Vagueness from the Breakdown of Fault-Tolerant Models

Siddharth Muthukrishnan

Department of History and Philosophy of Science, University of Pittsburgh
siddharth@pitt.edu

Feb 11, 2024

Abstract

Vague terms like "tall" and "bald" raise difficult questions in the philosophy of language. Attempts to understand such questions have focused mostly on ordinary language, with some recent focus on specialized languages such as legal language. I argue that it's worth looking farther afield to physics—specifically, modeling practices in physics. Physics may seem an unlikely place to find vagueness given its reputation for mathematical precision, but I show that it is shot through with vagueness. In particular, I argue that vagueness arises in physics when a model *breaks down*. I argue that ordinary descriptive language use can be seen as encoding implicit models of the world and that vagueness arises there too when these models break down. Further, I argue that vagueness has the characteristic features of sorites-susceptibility, borderline cases, and higher-order vagueness because these features are consequences of our models being fault-tolerant. Specifically, the characteristic features of vagueness result, respectively, from the characteristic features of fault-tolerance, namely, robustness, graceful degradability, and error-correctability. In sum, I argue that vagueness can be explained as the breakdown of fault-tolerant models, in both science and in ordinary language.

1 Introduction and Overview

Much of ordinary language is vague. Perhaps this is no great surprise; we have little need for extreme precision in everyday discourse. But what about discourses where extreme precision is sought after and, indeed, often achieved? Physics is one such discourse and, I will argue, it contains much vagueness. This suggests that vagueness is a symptom of something deeper, something valuable in our linguistic practices because it is retained even when we are seeking extreme precision. I will argue that this deeper valuable feature is *fault-tolerance*.

There is much to unpack here. In this introduction, I will sketch, and provide some context for, the main lines of argument that I will elaborate upon and defend in the paper.

To begin, by 'vagueness', I mean the phenomenon that many of our terms—such as "heap", "tall", and "bald"—seem to lack sharp boundaries and that such sharp boundaries seem impossible to draw non-arbitrarily. Hence, when I say that physics contains much vagueness, I am saying more than simply that physics traffics in coarse-grained descriptions—descriptions which don't distinguish between many different detailed ways the world can be, e.g., by assigning a glass of water a fixed temperature despite the constant restless movement of the water molecules. What I am saying is that such coarse-grained descriptions don't have sharp boundaries. The way physics is imprecise is more than just the imprecision that comes with underspecification of a target system, it is also imprecise in the boundaries it draws.[1] That is, I'm really looking at *vagueness* in physics; not just generality or ambiguity (Sorensen 2022, section 2).

In philosophy of language, vagueness is often thought of as a property of linguistic items such as predicates or sentences. But my focus isn't going to be specifically on vague *language* in physics. Rather, I will focus on the primary descriptive technology of physics—*models*.[2] The language of physics is secondary to the models employed in physics: the theoretical terms of physics are defined, at least partially, via the role they play in models: e.g., we understand the term 'viscosity' via the role it plays in the equations governing fluid flow.[3] Hence, I

---

[1] The importance of the lack of sharp boundaries in characterizing vagueness was emphasized by Sainsbury (1996).

[2] There is a large and complex philosophy-of-science literature on the nature and function of models in science; see, e.g., Frigg and Hartmann (2020) and references therein. To keep things tractable I will largely avoid engaging with this literature by not taking on any substantive view of models, except to note that they play a central role in descriptive and explanatory practices in physics. The notion of 'models' relevant for my argument should become clearer further into this paper.

[3] A useful recent framework to think about such theory-dependent understandings of terms

am interested in vagueness in the language of physics only insofar as it derives from the vagueness in the models of physics. My shift of focus from language to models in thinking about vagueness in physics is akin to the shift of focus from language to thought that is recommended by Bacon (2018) in thinking about vagueness in philosophy of language. Bacon argues that the vagueness of our language is really secondary to, and explained by, the vagueness of our thought, in part because our language functions to express our thoughts.[4] An analogous move is made, in the context of discussing vagueness in law, by Endicott (2000), who argues that it isn't just legal *language* that can be vague; the *law* itself is frequently vague. Similarly, in physics, language functions to express aspects of the descriptive and explanatory models we have constructed, and it inherits vagueness from the vagueness of models.[5]

So, in what sense do models in physics have boundaries? Models in physics typically come with regimes of validity. We can describe the behavior of a pendulum using a simple harmonic oscillator model—i.e., as a system with a definite oscillation frequency—only when the amplitude of oscillation is small.[6] Thus, the *boundaries* of the simple harmonic oscillator model for a pendulum are at those amplitudes at which this model *breaks down*. This is what I mean when I talk about the boundaries of a physics model: The boundary of a model is marked by the values of the parameters of the model at which the model breaks down, i.e., where it fails its descriptive-explanatory task. And my claim is that these boundaries are unsharp: physics doesn't offer principled ways to mark precise boundaries between where a model is valid and where it isn't.

On my characterization, to talk about the boundaries of a model, it seems necessary to compare the model to the behavior of a target system. But how do we talk about the behavior of the system without using the resources of a model? To avoid getting snared in this problem, and to give a clearer and more general characterization of the boundary regions of models, I will focus on the relation between models. Specifically, I will focus on different models that can be thought of as modeling a given system at different fineness of grain, and I will argue that there is no precise point at which these models change from agreeing with each other to disagreeing with each other. Crucially, one of these models

---

is the 'constitutive role functionalism' of Knox and Wallace (2023).

[4] More precisely, Bacon (2018) argues that vagueness is really a property of *propositions* instead of *sentences*, and that what is crucial to theorizing about vagueness is the role propositions play as the objects of thought.

[5] See Wallace (2022)'s notion of *predicate precisification* for a related idea of how we linguistically articulate aspects of models.

[6] See, e.g., (Nelson and Olsson 1986).

can be very thin: a model could simply be what might be considered *data*, i.e., an assignment of numerical values to certain measurable quantities. In this way, the focus on model-model relations subsumes model-world or model-system relations. Moreover, because there are many more model-model relations than model-world/model-system relations, and because such relations can be precisely characterized, model-model relations offer a framework to speak clearly about the boundaries where a model breaks down, for it allows us to precisely specify the set of situations where one model is applicable or inapplicable in terms of another model.

With all this in place, it'll be straightforward to see how vagueness arises when a model breaks down. I will focus on an example from celestial mechanics, where we model a solar system consisting of a large central star around which two planets orbit—one small inner planet and one large outer planet (akin, respectively, to Earth and Jupiter). We can construct different models for the inner planet's orbit: in particular, a two-body model that only represents a central star and an inner planet and a three-body model that also represents an outer planet. There will then be an unsharp boundary between those three-body models that agree with a given two-body model and those three-body models that disagree. And so, it is when the two-body model is breaking down that we see vagueness.

There is nothing special about this example; we can see this kind of vagueness throughout physics. *Prima facie*, we have every reason to believe that physics would work very hard to eliminate vagueness. After all, physics is known for its extremely precise successful predictions, such as predicting the values of some particle-physics quantities down to 12 significant digits.[7] But this penchant for precision doesn't seem to reach to drawing sharp boundaries for its models, which gives us reason to believe that vague boundaries is the result of something that is playing a valuable role in physics. What might that valuable thing be that is leading to vagueness?

I will argue that the valuable feature is *fault-tolerance*. This is a central normative requirement in engineering: roughly, it is the requirement that a system continue functioning despite encountering faults. Note that the focus on fault-tolerance is suggested by *where* we see vagueness arise in physics models: namely, where models *break down*. I will further argue that the standard characteristic features of vagueness—susceptibility to sorites paradoxes, the

---

[7] I'm thinking here of the magnetic moment of the electron; see, e.g., (Gabrielse 2013).

presence of borderline cases, and higher-order vagueness—can respectively be seen as consequences of three central sub-components of fault-tolerance—robustness, graceful degradability, and iterated error-correctability. Robustness ensures that our model continues functioning despite variability, and this yields the principle of tolerance (Wright 1975), which is the basis of the sorites paradox. Graceful degradability allows for borderline cases, because we want our models to degrade proportionally to the amount of errors, and so to avoid failing entirely and suddenly. And iterated error-correctability results in higher-order vagueness because we want to be able to introduce new models to handle the breakdown regime, and moreover we want those new models to be fault-tolerant in turn.

The focus so far has been on physics. If fault-tolerance is a good explanation for vagueness in physics, then perhaps it is also a good explanation for vagueness in ordinary language? I will argue that it can be. To carry the explanation over, we first need to see that the use of ordinary descriptive language also models the world in a way not dissimilar to how physics models the world. In particular, our descriptive terms can be seen as carving the world into categories which carry with them implicit or explicit 'theoretical' commitments—e.g., if we call an item 'solid', we are committed to it resisting some attempt at deformation. These networks of commitments are similar to the way models work in physics. Hence, ordinary language use encodes physics-like models.

But do we have reason to expect the models underlying ordinary language to be fault-tolerant? Unlike models in physics, they undergo much less deliberate construction. However, there is a sense in which our ordinary language is designed, for our language is an evolved adaptation, and hence *designed* inasmuch as the human eye is designed to see.[8] If ordinary language use constitutes models much like those we see in physics, and there are reasons why such models might be fault-tolerant, then fault-tolerance is a reasonable explanation of vagueness in ordinary language as well. Thus, we end up with a unified explanation for the presence of vagueness in our descriptive practices.

