On the nature of mistakes in nature

Global Philosophy (2025) 35:27
https://doi.org/10.1007/s10516-025-09762-5

ORIGINAL PAPER

On the Nature of Mistakes in Nature

Christopher J. Austin[1] · **David S. Oderberg**[1] · **Jonathan Hill**[2]

Received: 30 November 2024 / Accepted: 3 September 2025

Abstract
Some things happen of necessity, others merely happen to occur – but are there things that happen to occur, but *should not* have? The latter constitute *mistakes* and, prima facie, they are everywhere – from our setting the wrong cutlery at the dinner table to young turtles crawling in the wrong direction to the safety of the sea. As obvious and ubiquitous as they may seem, the question of whether mistakes are *real* is not an unfounded one. For inherent in the nature of mistakes is the core concept of *normativity*, as mistakes imply the existence of states of affairs that are *supposed* to occur, but which unfortunately do not. Whether normativity is a feature of the ontological fabric of our world, rather than an epistemological by-product of the heuristic framework we use to comprehend its denizens and their activities, is a question at the centre of a long-standing debate in the philosophy of science. In this paper, we will ask: what must the world be like if mistakes are really *out there*? In answering that question, we will highlight some central aspects of the nature of mistakes that any ontological foundation which purports to include them must somehow accommodate. After showing that even the most promising ontological framework that might do so – namely, a powers ontology – is seemingly not up to the task, we will propose a novel refocusing of the analysis of the nature of mistakes, one centred on the metaphysics of causal feedback and the concept of organismal flourishing.

Keywords Mistakes · Organisms · Agency · Causation · Powers · *Eudaimonia*

✉ Christopher J. Austin
Christopherja@gmail.com

David S. Oderberg
d.s.oderberg@reading.ac.uk

Jonathan Hill
Jonathan.Hill@mmu.ac.uk

[1] University of Reading, Reading, UK

[2] Manchester Metropolitan University, Manchester, UK

Published online: 22 September 2025

It is commonplace to assume that there are mistakes. The pre-theoretical vocabulary we employ in "everyday" language presumes that we, and other organisms make mistakes. We might mistakenly use the incorrect cutlery at a posh tea, and an owl might mistakenly flee from what it thinks is a snake, but is merely a defenceless caterpillar employing a clever bit of mimicry.[1] But mistakes are not always small, inconsequential errors – sometimes we attribute world-changing events to the occurrence of mistakes. To take a particularly instructive example, mistaken productions of particular protein chains are now widely understood to have kickstarted the evolutionary progression of organismal development in generating novel branches in the phylogenetic tree – these are mistakes upon which the existence of the whole human race crucially depend.[2] Mistakes, it seems, are everywhere: they occur in the inner-workings of the sub-systems that compose organismal life, and the lifeforms so composed – be they mammoth or minute – commonly make mistakes too. Indeed, 'mistake-making' (or, more precisely, the *potential for* mistake-making) might be one of *the* defining characteristics which demarcates the biological from the non-biological.[3]

But what are mistakes, exactly? Given their ubiquity, one would have thought that at least a few of the relevant literatures would be replete with various candidate definitions and explications of what 'mistakes' are and what 'mistake-making' consists in. Unfortunately, this is not so. Aside from their aforementioned importance in evolutionary frameworks, there are, as one might expect, a plethora of discussions about the *impact* of mistakes in the field of molecular biology in particular: cancers are widely regarded as the results of mistaken cellular expression profiles (Bertolaso 2016; Kaneko 2011), and disease in general is routinely characterised as a systemic cascade of dysfunction resulting from operational mistakes often caused by external actors (Matthewson and Griffiths 2017). Despite their importance however, finding a well-defined, systematic, and *ontological* account of what constitutes a 'mistake' is strangely absent from the philosophical literature.

Though we sadly lack rigorous candidate definitions, mistakes are so commonplace that we can at least answer the question of what 'mistake-making' consists in by starting with a formalisation of what we presume is the most intuitive and common layman's definition of 'mistake'. As a first pass, below is what we think we generally have in mind when we think of what mistakes are and how they are made.

(***M***) An agent ***x*** makes a mistake when the bringing about of a particular state **F** by ***x*** in a particular circumstance/causal context ***c*** is supposed to occur, but does not, either in virtue of ***x****'s* failing to bring about **F** in ***c***; or, more generally, ***x***'s bringing about some other state **G** in ***c***.

[1] We refer to the well-studied evolutionary ingenuity of *hemeroplanes triptolemus*.

[2] Standard duplication errors of course, but more recent research has viewed frame-reading errors as an important causal factor in evolutionary novelty (Drummond and Wilke 2009).

[3] We do not have the space here, and it would take the paper too far afield of its aims, to discuss or defend this claim in detail. For a more in-depth discussion, see Oderberg et al. (2023).

M captures the idea that inherent in the very concept of what a mistake is is the failure to bring about a state of affairs that *should have been* brought about. So, according to **M**, my not bringing my passport to the airport for my holiday was a mistake: I *should have* brought my passport (or "brought it about that my passport was in my possession when at the airport"), and I *failed* to do so, thus making a mistake. Or consider again the hungry owl who *should have* feasted upon the lone, defenceless caterpillar: having fled rather than fed, it made a mistake according to **M**, after being successfully deceived into thinking its prospective prey was in fact a predator. According to **M**, the transcription of DNA strands produced in the course of evolutionary innovation are indeed mistakes: polymerase enzyme catalysing the insertion of nucleotides it wasn't supposed to, or failing to insert the nucleotides it was supposed to, constitute failings to bring about the amino acids proscribed by the codon sequences in question. This picture of 'mistakes' that **M** paints is simple – and, we hope, rather uncontroversial – though it contains many distinct constituent conceptual components that we think are important to highlight.[4]

The first component that **M** immediately presents is the central concept of *agency*. Inherent in the formulation of **M** is the idea that mistakes are the purview of 'agents' – that is, entities that are the source of their own actions and activities. Note here that this construal of 'agent' is what is sometimes called a 'minimal concept of agency' – i.e. it has nothing at all to do with consciously initiating or deliberating over one's actions, so non-human organisms are not excluded from being agents in this "minimal" sense (Virenque and Mossio 2024; Ferrero 2022). In this minimal sense, for instance, rocks lack agency, but plants do not: the latter, but not the former, are capable of intrinsically initiating their own actions, rather than having any activity that might be ascribed to them being the result of extrinsic factors. Though they lack intelligence of the sort we're intimately familiar with and which we ascribe to other higher-order animals, the intrinsic environmentally responsive capabilities of plants are often characterised *agentially* (Maher 2017; Marder 2013; Dretske 1999), and this 'minimal' sense of 'agent' has also – rightly we think – recently been argued by Shapiro (2020) to extend to all living cells. There are of course many other components to what constitutes a 'minimal' biological agent[5], but this focus on being the (at least semi-) autonomous *source of activity* is essential in defining what a mistake *is* because mistakes seem to be necessarily the sort of things that are *made* by entities, rather than those that merely *happen to* entities.

As an illustration of this point, consider the case of an aspiring Everest climber who unfortunately perishes on his maiden trek of the mountain. Suppose that the climber was insufficiently prepared – he did not pack the proper protective gear, had neglected the necessary training, failed to hire a sherpa to guide him, and so on – and subsequently, for all of those reasons, he dies on the mountain. *Prima facie*,

[4] Although we are here starting with **M** as the intuitive definition of mistake-making, it's clear that it won't cover *all* instances of the sort of activities/end-states which we'd normally characterise as 'mistakes'. The reader is therefore encouraged to think of **M** as an instructive first-pass from which various lessons about the concept might be drawn, but which is ultimately to be kicked away like Wittgenstein's ladder.

[5] The concept of non-human, biological 'agency' as a general research framework to examine the biological world is being boldly explored by a number of authors, but see Sultan et al. (2021) for some recent excellent work.

the climber has *made a mistake* (a number of them, in fact): the end state of those actions – death – was brought about *by him*, originating in his own activities and decisions, and was *not* the end-state he *meant* to bring about – i.e. successfully scaling the mountain to its peak. As a contrast, suppose now that the climber *was* sufficiently prepared in all of the aforementioned ways and yet, when the day came, he was the unwitting victim of an unpredictable, sudden and swift avalanche which directly led to his death. Here we have the same end state – death – and yet, arguably, no plausible attribution of a mistake. The difference between these two cases is clear: though the outcome is the same, in the former case, but not the latter, that outcome is brought about *by* the entity *qua* an agent. Thus, only in the former case do we have a *pro tanto* instance of mistake-making. Call this the *Agency* requirement.

