

# CRYSTALLIZED REGULARITIES<sup>1</sup>

Verónica Gómez Sánchez

The nomic structure of our world spans many levels of description. The explanatory and predictive success of the ‘special sciences’ – biology, psychology, geology, and so on – reveals the existence of robust regularities (sometimes called ‘special science laws’) that knit non-fundamental phenomena into intelligible levels of description. There are two conceptions of how these robust regularities fit into the physical world. On a foundationalist conception, the physical laws (or physical properties) are the source of all other nomic facts, including the robustness of these macro-regularities. On an egalitarian conception, the physical laws are no more fundamental than the laws describing the behavior of genes, ecosystems, or societies.<sup>2</sup>

Egalitarians and foundationalists agree that generalizations about nonfundamental reality can differ in nomic status: Mendel’s laws of genetics and the laws of thermodynamics have an ‘elite’ status, which is not shared by merely true generalizations such as ‘All the coins in my pocket are silver’. But they disagree about how this elite status relates to physics. While egalitarians give a single account of eliteness, which applies both to physical and higher-level generalizations, the foundationalist's account of eliteness divides into two (roughly independent) parts. The first part is understanding what, if anything, generates the fundamental laws of physics. (Here the term ‘fundamental law’ refers to those regularities whose lawful status doesn’t derive from more basic lawhood facts – even if their being

<sup>1</sup> This paper has benefitted from discussions with many people, including David Albert, Laura Callahan, Sam Carter, Ruth Chang, Eddy Chen, Barry Loewer, Jill North, Jonathan Schaffer, Isaac Wilhelm, and Tyler Wilson. I’m especially grateful to Ezra Rubenstein, Ted Sider and an anonymous referee at this journal for extensive feedback on earlier drafts.

<sup>2</sup> Egalitarianism is not a very unified position. Fodor is egalitarian in holding that special science laws have their own metaphysical ‘oomph’, which they do not inherit from the physical laws. See J. A. Fodor, "Special Sciences (or: The Disunity of Science as a Working Hypothesis)," *Synthese* XXVIII, 2 (October 1974): 97–115. Conversely, Nancy Cartwright, Craig Callender, and Jonathan Cohen arrive at an egalitarian conception by downgrading the metaphysical status of physical laws. For them, all lawful generalizations earn their nomic status in the same way: by capturing patterns that our experience of the world presents us with, in a vocabulary that is useful for us. See Nancy Cartwright, *The Dappled World: Essays on the Perimeter of Science* (New York: Cambridge University Press, 1999); Craig Callender and Jonathan Cohen. "Special Sciences, Conspiracy and the Better Best System Account of Lawhood," *Erkenntnis*, LXXIII, 3 (November 2010): 427–47; and Jonathan Cohen and Craig Callender, "A Better Best System Account of Lawhood," *Philosophical Studies*, CXLV, 1 (July 2009): 1–34.

laws is not fundamental).<sup>3</sup> The second part is understanding how those fundamental physical laws generate nomic facts spanning all levels of reality.

This paper aims to advance the second part of the foundationalist project. I develop and defend a reductive account of ‘robust generalizations’ or ‘special science laws’ which vindicates various aspects of scientific practice that may initially seem in tension with foundationalism. This account accommodates physically contingent, non-strict robust generalizations that have privileged formulations in non-fundamental vocabularies. The basic idea is that robust generalizations earn their elite status by featuring in the ‘ideal scientific summary’ of the nearby physical possibilities.

## I. PRELIMINARIES

*I.1. The theoretical role of robustness.* When we’re justified in attributing robustness to a generalization, we’re thereby entitled to rely on it in characteristic ways when engaged in predicting, reasoning hypothetically, and explaining. Before putting forward a metaphysical account of robustness, it helps to rehearse these conceptual connections. This will help clarify the target concept, as well as its significance to our theorizing.

Consider firstly the connection between robustness and prediction. Encountering various black ravens warrants the conclusion that the next encountered raven will be black. But observing that some of the coins in someone’s pocket are silver doesn’t similarly justify the expectation that unobserved coins in her pocket are also silver. This intuitive asymmetry plausibly results from the fact that the generalization ‘All ravens are black’ seems like a good candidate for a robust regularity, while the generalization ‘All coins in my pocket right now are silver’ doesn’t.

Secondly, robust generalizations support counterfactuals/hypotheticals. If you take ‘All ravens are black’ to be robust, you probably also endorse (and should endorse) a set of associated counterfactuals: ‘If a raven had been born a minute ago, it would have been black’, ‘If I were to fill this room with ravens, they would all be black’, and so on. In evaluating such conditionals, you consider alternative ways things might have been while holding fixed the raven-black connection. Contrast this to an accidental generalization such as: ‘All the people in this room are tall’. I might believe this to be

<sup>3</sup> I’ll be neutral on whether fundamental laws are metaphysically primitive, summaries of the world’s actual history, or generalizations that flow from the essences of properties.

true, but I do not endorse counterfactuals such as: ‘If Anne were in this room, she would be tall’, ‘If someone were to enter this room, they would be tall’, and so on.<sup>4</sup>