What sort of explanation am I offering in this paper? Let me first clarify what I am not doing. My goal here isn't to offer an account of vagueness, of the kind that leading accounts like epistemicism (e.g., Williamson (1994)), supervaluationsm (e.g., Fine (1975)), contextualism (e.g., Shapiro (2006)) and others aim to be. In particular, I do not aim to offer a systematic or formal semantics for vague terms. Nor do I offer an answer to the question of what the

---

[8] For a sustained defense of the view that our language is an evolved biological entity, see the work of Millikan (1984).

nature of vagueness is (a question articulated by Eklund (2005)), which would be answers such as *vagueness is incompleteness of meaning* (Fine 1975), *vagueness is semantic indecision* (MacFarlane 2016), *vagueness is boundarylessness* (Sainsbury 1996), *vagueness is ignorance* (Williamson 1994), and "the vagueness of an expression consists in it being part of semantic competence to accept a tolerance principle for the expression" (Eklund 2005). Neither am I attempting to show how vagueness might emerge from the likes of signalling and representation games.[9]

So what *am* I doing? I'm providing an explanation for why our descriptive practices are vague, namely that they result from fault-tolerance. But I don't want to go so far as to say that vagueness just is fault-tolerance (in appropriate circumstances). One reason for this is that I'm only concerned with vagueness of terms that attach to practices of modeling and describing the world. This then leaves open the question of how to explain vagueness that attaches to normative practices, such as vagueness in the law or moral vagueness.[10] But it is worth noting that in the context of vagueness in law a project of a character similar to mine has been carried out by Endicott (2011). One might think that law, given that its function is to specify norms of behavior, would be entirely disadvantaged in being vague. But Endicott argues that leaving the law vague protects against arbitrary precision. In particular, it allows judges to exercise appropriate judgment in varied contexts, and it allows subjects to conform to laws in varied ways.[11] My project starts off broadly along these lines: I attempt to answer how and why vagueness might arise in a particular discourse where the presence of vagueness might be *prima facie* surprising. I further aim to show that such answers might help us better understand vagueness outside that discourse as well.

Another, and more central, reason why I'm not defending a stronger thesis such as 'vagueness is fault-tolerance' is because I think that the questions of vagueness are better approached in a piecemeal fashion, instead of trying to give a unified overarching theory that tries to handle all aspects and all cases. An analogy will be useful. Say we want to explain *color*. A complete explanation of color would require accounting for many aspects of color phenomena: how visible

---

[9] See O'Connor (2020, p. 24) and references therein.

[10] For vagueness in law, some starting points are (Endicott 2000; Keil and Poscher 2016), and for moral vagueness, see (Schoenfield 2016; Sud 2019; Hawthorne 2022) for one thread of the discussion.

[11] In this context, see also recent monographs by Lanius (2019) and Asgeirsson (2020). They clarify the kinds of indeterminacy found in law, its relation to vagueness, and how such vagueness might be valuable.

light behaves, how materials reflect and absorb light, how the human eye works, how the brain integrates information from the eye, how culture and environment affects all this, and so on.[12] Colors are certainly not phenomena that admit a simple explanation that's constrained to one domain. I think vagueness is similar. One can ask many different kinds of questions about it, such as: What is its nature? Is it a property of our language or our concepts or our thought? Is there vagueness in the world? What is the right logic to deal with it? How could it have emerged through our communicative practices? Does it confer any advantages or disadvantages?... the questions can be multiplied. Hence, in a manner similar to color, I don't think vagueness will admit of a simple explanation constrained to one domain.

However, much like in color, we can make progress on this complex of questions by tackling parts of it. For instance, in the case of color, it is a valuable and hard-won part of the (as yet unavailable) complete explanation of color that our attributions of color to an object are influenced in an important way by the spectrum of light reflected off the object. Similarly, and more modestly, I aim for the explanation offered in this paper to be a useful component in the (as yet unavailable) complete explanation of vagueness.

Plan. I very briefly introduce vagueness in Sec. 2, and isolate those features of vagueness that I take as explanatory targets. Next, in Sec. 3 I give examples from physics—focusing mainly on an example from celestial mechanics—which show how vagueness arises in physics. Here, I display vagueness arising in the *relation* between different kinds of models. Next in Sec. 4, I articulate several puzzles that might arise from the way I have framed issues and explain why I have chosen to frame things in the way I have done. Following this, in Sec. 5, I argue that our ordinary language use encodes models of the world that are similar in important ways to the kinds of models we use in physics, making it reasonable to use the same kind of explanation for vagueness in both cases. Then, I show how vagueness arises from the break down of fault-tolerant models. I'll first show, in Sec. 6, how the characteristic features of vagueness arise *when* models break down. Next, I'll argue, in Sec. 7, that we can explain why these features arise during break down by assuming that the models we are using are fault-tolerant. I conclude in Sec. 8.

---

[12] See, e.g., Chirimuuta (2015) for one attempt at a (partial) synoptic account.

2 What is vagueness?

There's no univocal definition of vagueness since part of what's at stake in disputes about vagueness is what is definitive of vagueness. We can, however, state some characteristic features of vagueness and display them in examples.[13] This is enough to get a working handle.

To see these characteristic features, consider the following sentence.

(B) Jack is bald.

There could be situations in which it is clear that Jack is bald, and hence clear that (B) is true. And there could be situations in which it is clear that Jack is not bald, and hence clear that (B) is false. But, crucially, there could also be situations in which there seems to be no principled way to decide whether or not Jack is bald—even if we know all there seems to be to know about Jack's hair, including how many hairs he has on his head, their length, and their arrangement. These are *borderline cases* of baldness. This is the first characteristic features of vagueness: a predicate is vague only if there are borderline cases—cases in which it is neither clear that the predicate applies nor clear that it doesn't, and moreover, there seems to be no principled way in which we can decide whether it applies or not.[14]

The second characteristic feature of vagueness is *sorites susceptibility*, i.e., susceptibility to sorites paradoxes. We cannot simultaneously endorse all three of the following statements though they're individually attractive: (i) A man with zero hairs is bald; (ii) A man with a hundred thousand hairs is not bald; (iii) If a man with k hairs is not bald, then so is a man with $k - 1$ hairs.

Borderline cases tempt us to cordon them off into a category of their own, and to then argue that vague terms such as *bald* don't just have an extension and an anti-extension, but also a borderline extension that we might call *borderline bald*. But we can't dispense with borderline cases or the sorites paradox just with this move because, if we do so, there will still be cases which we can't

---

[13] The features I use below to characterize vagueness are almost the same features as used by Keefe and Smith (1996, pp. 2-3), but with one small difference. At the end of this section, I'll briefly discuss this difference and say why I chose my particular characterization.

[14] The term "borderline" can be misleading because "line" suggests that the border is sharp. But that's precisely what isn't the case for vague terms. A better phrase might be "border-zone" or "border-region", since there might be quite a large region of cases within which it is unclear whether or not a predicate applies. Indeed, this is important to keep in mind when discussing higher-order vagueness, for we need to be able to subdivide border-zones into further subcategories. However, since "borderline" is standard terminology, I will stick with it.

categorize, in any principled way, as *bald* or *borderline-bald* and cases which we can't categorize, in any principled way, as *borderline-bald* or *not bald*. So we will have new borderline cases between the intermediate category and the old categories. And we can raise Sorites paradoxes here as well, by transitioning, small step by small step, from *bald* to *borderline bald* and from *borderline bald* to *not bald*. This is second-order vagueness. This can be iterated to third-order, to fourth-order, and to higher orders. This is *higher-order vagueness*, the third characteristic feature of vagueness.

I will take these three features of vague terms—the presence of borderline cases, sorites susceptibility, and higher-order vagueness—to both be identifying features of vagueness and the explanatory targets for my proposed explanation of vagueness.

A brief comparative remark. Keefe and Smith (1996) characterize vagueness using the following features: the possibility of borderline cases, the presence of fuzzy boundaries, and sorites susceptibility. My characterization swaps out fuzzy boundaries for higher-order vagueness. But these are interchangeable. One has vagueness up to arbitrarily high orders if and only if one has fuzzy boundaries. If a term has fuzzy boundaries, then that means any attempt to draw sharp boundaries around the applicability of a term, or around the borderline-applicability of the term, or around the borderline-borderline applicability of the term... will all fail. Thus fuzzy boundaries lead to higher-order vagueness. Conversely, if no matter how many penumbral categories we draw between the extension and anti-extension of a term, we keep encountering new borderline cases for these categories, then we will fail at ever drawing sharp boundaries—and so we might aptly described the boundary between the extension and anti-extension as a *fuzzy* boundary.[15] I prefer higher-order vagueness over fuzzy boundaries in characterizing vagueness because it is a clearer target for explanation. Moreover, the most serious objections to most accounts of vagueness are phrased in terms of higher-order vagueness (Williamson 1994), and so it seems worth focusing on that way of posing the issues.