The second important component of **M** is *normativity*. The very possibility of mistake-making entails, according to **M**, that some states brought about by agents are *meant* to occur (in particular circumstances/causal contexts) and correspondingly, that some such states are *not meant* to occur (in particular circumstances/causal contexts). This, and likely all instances in the biological realm, is a form of *contextual* normativity that is quite clearly 'parameter-dependent': those states that are *meant* to occur – which *ought* to occur – are meant to do so *only* in tandem with certain values of contextual particular parameters being met. Typically these parameters will specify both the intrinsic state of an organism (its current chemical composition that describes its hunger/mating/danger dilemma, its physical constitution and its integrity, etc.) and the state of that organism's environment (which flora and fauna are in proximity, whether geological obstacles are present, etc.).[6]

Mistakes inherently represent the failure to achieve a goal-state, and typically one that is a normally, or regularly produced state.[7] For many, the normativity inherent in biological agents is typically understood as being grounded in evolutionary adaptation: when a heart fails to pump blood, it has failed to perform the regularly produced function for which it was selected.[8] But irrespective of how one analyses the source of biological normativity, what matters for our present discussion is that we are realists about that normativity: there are *oughts* in nature, and they are proscribed by the semi-law-like regularities which govern the chemical-*cum*-kinetic activities which constitute the behaviour of biological agents (Veit 2021; Matthewson and Griffiths 2017). And this biological normativity is essential to mistake-making. The reason is simple: if there were no goal-state for a particular activity that is brought about by an agent, that activity would be incapable of being a mistake. The popular Tolkien

[6] Thus on a metanormative analysis, the characteristic *ought* claims involving organisms and other biological agents, they would likely necessitate being read and evaluated using a broadly 'contextualist' semantics (Henning 2014; Dowell 2013).

[7] We know that there is a mistake that has occurred in our lamps if we flip the switch on the wall and the room does not become illuminated, as that is what normally, or regularly happens. However, for intentional agents like Humans, regularity is not always an indicator of goal-directed activity: we can form intentions with particular goal-states – and thus have actions which are liable to constitute mistakes – which are completely novel.

[8] It would take this discussion too far afield to discuss the intricacies of the long-standing debate surrounding the concept of 'function' as "selected effect" in biology and its relation to normativity. For an excellent overview, see Garson (2016).

saying that "not all who wander are lost" is true, but actually incomplete. The saying should be: any and all who (truly) wander *cannot* be lost. If you are, for instance, genuinely ambling through the forest with no goal, or end-state in mind – not aiming toward arriving at a particular campsite, not trying to find a specific mushroom patch, etc. – then *wherever* you end up, the result of your walk *cannot* constitute a mistake. Thus only actions brought about by agents which have proper end-states toward which they are "directed" are candidates for being *mistakes*. Call this the *Normativity* requirement.

The third aspect of mistakes that is inherent in **M** is an important one, and it is this: in order for a given state to be a mistake (or even *possibly* be a mistake), the agent in question must be capable of both being and *not* being in, or producing that state – either by it being possible for it to fail to produce that state, or else by the possibility of its producing some alternate state. Put more simply, if some state **F** is to be a candidate mistaken state, either ¬**F** or some other state **G** must be possible states. This is because, *prima facie*, making a mistake implies that it might not have been made; or, attributing a mistake to an agent implies that the agent could have done differently. So, for instance, consider the pitcher plant, which has a large aperture that lures and captures its food for nourishment. Pitcher plants' apertures aren't always open, and indeed, it would be disastrous if they were: having those apertures closed in low humidity environments is essential to maintaining their internal moisture – if they were open in those environments, that life-sustaining moisture would quickly evaporate.[9] Plausibly then, a pitcher plant having an open aperture in low humidity environments is a mistake. But if pitcher plants did *not* have lids, and were thus *incapable* of closing their apertures, then their being open in low humidity environments could *not* be a mistake. Since there is, *ex hypothesi*, simply no other way the pitcher plant *could* be, it could not be making a mistake being the way it is (i.e. the way it *must* be). So, in order for any state of an agent to be a candidate mistake, the agent must be capable both of being and not being in that state. Call this the *Alternative State* requirement.

The fourth aspect of mistakes that **M** implicitly enshrines is equally important: the crucial role that circumstance and context play in the definition of a mistake. Note that in **M**, the target state **F** is always flagged with a specific context **c**, and the failure to bring **F** about (or the bringing about of an alternate state **G**) in *that particular context* constitutes a mistake. This is no accident as mistakes are always *relative to* specific circumstances or causal contexts – that is, specific sets of entities, activities, and their arrangement into complex states of affairs; more generally, you might say, the 'environment' in which the state is produced. So particular states are only possibly candidate mistakes in the context of certain circumstances. For instance, consider a state one might produce – 'swinging my cricket bat wildly'. Plausibly, without indexing this state to a particular context, it is impossible to get any conceptual grip on whether producing this state might be a mistake on my part. Swinging my cricket bat wildly in the context of being at the wicket with a ball hurdling toward me is

[9] The functional viability of the aperture mechanism in capturing prey also crucially depends upon the humidity level of its environment (Bauer et al. 2009), and the same point could be made using this facet, though the details would be more complex.

generally not a mistake, but producing that *exact same state* (of swinging the bat wildly) when I'm at the dinner table generally *will* be a mistake. So, circumstance and context – and not just the state itself, as we have just seen – are vitally important to characterising any mistake state. Call this the *Circumstance-Index* requirement.

We hope that, on reflection, **M** and the implicit requirements it contains for mistake-making just highlighted are not especially surprising. **M** is meant to express in a simple formalisation the pre-theoretic concept of *what it is for* an agent to make a mistake with which we are all intimately familiar. That familiarity is enshrined by our continued use of the concept of 'mistakes' in an attempt to accurately describe the causal structure of the world around us: "missing the mark", "failing to do what should be done", and generally, "getting it wrong" are all ways in which we think about both our actions and those of the other living denizens of the natural world; and plausibly, when we employ this sort of descriptive language, it is **M** (or something very close to it) that is implicitly in operation. However, sociologically interesting though it may be, our task in this paper is not merely to give a descriptively sufficient account of our mistake-making linguistic practices. Rather, our goal is to attempt to locate an *ontological* foundation for the existence of mistakes in nature.

Getting clear on *what it is to be* a mistake is an important first step toward achieving that goal, since in the metaphysics *of* science, *pace* Meno, one needs to know what it is they're looking for if they are to have any hope of finding it. But whether there is anything matching that description out there *to be found* is another question entirely, and it is indeed an open one. For however widespread the practice, and however natural it may seem to utilise the concept of mistakes to explain the goings on of organismal world around us, as most metaphysicians know, linguistic conventions and the artefacts of our shared vocabulary do not always – and very often do not – reflect reality. It's very common to say, for instance, that one's eyes were bigger than their stomach, or that "time is money", or that plants *love* the sun – all phrases which certainly have an established meaning, but to which we rightly attribute varying degrees of what we might call 'ontological sincerity'. There are, in fact, endless ontological assertions which we all regularly make altogether unthinkingly that are by now deeply rooted linguistic conventions, though nevertheless entirely erroneous: we refer to the Solar System as a singular, discrete object, and we treat time as a dimensionless, unidirectional river, for example; and if some more radical metaphysicians are to be believed, even our most seemingly innocent references to ordinary objects like tables and chairs constitute misplaced attributions of ontological 'objecthood' (van Inwagen 1990; Brenner 2018). Given the general unreliability and unsuitability of our language to accurately track ontology we ought not to assume then that, as prevalent as 'mistake-making' language is as a descriptive device we use to understand how the world works, that the world itself really works in that way.

