA according to the definition, we can use the conceptual toolbox of teleology, of natural selection, or of whatever best picks out the distinctively biological realm—but it is the criteria in the definition that identify the agents. The only exception to this idea is our explicit appeal to goals in clause 3, but here again we presuppose that the goal is a biological one, and we interpret 'goal' broadly. The idea is simply that there is an external target of interest for the intransitive MBA, such that the agent does whatever it does transitively to itself in furtherance of interests in respect of some outside goal—once again, activities such as identifying or avoiding a potential predator, attracting a mate, and locating a potential prey come to mind.

5. Objections and Replies

We now consider some of the objections (and demands for further explanation) most likely to occur to an open-minded sceptic—by which we mean someone who does not dismiss *a priori* an ontological demarcation between physics and biology, or life and non-life, and does not reject out of hand the teleological phenomena presupposed by or implicit in our account. We do not pretend that such objections, or others that will arise, can be definitively dealt with here; rather, we see our responses as the launchpad for further discussion and exploration of the issues.

5.1. DNA error

The first thing most biologists would think of when considering mistakes are DNA errors, for example, DNA lesions caused by UV light, copying errors in mitosis or meiosis, and errors in chromosome segregation. How does DNA error fit into the framework? In particular, we can start with the question of whether DNA errors are mere failures or mistakes. We first remind the reader that mere failures, as asserted earlier, are not mistakes in the strict sense used here. They can be spoken of as 'mistakes' in the 'loose and popular' sense and can be investigated in their own right. If a DNA error is a mere failure, it should be researched in those terms. That said, we might wonder whether a DNA error is a mere failure or a strict mistake at all: if it leads to a mutation then why should this be a mistake in any sense? Orthodox Darwinism teaches that mutation drives novel adaptations (and speciation), so where is the mistake?[26]

For this reason, mutations that are often called 'copy errors' should, we submit, be considered neutral in themselves. Whether a variation in genetic processes should be

[26] One possible suggestion is that the mistake would be in the failure of individual organisms to perpetuate the species (and perhaps also a group-level mistake by the species to perpetuate itself); further analysis is not possible here.

called a literal error or mistake, whether loose or strict, depends on its implications for action in the environment, which is just another way of referring to what it means for the system's functioning. We might call DNA transcription mistaken because it is known to lead to harmful mutations (perhaps with high probability), but then it is the system subserved by (or, better: produced by) the DNA that determines whether there is a mistake. Most mutations are harmful, so we reflexively speak of DNA errors even though, considered in isolation from the superservient system, they are neutral.

Would, though, any DNA errors be mistakes in the strict sense? This hinges primarily on the agency question discussed earlier: what makes the mistake? If DNA is just information embodied in the double helix nucleotide chains, then since information is not an agent, neither will DNA be an agent. Whether DNA is just information, and how it interacts with a system's causal structure is, however, controversial (Wills 2016). But DNA and RNA polymerases look like agents: DNA replication errors are attributed to the former, and DNA–RNA transcription errors to the latter. Why are they agents? Because they are enzymes, and enzymes catalyse chemical reactions (recall the Kuykendall example above). For example, RNA polymerase unwinds DNA, guides nucleotides into position for RNA synthesis, builds nucleotide chains, and proofreads its own activity. It also pauses transcription, backtracks, and splits misincorporated nucleotide pairs. The best explanation of this behaviour—and we can leave aside crude circularity objections based on the terminology used here, which is ubiquitous in biology—is that the behaviour is agential error correction. Ultimately, though, the details need to be measured against our definition of the MBA.

5.2. Metaphorical or literal?

How literal is mistake talk anyway? A critic might note that biology is littered with metaphor, for example, mentalistic language literally true only of either humans alone, higher mammals, or some other subset of organisms: thinking, deciding, choosing, imagining, preferring, and so on. How do we know whether any such talk is literal or where the boundary between the literal and the metaphorical lies? And if this foundational problem is unsolved, how can we hope to convince anyone that the language of mistakes (and perhaps agency) is clearly on the side of the literal?