---

[15] I take this to be one of the imports of Sainsbury (1996), where he argues for the claim: "A vague concept is boundaryless in that no boundary marks the things which fall under it from the things which do not, and no boundary marks the things which definitely fall under it from those which do not definitely do so; and so on" (p. 257).

3 Vagueness in physics

The primary example we'll be working with in this paper concerns different models for the orbit of a planet. Say we are modeling the following sort of solar system: a central star around which two planets stably orbit. (You might imagine a system with just the Sun, the Earth, and Jupiter.) And say we are interested in the orbit of the inner, Earth-like, planet. We can construct many models for the inner planet's orbit, at increasing levels of detail. At the simplest level, one can model the inner planet as being on a circular orbit. At a more sophisticated level, we can construct a model that ignores the outer planet and treats the star as stationary. This yields a standard *two-body model* in Newtonian gravitation.[16] Going even further, we can consider *an intermediate model* which includes the outer planet as a perturbation to the two-body model. An even more detailed model is the full *three-body model* of the star and the two planets.[17] One can construct even more detailed models: models with general-relativistic corrections, full general-relativistic models, models that include the size and rotation of the bodies, and so on.

In what follows, we'll largely focus on the *two-body* model and the *three-body* model and later bring in the *intermediate* model. These models are systems of equations that take as input the masses of the bodies and their positions and velocities at some point in time and output a solution for the trajectory of the inner planet (this trajectory is what we've chosen to focus on). For some of these inputs, the two-body model will say the inner planet is on an elliptical orbit. Meanwhile, on the same inputs, the full three-body model will say the inner planet is on a more complicated, fluctuating orbit that's not a perfect ellipse. (Note that the inputs that enter into the two-body model will be a strict subset of the inputs that enter into the three-body model, since the three-body model also has an outer planet.)

When do these two models *agree* about the inner planet's orbit? The natural criterion for agreement is some scheme of *approximation*. If a three-body model (with appropriate input parameters) produces an orbit for its inner planet that is *approximately* the same as the orbit produced by the two-planet model for its inner planet, then we can take the two models as agreeing about the motion of the inner planet.

---

[16] Such a model is commonly analyzed by mapping onto a *central-force* model. See, e.g., Kleppner and Kolenkow (2014, Ch. 10) for more details.

[17] While such a model resists a closed-form solution—the famed three-body problem (Barrow-Green 2010)—it can be numerically solved on computers.

Vagueness enters when we ask *which* three-body models agree with a given two-body model. To see this, take a specific two-body model (i.e., a model with its parameters fixed at some values); this model will make a claim about the orbit of the inner planet—say, that it follows a particular elliptical orbit. Now note that there are many different three-body models consistent with the specified parameters of the two-body model: i.e., there are many different three-body models which have the same mass, initial velocity, and initial position for their inner planet and central star as the two-body model, but all these three-body models differ amongst each other on the mass or initial velocity or initial position of their outer-planet. Each such three-body model will make different predictions about their inner planet's orbit. Some of these three-body models can be taken as *agreeing* with the two-body model because they approximate the two-body model's inner planet orbit; contrariwise, some three-body models will *disagree*. What is vague here is that there is no way to draw a sharp boundary, in a principled way, between those three-body models that agree with the given two-body model and those that don't.

More precisely, we can exhibit in this example the characteristic features of vagueness (specified in Sec. 2 above).

(i) Borderline cases: There will be three-body models such that it is neither clear that they agree with the given two-body model nor clear that they disagree, and moreover, there are no principled reasons that allow us to decide whether or not they agree.

(ii) Sorites-susceptibility: There will be sequences of three-body models—adjacent models differing ever so slightly (say, the adjacent models' outer planets' mass differs by grams)—that start out clearly agreeing with the two-body model but end up eventually disagreeing.

(iii) Higher-order vagueness: Separating the three-body models that are in borderline agreement with the given two-body model into a new, distinct category will result in new borderline cases between the new and the old categories; and this will continue at higher-orders if we carve out further intermediate categories.

The example I've given above isn't by any means cherry-picked. It is easy to generate many examples in physics where one constructs two different models for a system or a phenomena, and where the relation between the two models is vague. Let me list a few further examples (one can generate many more):

- *Temperature of a gas.* Consider a box of gas. We can model it using statistical mechanics, and attribute a certain temperature T to it. We can also model it as a swarm of particles, with definite positions and momenta at any given time, obeying Newtonian equations of motion. There are uncountably many Newtonian models—where a Newtonian model is an assignment of particle positions and a Newtonian dynamical time-evolution of that configuration of particles—that are compatible with the same statistical-mechanical description, and hence the same temperature attribution. (It might help to imagine ever so slightly disturbing the position of just one particle in a box with $\sim 10^{23}$ particles and seeing how that won't make a difference to the overall temperature attributed to the box.) There is no principled way we can draw a sharp boundary between those Newtonian models that count as agreeing with the statistical-mechanical model and those that don't, leading to vagueness.

This example can be extended: Any macrostate defined by some macroproperty, or collection of macroproperties, will bear a vague relation to microphysics.[18] More precisely, there is no principled way we can draw a precise boundary between those microstates that count as realizing a particular macrostate and those that don't.[19]

- *Chemical potential.* Consider a statistical mechanical system, such as a box of gas, which is allowed to exchange particles with its environment. We can attribute to such systems a quantity called *chemical potential*; roughly, it quantifies how easy it is for the particles to enter or exit the system. The same system can be studied with Newtonian models (as in the previous example), but these models won't have a fixed number of particles (unlike the previous example). As in the previous case, there's no sharp boundary between those Newtonian models that count as having a certain chemical potential and those that don't.

Why bring up the example of the chemical potential if it is structurally so similar to the above example of temperature? I do so because it highlights the point that the objects we attribute to the system can be quite dependent on the model under consideration. (This point will be discussed further in the next

---

[18] See Albert (2000) for a classic philosophical discussion concerning the definitions of and the relations between macrostates and microstates.

[19] The vagueness of the relation between macrostates and microstates is the starting point of Chen (2022). He uses this vagueness to argue that the fundamental laws might be vague. In contrast, my focus stays on the non-fundamental, and my goal here is to use such examples to construct a novel explanation of vagueness instead of arguing for the existence of a novel kind of vagueness as Chen (2022) does. See also, Miller (2021), for a different argument for imprecision at the fundamental level.

section.) The different Newtonian models realizing the chemical potential can have different numbers of particles, so it's not as if it always makes sense to speak of our models as concerning a target system that has a fixed collection of objects.

- *Viscosity of a fluid.* We can model a fluid using the Navier-Stokes equations,[20] which requires attributing a certain viscosity η to it. We can also model the fluid as a collection of a large number of molecules, dynamically interacting with each other via some potential energy function. There will be no principled way to draw a sharp boundary between those molecular-dynamical models that count as agreeing with the Navier-Stokes model—and hence attributing to it the viscosity η—and those that don't.

- *Effective theories.* There are systematic recipes available in physics (so-called renormalization group methods) for constructing a simple model (with a small number of parameters) starting from a much more complex model (with a large number of parameters). Under these recipes, many different complex models will 'flow' (i.e., be taken by the renormalization procedure) to the same simple model.[21] Such simpler models, often called *effective theories*, come with *cut-offs*, which determine a regime of validity—i.e., these models/theories will only be applicable when the appropriate inputs to these models are not beyond the cut-off values. These cut-offs exist because the simple model, by construction, is insensitive to the precise details of the goings-on beyond the cut-off—goings-on which would require the use of the more complex models. The renormalization group procedure folds-in the beyond-cut-off details into the parameters and structure of the simple model.

That said, the renormalization group recipes offer no principled way to draw a sharp boundary between those complex models that flow to the same simple model and those that don't, i.e., there is no way to determine a precise value as *the* cut-off.[22] And what's more, that there be no precise cut-off value for the validity of an effective theory is a central normative requirement of renormalization group methods: one explicitly works to get rid off any dependence on the precise cutoff value when using the renormalization group.[23]

---

[20] See, e.g., Thorne and Blandford (2017, pp. 711-713).

[21] See Williams (2023, Sec. 4.3) for an philosophical introduction to these ideas.

[22] Note that effective theories have imprecision in them well-before the cut-off scale. This is a well-known point in physics; see, e.g., Miller (2021, pp. 2905-2907) for a place in the philosophical literature where this is discussed.

[23] One plausible inference from the arguments of my paper is that fault-tolerance might

4 Puzzles about my framing

In this section, I'll discuss six puzzles that might arise about how I am framing things in this paper. Below, after expressing what each puzzle might be, I provide my reply to it. Hopefully, by working through the following dialectics, the way I'm thinking about vagueness in physics becomes clearer.

1. *Derivative from ordinary language vagueness?* Is the kind of vagueness I have identified here, between physics models, just derivative from the vagueness of the relation expressed by the ordinary-language term "is approximately the same as"? After all, in the celestial mechanics example above, vagueness arose because the criterion of agreement between a two-body and a three-body model was that the orbits that they each described for their respective inner planets *approximately* agreed. And if so, isn't the kind of vagueness I've identified here just a species of the vagueness found in ordinary language?