Well, why *wouldn't* it work that way? What reasons might we have for thinking 'mistakes' are *not* real? One important reason we might be sceptical of the reality of mistakes stems from the fact that they simply don't appear in our most developed picture of 'fundamental' reality – e.g. in the ontological frameworks which characterise the Standard Model of particle physics and Quantum Field Theory. We don't say – nor *should* we – that electrons, for instance, are able to make mistakes. Why not? *Prima facie*, they just aren't the sort of entities to which **M** might reasonably apply:

the activities that electrons engage in are causal resultants of their being negatively charged, and according to the laws of nature that enshrine the essence of electrons, this faculty is *not* context-dependent, *nor* capable of occupying alternate states and so on. Furthermore, mistake-making doesn't feature in any of the explanatory strategies we utilise to understand the behaviour of electrons (or of systems in which electrons are involved). We have no need to appeal to mistakes occurring in the activities of electrons in order to understand *how* or *why* those entities brought about the states they did in some experimental set-up; and plausibly we would find no purchase in making such an appeal, even if we for whatever reason thought it necessary.

It is certainly true that there are mistakes "in physics", in that we have mistakes present in the operations of our experimental set-ups in which electrons (for instance) are involved. But the faults here – the mistakes – are our own: *we* have incorrectly dialled the voltage gauges, *we* have incorrectly calculated the optical trajectories, etc. And in general, these sorts of mistakes involving electrons are present in cases wherein our *expectations* with respect to the predicted or desired outcome(s) of our experimental set-ups fail to match-up to the actual outcome. Importantly, in other words, these mistakes are grounded in *epistemology*, rather than *ontology*, as the failure of reality to match-up to our expectations in our experiments says nothing at all about the 'state of the world', except insofar as that state is not the one we previously predicted or had hoped for.[10] The entities involved in those set-ups have not *made any mistakes*, properly speaking.

With all this in mind, a natural question is: whence mistakes? If there are not – and seemingly *cannot* be – mistakes at the most 'fundamental' level of ontology, one doesn't have to be a dyed-in-the-wool reductionist to see why the claim that mistakes are a genuine feature of the world might be on shaky ground. That 'level' of reality is seemingly populated by entities which do what they do because they *could not do otherwise* and the events they participate in occur precisely the way they *must occur*, and so their activities are wholly independent of any contextual variation in which any plausible sort of 'error' might arise. And so it is understandably difficult to see *how* mistake-making might 'emerge' from a base where they are *prima facie* impossible. There are unquestionably agents and states which satisfy **M** the higher up we get up the 'ontological ladder' (and certainly in the anthropomorphic realm), but the question is: are these mere mirages and only epistemological artefacts of our expectations based on incomplete information about those agents and their relevant activities?

Given that mistake-making is, as we've already said, a ubiquitous phenomena, we'd ideally like to be able to have an ontological framework that is able to sufficiently capture the metaphysical structure of the world which accommodates for the emergent existence of mistakes. We might have little qualms about admitting that quarks and electrons make no mistakes, but if we're thereby forced to deny that

[10] Even if *probability* is a genuine, "ground-floor" feature of fundamental reality – as some interpretations of Quantum Mechanics dictate – it is not at all clear that *mistakes* are possible at that 'level'. For one, indeterminism is not enshrined in **M**'s specification of 'mistake' in any significant respect. Furthermore, here again, any potential mistakes we might declare such systems to make look more to be a case of our (perhaps inescapable) inability to *know* which state the wavefunction will collapse into (for instance) which isn't tantamount to the wavefunction *mistakenly* producing F rather than G (or any other criteria from **M**).

mistakes are *in principle possible*, our understanding of the biological world in particular would be radically affected: just as an organic world without goal-directedness is explanatorily deficient, so too such a world without mistake-making is bereft of descriptive depth. Can we have a philosophically rich account of the nature of mistakes that salvages our pre-theoretical view of the natural world?

1 A Powerful Proposal

The first, and most important step in providing such and account must consist in having an ontology in hand that is "mistake-ready" – that is, one that is capable of supporting, or grounding the requirements laid out in the previous section. Given those requirements, a compelling candidate immediately comes to mind: a powers ontology. An ontology of causal powers, by now a well-known philosophical framework, should naturally strike us as suitably "mistake-ready" for a number of reasons.[11]

Powers – sometimes called *dispositions*, or *capacities* – are, first and foremost, *teleological* properties: they are defined and individuated by their production of particular end-states (their 'manifestations') upon the occurrence of particular trigger conditions ('stimuli'). The philosopher's favourite toy example is 'fragility' – the power possessed by, for instance, the vase the flowers on my desk are in, which is defined by the subjunctive conditional <if sufficient force is applied, then will break>. On a metaphysical analysis of this seemingly mundane fact, we say that properties like 'fragility' exhibit *directedness*: they are *directed toward* the production of certain states of affairs (like 'breaking') given the occurrence of certain circumstances (like 'applied surface tension force'). These sorts of properties are everywhere: from chlorophyll's power to produce the photosynthesis which powers plant life to neuronal axons' prowess in transmitting the action potentials which power our minds.

As a consequence of the very nature of such properties, so the power ontology posits, the natural world consists of entities which are *pointed* toward particular states and *poised* to produce them. In this way, powers are commonly understood to function as the metaphysical seat of causal agency for the entities which possess them. According to a powers ontology, the *activity* that characterises such properties just described is a non-accidental and intrinsic feature of them. Indeed, its ability to offer an analysis of properties whose nomic nature is not metaphysically contingent upon their relation to *extrinsic*, higher-order laws of nature has historically been one of the central reasons that many have adopted a power ontology (Cartwright 1994; Mumford 2004). As sources of their own activity, powers are often construed as playing an ineliminably central role as the causal instigators and architects of the various goal-directed activities that entities engage in.[12] In this way, powers are the paragons of the sort of *agency* and *normativity* required by **M**.

[11] We will here assume familiarity with the basic theoretical framework of a powers ontology. For a more in-depth look at what powers are and for what purposes they have been employed in metaphysics in the last decade, see Damschen, Schnepfand and Stüber (2009).

[12] Powers are often understood as the ground of the *impetus* that would otherwise be lacking in the non-powers populated, Humean "dead world of mechanism" (Ellis 2001).

An important aspect of powers is the *context sensitivity* of their manifestations. The manifestation of powers, as highlighted above, is stimulus-state dependent, in the sense that their production requires activation by a specific local, or environmental cause (or set of causes). The normativity with which these properties are imbued therefore is only causally relevant in particular situations, given that powers are *directed toward* the production of certain end states (their manifestations) in specific contexts (where their stimulus states obtain), and *not* others. This complexity of powers' conditions of activation also extends to their manifestation states themselves, in two ways. Firstly, powers are not *always* manifest: as they do not always, or at all times, receive their stimuli, they are capable of being possessed by some entity and yet not producing their manifestation.[13] So, for instance, the 'fragility' of a vase which is locked away safe in a vault would still exist, even if it never manifested vase being broken. Secondly, powers often – if not always – have more than one manifestation state. Powers are capable of producing many "variations on a theme" – the fragile vase can *crack* when struck, or *break*, or *explosively shatter*, and so on; this can be called *quantitative variation*. Some powers even exhibit *qualitative variation* in their manifestation state: 'negative charge', for instance, produces *repulsion* (when met with a like charged particle) and *attraction* (when met with an unlike charged particle).[14] Taken together, the *context sensitivity* of the manifestations of powers and their corresponding ability to exhibit *alternate states* makes them ideal properties for meeting the aforementioned requirements of **M**.

For these reasons, a power ontology seems perfectly placed to provide an ontological ground for mistakes in nature: the metaphysical nature of a power provides, as we have just seen, all of the necessary functionality a biological entity would require in order to (potentially) make genuine mistakes, according to **M**. But that is only half the story of course. In order to have a proper account of mistakes, we need to understand not only how our proposed ontology *supports* mistake-making, but also precisely *how* the entities of that ontology actually *make* mistakes. With that in mind, we want to know: what constitutes mistake-making from the perspective of the properties themselves?