We accept that mistake talk is a kind of teleological talk, and that there is a general worry about teleological language in biology. This is a large, ongoing debate to which our contribution is limited.[27] We take standards of correctness to be real phenomena in biology, and hence departures from them are also real. We also take it that some organisms make mistakes—that the examples given earlier allow of no other description. The more urgent question for us is how far this can be extended: which

[27] Crawford (2020) has a very useful discussion, landing firmly on the side of teleological realism.

organisms, which systems? Should groups really be included? Our focus is the individual organism and its subsystems. We offer mistake-making as a plausibly real phenomenon in need of definition and analysis, not pretending that this is an irrefutable defence of teleological realism. We do submit, however, that it is not enough to insist that mistake talk must be metaphorical: the critic needs to offer an alternative account that better explains the phenomena. Moreover, our ultimate aim is to operationalize the concept of teleology by examining mistake-making and looking ultimately for testable hypotheses of use to the working biologist. We say more in the next section, noting for now that if such operationalization is fruitful at the lab bench, teleological realism will have been given an indirect buttress.

5.3. Anthropomorphism?

Mistake, insists the objector, is a human concept. It carries with it ideas of freedom, responsibility, morality, care and negligence, recklessness, and so on. It is simply illegitimate to apply it to non-human organisms, let alone parts and subsystems.

Our reply is already explicit in what we said at the start of this article. We add here that the same charge could be applied to many phenomena, such as action, consciousness, attention, and desire, among others. There are distinctively human characteristics associated with all of these, yet this does not prevent us from routinely and properly applying such concepts, albeit in qualified ways, to at least some regions of the non-human organic world. Speaking more generally, just because we find our way into a concept from the human side outward, it does not follow that the concept is confined to the human case and cannot be properly extended beyond the human boundary. We will, however, nearly always need to qualify heavily our extensions so as to avoid various pitfalls, such as homuncularism, the pathetic fallacy, confusing consciousness with self-consciousness, misattributions of free will or responsibility, and so on.

5.4. The relativity of mistakes

A critic might insist that there is something unstable, capricious, or objectionably relativistic in the attribution of mistakes outside the human realm. For example, is a polar bear that is transferred to the desert and maintains the same level of energy expenditure (activity) making a mistake? It was not a mistake in the Arctic. Again, suppose natural mutation produces a two-headed *Drosophila*: is this a DNA error? We saw earlier that this depends on the system subserved by the DNA, in the environment of the system. A two-headed *Drosophila* will do badly in the current environment. But suppose it is cryogenically preserved for a thousand years, by which time the environment has radically changed and is now hospitable to two-headed *Drosophila*. It is no longer a mistake. The mutation will be preserved. Was it ever a mistake given the outcome for this organism?

Our reply to this important objection is that we cannot prescribe in advance what the spatiotemporal boundaries are within which something is properly called a mistake or correct operation. Moreover, many human mistakes are no different. Perhaps it is stretching the imagination to propose an environment in which adding up a restaurant bill in a way that was incorrect in our environment would no longer be a mistake in that environment. But we can easily imagine two environments—think UK and USA—in which driving on the left was a mistake in one but not in the other.

Here, the critic might think they can land a killer blow, namely, that on our theory we can never know whether an operation is a mistake—not until the end of the cosmos. Our general reply to this rejoinder is: do not confuse epistemology and metaphysics. We may not know whether something is a mistake in some unobserved (perhaps future) environment, but this does not mean there will be no truth of the matter. More specifically, we add, whether something is a mistake is not an ultimate question, and so we do not need to wait until the end of the cosmos to find out. The characterization of an operation as mistaken will vary with the environment. It is not a decision for the end of time, nor is it an absolute matter. It is relative to environment.

The critic will likely not give up at this point, launching the further rejoinder that standards of correctness are, then, hopelessly unstable; do they always vary? Our reply is that some standards will indeed vary, for example, correct calorie expenditure in a warm or cold environment. Some, by contrast, will not: basic vegetative functions, for a start, such as growth and maturation, nutrition, health, and ultimately survival. The border between variant and invariant standards is a fascinating question begging for further analysis.