Reply: To assume so would be to beg the question against the kind of project that I'm attempting in this paper. It's true that we relied on our ordinary notion of approximation to get a handle on vagueness in physics, but that doesn't imply that ordinary-language vagueness is explanatorily prior to the vagueness of the relation between models. Given what we have seen so far, we might just as plausibly describe the vagueness of the term "approximation" as being explained by the vagueness of the relation between models. Articulating this is the project I'm attempting in this paper: To explain the broader phenomenon of (descriptive) vagueness by starting with vagueness in physics modeling.

2. *Whereof predicates and sentences?* Following up on the previous puzzle, given that, vagueness is typically construed as a property of predicates and sentences, how does one then connect the vagueness of model-model relations to the vagueness of predicates and sentences?

Reply: We can frequently identify sentences, which employ predicates, that will be vague when we think of their meanings in terms of the model-model relation. For example, in the context of the celestial mechanics example discussed above, consider the sentence:

> (S) The inner planet has an elliptical orbit of eccentricity 0.167.

If we interpret (S) as simply expressing a statement about just a two-body model, then this sentence will not be vague: it is either the case that a given two-body

---

explain why we have this normative requirement in renormalization group.

model predicts an elliptical orbit of eccentricity 0.167 for its inner planet or it is the case that it doesn't; there's no room for borderline cases here. However, if we interpret (S) as expressing a statement about a two-body model (namely one that predicts an elliptical orbit of eccentricity 0.167) that is obtained by approximation from a three-body model, then (S) will be vague. For, there will be no sharp boundary between those three-body models that yield a two-body model under which (S) is true and those three-body models that don't.

This latter reading of (S) isn't perverse. One way to see how such a reading might be natural is to imagine that we are in a world where a three-body model describes the ground truth—i.e., it is the fundamental theory of that world. In that world, then, whether or not (S) can be justifiably asserted will depend on the extent to which a two-body model predicting an elliptical orbit of eccentricity 0.167 is a good approximation to the correct three-body model of the world. And if the agreement between the two-models is borderline, then the assertibility of (S) will also borderline. Then, assuming that warranted assertibility entails truth, then that will be enough to establish the vagueness of the above sentence: if (S) is a borderline case of warranted assertibility, then we will judge it to be borderline true or false as well.

Having done this exercise in imagining that the world is fundamentally described by a three-body model, we can now note that the moral of the above argument will carry over to worlds like ours, where the fundamental theory (if there's one) is far removed from any Newtonian theory. The point is that the warranted assertibility of any statement we utter that employs terms of some higher-level model or theory—such as a two-body model—will depend on the degree to which that model or theory is a good approximation to the fundamental theory of our world. Further, this approximate relation of the higher-level model to the fundamental theory will be spelled out in terms of a large chain of approximate relations between a whole sequence of intermediate models; e.g., we'd start from the two-body model and work our way down, via approximation relations, through the three-body model, through relativistic theories, through quantum theories, and presumably terminating at the fundamental theory. Hence, in this way, we can see that many statements we utter in scientific contexts will be vague, but that vagueness will be derivative from the vagueness of model-model relations.

3. *Why multiple models?* In order to raise the issue of vagueness in this context, I needed to allow for there to be multiple models with parameters compatible with

a given model. To elaborate: in the example above, I assumed that there would be multiple three-body models which aren't ruled out simply by the parameters (masses and initial positions and velocities) of a given two-body model. That is, holding fixed the masses and initial positions and velocities for the central star and the inner planet, we can still specify (uncountably) many three-body models with those parameters by varying the mass or the initial position and velocity of the outer planet. But it might be puzzling why we take all but one of these multiple models into consideration. Given a target system, isn't there precisely *one* three-body model truly compatible with it? That is, wouldn't the appropriate three-body model simply be picked out by setting the mass and initial position and velocity of the outer planet equal to the actual values of the outer planet in the target system?

Reply: First, note that even if there is precisely one three-body model corresponding to the relevant context, we could still ask the question whether, for *that* three-body model, the two-body model is in clear agreement with it or not, and to answer that question we have to at least imagine the possibility of varying the underlying three-body model. In this, this situation is no different from how we go about investigating vagueness in ordinary language. Given an actual man, there is a fact about how many hairs he has. Nevertheless, to ask whether he's clearly bald or not, it is useful to imagine varying the number of hairs on his head and thinking through how that might change our judgment of his baldness.

But a different response to this concern, and one that is perhaps more apropos of the discussion here, is that we cannot assume that there is precisely one finer-grained model that captures the actual system. Take the celestial mechanics case. To assume that there is a single three-body model that corresponds to the system under consideration is to assume that we can specify the values of the parameters in the system to arbitrary precision. But this is implausible. For instance, there is no single real number that we can say is exactly equal to the mass of Jupiter, for we cannot draw identify a precise set that consists of all and only the constituents of Jupiter.[24] And even if we could identify such a set, it isn't clear that the concept of *mass* is well-defined beyond the Planck scale, and an infinitely precise specification of the mass of an object, such as Jupiter, would require commitments about the Planck scale and beyond. Consequently, for even slightly realistic scenarios we will have multiple models corresponding

---

[24] This can be seen as an instance of the Problem of the Many (Weatherson 2016).

to a given system.

4. *Don't models almost always disagree?* The way I introduced the issue of vagueness was by focusing on the question of when different models of a system *agree*. But isn't there a trivial answer to this question: namely that the two different models *never* agree? After all, they never make *exactly* the same claims about the system—for instance, no three-body model will attribute precisely the same orbit to the inner planet as the two-body model does?

Reply: If we don't have a notion of agreement that allows for nontrivial agreement, then it is difficult to make sense of how different models of a system might be speaking of the same system. To see this, note that while we can, and naturally do, identify the planet mentioned in the two-body model with the inner planet mentioned in the three-body model, nothing *necessitates* we do so. We *could* just as well take the models to be about different planets. Consequently, to retain a sense that the two models are talking about the same system, we employ criteria of cross-model identification, and such a criterion cannot demand perfect agreement between models, on pain of making no cross-model identifications at all.

This becomes point becomes particularly salient in the temperature, chemical potential, and viscosity examples above (Sec. 3), where the kinds of objects and properties of one model (gases with temperatures or chemical potentials and liquids with viscosities) are very different from the kinds of objects and properties of the other model (particles with positions and velocities, possibly interacting via forces). So we clearly need some criterion of approximate agreement to make sense of the fact that these two very different models can be about the same target system.

5. *Why relativize objects to models?* In the way I set things up, I have relativized objects to models; I was speaking of things like "the inner planet of the three-body model" or "the molecules of a Newtonian dynamical model". But this way of relativizing objects to models might raise a puzzle: Insofar as we think of the different models whose relations we are analyzing here as modeling the same target system, then why not simply have the terms in the models refer directly to the objects in the system?

Reply: The main reason I relativize to objects to models is because of the theory-ladenness of objects. This theory-ladenness can happen in two ways. First, the meaning of a potentially object-referring term in a theory or model will depend, at least in part, on the theory in which the term features.[25] Second, what counts as objects will depend on the theory in question. So, e.g., if we have a particle theory, then according to that theory, the objects are particles, but if we have a field theory, then according to that theory, the objects are fields. These two theories may well apply to the same system at the same time, but attribute different objects to the same system.[26]

6. *Whereof truth?* In typical examples of vagueness in philosophy of language, such as "Jack is bald", we consider statements which are at least *prima facie* truth evaluable. Further, their truth value depends (at least partially) on what the world is like. Thus, whether "Jack is bald" is true depends (at least partially) on the number and pattern of hairs on Jack's head. However, in the examples I'm considering, I don't really focus on model-world relations; instead, I stick to model-model relations. Why?

Reply: This is to avoid getting entangled in the realism/anti-realism debate. For, if I commit myself to models being true or approximately true (however that notion is spelled out), then I commit myself to some variety of scientific realism. And even if one readily embraces realism, our models will almost always be only approximately true, and we would need a handle on the notion of approximate truth in play here before we can investigate vagueness. Focusing on model-model relations allows me to sidestep the thorny issues surrounding approximate truth. More generally, as I articulated in Sec. 1 and also in the discussion above in this section, we can get a handle on model-world relations by treating as a special case of model-model relations, where either one of the models is *data* or where one of the models is a fundamental description of the world.

5 Ordinary language use as scientific modeling

So far we have been focused on models in physics. Below, I will explain how vagueness arises when models break down. I want my explanation to extend beyond models in physics and to explain vagueness in ordinary descriptive language as well. For the explanation to be extendable in that way, I need to

---

[25] For further explication of the idea of semantic dependence of theoretical terms on the theories they feature in, see Andreas (2021, Sec. 2) and references therein.

[26] This point is especially clear in quantum field theory where particles are emergent from fields. More broadly, whenever we can describe a system with a higher-level theory and a lower-level theory, we may well have different ontologies at those two different levels. For more, see, e.g., Wallace (2022) and Guo (2023).

argue that when we use ordinary descriptive language, we model the world in a way that is relevantly similar to the way in which we model the world in physics. This will allow us to give a relatively unified explanation for vagueness in both physics and ordinary language use, for they can both be seen as vagueness that attaches to modeling in general.