With **M** in view, a framework for mistake-making – one drawn from the powers literature itself – rather naturally suggests itself: masking phenomena. Originally proposed by Johnston (1992) and Bird (1998), the phenomenon of masking refers to cases where although the stimulus of a power is present, due to the causal influence, or interference of some other entity (the 'mask'), the manifestation of that power fails to come about.[15] As earlier characterised, the existence of a power entails the truth of a subjunctive conditional concerning the entity which possesses it – so, to return to our earlier example, an entity's possessing 'fragility' entails the truth of

[13] Or, on some accounts of powers, of not themselves being in the manifested state – see Marmodoro (2018).

[14] This qualitative variation is sometimes referred to as a power being 'multi-track'. See Heil (2003), Martin (2007), and Williams (2011).

[15] N.B. there are other, increasingly strange ways that the subjunctive conditionals associated with powers might fail that have been dreamed up by philosophers – the literature is full of examples from *finks* to *wizards* - Martin (1994), Lewis (1973). However, these conditions/causes aren't relevant to the current discussion, so we have omitted them here.

the conditional < if sufficient force is applied, then will break>. But now suppose that one were to apply sufficient force to a fragile vase which is covered in a thick layer of protective bubble-wrap: even though 'fragility' remains present (the vase still possesses the power), and the stimulus occurs (the requisite force is applied), the manifestation of that power (the breaking) will not occur due to the causal influence of the bubble-wrap (the mask).

It's not difficult to see how utilising masking phenomena might neatly explicate what mistake-making, according to **M**, consists in within a power ontology, as cases of masking represent situations wherein powers *fail* to *actively* produce the end-states they are *directed toward* when they (seemingly) *should*. Not only does masking phenomena provide a satisfactory theoretical account of mistake-making, but the idea is seemingly easily applicable to a wide variety of mistakes we might want to capture in the biological world. For example, consider a hungry frog waiting on a lily pad. In virtue of the powers of its perceptual system to detect prey, gauge its distance, velocity, and position, etc., the frog is *disposed toward* catching a fly on its tongue (manifestation) when one is in range and it is hungry (stimulus conditions). But now suppose that a fly is in range of the frog and the frog is suitably hungry – so the appropriate stimulus conditions have been met – but the sunlight reflecting on the pond it is sitting in is very intense and nearly blinding at the angle relative to the frog's position on its pad. As a result, when the frog goes to catch its prey, it aims too high and too short with its tongue, missing the fly completely. Here we can say what the mistake the frog has made consists in: its perceptual system is *supposed* to bring about a particular end-state (catching a fly) in these particular circumstances (these hunger levels, this sort of insect in this general proximity), and yet, due to the influence of a 'mask' (the sunlight at a particular intensity and angle), that state does not come about.

The upshot of this section's discussion should by now be clear. By utilising a powers ontology we not only have a source of genuine agential normativity in the natural world, but with the framework of masking phenomena we are also afforded a way of modelling how changes in the surrounding causal circumstances of these properties can result in genuine failures of normativity – that is, according to **M**, mistakes. We have seemingly arrived then at an ontological foundation for the emergent existence of mistakes in the biological world – but our work is not done. For while we can be reasonably sure that an ontology of powers is "fit for purpose", as it were, and have seen the promise of the mistake-making account *via* masking phenomena, we have yet to scrutinise that account.

2 A Mistaken Analysis

Unfortunately, as intuitive and promising as the masking account of mistakes is, it has two central problems. One of them is relatively minor, though pertinent to the desiderata that the paper highlighted in the first section, but the other is rather major and raises a substantial problem with the account. Understanding *why* the second problem *is* a problem will however be an instructive step in our search for an adequate account of mistakes in nature.

The first problem with the account, as careful readers might have already noticed, is that it is simply *too permissive*: it allows mistakes to be much more prevalent than we want them to be. More specifically, it permits mistakes to exist in places we would all likely agree that they *do not* exist and, as a consequence of this (and more relevant to the aims of this paper) it thus fails to be able to give an account of mistakes as a phenomenon unique to the biological realm. The 'permissiveness problem' here is rooted in the ubiquity of powers themselves, as they are of course not limited to the biological realm. As we have already said, for instance, electrons have powers: in the philosophical literature their signature causal contribution 'negative charge', in particular, is very commonly rendered dispositionally *via* subjunctive conditionals. There is of course nothing wrong, as far as we're concerned, with one's ontology being *powerful* "all the way down". The problem however is that the phenomenon of masking on which our current theory of mistakes is grounded is not the purview of a particular and limited class of power, but is instead a generalisable phenomena for any power *qua* causal property whose operation can be rendered conditionally.[16]

Thus, since electrons have powers which are a genuine source of normativity in nature (it is in virtue of the property of 'negative charge' that electrons regularly and actively engage in repulsion/attraction activities), and those powers are susceptible to masks, when that power is *masked*[17], according to the current account of mistakes, those entities are *making mistakes*. This is a problem because, as we've already emphasised in the first section of this paper, pre-theoretically, electrons and their fellow denizens of the 'groundfloor' of our ontology just aren't proper candidates for mistake-making. But more than this, if our account of mistakes allows mistakes to exist at this "level" of ontology, then our previous "whence mistakes?" question (in the first section of this paper) will of course be rendered meaningless, as there will be no genuine metaphysical separation between the *living* and the *non-living* realms with respect to the presence of mistakes.[18] What we want, instead, is an account of mistakes that not only respects this boundary, but explains why it exists.

The second problem with this account is much more important. Leaving aside the worry that the account might render mistakes implausibly ubiquitous, there is a larger problem that looms – one that threatens any account of *what it is to be a mistake* that is based on masking phenomena. The problem is, in short, that a sufficiently deep understanding of the conceptual mechanics of masking phenomena has the potential to render those phenomena fundamentally illusory. And of course, if this is true,

[16] Of course, the 'masks' at different "levels" of our ontology might be more or less prevalent – plausibly, there are less opportunities for regular disruptions of the causal consequences of powers at the more 'fundamental' levels of reality – and what constitutes effective masking will likely vary in complexity at different levels, insofar as the causal processes at the lower level – that is, those whose operation is more closely mapped to 'laws of nature'. It should be noted that there are some philosophers who hold that powers at the most fundamental levels *cannot* be masked, which they sometimes call 'surefire' powers – see especially Williams (2019).

[17] Suppose that two electrons A and B are separated by a thin, solid magnetic barrier C. A should exhibit repulsion (manifestation) when met with a like charge (B) at a certain distance (stimulus), but it does not, due to the presence of C (the masker).

[18] This is an adoptable, perfectly consistent position: one could hold that 'mistake-making' *doesn't* cut across the ontological divide as we've characterised it here. However, this position is, for the purposes of this paper, off limits *per hypothesis*.

given our current account, the same fate will await 'mistakes' in nature. To understand this problem, we need to (briefly) examine the philosophical context in which masking phenomena were first raised.

Although we have thus far been using them as a useful conceptual tool in the context of this paper, masking phenomena have long been the bane of philosophers who wish to endorse a power ontology. Indeed, the *raison d'etre* of 'masks' when introduced in the philosophical literature was to function as a serious objection to the purported powerfulness of powers. Masks, it was argued, have the potential to sever the conceptually tight connection between *property* and *conditional* which lies at the core of a powers ontology. Recall that the teleological directedness of powers – in which their normative 'force' consists – is cashed-out in terms of their *making-true* certain conditionals and counterfactuals concerning the casual activity of the entities which possess them: it is in virtue of the vase's possession of the property of 'fragility', for instance, that the conditional < if struck, then breaks>is true of the vase. This connection between *property* and *conditional* – one that states what it *will* do, or what it *normally* does, in certain circumstances – is the very idea of *what it is to be a power*, and thus a metaphysical prerequisite for any property *being* a power.[19]

When we introduce masking phenomena however, the problem which presents itself is that they represent cases where a particular property (a power, *per hypothesis*) is, although present, fundamentally *powerless* inasmuch as the conditional associated with that property *is not true*. As masks sever this crucial connection between property and conditional, and insofar as that connection is the foundation of the role which powers are meant to play in one's ontology *qua* the ontological underpinning of normativity and regularity in the causal structure of the world, the fact that any and all powers are even *potentially* susceptible to them spells serious trouble for a powers ontology. Naturally then, there is a large literature that consists of attempts to salvage a powers ontology from the problem that masking phenomena introduce.[20] The most prominent of these strategies is especially important here, as it has implications for our current examination of the masking account of mistakes.