Considered generically, some (but not all) standards vary. With the indexing present in our definition, however, no standard strictly varies; an absolute standard (for example, vegetative) is invariant-indexed to all environments. A relative standard (for example, calorie expenditure) is invariant-indexed to a set of environments. The critic might at this point make a further attempt, asking rhetorically whether trivial microsecond-long changes in a highly unstable environment are also covered by the definition. If so, we are purchasing invariance at the cost of triviality. Our final reply to this last shot—noting that there is plenty more to say—is simply that there is nothing trivial about it. Highly disruptive environments can involve radical changes of standards of correctness, which will have serious implications for the organism. That is precisely why relatively stable environments are good for living things.[28]

5.5. Nothing more to mistake than malfunction—or not much more

A critic might claim that what we are calling mistakes are just another kind of malfunction souped up to look like a distinctive phenomenon. If Jim adds up the restaurant

[28] We recognize, of course, that some instability can drive novel and beneficial adaptations.

bill incorrectly, this will, it might be thought, simply be a case of malfunction of, say, the parietal lobe. So what is special about just another malfunction? Our reply is that of course Jim may have a brain lesion wholly responsible for the mistake, in which case Jim has suffered a mere failure, and that failure causes a calculation mistake, but the mistake is superficial and of no intrinsic interest: all that matters is the lesion itself. Not all human mistakes are like that, however, nor are mistakes by other organisms. In other words, it is not an ontologically innocent thought that mistakes are nothing but system malfunctions; on the contrary, it is a highly reductionistic thought, contrary to our anti-reductionist assumptions. To regard biological mistakes as no more than mere system failures in all cases both pre-empts empirical investigation and is itself a controversial philosophical claim in need of defence. We submit that there seems to be a clear difference between the phenomenon of mistake and that of malfunction, even though they might coincide in a particular case.

More interesting, however, is the idea that tying mistakes too closely to malfunction misses the nature of mistakes and their role in biology. Mistakes can equally be tied to normal function, since it is the much vaunted 'plasticity' of living systems—trial and error, experimentation, learning, persistence in the face of obstacles, being organized around a goal or set of goals, and so on—that gives rise to the potential for many kinds of mistake. If the organism can learn from the mistake, its stock of information is updated and corrected, enabling more effective action in its environment) or perhaps in new environments) than previously. Another feature of organismic behaviour is the all-but-ubiquitous need to select or respond only to certain features of the environment, and not others, in order for its actions to be timely—which includes being rapid enough for likely success. This means a trade-off between speed and accuracy, hence the possibility for mistakes. A fish that needed to respond to every feature of an edible prey before striking would not last long in its environment.[29] But its necessary response only to select features of prey (such as length, colour, and shape) is precisely what can lead to the mistakes that enable fishing to be a sport. Not only, then, is the trade-off needed by the organism, but the mistakes to which this can give rise are exploitable by other species. Thinking of mistakes, then, as simply an aspect of malfunction rather than as also being—perhaps more importantly—a feature of function is to miss a significant portion of what makes biological mistakes both important and interesting.

Given these considerations, it is consistent with our broad account of function as behaviour that supports, promotes, and protects the overall flourishing of the organism, that function counting as 'normal' according to a given theory might still involve a mistake. The influential selected effects theory of function, noted earlier, has it that a trait's function is whatever it was selected for: the 'normal' environment is plausibly thought of as the actual environment in which the trait was selected, and

[29] See (Neander 2017, pp. 100–105) on the stimuli of predatory behaviour by toads.

perhaps also relevantly similar counterfactual environments (according to some criterion of relevant similarity).[30]

A selected effects theorist might plausibly argue that thermoregulation by a polar bear in the desert was normal functioning (assuming it to be the same level of regulation as in the Arctic) because it was working in the same way as in the environment for which the trait was selected. The contrast would be with, say, artificial interference with the trait through genetic manipulation so as to get it to work in a different way, whether suited to a hot climate or not. However, the theorist might claim that the thermoregulation was malfunctioning precisely because it was operating in an environment for which it was not selected.[31] The most plausible judgement will likely depend on the details of the case, but it should be pointed out that the fourfold framework of 'going wrong' for organisms, laid out by Matthewson and Griffiths (2017), does not easily classify the polar bear in the desert. Their first way of going wrong, echoing the selected effects theory, is constituted by a 'broken mechanism' that 'fails to perform its function' because it is 'unable to fulfil the causal role for which it has been selected in the recent evolutionary past' (2017, p. 453). Their second, reflecting Millikan's particular view, involves an 'abNormal' environment where 'the mechanism is operating in accordance with its [naturally selected] design but outside the operating parameters for that design' (p. 454). Their third involves an 'inhospitable environment' (p. 455).