Let me now argue that our use of descriptive language—especially the kind of language that we prototypically take to be vague—is at least partially aimed at modeling the world. Consider declarative sentences, using which a speaker says something is such-and-so, e.g., "Jack is bald" or "Anne is tall". This kind of language models the world by *categorizing* objects—i.e., by slotting objects into different categories. We have to be careful with this statement though. If we take categorizing an object to be the same thing as describing a *set* which contains the object, then this leads to the question whether vague predicates like "tall" and "bald" pick out a well-defined set. To avoid this concern, I'll appeal to a notion of categorization that is broader than placing objects into mathematically respectable sets—the notion I'm appealing to being what we tend to employ in most acts of collecting, classifying, organizing, taxonomizing, and so on: an activity that doesn't require appeal to the machinery of set theory.

These ordinary acts of categorization are continuous with scientific taxonomy. These categories frequently contain implicit theories: as in, by categorizing an object as such-and-so, we commit ourselves to categorizing that and other objects in certain ways. So, for example, if I say "John is tall", I'm committed to thereby categorizing anybody who is taller than John as also tall and to not categorizing John as short. This is akin to how, in chemistry, if I categorize a particular sample as metallic copper, then I'm committed to it being conductive.[27] That categorization comes with a network of commitments strengthens the analogy between ordinary description and scientific modeling.[28]

I'll now consider three potential points of differences between models in science and models behind ordinary language use. To each point, I will try and argue that the differences aren't as great as one might have thought.

The first potential point of difference is *explicitness.* Models in science often

---

[27] The latter example is from Brandom (2019). One might develop this line into an inferentialist account of language—as Brandom does—on which the meanings of terms just is a network of inferential relations. I do not however need to endorse such an account of language.

[28] Similar ideas have been explored by some psychologists, who have argued that concepts can be thought of as *theories*, akin to scientific theories, and that these concepts change in child development in much the same way theories change as science develops. See, in particular, the *theory theory* of concepts defended by Gopnik and Meltzoff (1998). See also Carey (2009) and Keil (1992).

explicitly specify what inferences are permissible and impermissible. Many models in science, especially in physics, are explicitly *mathematical*, such as the models we saw above, and by being mathematical they provide a sharp and explicit delineation of which inferences are allowed and disallowed. E.g., in thermal physics, there is often an explicit specification of how temperature is constrained by other quantities, such as pressure or volume: e.g., $PV = nRT$ for ideal gases. Whereas, in ordinary language, permissible inferences and constraints are rarely explicitly specified, leave alone mathematically. The Oxford English Dictionary (2023) entry for *bald* is "Having no hair on some part of the head where it would naturally grow; hairless". So this explicitly specifies that uses of "bald" typically entails claims about "hair" and "head". But such specification falls well short of any kind of detailed regimentation.

While this is true, the amount of explicitness available in science should not be overstated. In science: (a) there are many potentially undiscovered and unarticulated relations in science; e.g., the temperature of a gas is related to the colors of spectral lines it emits, but this observation wasn't articulable with classical statistical mechanics but required the advent of quantum mechanics; and (b) scientific terms bear a host of relations to ordinary-language concepts which are rarely made explicit; the relation between the ordinary-language notion of temperature and the scientific concept of temperature is complex and rarely fully articulated.[29] More generally, a great deal of scientific knowledge and procedure is unsystematized and implicit, available only to those embedded in research communities.[30] This makes many aspects of scientific models more implicit, and hence closer to ordinary language, than one might initially think.

A second potential point of difference between scientific models and the models underlying ordinary language is that models in science are socially constituted, i.e., supported and underwritten by a large community of scientists who determine the conditions under which a model is correctly or incorrectly employed. By contrast, the models underlying ordinary language use might seem more individualistic—i.e., it might seem I'm mostly just expressing aspects of my personal mental model (in my head, so to speak) of the world when I use descriptive language, such as when I judge John to be tall.

However, here too, the differences should not be overstated. First note

---

[29] See Wilson (1982, pp. 564-566) for a brief discussion of this point. The historical and conceptual complexities of the scientific notion of temperature are discussed in Chang (2004).

[30] This is a frequent theme in philosophy of science. A couple of useful entry points here are Thomas Kuhn's notion of a *disciplinary matrix* (Kuhn and Hacking 2012) and Michael Polanyi's notion of *tacit knowledge* (Polanyi 1958; Polanyi and Sen 1966).

that the correctness conditions for our ordinary language use are governed by social conventions. After all, we learn our language from the community, and the community imposes the standards that govern correct and incorrect usage. Secondly, and more to the point, it's widely accepted that the contents of our mental states when we use ordinary language isn't simply a matter of our internal psychological states, but that they can depend, constitutively, on the linguistic community we are a part of.[31] So even with vague predicates like "tall", the question of whether I correctly judge John to be borderline tall need not be entirely dependent on my internal mental state—my community may well have a say.[32] Hence, the way in which the content of our mental states are constituted when we use scientific terms may not be quite so different from the way in which content of our mental states are constituted when we use ordinary language.

The third point of potential dissimilarity is that while scientific models are deliberately constructed to describe and explain certain systems, ordinary language emerges from a web of social practices and so has a structure that is rarely the outcome of conscious deliberation and decision. This point is particularly relevant for the arguments to come since below I will argue that vagueness arises from a kind of fault-tolerance that is enjoyed both by models in science and by ordinary language. And appeal to ideas like fault-tolerance shows that we are saying our models possess the kind of virtues one would expect of engineered artifacts. While scientific modeling might plausibly be argued to be deliberately engineered in certain ways to embody certain virtues[33], it is perhaps harder to see why our ordinary language would also embody such engineered virtues if they only emerge out of a largely unconscious web of social practices.

To address this point, first note that scientific models aren't quite so deliberately engineered as one might think. Such models are also, in significant part, the outcome of a complex social and evolutionary process.[34] Similarly, languages and the cognition supporting it can be seen as evolved. This could be biological evolution (see, e.g., Pinker (1994/2007)), cultural evolution (see, e.g., Everett

---

[31] Particularly relevant here is Burge (1986). See also Putnam (1975).

[32] This observation is central to Williamson (1994)'s epistemicist account of vagueness, which claims that vague predicates pick out sharp boundaries but that we are unaware of these boundaries. It is the presence of wide content that underwrites how we can be thinking of sharp boundaries unbeknownst to ourselves when we use vague terms. However, I do not need to subscribe to epistemicism; I'm simply emphasizing how the content of vague terms need not be so individualistic.

[33] See, e.g., Wimsatt (2007) for, *inter alia*, an extended defense of the use of engineering metaphors in the philosophy of science.

[34] See Hull (1988) for an articulation of how scientific development can be seen as an evolutionary process. See also Wilson (2006) for similar themes.

(2012) and Heyes (2018)), or some combination of the two thereof (see, e.g., Deacon (1997)).[35] Once we see that language (and science) are evolved—their features selected for on the basis of some kind of 'fitness'—then we have a plausible route for the emergence of design virtues, such as fault-tolerance.[36] (Beyond these brief remarks, I will not further argue in this paper for *how* exactly fault-tolerance might emerge in linguistic/scientific models.)

6 Model breakdown and vagueness

So we have seen that vagueness arises in models in physics. But what is characteristic of the situations in which such vagueness arises? And why does it arise there? These are questions we now turn to.

I will first argue, in this section, that the characteristic features of vagueness arise *when* a model breaks down. The standard by which we evaluate whether a model breaks down or not is given by the more fine-grained model (which, recall, could even be data) that we are comparing the model to. Following this, in the next section (Sec. 7), I will consider *why* model breakdown carries with it the characteristic features of vagueness; my answer will be that they occur as products of fault-tolerance.

Let's see how entering a borderline zone signals model breakdown. Let's return to the solar system example presented in Sec. 3. Why couldn't we draw a sharp boundary between those three-body models that agree with a given two-body model and those that don't? It was because there were three-body models such that we couldn't find any principled reasons which allowed us to decide whether or not they agree with the two-body model. This will happen when the three-body model is predicting an orbit for the inner planet that isn't entirely well-modeled by the given two-body model. Because if it were, then we'd have a reason to say that the two models agree. At the same time, the two-body model isn't so bad a fit that we have a clear reason to say that the three-body model does *not* agree with the two-body model. Consequently, the two-body model is *breaking down* in describing the behavior that is predicted by the three-body model precisely when we have borderline agreement between the three-body and the two-body models.

---

[35] Also, relevant here is the work of Ruth Millikan (see, e.g., Millikan (1984, 2017)). See also Richard (2019).

[36] See, e.g., the arguments of Dennett (2017) as to how evolution supplies design without designer, including for features of language.

This observation about when borderline cases arise—that they arise when models break down—also fits ordinary language vague terms. E.g., consider "bald". We can model a person's hair situation using two kinds of models. We can classify them as "bald" or "not bald" (see Sec. 5) or we can model them by ascribing a certain number of hairs to their head. The person in question will be borderline bald just if it isn't clear if the *bald/not bald* model agrees with the *number of hairs* model. And this will happen precisely when the bald/not bald model starts to break down.