Far and away the leading strategy for dealing with 'the problem of masks' for a powers ontology is to "reform" the conditionals associated with powers by employing *ceteris paribus* clauses.[21] The inclusion of a *ceteris paribus* (CP) clause – which declares that "all else is equal" – in the subjunctive conditionals which characterise a power is meant to play an important function, namely *restricting* the states of affairs in which the relevant stimulus occurs and, in doing so, effectively exclude masks

[19] This idea is often called *the Conditional Analysis* of powers.

[20] The majority of these attempts consist in "reforming" the analysis of powers *via* conditionals (Mellor 2000; Gundersen 2002; Choi 2006), though some more recent attempts instead opt to decouple the 'conditional analysis' *of* powers from the *ontology* of powers by characterising powers independently of such conditionals – see especially Vetter (2015), Williams (2010), and Mumford and Anjum (2011). One of us has their own preferred, admittedly radical solution to this (and associated) problems which decouples conditionals from powers, though it isn't relevant here – see Austin and Roselli (2021).

[21] Whether or not the inclusion of *ceteris paribus* clauses in one's philosophical account of the world is acceptable is a matter of great debate, an examination of which would take this paper too far afield. However it should be noted that many philosophers have thought that the formulation of scientific laws cannot be had without them, and that any philosophical framework which hopes to capture the natural world must include them - see especially Cartwright (2012).

from entering the causal equation. For instance, consider again the masking case of the fragile vase wrapped in bubble wrap: the power (fragility) is present, and the stimulus (striking) is present and yet the manifestation (breaking) does not occur. Recall again the conceptual problem here – the fact that the stimulus occurs but the manifestation does not clearly shows that, *contra* our account of the nature of powers, the existence of 'fragility' *does not* entail the truth of its associated conditional < if struck, then breaks>. The CP strategy's response to masking phenomena is simple: the conditional's failing to be true is no problem, since the *proper* stimulus never occurred.

According to this response, the *correct* stimulus for 'fragility' isn't merely 'striking', but rather 'striking *and all else being equal*'. In other words, 'fragility' produces its manifestation (breaking) when struck *and* there isn't a massive amount of bubble wrap, *and* there isn't a sudden change in the molecular constitution of the vase, and there isn't a mischievous wizard casting an "anti-break" spell, and…*et cetera*.[22] The CP strategy for "solving" the problem of masks is then to reinforce the 'stimulus' conditions of powers by redefining their causal boundary, expanding it in a way that ensures that there aren't any relevant "stimulus-defeating/nullifying" causal factors *also* at play within it. In doing so, we preserve the truth-making link between powers and their associated subjunctive conditionals: all fragile vases will break if struck *and all else is equal*, and in cases where fragile vases are struck *but do not break*, we do not have a failure on the part of the power of 'fragility' to produce its manifestation, but rather the simple and expected case of a power not manifesting due to its proper stimulus – one including a CP clause – not occurring.[23]

This intuitive solution to the 'problem of masks' which powers theorists have championed is however problematic for a powers-based account of mistakes. The reason is simple: if we take it – or any conceptually similar solution – on board, we fundamentally remove the existence of 'fail-states' from causation *via* powers. As a result of this, mistakes as we know them, are rendered essentially illusory. Why? Because according to the CP-strategy, purported cases of masking phenomena do not reflect the ontological limitations of powers themselves with respect to *their ability* to produce their characteristic end-states, but are rather an artefact of our own *epistemic* limitations with respect to *our ability* to correctly or sufficiently capture the nature of a power *via* conceptual analysis. When something appears to have "gone wrong" in the causal activity of a power, this is merely our erroneous formulation or understanding of the *correct* stimulus conditions for that power.

So to have an acceptable account of powers – one that isn't plagued by the 'problem of masks' – we arrive at an account wherein no power can truly *err*. Every state a power seemingly "fails to produce" is ultimately, ontologically perfectly expected and accounted for. There will certainly be "unexpected results" and "unpredicted failures" in any experimental set-up in which we might be interested, but there will be no

[22] In this way, CP clauses function a bit like Armstrongian (2004) 'totality facts': they declare that the stimulus includes some particular causal factor (or set of such factors) *and nothing else*.

[23] This might neatly solve the problem of masks, but the concern remains that the inclusion of CP clauses in a subjunctive conditional renders it essentially trivially true, and hence, uninformatively ad hoc. Specifying CP clauses in a non-circular fashion without sacrificing their theoretical utility is no easy task.

ontological errors in powers – everything happens just as it should, given the *specific* and *proper* stimulus conditions at play. Fragile vases wrapped in bubble wrap which do not break upon being struck, for instance, are not exhibitions of ontological errors in manifestation, but rather errors in epistemic calculation: not only are such vases not *supposed* to break in these circumstances, but they're *supposed to not break*. Or consider again the frog who misses the fly. We want to say, intuitively, that the frog has made a mistake, but on our current account, we cannot. We should instead, apparently, look more closely at the stimuli and circumstance of the relevant power(s) and we will find – so goes the claim – that we have miscalculated: we will find that *ceteris* were not *paribus*, and that we had missed something relevant from the causal set-up. Where we will end up, ultimately, is saying something like "Actually, given *this* exact angle of sunlight shining with *this* level of intensity at *this* distance relative to the frog's eye, and given *this* precise condition of the frog's legs, etc., the frog's missing the fly is *exactly* what *should* have happened in this particular circumstance".

Thus our utilisation of powers to ground the possibility of mistake-making *via* the causal framework of masking phenomena is, on closer analysis, ultimately self-defeating, and by employing it we are left with the unwanted result that mistakes are not possible. Where we end up if we take such an approach must seemingly be to admit that while mistake-making may be a common way in which we understand the causal operation of the world around us, the concept has no true ontological correlation, and there are no mistakes being made *in nature*. If we think, as we suggested at the outset of this paper that we ought to, that mistake-making *is* a real feature of the natural world, we must then abandon any account of *what it is to be a mistake* that relies upon the ontologically superficial 'errors' inherent in apparent cases of masking.

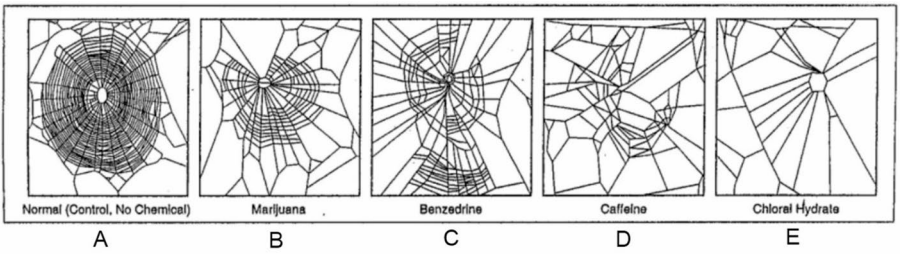
3 What Makes Mistake-Making

Let us take stock. Having adopted **M** as our intuitive starting point to define 'mistakes', we've said that an ontology of powers is our most promising framework to capture *what it is* to make a mistake as these properties function as the ground of intrinsically teleological and thus normative causal pathways of the sort which are subject to failure. However the way in which the failure of normativity that mistakes consist in has been thus far understood – *via* cases of masking – brings us back to our initial problematic position that, contrary to appearances, mistakes are merely an epistemic, rather than an ontological, phenomenon. Where exactly then might we have taken a wrong turn? Is there a way to salvage an account of mistakes as genuine ontological events? We could of course attempt to further explicate how causal powers might engender or suffer failures in normativity and reformulate our account accordingly. However, we do not think this tactic holds much promise. What we will suggest instead is that we take a new approach to defining mistakes – one that departs substantially from **M**.