The polar bear in the desert does not exactly have a 'broken mechanism' for temperature regulation (as it would if, say, its fur had fallen off through disease), though it is unable to perform the role for which the mechanism was selected, namely, maintaining the right body temperature for health and survival in a polar environment. The desert is an 'abNormal' environment, but it is also simply inhospitable. So is the polar bear malfunctioning—'going wrong'—or not? Take polar bears *A* and *B*. Polar bear *A* is flown to the desert and left there to fend for itself. Polar bear *B* remains at the pole, but the pole undergoes a sudden warming event within its lifetime, taking it to decidedly warmer temperatures. Both bears fail to thrive, yet the first way of going wrong does not determine clearly whether either is malfunctioning; the second way applies to *A* but arguably not to *B* since *B* is still, technically, in its arctic environment; and the third way applies to both, since any environment in which an organism fails to thrive due to environmental conditions is *ipso facto* inhospitable. Yet we still have two bears in, *ex hypothesi*, exactly the same physiological condition for exactly the same reason—they cannot maintain a body temperature conducive to health and survival. Intuitively, they are both malfunctioning for the same reason.

[30] See also (Millikan 1984).

[31] It is not clear which way Millikan (1984) would go on this question. According to Matthewson and Griffiths (2017, p. 454), her view is that function is correct if it is in accord with what the trait was selected for. On the other hand, in (Millikan 1989, p. 295), malfunction is tied to such phenomena as being 'diseased, malformed, injured, broken', just as (Neander 1991, p. 180) ties it to 'disease, deformity, lack of use', though she also speaks of an organism's being 'exiled from their natural environment'.

One problem is that function language is now so bound up with various theories as to be employable with diverse meanings according to the exigencies of those different theories. Considering 'function' as a hyper-technical term whose meaning is relative to a given theory, one might wonder what of substance hangs on whether or not we say that the polar bear in the desert is malfunctioning. On our broad view of function, it is plausible to say that although the bear malfunctions inasmuch as it is unable to maintain its health and well-being (it cannot act effectively in its actual environment), it does not malfunction in the sense of doing anything other than what is in accord with the standard of correctness for that kind of organism—what its nature is, given the species to which it belongs and in light of our best observation and understanding of how that species is built and behaves. What is clear, however, is that given only the facts specified—an arctic creature in a hot environment—neither polar bear imagined above is making a mistake, according to our definition. Mistakes would only enter into the picture if, say, the lack of proper homeostasis caused the polar bear to act in ways that accelerated its demise. These, on our theory, would be unavoidable mistakes.

A critic, rather than simply equating mistake with malfunction, might assert that what we call mistakes are simply malfunctions plus agency. Mistakes are still malfunctions, albeit of a special kind. Our response is that this still misses the distinctive nature of mistakes. The making of a mistake can be a kind of malfunction: if I am distracted and walk straight into a lamp post, then I have cognitively malfunctioned, and my agency in the matter means I have made a mistake (as opposed to the mere failure of being hurled into a lamp post by a hurricane). But a mistake might not be any kind of malfunction at all: the fish that takes the bait does not malfunction, nor does the hen that tries to hatch a golf ball. It's just not in their natures to be able to make the requisite discriminations, which means that treating such behaviour as a malfunction is to miss what is going on and, more practically, can result in wasted time and effort trying to make the organisms get it right.

6. Case Studies for Novel Hypotheses

As indicated above, we see the mistakes framework as a new way of operationalizing teleological concepts, putting them to work in the generation of novel and testable hypotheses or research questions of interest to the working biologist. No doubt some biologists already find the concept of mistake useful in their work. In addition, whether this or that particular hypothesis has already been raised by biologists is less important than that the framework itself be capable of generating novel hypotheses and ways of looking at old phenomena as well as searching for new ones. In this spirit, we propose some specific hypotheses generated by thinking in terms of mistakes. In general, we expect typical mistakes (actual or potential) to involve place, time, quantity, and quality. Typical actions involving these would include mislocation,

mistiming, mismeasurement, misidentification, misclassification, misregulation, and misremembering.