To fully see that vagueness manifests when models break down, it isn't enough to just see that borderline cases arise when models break down. We also need to see the characteristic boundarylessness of vague terms—we need the borderline zone to not have sharp boundaries. We already saw in Sec. 3 how we can construct sorites sequences for cases of model agreement. Let me elaborate on that point here; specifically, let me delineate the proximate reasons[37] as to why model breakdown is typically sorites-susceptible. Sticking with the planetary orbit case, we can imagine changing the parameters (masses of the planets, interplanetary distances) of the three-body model of the system so that the orbital trajectory of the inner planet transitions slowly but surely from being well-modeled to being ill-modeled by a two-body model. This transition is soritical because these kinds of models don't display sharp transitions in their behavior at some parameter value. More generally, the models we employ rarely provide *precise* information about where they will fail. For instance, neither the two-body model, nor the three-body model, nor anything in the methods for deriving one from the other within the framework of Newtonian physics, provides a precise quantitative cut-off at which one can take the two-body model to have failed. This doesn't mean one cannot provide any quantitative characterization of where the breakdown occurs. For instance, the two-body model agrees well with the three-body model when the outer planet isn't too massive relative to the inner planet and is sufficiently far away from the inner planet, and disagrees otherwise. But clearly this characterization of when model failure occurs, while quantitative, isn't *precise*.

To understand this point better, it is useful to contrast sorites-susceptible model relations—such as the ones we have looked at here—with model relations that aren't sorites-susceptible. Such frameworks plausibly arise when we consider phase transitions. If a physical system displays a phase transition, then there will

---

[37] As opposed to deeper reasons, having to do with fault-tolerance, as we will see in the following section.

be sharp discontinuities from one model to another at a specific parameter value. For instance, in modeling the phase-transition from solid ice to liquid water, we can supply a perfectly precise value at which we must abandon modeling the system as a solid in favor of modeling the system as a liquid—namely at the temperature of 0 °C. Note that it might still be vague what the temperature of the system is when compared to a Newtonian-mechanical model (as remarked in Sec. 3). Moreover, the sharpness of the transition will never be physically realized given that the sharpness only obtains in idealized infinite systems. Such infinite idealizations are philosophically quite contentious.[38] Consequently, I advance the phase-transition example only for illustrative purposes. Indeed, it is actually quite difficult to come up with a non-idealized model that can capture physical phenomena and yet not be vague in some sense. All that said, if a statistical-mechanical model in the infinite limit which ascribes temperature is being checked for agreement with a fluid/solid-mechanical model that ascribes shapes, volumes, compressibility, rigidity, etc. (properties that distinguish solids from fluids), then we will have sharp transitions, and hence a model breakdown that isn't sorites-susceptible. And so sorites-susceptibility arises in those kinds of systems which have models that break down slowly instead of suddenly.

In a similar manner, the models underlying our ordinary language don't carry information about precise points of break down—it is not part of our practices of representing or thinking of or uttering statements about patterns of hair that there is a precise point at which we switch from modeling a person as bald to not bald. Usually, we introduce precise cut-offs, and concordant phase-transition-type models, only in specialized contexts, such as in the law: for example, when a legal system defines "adult" as "greater than 18 years of age" (say, in a context concerning voting rights). Outside of such specialized contexts we don't use models with precise cutoffs and so much of our ordinary language is sorites-susceptible.

So we have connected two of the three characteristic features of vagueness specified in Sec. 2—*viz.* the presence of borderline cases and sorites-susceptibility—with model breakdown. The absence of higher-order vagueness might be conspicuous, but I defer its discussion to the next section, where it is more naturally discussed and where it too will be connected with model breakdown.

It's worth remarking before we move on that I don't intend model breakdown

---

[38] See, e.g., Fletcher et al. (2019), for an introduction to and overview of this debate.

to be a non-circular definition of vagueness. After all, what it is for a model to start breaking down can't really be specified without vagueness. More generally, it is unlikely that one can define what it is to be borderline without appeal to vague terminology.[39]

7 Fault tolerance and vagueness

In the previous section, we saw that the characteristic features of vagueness appear as a model breaks down (relative to another model). Now we turn to the question of *why* these features emerge during model breakdown. *Why* are our models this way? Wouldn't any systematic science—especially one so enamored with precision as physics is—*want* models with clear, sharp boundaries, with no room for ambiguity about applicability? The fact that vagueness is associated with model-breakdown gives us a clue. If vagueness serves some sort of purpose, then it seems plausible that that purpose is best served when our models of the world start breaking down. I propose that vagueness is a signature of the fault-tolerance of our models—a signature that appears as a fault-tolerant model breaks down.

Fault-tolerance is a central normative principle in engineering. One thing we want of a fault-tolerant system is that it continue functioning, insofar as possible, despite encountering variability. That is, we want a system that is *robust*. Moreover, when smooth functioning is no longer possible—say if the errors are too severe—then we want the system's performance to degrade proportionally to the amount of errors encountered. That is, we want the system to *degrade gracefully*. Finally, we want the system to allow itself to be *patched* in appropriate ways as it is degrading. That is, we want the system to be *error-correctable*.[40,41]

Let's now see how these three desiderata of fault-tolerance—robustness,

---

[39] See, e.g., Sainsbury (1996) for argument on this point. See also Cook (2002) for a careful discussion of the consequences of such arguments for formal semantical approaches to vagueness.

[40] There are different ways the notion of fault-tolerance is presented in the engineering literature. For some entry points, see, e.g., Pradhan (1996) and Dubrova (2013). I have chosen three aspects that I think are relevant to explaining vagueness. Let me also emphasize that in engineering contexts, one can construct mathematical models that describe or aid in the design of fault-tolerant systems. In the paper, I do not engage with the specific mathematical details and rely more on the broader qualitative insights of that discipline, partially because the mathematical details will depend on the system in question.

[41] My notion of fault-tolerance is related to Wimsatt (2007)'s notion of error-tolerance, but the value he sees in error-tolerance is somewhat different from the value I'm identifying, so I stick with my terminology. Moreover, 'fault-tolerance' is a well-established notion in engineering, which I'm drawing on.

graceful degradation, and error-correctability—when applied to modeling practices, generate, respectively, the phenomena characteristic of vagueness—sorites-susceptibility, the presence of borderline cases, and higher-order vagueness. Consequently, if we assume that our models are fault-tolerant—which they plausibly are since fault-tolerance is a valuable asset to any system—then we can see why vagueness will likely arise in both models in physics and models underlying our ordinary language.

Start with robustness and sorites-susceptibility. Robustness requires that if the situations we encounter differ only slightly, then we should be able to continue using the same model, much like how we want a bridge to continue being stable if just one more person steps on to the bridge. Thus, if the slightest change in a situation that we are modeling necessitated the use of a new model, then our models would not be robust. This explains the *principle of tolerance* that attaches to prototypically vague predicates, which states that there is "a degree of change too small to make any difference" (Wright 1975, p. 333). The principle of tolerance endows our models with robustness. However, it also makes them sorites-susceptible. Thus, sorites-susceptibility is the price we pay for robust models.

Let us bring out more clearly how robustness explains sorites-susceptibility, both in the context of physics and in ordinary discourse. Let's start with physics. Say I'm predicting the motion of a projectile using Newton's equations. Say I'm trying to calculate the final velocity of the projectile starting with its initial velocity (and other data). Even if I could measure, very precisely, the initial velocity of the projectile, I wouldn't be justified (in most realistic scenarios and without adding many caveats) in predicting the projectile's final velocity to an equally great degree of precision. This is because such a prediction would be fragile: a small perturbation or injection of noise (say a stray wind current or a local inhomogeneity in the gravitational field) might be enough to render the prediction inaccurate. So if I want a robust prediction, then it'll be better if I make a less precise prediction. However, this will come at the cost of sorites-susceptibility, for a robust prediction is, by design, tolerant to small changes in the target situation. But we can always chain together many small changes (each individually small enough to preserve the accuracy of the prediction) to create changes large enough to make even the robust prediction fail.

Turn now to more ordinary contexts. In the context of ordinary conversations, it is a norm that we ought not to make our statements more precise than necessary. This is one part of Grice's Maxim of Quantity, the part which exhorts us to make our contributions to a conversation no more informative than necessary (Grice 1989, p. 26), and that entails that we shouldn't make our contributions more precise than necessary.[42] So, e.g., if someone asks me, in an informal context, how much a cup of coffee costs at the coffee shop nearby and I say that it's \$3.46 and there's a 6.1% tax and at least 12% tip is expected, then I would have given an unnecessarily precise answer. One reason why such a precise answer can be criticized is that the answer is not robust. If the coffee shop decides to change the price tomorrow to \$3.55 or the state's sales tax rate changes to 6.25% or the norms around tipping change, my answer would have been invalidated. However, if I had simply said, in response to the original query, that the price is about \$4, then my answer would have been much more robust to such sources of noise. But now, it is unclear how far from \$4 the price of the coffee can drift before the initial statement counts as false or misleading. So, as above, robustness has come at the price of sorites-susceptibility.