Where exactly have we gone wrong in our analysis? As we see it, the problem is not in utilising powers to capture the "metaphysical machinery" of mistake-making. As causal difference-makers, powers are a fundamental part of grounding structural

change in the natural world which possess the most central features of the mistake-making process: agency, normativity, alternative states, and context/circumstance dependence. Rather, we want to suggest that where we've gone wrong is in our focus on the *manifestation* of powers – the *process* of manifesting, but more importantly the *end-state* itself – as the locus of mistakes. It's clear that the manifestation produced by a power *is* important here. The power's producing state S*, rather than S, is where the mistake lies, as it is the 'mistaken state'. In defining *what it is to be* a mistake then, the natural thought is that we should focus on the *fact that the state* (S) *did not come about*. Thus our focus on the 'failure of normativity' enshrined in **M** as the foundation of mistake-making. But as we've just seen in the previous section, in locating the "mistakenness" in the process of manifestation of a power we will, upon closer analysis, inevitably render those processes that we *want* to turn out as mistakes as perfectly executed, non-erroneous ones. If the focus on the manifestation or manifestation process is going to give us the wrong result – namely, false positives – where *should* we focus?

Let us return to the fundamentals of mistake-making by looking at a new and illustrative example case where we can employ a walkthrough of our reasoning in identifying mistakes which we think will help us refocus our examination. In 1954, pharmacologist Peter Witt thought he might be able to help his colleagues who were attempting to film a spider creating its web – they could not predict or control when it would do so, and so would regularly miss the moment. Witt thought he might be able to control the spider's spinning process by employing a particular chemical drug to induce the behaviour on demand. After a series of careful experiments with various drugs, what he found, and later published in *Scientific American* (Witt 1954), was an unexpected and rather interesting result. Witt could not help his colleagues by inducing spinning on demand, but what he could do was manipulate the shape and structure of the spun webs with the application of different chemical compounds – see the image below.[24]



An interesting result on its own, but for our purposes, the point we want to make here is that, if we were to observe any of the non-control webs in nature, we would – rightly, we think – designate them as *mistakes*: they look like mistakes in web-spinning and not the *intended* result that spiders are *supposed* to bring about. If we had to give an analysis of these non-control webs, we might want to say something like "webs B-E are *mistakes* because they are non-normative – their shapes and structures

[24] This image is not from Witt directly, but from NASA (1995), who conducted further experiments with different chemical substances. The lettering (A-E) is our own addition.

are *not* the typical and regular results of web-making". In the framing of our previous analysis based on **M**, we might say that spiders have a particular power – call it web-making (**WM**) – that is supposed to produce (upon the satisfaction of specific initial conditions) a particular manifestation (**W**) which consists in a web with a particular shape, structure, and pattern (**A**), and these webs (**B** – **E**) are mistakes precisely because they are instances where **WM** has been *masked*: the power has failed to produce **A**, but because of a disruptive influence, it has instead produced ¬**A** (**B** – **E**).

But we've now seen the folly in this sort of reasoning, and with the aid of this example, we can even more precisely state it. Much as we'd want to, on this reasoning we could not call **B** – **E** mistakes because the web formations they represent *are not* erroneous anomalies. Instead, they are each individually *exactly the sorts of webs we would expect*, given their particular stimuli (the specific drug exposure). None of them might be the sort of web we'd normally *expect* the spider to make (**A**), but none of them would be *errors*, or *mistakes*: they are each *normatively* produced, and the relevant power (**WM**) is doing, in each of these circumstances, precisely what it's *supposed* to do. Interestingly, Witt's (1954: 84) initial study specifically highlights the regularity and normativity inherent in these non-control variations quite well, noting that exposure to Marijuana "produces no disturbance of the sense of direction, but it does cause the spider to omit the first part of the spiral: the animal starts closer to the center and leaves the outer part of the web uncovered by cross members…This effect is peculiar to marijuana – *it is always produced by that drug*, *and only by that drug*, *as far as I know*"; emphasis mine.[25]

The spiderweb example usefully highlights what's mistaken, as it were, about **M**. But if we want to maintain that spiders which produce webs of this sort (**B** – **E**) are, in fact, making mistakes – which we think we do – then what do these mistakes consist in? We know, from the discussion of the previous sections, that the relevant mistakes here are not to be found in any of the particular webs that are produced – that is, not in any of the particularised manifestations of **WM** that **B** – **E** represent – or in the process itself of producing any of those webs, nor in the fact that the web that is *supposed* to be produced (**A**) is *not*. Our suggestion is that the locus of the *mistakenness* of **B** – **E** is to be found in the *feedback effect* that those productions have upon the spider itself *qua* agent. In other words, the reason why webs **B** – **E** are mistakes isn't simply because they aren't the *typical* or *regular* structures that spiders produce, but rather because those web patterns negatively affect the viability of the future effective action of the spiders themselves.

Putting aside **M**, consider again why and when we're apt to identify some action or end-state of an agent as a 'mistake'. One of the reasons has to do to with the failure of normativity enshrined in **M**, but another, equally (if not more) important reason has to do with the *consequences* that the performance of that activity or the production of that end state has on the agent. We say, for instance, that eating an unhealthy meal is a mistake – but why? If I *choose* to eat that unhealthy meal, and it is my *goal* to do so,

[25] Also noteworthy is the fact that the follow-up experiments done by NASA (ibid.) were done explicitly to show that spiderweb typology might be a practical and useful way to reliably detect the exposure levels of particular drugs in the wider environment – a practice that would obviously rely upon the regularity and normativity of such results.

and I do, how could I be 'making a mistake'? What we mean when we call such activity a mistake isn't that the agent who eats that meal has failed to achieve their goal, but rather, that the goal thus achieved has, or will have, negative consequences on the future activities of the agent: their bodily health will suffer, their cognitive acuteness might be dulled, and their ability to effectively perform future actions might thereby ultimately suffer as well. Plausibly, it is this sort of negative *feedback* that the action/end-state produced by the agent has *upon the agent itself* which renders such actions mistakes. Likewise, for the spiders in our example, webs formed in these ways (**B** – **E** and the like) are not efficient, or adept structures for trapping and retaining food for the spider's nourishment, and are thus ineffective at allowing it to replenish the energy levels it needs to continue to flourish in its environment.[26] Such web formations might be typically and regularly produced in particular circumstances (i.e. on the occasion of exposure to specific chemical compounds) – and so be the expected, and seemingly "correct" products of *these* spiders in *these* circumstances – and yet, in the ways just mentioned, have *consequences for* the spiders which produce them which effectively threaten those creatures' future flourishing. It is this, we claim, in which the mistakenness of such webs consists.

A central notion involved in this idea of what it is to be a mistake is the concept of 'flourishing'. What does an organism's flourishing consist in? This is a complex question, but a simple place to start is by showing what it is *not*. Mistakes occur, we've claimed, when an organism's activity threatens its flourishing. In the basic, most general form, we might construe this simply as a threat to agency, or the ability of an organism to continue to carry-out goal-directed action *via* the activation of their capacities. Activities which threaten to, or actually do diminish their future ability to carry-out such actions by limiting their capacity to effectively translate their environment (both intrinsic and extrinsic) into future goal-directed states – and thus, diminishing their ability to express their own *agency* – are antithetical to an organism's flourishing. Clearly, on this definition, an organism performing and action which leads to it being physically *disabled* in any respect is a threat to its agency, and its performing an action which ultimately threatens its very survival – its persistence as a *living* being – counts as a mistake as a kind of edge case: death is the ultimate 'lack of agency'.

Defining 'flourishing' by what it is *not* is certainly an intuitive way to capture *part* of the concept, but it is vitally important to also have a good grip on what flourishing *is*. As we conceive of it, flourishing should be understood as a state of being whose conditions for satisfaction are unique to the *kind* of organism in question.[27]

[26] There are a variety of other reasons one could pick-out that such webs have a negative feedback effect upon the flourishing of the spider. Some web structures might be perfectly adequate for *catching* prey, but feature strands which lack the appropriate tensile strength for the spider's traversal and manipulation of the web to do anything about the captured prey, or, due to the specific structural features of those webs, they might be unable to sufficiently *retain* the prey which have been trapped in them.

[27] Perhaps flourishing is even a *species-specific* state. We won't be stating any upper or lower bounds of the taxonomic tree here, for two reasons. Firstly, 'kind' is not being used in a classificatory way, as in some way above or below any other taxonomic rank. Secondly, it seems to me that 'flourishing' might be specific to various "levels" of being in which mistakes can occur, and a single organism like a spider might be capable of making a mistake *qua spider* but also be capable of making entirely distinct ones *qua arachnid*.