6.1. Case study: *C. elegans*

The much-studied nematode *C. elegans* is capable of associative learning, in particular the association of food with environmental cues such as temperature, smell, and taste. If they are given two cues, one paired with food and one not, then in a 'choice test' (note the language), the worms will choose the cue previously associated with food (Lin and Rankin 2010). A hypothesis generated by our framework would be: *C. elegans* can mistakenly identify the cue.

The short- and long-term memory of *C. elegans* is demonstrated by their being cultivated with food at a specific temperature and then challenged with a temperature gradient where food is absent. Still, they will navigate toward the cultivation temperature and remain there for several hours (significant in a lifespan of one or two weeks). Moreover, the association can be reversed rapidly by presenting a pairing of food with a different temperature (Lin and Rankin 2010). Given that rapid modification of memory demonstrates plasticity of learning and behaviour, our research question would be: can *C. elegans* misidentify temperature or even misremember (forget) previously remembered temperatures?

6.2. Case study: cellular false alarms and miss events

Using striking language, researchers have described working on cell 'decision making errors' resulting from noise and signalling failures (Habibi et al. 2017). Using 3T3-immortalized mouse embryonic fibroblasts, they found that a cell can respond differently to the same input, leading to incorrect cell decisions and responses. One kind of mistake is a 'false alarm event [. . .] where there is no object but noise misleads the system to falsely declare the presence of an object', in this case tumour necrosis factor (TNF) (Habibi et al. 2017, p. 2). Another is a 'miss event', where the system 'will miss the presence of the object' (p. 2). The likelihood of a false alarm or miss event can be computed using probability distributions, where 'decision error probability is a metric defined such that it directly reflects departure of the pathway from normal behavior and its expected response' (p. 7). As the authors note, this kind of mistake—incorrect decisions and choices in the presence of noise—can be found in all kinds of organisms from bacteria to mammals, in addition to the sub-systemic cellular parts they studied directly. In the light of our mistakes framework, two research questions suggest themselves: Is there a cellular mechanism for distinguishing signal and noise, thus enabling direct investigation rather than indirect via probability distributions? And do cells mismeasure the amount of TNF and, if so, what is the underlying mechanism?

6.3. Case study: haemostasis (blood clotting system)

Of particular interest to us is the blood clotting system (Hill et al. 2022). Damage to the endothelial cells of blood vessels stops clotting inhibitors, and the cells secrete von Willebrand factor, initiating haemostasis (vasoconstriction, platelet plug formation, and clotting). At the first stage, platelet aggregation and adhesion at the injury site are activated by exposure to collagen from the damaged area. This suggests the research question: can platelets misidentify collagen and begin unnecessary aggregation and adhesion due to being fooled by collagen-like proteins? Further, if so, what mechanism is involved? A more specific hypothesis to narrow the focus is: the GPVI receptor can bind to collagen-like proteins.[32]

More general timing questions also present themselves, since clotting has to be timely: too early (in advance of injury) leads to unnecessary clotting; too late and the organism can bleed to death. Either way, the question of how the onset and termination of clotting are determined by the system is crucial but underexplored. In addition, the aggregation and adhesion of platelets at the injury site need to be sufficient in order for termination to occur. Timing of termination and measurement of sufficiency are almost certainly linked, but how? Once we know how, we can explore the various ways in which the relevant operations can go wrong, in other words, can be mistaken. Further research questions are, then: How do platelets time the onset and termination of aggregation and adhesion? How do they determine that there is sufficient aggregation and adhesion in order for termination to occur?

We speculate that there are no 'magic' quantities in the offing here: it would be remarkable (a significant discovery) were there to be a specific quantity, or precise range of quantities, such that for a given system clotting terminated simply in virtue of a thrombus falling within that range—whether in terms of size, mass, density, and so on. It is, we suggest, as unlikely as supposing that there was a specific time at which clot formation had to begin; it simply has to be not too early and not too late. Similarly, thrombus size has to be sufficient—sufficient, that is, for the survival and health of the organism. If such a process as blood clotting is to be modelled effectively, then the goal-directedness of the process will almost certainly have to be built into the model from the outset, rather than 'read off' an underlying non-teleological structure.