Next, let's turn to how graceful degradability explains the presence of borderline cases. A system degrades gracefully just if the degree to which it degrades (when it does degrade) is proportional to the degree to which errors have been encountered. That is, it retains partial functionality instead of failing catastrophically. E.g., if a bridge is slowly getting overloaded, we'd like the bridge to degrade by slowly showing cracks instead of collapsing altogether. I claim that if our models degraded gracefully, then they would allow for borderline cases. As we saw in the previous section, we can interpret borderline cases as cases where a model is breaking down. If we want such break down to be graceful, then what we need is that the model continue being useful despite its limitations. But this is what we see in a borderline case: It is a case where a model that works well for nearby cases is failing but hasn't failed so badly that we clearly need to abandon it altogether. Let's first see how this works in the context of modeling in physics and then turn to models in ordinary language.

In the physics context, in the celestial mechanics example, when we see the system deviating from a two-body model, it is still useful to describe the system as one that is close to but deviating from a two-body model. (As mentioned in Sec. 3, we can construct new intermediate models based on this observation; intermediate models will be central to our discussion below on error-correction.) Indeed, the theoretical framework in which we state these models come equipped

---

[42] However, Grice does not discuss robustness as a potential justification for this norm. His focus is more on how unnecessary precision can be distracting.

with resources to quantify the degree to which the behavior of the system[43] is deviating from the behavior predicted by the two-body model: we can define appropriate distance measures between the two-body orbit and the three-body orbit, or specify the deviation of orbital parameters such as eccentricity, or quantify the amount of orbital precession, and so on.[44] Thus, we see that borderline cases—cases where we have no principled reasons to accept wholesale or reject wholesale a relevant model—can be seen as a consequence of graceful degradation: the fact that a case is borderline allows the use of a nearby model that is still useful despite failing. In the context of ordinary language, we see a similar phenomenon. When we classify someone as borderline bald, we are neither accepting nor rejecting the classificatory model of *bald/not bald* wholesale, as we might do if we are asked whether a building is bald or not bald (in non-metaphorical terms), or as we might do with someone who has an unusual hairstyle that leaves exactly one half of their scalp without hair. In contrast, for a borderline bald person, the bald/not bald model isn't entirely useless; it establishes relevant classes of cases with which to compare the case at hand. For such cases the bald/not bald model, though degraded, is gracefully degraded.

So we have dealt with borderline cases and sorites-susceptibility and seen how each of those can be explained, respectively, by graceful degradability and robustness. Let us now turn to higher-order vagueness, and see how that is explained by error-correctability, the third feature of fault-tolerance I have listed above. A system is error-correctable just if it has affordances that allow one to repair the system when it is encountering faults. When a bridge begins to display cracks, we want to be able to patch those cracks. In the context of our discussion, error-correctibility is realized by affordances of our models which allow for the construction of intermediate models, models that work better than the initial models, which are now breaking down (as discussed in Sec. 6). I claim that higher-order vagueness is a consequence of the iterated error-correctability of our models: i.e., to error-correct not just the initial failing models, but also the new intermediate models we construct because they too will inevitably fail in certain circumstances.

Let's see how the process of constructing intermediate models works, first in physics and then in ordinary language. Suppose we are in a situation where a model is breaking down: Say that according to a three-body model, the inner

---

[43] As clarified in Sec. 4, when I say "the behavior of the system", I mean the behavior of the system according to the three-body model.

[44] See, e.g., Fitzpatrick (2012) for details.

planet is in an orbit which is deviating from an ellipse. In this regime, a two-body model isn't in adequate agreement with the three-body model. A natural and standard move here is to consider a new, corrected model. So let's say the orbit of the inner planet according to the three-body model is deviating from an ellipse because the outer planet is too massive to have its influence approximated away. This would perturb the orbital parameters (such as the eccentricity and inclination of the orbit) of the inner planet so that they are no longer constants in time, as they would be under a two-body model. To incorporate this, we can generate a new model that supplies equations that tell us *how* the orbital parameters of the inner planet change due to the perturbation.[45] These equations would constitute an *intermediate* model: a model on which the deviation from an elliptical orbit that we earlier deemed a symptom of the breakdown of the two-body description, is now a clearly explicable case according to the new model.

We can construct similar intermediate models in ordinary language as well. When we encounter situations that our categories—which carry with them their implicit models—cannot capture, we are able to add *corrections* to our pre-existing categories to create new penumbral categories. Our language contains tools to do that. These are adverbial phrases such as *kind of*, *sort of*, *almost*, *nearly*, and so on. So if someone is above-average height but not clearly tall, we often say "they're *kind of* tall". Such linguistic tools are of a piece with perturbation theory in physics: they offer us the ability to *extend* our current models to nearby domains where our current models don't work so well.

So we see how we can extend our pre-existing models to handle their breakdown regimes. But what happens when the newer, extended models start breaking down in their own turn, when faced with certain situations? So, e.g., maybe the variation in the orbital parameters according to the three-body model is deviating even more so than what is predicted by the intermediate model. Or, e.g., if we encounter someone for whom we cannot clearly decide whether or not they are kind of tall. These situations where intermediate models break down are higher-order borderline cases. To deal with such situations, we can iterate the process of error-correction again and generate newer higher-order intermediate models to deal with the novel situations. In the physics case, the standard way to do this is to add more terms to one's perturbation series yielding new models that can capture finer detail in the behavior of the orbit. In ordinary

---

[45] This would be a perturbation theory in the ratio of the mass of the outer planet to the mass of the central star. See Fitzpatrick (2012, Ch. 9) for the gory details.

language we can do this by chaining the adverbial phrases *kind of*, *sort of*, and so on. So if I say, “He’s kind of tall”, someone can reasonably respond to me and say, “Well, he’s only kind of sort of tall”, and we might both agree with that characterization.

As a brief aside, it is worth remarking here that judgments about how well a model vs. its extended version fits a given situation depends on the context of use. So, if we are in a context where sorting a building as *tall* vs. as *kind of tall* makes a significant difference (say because it is relevant to the decision of a legal dispute), then we will may very well spend quite a lot of time debating whether or not the relevant building is tall or only kind of tall; however, if the context is just that of a casual conversation, then I might easily concede that the building is only kind of tall. Similarly, I might demand a much more precise fit, and hence demand an extended version of a model of a solar system, if I’m planning on launching a spacecraft from one planet to land on the other planet, for then, I will need to be very confident in the location of both planets at certain times; however, I might not demand as much precision, and hence settle for a simpler model, if I’m only interested in some coarser patterns of the solar system.[46]

Returning to the main thread, to get vagueness at higher and higher orders, all we need to do is see that models, including new intermediate or penumbral models will also break down, and to note that these new models are in turn fault-tolerant, i.e., intermediate models too are robust, degrade gracefully, and are error-correctable. Insofar as we can keep extending our fault-tolerant models in fault-tolerant ways, we will keep getting higher and higher orders of vagueness.[47]

To sum up, we have seen how the characteristic features of vagueness—borderline cases, sorites-susceptibility, higher-order vagueness—arise *when* models breakdown, both in physics and in ordinary descriptive discourse. Further, we have seen that a plausible explanation for *why* these features arise during breakdown is that they are the consequences of these models being fault-tolerant, in particular, these models degrading gracefully, being robust, and being error-correctable.

---

[46] Note though that while fixing the context might help in the selection of a relevant model, this wouldn’t eliminate vagueness, which concerns the boundaries of these models.

[47] There is a debate in the literature concerning whether vagueness will always carry through to arbitrarily high orders or whether it will terminate at some order or other: see Williamson (1999), Mahtani (2008), and Dorr (2015) for a particular thread of this debate. I don’t need to, and I don’t, take a stand on this issue here.

8 Conclusion

The world is a complicated place, and we have to construct many different models to navigate this complicated place. Different models have different regimes of validity, i.e., different kinds of situations for which they can serve as a good model. It turns out that there are no sharp boundaries between those kinds of situations that a model is good model for and those kinds of situations that it isn't.

The characteristic features of vagueness arise when models break down. If our models are fault-tolerant, then they would display the characteristic features of vagueness during break down. Our ordinary descriptive language can also be seen as modeling the world much as models in physics model the world. Putting these things together, we have a unified explanation for vagueness as it arises both in physics and in ordinary descriptive language—namely that it is the consequence of the breakdown of fault-tolerant models.

Acknowledgments—Special thanks to David Wallace, Kate Stanton, and Bob Batterman for detailed feedback on drafts and many conversations at different stages of this project. Thanks also to Colin Allen, Vera Matarese, Marian Gilton, Bixin Guo, Travis McKenna, John Norton, Jeffrey Russell, and Mark Wilson for feedback and conversations.

References

Albert, David Z. 2000. *Time and Chance.* Cambridge, Mass.: Harvard University Press.

Andreas, Holger. 2021. "Theoretical Terms in Science." In *The Stanford Encyclopedia of Philosophy,* Fall 2021, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University.

Asgeirsson, Hrafn. 2020. *The nature and value of vagueness in the law.* Bloomsbury Publishing.

Bacon, Andrew. 2018. *Vagueness and Thought.* Oxford, England: Oxford University Press.

Barrow-Green, June. 2010. "The Three-Body Problem." In *The Princeton companion to mathematics,* edited by Timothy Gowers, June Barrow-Green, and Imre Leader. Princeton University Press.

Brandom, Robert B. 2019. *A spirit of trust: A reading of Hegel's phenomenology.* Harvard University Press.