So a "threat to flourishing" will be an action/activity/state which impedes the expression of the *kind-specific* capacities that constitute an organism's particular way of life. This conception of 'flourishing' is roughly Aristotelian, akin to his notion of *eudaimonia*.[28] For Aristotle, an organism's "living well" (*eudaimonia*) requires its successful activation of the capacities that are particular to it – those that constitute its specific *functions/goals* as a member of the kind to which it belongs. A consequence of this is that what it is to "live well" for one organism *will not* in most cases be what it is to live well for another. For instance, to return to our example, constructing certain web patterns (e.g. **B** – **E**) is a mistake *for spiders* because they result in those organisms not being able to carry out the activities which are particular to that kind of organism's lifestyle: such patterns are inadequate at trapping insects for nourishment, or disallow nimble traversal for reaching prey, or are insufficiently braced for harsh winds, and so forth. All of these conditions clearly threaten the proper activation of the spider's capacities *qua spider*, and none of them would have any relevance whatsoever to, say, a turtle, for whom flourishing consists in burrowing for shelter, engaging in migratory hatching rituals *via* magnetic navigation, and so on.

With all of the above in mind, this novel account of what it is to be a mistake is what we will call the *Flourishing Feedback* (**FF**) account. According to **FF**, an organism makes a mistake when it produces a state or engages in activity which threatens to, or actually does diminish its ability to flourish – that is, its ability to effectively translate its environment (both intrinsic and extrinsic) into future kind-specific goal-directed states. What separates **FF** from our earlier attempts to define mistakes ought to by now be clear. Firstly, on **FF**, mistakes are not constituted by a simple *failure* of a particular state to be brought about by some agent – so cases of masking will not be *ipso facto* cases of mistakes. While the failure to bring about a state toward which an agent is goal-directed *may* constitute a mistake for other reasons (namely, those listed in **FF**), it is not the simple *lack* of such a state coming about in which the 'mistakenness' consists. Secondly, according to **FF** mistakes don't consist simply in the fact that the state which an agent produces is not the *normal/typical* state produced in some particular circumstance. Failures of normativity of this sort – *true* failures of normativity – might reasonably reliably track the occurrence of agential mistakes on **FF**, but they also might not: we have just seen the potential perils of equating the two above in this section (with our example of web variations), and **FF** ensures that our theory of mistakes is not apt to produce *false negatives* in cases wherein normativity is maintained; in such cases, flourishing may yet be threatened, which is the true test.

Thirdly note importantly that according to **FF** whether an action/activity/state is a mistake is *not* an *intrinsic* affair – that is, not intrinsic to the action/activity/state itself. In other words, there is nothing *about any particular activity*, *state*, *etc. in and of itself* that has any bearing whatsoever on whether it constitutes a a mistake. Rather, on **FF**, mistakenness is an extrinsic and primarily *relational* affair in two important ways. In the first place, as we have already seen, the mistakenness of a particular state/activity consists in the *causal feedback loop* that holds between the bringing about of that state/activity on the one hand and the organism itself on the other; or

[28] A central concept in his *Nichomachean Ethics (Aristotle* 1984*)*. For a contemporary discussion, see Bielskis, Leontsinia, and Knight (2020).

more specifically, the organism's kind-specific agential capacities. Because of **FF**'s emphasis on the consequences of the interplay of the elements of this feedback loop, the mistakenness of any state/activity is determined, ontologically, not on the character or quality of that state/activity itself, but instead on whether that state/activity negatively affects the flourishing of the organism in question. So, for instance, if I were to jump from a high cliff and flap my arms wildly, that activity would doubtless be a mistake – this is an action which would certainly affect my future flourishing *qua* Human, as my days as an agent would be over rather shortly. However, that very same activity, when carried out by an eagle, would *not* be a mistake: indeed, it is paramount to the flourishing of that kind of organism that it be able to take flight in such ways; for travel, for hunting, etc. The difference between these two judgements with respect to mistakenness is capture by **FF** wherein the activity itself – its intrinsic character *qua* action – is *mistake-neutral*, and is only a mistake according to its causal relation to a particular agent and its kind-specific manner of flourishing.[29]

In the second place, the extrinsic nature of mistakenness is evident in mistakes being *environmentally dependent*; a requirement for any satisfactory account of mistakes, as described in the first section of the paper as the *circumstance-index* requirement. Whether an activity is a mistake depends not only upon the kind of organism engaging in that mistake – as just detailed above – but also upon the environment that activity takes place in. Why think this is the case? Consider that a single activity might be detrimental to an organism's flourishing in its normal/typical environment, but be greatly conducive to its flourishing in a novel niche it might find itself in – and thereby *not* be mistakes therein, according to **FF**. To take our earlier example, a spider's strange webs whose production would normally be mistakes (**B** – **E**, from above) might, in different environments, be better at catching new types of insects that the spider would not normally encounter, or be structurally more sound in the lower average wind speed of a novel environment, and so forth. In other words, those same webs, normally (and rightly) deemed mistakes might, in other environments, actually contribute to a spider's flourishing, rather than threaten it, and hence not be mistakes at all. So even when a particular action/activity/state *is* a mistake, it is so not because of the intrinsic character of that action/activity/state, but rather due to the relational structure that holds between *it* and the organism (or more specifically, the *kind* or organism it is) *and* the environment that it occurs in.

For all of these reasons then, **FF** represents a marked improvement over **M**: not only does it more sufficiently meet the earlier mentioned desiderata for a suitable theory of mistakes, but it also solves a number of problems that **M** (and any theory of its ilk) engenders.

[29] One could also think of hypothetical cases which illustrate the same point. Suppose we observe novel, extra-solar 'aliens' engaging in activities we *know* are harmful for 100% of the Earth's population. We would perhaps be quick to label such activities mistakes, though we may in fact be mistaken, as we do not know the conditions for flourishing for them, and *that* is what counts, according to **FF**.

4 The Ontological Divide

Let us return then to the dialectic we highlighted at the beginning of this paper. There we asked whether a philosophically coherent and satisfying account of *what it is to be a mistake* according to which (**a**) mistakes are not merely projections of our own epistemic judgements, grounded in our having incomplete data or general ignorance and which is (**b**) capable of justifying an objective demarcation of a 'level of reality' at which mistake-making appears. What we want to do now is to illustrate that **FF** satisfies both **a** and **b** in a compelling fashion.

Take **b** first. Due to the two-pronged aspects of **FF**, we can clearly draw a line above which mistakes occur (or are capable of occurring) in the 'ontological ladder' of reality and below which they do not (or cannot) – one that depends upon the way the world actually is rather than the way we, correctly or incorrectly, perceive it to be. Firstly, consider *feedback*. While there are innumerable biological examples – a few of which we've examined earlier in this paper – in the biological realm, the activities of the entities of physics have no integrated causal feedback loops from the manifestation of their powers back into their future activities. There is no genuine sense, for instance, in which the *attraction* of an electron to another electron somehow affects the former's future activities in a positive or negative fashion. There is no *historicity* in either of the electrons in which such causal feedback might be integrated: both entities unfailingly continue on, doing what they do, irrespective of their causal interactions.

Furthermore, consider the other half of **FF** – *flourishing*. Even if there somehow were substantial causal feedback loops of the sort described by **FF** in the realm of physics, it's clear that the entities therein have no kind-specific conditions for 'flourishing'. Electrons certainly have kind-specific activities in which they activate kind-specific capacities, but there is no associated manner of activation of those capacities according to which 'electron-flourishing' occurs, or fails to occur. [30] In other words, there is no sense in which any instance of 'attraction' or 'repulsion' could count – under some suitably objective measure – as being *done well*, or being done *in the service of the (future) good of the election*. The reason for this is simple: one electron cannot 'attract' protons any *better* than any other electron since, as we have already seen, given their lack of *historicity*, there is no causal feedback loop that exists between *agent* and *activity* in which a comparative judgement between two electrons might be reasonably grounded. According to **FF** then, there exists an *ontological* demarcation line which is drawn in the world whenever there are entities which (1) can exercise capacities whose 'results' can causally feedback into the exercise of its future capacities and are such that (2) there is a kind-specific standard for the well execution of those capacities.