7. Conclusion

The mistake-making framework is not designed to undercut the existing, highly productive frameworks within which biological research is conducted. Instead, it brings to the surface a way of looking at living systems that is implicit in much biological thinking. It is non-reductionist in spirit and letter, though merely setting out the framework does not prove non-reductionism; this requires further exploration of the very notion of standards of correctness in biological function and how they relate to physics.

[32] The GPVI receptor is a glycoprotein receptor for collagen that is expressed in platelets.

Further questions, that we have only touched on here, await future work. For example, mapping the precise relations between mistake, failure, and malfunction will, we speculate, help to progress the debate over functions. Development of the concept of the MBA will add shape to the overall theory we are defending. What is particularly distinctive in the theory of mistakes as we outline it, though, is that it is designed to yield research questions and associated testable hypotheses—to operationalize teleology in a way that is of direct interest to the working biologist. Any given biologist might still devise a hypothesis in this area without even thinking of mistakes, or only doing so implicitly. There is, however, an advantage in hypothesis generation if the underlying conceptual scheme is spelled out, thereby stimulating biologists to conduct their research on a given system according to the mistakes framework. Whether new and interesting discoveries will result from using the framework is for future determination.

Acknowledgements

The authors gratefully acknowledge the financial support of the John Templeton Foundation (grant no. 62220). The opinions expressed in this article are those of the authors and not those of the John Templeton Foundation.

David S. Oderberg
Department of Philosophy
University of Reading
Reading, UK
d.s.oderberg@reading.ac.uk

Jonathan Hill
School of Psychology
Manchester Metropolitan University
Manchester, UK
jonathan.hill@mmu.ac.uk

Christopher Austin
Department of Philosophy
University of Reading
Reading, UK
christopherja@gmail.com

Ingo Bojak
School of Psychology and Clinical Language Sciences
University of Reading
Reading, UK
i.bojak@reading.ac.uk

François Cinotti
School of Psychology and Clinical Language Sciences
University of Reading
Reading, UK
francois.cinotti@gmail.com

Jonathan M. Gibbins
School of Biological Sciences
University of Reading
Reading, UK
j.m.gibbins@reading.ac.uk

References

Bauer, W. D. and Mathesius, U. (2004). 'Plant Responses to Bacterial Quorum Sensing Signals', *Current Opinion in Plant Biology*, **7**, pp. 429–33.

Brandon, R. N. (2014). *Adaptation and Environment*, Princeton University Press.

Crawford, A. L. (2020). 'Metaphor and Meaning in the Teleological Language of Biology', *Communications of the Blyth Institute*, **2**, pp. 5–24.

Cummins, R. (1975). 'Functional Analysis', *Journal of Philosophy*, **72**, pp. 741–65.

Dretske, F. I. (1999). 'Machines, Plants, and Animals: The Origins of Agency', *Erkenntnis*, **51**, pp. 523–35.

Garson, J. (2016). *A Critical Overview of Biological Functions*, Springer.

Garson, J. (2019). *What Biological Functions Are and Why They Matter*, Cambridge University Press.

Habibi, I., Cheong, R., Lipniacki, T., Levchenko, A., Emamian, E. S. and Abdi, A. (2017). 'Computation and Measurement of Cell Decision Making Errors Using Single Cell Data', *PLOS Computational Biology*, **13**, available at <https://doi.org/10.1371/journal.pcbi.1005436>.

Heil, J. (2012). *The Universe as We Find It*, Clarendon.

Hill, J., Oderberg, D. S., Gibbins, J. M. and Bojak, I. (2022). 'Mistake-Making: A Theoretical Framework for Generating Research Questions in Biology, with Illustrative Application to Blood Clotting', The Quarterly Review of Biology, 97, pp. 2–13.

Kuykendall, D. W. (2024). 'In Defense of the Agent and Patient Distinction: The Case from Molecular Biology and Chemistry', *British Journal for the Philosophy of Science*, **75**, pp. 443–63.

Lee, D. F., Lu, J., Chang, S., Loparo, J. J. and Xie, X. S. (2016). 'Mapping DNA Polymerase Errors by Single-Molecule Sequencing', *Nucleic Acids Research*, **44**, available at <https://doi.org/10.1093/nar/gkw436>.