Burge, Tyler. 1986. "Individualism and Psychology." *Philosophical Review* 95 (January): 3–45. https://doi.org/10.2307/2185131.

Carey, Susan. 2009. *The Origin of Concepts.* New York: Oxford University Press.

Chang, Hasok. 2004. *Inventing temperature: Measurement and scientific progress.* Oxford University Press.

Chen, Eddy Keming. 2022. "Fundamental Nomic Vagueness." *Philosophical Review* 131 (1): 1–49. https://doi.org/10.1215/00318108-9415127.

Chirimuuta, Mazviita. 2015. *Outside Color: Perceptual Science and the Puzzle of Color in Philosophy.* Cambridge, Mass.: The MIT Press.

Cook, Roy T. 2002. "Vagueness and Mathematical Precision." *Mind* 111 (442): 225–247. https://doi.org/10.1093/mind/111.442.225.

Deacon, Terrence William. 1997. *The symbolic species: The co-evolution of language and the brain.* 202. WW Norton & Company.

Dennett, Daniel C. 2017. *From bacteria to Bach and back: The evolution of minds.* WW Norton & Company.

Dorr, Cian. 2015. "How Vagueness Could Cut Out at Any Order." *Review of Symbolic Logic* 8 (1): 1–10. https://doi.org/10.1017/s175502031400032x.

Dubrova, Elena. 2013. *Fault-tolerant design.* Springer.

Eklund, Matti. 2005. "What Vagueness Consists In." *Philosophical Studies* 125 (1): 27–60. https://doi.org/10.1007/s11098-005-7773-1.

Endicott, Timothy. 2011. "The Value of Vagueness." In *Philosophical foundations of language in the law,* edited by Andrei Marmor and Scott Soames. Oxford University Press.

Endicott, Timothy A. O. 2000. *Vagueness in Law.* New York: Oxford University Press UK.

Everett, Daniel L. 2012. *Language: The cultural tool.* Vintage.

Fine, Kit. 1975. "Vagueness, Truth and Logic." *Synthese* 30 (3-4): 265–300. https://doi.org/10.1007/BF00485047.

Fitzpatrick, Richard. 2012. *An introduction to celestial mechanics.* Cambridge University Press.

Fletcher, Samuel C., Patricia Palacios, Laura Ruetsche, and Elay Shech. 2019. “Infinite Idealizations in Science: An Introduction.” *Synthese* 196 (5): 1657–1669. https://doi.org/10.1007/s11229-018-02069-6.

Frigg, Roman, and Stephan Hartmann. 2020. “Models in Science.” In *The Stanford Encyclopedia of Philosophy,* Spring 2020, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University.

Gabrielse, Gerald. 2013. “The standard model’s greatest triumph.” *Physics Today* 66, no. 12 (December): 64–65. ISSN: 0031-9228. https://doi.org/10.1063/PT.3.2223. eprint: https://pubs.aip.org/physicstoday/article-pdf/66/12/64/15826275/64_1_online.pdf. https://doi.org/10.1063/PT.3.2223.

Gopnik, Alison, and Andrew N. Meltzoff. 1998. *Words, Thoughts, and Theories.* The MIT Press, September. ISBN: 9780262274098. https://doi.org/10.7551/mitpress/7289.001.0001. https://doi.org/10.7551/mitpress/7289.001.0001.

Grice, Herbert Paul. 1989. *Studies in the Way of Words.* Cambridge: Harvard University Press.

Guo, Bixin. 2023. “Two Approaches to Reduction: A Case Study from Statistical Mechanics.” *Philosophy of Science,* 1–36. https://doi.org/10.1017/psa.2023.52.

Hawthorne, John. 2022. “Moral Vagueness and Epistemicism.” *Journal of Ethics and Social Philosophy* 23 (2). https://doi.org/10.26556/jesp.v23i2.1602.

Heyes, Cecilia. 2018. *Cognitive gadgets: The cultural evolution of thinking.* Harvard University Press.

Hull, David L. 1988. *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science.* University of Chicago Press.

Keefe, Rosanna, and Peter Smith, eds. 1996. *Vagueness: A Reader.* The MIT Press.

Keil, Frank C. 1992. *Concepts, Kinds, and Cognitive Development.* The MIT Press, January. ISBN: 9780262276900. https://doi.org/10.7551/mitpress/2065.001.0001. https://doi.org/10.7551/mitpress/2065.001.0001.

Keil, Geert, and Ralf Poscher, eds. 2016. *Vagueness and Law: Philosophical and Legal Perspectives.* Oxford: Oxford University Press.

Kleppner, Daniel, and Robert Kolenkow. 2014. *An introduction to mechanics.* Cambridge University Press.

Knox, Eleanor, and David Wallace. 2023. *Functionalism Fit for Physics,* October. http://philsci-archive.pitt.edu/22631/.

Kuhn, Thomas S., and Ian Hacking. 2012. *The Structure of Scientific Revolutions: 50Th Anniversary Edition.* University of Chicago Press.

Lanius, David. 2019. *Strategic Indeterminacy in the Law.* New York: Oxford University Press.

MacFarlane, John. 2016. "Vagueness as Indecision." *Aristotelian Society Supplementary Volume* 90 (1): 255–283. https://doi.org/10.1093/arisup/akw013.

Mahtani, Anna. 2008. "Can Vagueness Cut Out at Any Order?" *Australasian Journal of Philosophy* 86 (3): 499–508. https://doi.org/10.1080/00048400802001954.

Miller, Michael E. 2021. "Worldly imprecision." *Philosophical Studies* 178 (9): 2895–2911.

Millikan, Ruth Garrett. 1984. *Language, Thought, and Other Biological Categories: New Foundations for Realism.* MIT Press.

———. 2017. *Beyond Concepts: Unicepts, Language, and Natural Information.* Oxford, United Kingdom: Oxford University Press.

Nelson, Robert A, and MG Olsson. 1986. "The pendulum—Rich physics from a simple system." *American Journal of Physics* 54 (2): 112–121.

O'Connor, Cailin. 2020. *Games in the Philosophy of Biology.* Cambridge University Press.

Oxford English Dictionary. 2023. *s.v. "bald (adj.), sense I.2.a".* Accessed September 2023, September. https://doi.org/10.1093/OED/9473018362.

Pinker, Steven. 1994/2007. *The Language Instinct: How the Mind Creates Language.* Harper Perennial.

Polanyi, Michael. 1958. *Personal Knowledge.* Chicago, University of Chicago Press.

Polanyi, Michael, and Amartya Sen. 1966. *The Tacit Dimension.* Chicago, IL: University of Chicago.

Pradhan, Dhiraj K. 1996. *Fault-tolerant computer system design.* Vol. 132. Prentice-Hall Englewood Cliffs.

Putnam, Hilary. 1975. “The Meaning of ’Meaning’.” *Minnesota Studies in the Philosophy of Science* 7:131–193.

Richard, Mark. 2019. *Meanings as Species.* Oxford, United Kingdom: Oxford University Press.

Sainsbury, R. M. 1996. “Concepts Without Boundaries.” In *Vagueness: A Reader,* edited by Rosanna Keefe and Peter Smith, 186–205. MIT Press.

Schoenfield, Miriam. 2016. “Moral Vagueness is Ontic Vagueness.” *Ethics* 126 (2): 257–282. https://doi.org/10.1086/683541.

Shapiro, Stewart. 2006. *Vagueness in Context.* Oxford University Press.

Sorensen, Roy. 2022. “Vagueness.” In *The Stanford Encyclopedia of Philosophy,* Winter 2022, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University.

Sud, Rohan. 2019. “Moral Vagueness as Semantic Vagueness.” *Ethics* 129 (4): 684–705. https://doi.org/10.1086/702976.

Thorne, Kip S, and Roger D Blandford. 2017. *Modern Classical Physics: Optics, Fluids, Plasmas, Elasticity, Relativity, and Statistical Physics.* Princeton University Press.

Wallace, David. 2022. “Stating Structural Realism: Mathematics-First Approaches to Physics and Metaphysics.” *Philosophical Perspectives* 36 (1): 345–378. https://doi.org/10.1111/phpe.12172.

Weatherson, Brian. 2016. “The Problem of the Many.” In *The Stanford Encyclopedia of Philosophy,* Winter 2016, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University.

Williams, Porter. 2023. *Philosophy of Particle Physics.* Elements in the Philosophy of Physics. Cambridge University Press. https://doi.org/10.1017/9781009205382.

Williamson, Timothy. 1994. *Vagueness.* Routledge.

Williamson, Timothy. 1999. “On the Structure of Higher-Order Vagueness.” *Mind* 108 (429): 127–143. https://doi.org/10.1093/mind/108.429.127.

Wilson, Mark. 1982. “Predicate meets property.” *The Philosophical Review* 91 (4): 549–589.

———. 2006. *Wandering significance: An essay on conceptual behavior.* Oxford University Press.

Wimsatt, William C. 2007. *Re-engineering philosophy for limited beings: Piecewise approximations to reality.* Harvard University Press.

Wright, Crispin. 1975. “On the Coherence of Vague Predicates.” *Synthese* 30 (3/4): 325–365. ISSN: 00397857, 15730964, accessed August 29, 2023. http://www.jstor.org/stable/20115035.