[30] On another way of thinking, one might claim that electrons – and their ilk – are in fact the *most* flourishing entities of all, since they *always* and *perfectly* exercise their kind-specific activities, and hence do so *well*. However, this isn't the concept of 'flourishing' at issue in **FF**, and even if we were to accept this line of thought, electrons still would be incapable of making mistakes, as their activities could *never fail* to be conducive to their kind-specific flourishing. Thus our judgement on the matter would be the same: mistakes are not possible at that 'level' of reality.

Now onto **a**. Taking into consideration its two aspects just discussed, we think **FF** also provides an account of *what it is to be a mistake* that is suitably *objective*, rather than merely subjective. In other words, **FF** offers an account of mistakes that picks out facts *about the world* whose character and qualities are independent of the ways in which we choose to conceptualise them. To put it briefly, this is because whether feedback of the sort enshrined in **FF** occurs, and whether that feedback contributes to organismal flourishing are not matters on which our subjective judgements have any say. It is not "up to us", so to speak, whether an entity's activities establish a causal feedback loop which negatively or positively affects their future capabilities for effective agency – and certain actions/activities undertaken by specific organisms either *do so* or they *don't*, irrespective of our judgements on the matter. Likewise, the conditions under which specific organisms *flourish* is also not "up to us": whether 'catching insects' is beneficial to spiders, or whether *this sort of web structure* is bad or good at catching insects are not matters on which our personal, perspectival judgements have any say.

Lastly, and relatedly, even though according to **FF** whether an action/activity is a mistake is an *extrinsic* and *relational* affair (as described in the previous section), that does not entail that its account of mistakes mind-dependent. Whether something is a mistake being essentially a fact dependent upon that action/activity's relation to a particular context is *not* the same as it being a fact upon which we, as conscious agents, have any say: it is not "up to us", for instance, whether *in this environment*, *this sort of web* is better or worse at catching insects for a particular spider. Whether feedback occurs, whether that feedback contributes or hinders flourishing, *and* in which environments it does so are all matters of empirical investigation – facts about the world that are discovered, not dictated.

Thus, in light of the above discussion, we claim that **FF** delivers a theory according to which mistakes not only *really* exist – as objective features of the world – but also exist in some strata of reality and not others. For that reason **FF** satisfies our central desiderata with which we began this entire examination: explicating a philosophically rich account of the nature of mistakes that salvages our pre-theoretical view of the natural world. If the arguments of this paper have been successful, this satisfaction need not only be for its own sake – it might also teach us a lesson. Namely that if **FF** is correct that mistake-making is a genuine feature of the biological world, and making mistakes requires *agency*, *normativity*, and *flourishing*, failing to include and ontologically account for those concepts in our metaphysical models of the world – as many contemporary philosophers are wont to do – would indeed be a rather great mistake.

Acknowledgements The authors gratefully acknowledge the financial support of the John Templeton Foundation (#62220). The opinions expressed in this paper are those of the authors and not those of the John Templeton Foundation.

Author Contributions C.A., D.O., and J.H. all equally wrote and revised the manuscript text.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing Interests The authors declare no competing interests.

References

Aristotle (1984) The Complete Works of Aristotle (Vol. I & II). (J. Barnes, Trans.) University Press, Princeton
Armstrong D (2004) Truth and truthmakers. Cambridge University Press, Cambridge
Austin C, Roselli A (2021) The dynamical essence of powers. Synthese. https://doi.org/10.1007/s11229-021-03450-8
Bauer U, Federle W, Willmes C (2009) Effect of pitcher age on trapping efficiency and natural prey capture in carnivorous *Nepenthes rafflesiana* plants. Ann Bot 103(8):1219–1226
Berolaso M (2016) Philosophy of cancer: A dynamic and relational view. Springer, Dordrecht
Bielskis A, Leontsini E, Knight K (2020) Virtue ethics and contemporary aristotelianism. Bloomsbury, London
Bird A (1998) Dispositions and antidotes. Philos Q 48(191):227–234
Brenner A (2018) Science and the special composition question. Synthese. https://doi.org/10.1007/s11229-016-1234-6
Cartwright N (1994) Nature's capacities and their measurement. Oxford University Press, Oxford
Cartwright N (2012) The dappled world: A study of the boundaries of science. Cambridge University Press, Cambridge
Choi S (2006) The simple vs. reformed conditional analysis of dispositions. Synthese 195:369–379
Damschen G, Schnepf R, Stuber K (eds) (2009) Debating dispositions: issues in Metaphysics, epistemology and philosophy of Mind. Walter de Gruyter, Berlin
Dowell J (2013) Flexible contextualism about deontic modals: a puzzle about information sensitivity. Inquiry. https://doi.org/10.1080/0020174X.2013.784464
Dretske F (1999) Machines, plants and animals: the origins of agency. Erkenntnis 51:19–31
Drummond D, Wilke C (2009) The evolutionary consequences of erroneous protein synthesis. Nat Rev Genet 10(10):715–724
Ellis B (2001) Scientific essentialism. Cambridge University Press, Cambridge
Ferrero L (2022) The Routledge handbook of philosophy of agency. Routledge, London
Garson J (2016) A critical overview of biological functions. Springer, New York
Gundersen L (2002) In defence of the conditional account of dispositions. Synthese 130:389–411
Heil J (2003) From an ontological point of view. Clarendon, Oxford
Henning T (2014) Normative reasons contextualism. Philos Phenom Res. https://doi.org/10.1111/j.1933-1592.2012.00645.x
Johnston M (1992) How to speak of the colors. Philos Stud. https://doi.org/10.1007/BF00694847
Kaneko K (2011) Characterization of stem cells and cancer cells on the basis of gene expression profile stability, plasticity, and robustness. BioEssays 33(6):403–413
Lewis D (1973) Counterfactuals. Blackwell, Oxford
Maher C (2017) Plant minds: A philosophical defense. Routledge, New York
Marder M (2013) Plant-Thinking: A philosophy of vegetal life. Columbia University, New York
Marmodoro A (2018) Potentiality in Aristotle's Metaphysics. In: Engelhard K, Quante M (eds) The handbook of potentiality. Springer, pp 15–43

Martin C (1994) Dispositions and conditionals. Philos Q. https://doi.org/10.2307/2220143
Martin C (2007) The Mind in nature. Oxford University Press, Oxford
Matthewson J, Griffiths P (2017a) Biological criteria of disease: four ways of going wrong. J Med Philos 42:447–466
Mellor D (2000) The semantics and ontology of dispositions. Mind. https://doi.org/10.1093/mind/109.436.757
Mumford S (2004) Laws in nature. Routledge, London
Mumford S, Anjum R (2011) Getting causes from powers. Oxford University Press, Oxford
NASA (1995) Using spider-web patterns to determine toxicity. NASA Tech Briefs 19(4):82
Oderberg D, Hill J, Austin C, Bojak I, Cinotti F, Gibbins J (2023) Biological mistakes: what they are and what they might mean for the experimental biologist. Br J Philos Sci. https://doi.org/10.1086/724444
Shapiro J (2020) All living cells are cognitive. Biochem Biophys Res Commun 564:134–149
Sultan S, Moczek A, Walsh D (2021) Bridging the explanatory gaps: What can we learn from a biological agency perspective? BioEssays, 1-14
van Inwagen P (1990) Material beings. Cornell University Press, Ithaca
Veit W (2021) Biological normativity: a new hope for naturalism? Med Health Care Philos. https://doi.org/10.1007/s11019-020-09993-w
Vetter B (2015) Potentiality: from dispositions to modality. Oxford University Press, Oxford
Virenque L, Mossio M (2024) What is agency? A view from autonomy theory. Biol Theory. https://doi.org/10.1007/s13752-023-00441-5
Williams N (2010) Dispositions and the argument from science. Australasian J Philos, 1–20
Williams N (2011) Putting powers back on multi-track. Philosophia 39(3):581–595
Williams N (2019) The powers metaphysic. Oxford University Press, Oxford
Witt P (1954) Spider webs and drugs. Sci Am 191(6):80–87