Lin, C. H. and Rankin, C. H. (2010). 'Nematode Learning and Memory: Neuroethology', in M. D. Breed and J. Moore (*eds*), *Encyclopedia of Animal Behavior*, Elsevier, pp. 520–26.

Locke, J. (1975). *An Essay Concerning Human Understanding*, Clarendon.

Lowe, E. J. (2013). 'Substance Causation, Powers, and Human Agency', in S. C. Gibb, E. J. Lowe and R. D. Ingthorsson (*eds*), *Mental Causation and Ontology*, Oxford University Press, pp. 153–72.

Martin, C. B. (1993). 'Power for Realists', in J. Bacon, K. Campbell and L. Reinhardt (*eds*), *Ontology, Causality, and Mind: Essays in Honour of D. M. Armstrong*, Cambridge University Press, pp. 175–86.

Matthewson, J. and Griffiths, P. E. (2017). 'Biological Criteria of Disease: Four Ways of Going Wrong', *Journal of Medicine and Philosophy*, **42**, pp. 447–66.

Mikkelson, D. (1996). 'Did Disney Fake Lemming Suicide for the Nature Documentary "White Wilderness"?', Snopes, 27 February, available at <www.snopes.com/fact-check/white-wilderness-lemming-suicide/>.

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*, MIT Press.

Millikan, R. G. (1989). 'In Defense of Proper Functions', *Philosophy of Science*, **56**, pp. 288–302.

Mumford, S. and Anjum, R. L. (2018). 'Powers and Potentiality', in K. Engelhard and M. Quante (*eds*), *Handbook of Potentiality*, Springer, pp. 261–78.

Neander, K. (1991). 'Functions as Selected Effects: The Conceptual Analyst's Defense', *Philosophy of Science*, **58**, pp. 168–84.

Neander, K. (2017). *A Mark of the Mental: In Defense of Informational Teleosemantics*, MIT Press.

Oderberg, D. S. (2026). Investigating Biological Mistakes, Springer.

Okasha, S. (2018). *Agents and Goals in Evolution*, Oxford University Press.

Potapova, T. and Gorbsky, G. (2017). 'The Consequences of Chromosome Segregation Errors in Mitosis and Meiosis', *Biology*, **6**, available at <doi.org/10.3390/biology6010012>.

Ramamoorthy, H., Abraham, P. and Isaac, B. (2014). 'Mitochondrial Dysfunction and Electron Transport Chain Complex Defect in a Rat Model of Tenofovir Disoproxil Fumarate Nephrotoxicity', *Journal of Biochemical and Molecular Toxicology*, **28**, pp. 246–55.

Schneider, M. C., Prosser, B. E., Caesar, J. J. E., Kugelberg, E., Li, S., Zhang, Q., Quoraishi, S., Lovett, J. E., Deane, J. E., Sim, R. B., Roversi, P., Johnson, S., Tang, C. M. and Lea, S. M. (2009). '*Neisseria meningitidis* Recruits Factor H Using Protein Mimicry of Host Carbohydrates', *Nature*, **458**, pp. 890–93.

Stacy, A., Andrade-Oliveira, V., McCulloch, J. A., Hild, B., Oh, J. H., Perez-Chaparro, P. J., Sim, C. K., Lim, A. I., Link, V. M., Enamorado, M., Trinchieri, G., Segre, J. A., Rehermann, B. and Belkaid, Y. (2021). 'Infection Trains the Host for Microbiota-Enhanced Resistance to Pathogens', *Cell*, **184**, pp. 615–27.

Tinbergen, N. (1969). The Study of Instinct, Oxford University Press.

Tymoczko, J. L., Berg, J. M. and Stryer, L. (2015). *Biochemistry: A Short Course*, W. H. Freeman.

Walsh, D. M. (2015). *Organisms, Agency, and Evolution*, Cambridge University Press.

Wildner, G. (2023). 'Antigenic Mimicry: The Key to Autoimmunity in Immune Privileged Organs', Journal of Autoimmunity, 137, available at <doi.org/10.1016/j.jaut.2023.102942>.

Wills, P. R. (2016). 'DNA as Information', *Philosophical Transactions of the Royal Society A*, **374**, available at <https://doi.org/10.1098/rsta.2015.0417>.