Thirdly, robustness connects to explanation. This connection has two aspects. On the one hand, robust generalizations explain their instances. ‘All ravens are black’ explains why the particular ravens I happened to see today were black, whereas ‘Everyone in this room is tall’ doesn’t explain why Anne, who happens to be in this room, is tall. On the other hand, robust generalizations are capable of backing explanations of one particular state of affairs in terms of another. Why am I sleepy? Because I skipped my morning coffee. Here, the generalization ‘Coffee-drinkers who are caffeine-deprived get sleepy’ serves as a linking principle between two properties – *sleepiness* and *caffeine-deprivation* – licensing an explanation of an instance of the first in terms of an instance of the second. Only robust generalizations link properties in this way. I cannot explain why Anne is tall by citing the fact that she entered the room a minute ago. In a nutshell: explanations trace patterns of dependence that are underwritten by robust regularities. (The robustness of the generalization does not, however, suffice for the success of an explanation: not any sound argument containing a robust generalization as a premise is explanatory.)<sup>5</sup>

Because of these connections to predictions, hypotheticals, and explanations, robust regularities may deserve to be called ‘laws’. I prefer to use a different term, to avoid connotations that the term ‘law’ has acquired. We should not assume at the outset that robust regularities are physically necessary, perfectly accurate, unrestricted in scope, and so on. Moreover, I want to leave open the possibility that robust regularities are importantly different from physical laws, even if they play some of the same roles.

*I.2. Robustness vs. Physical Necessity.* Providing a foundationalist account of robustness may initially seem like an easy task. The foundationalist can recognize a basic nomic distinction among macro-regularities: some are physically necessary (they are necessitated by the physical laws), and some are physically

<sup>4</sup> As Michael Strevens points out in "Physically contingent laws and counterfactual support," *Philosopher's Imprint*, VIII, 8 (August 2008): 1–20, contingent generalizations like ‘All ravens are black’ are not held fixed under *all* counterfactual suppositions. I follow Strevens in thinking that counterfactual support is a matter of degree: robust regularities have a much higher degree of counterfactual support than accidental generalizations. (I discuss this further in § 4.2).

<sup>5</sup> The claim that only robust regularities back explanation is compatible with causal/difference-making accounts of explanation that attribute a central role to causal generalizations. We can think of causal generalizations as constituting a special sub-class of the robust regularities, whose membership conditions are to be specified by a theory of causation/causal explanation.

contingent (they fail to hold in some physically possible scenarios). But more careful reflection reveals that the notion of physical necessity does not coincide with the notion of robustness that I seek to understand. Many physical necessities are not regarded as robust from the perspective of the higher-level sciences, and some physically contingent regularities are so regarded.

Grant, for argument's sake, that 'All emeralds are green' is a physically necessary robust generalization. The truth of this statement metaphysically entails the truth of 'All emeralds are green-or-blue', so the latter will be physically necessary as well. However, the generalization involving the disjunctive property *green-or-blue* does not thereby acquire the elite explanatory status that the former statement has. This suggests that the class of robust regularities is not closed under metaphysical necessitation: a statement's robustness is not automatically transmitted to the statements that it necessitates.

For familiar reasons, we cannot deal with this by associating robustness with a particular syntactic form. Compare 'All emeralds are green' and 'All gremeralds are grue' (where *x* is a gremerald if and only if it is an emerald and observed before 3000, or a sapphire and not observed before 3000). The two generalizations are syntactically alike, and have the same modal strength. Nonetheless, they clearly differ in explanatory power.

Note that this issue cannot be solved by imposing a general ban on gerrymandered properties. Take 'Consuming caffeine helps people stay awake'. While this statement is plausibly robust, it entails non-robust generalizations that involve no disjunctive or gerrymandered predicates. Consider a variant of an example due to Henry Kyburg<sup>6</sup>: let an object be 'hexed' if and only if some hexing gestures are performed over it. 'Consuming caffeine helps people stay awake' metaphysically necessitates 'Consuming caffeinated beverages which have been hexed helps people stay awake'. But the latter generalization does not inherit the explanatory status of the original one. If I'm trying to explain why I'm more likely to be up late when I drink coffee in the evening, the right regularity to cite is the one about caffeine, not the one about hexed caffeinated beverages.

One could insist that the above differences are not differences in nomic status, and argue on this basis that explaining them is not a task for a theory of robustness. One can pass the task over to the theory of explanation, which could then pass it along to the pragmatics department. It should be apparent, however, that other things being (roughly) equal, we should prefer a theory of robustness which helps us capture the above differences in explanatory power. Thus, I'll be reserving the term

<sup>6</sup> Philip Kitcher, "Explanatory Unification," *Philosophy of Science*, XLVIII, 4 (December 1981): 507–31.

‘robustness’ for a class of explanatorily elite statements, which is not closed under metaphysical necessitation. It will not follow from the claim that ‘Fs are Gs’ is robust that ‘Fs are Gs-or-Hs’ is robust, even if the former necessitates the latter.

I have presented reasons to think that physical necessity does not suffice for robustness. There are also reasons to doubt that physical necessity is necessary for robustness. There is a growing consensus that many physically contingent generalizations are capable of playing the roles associated with robustness.<sup>7</sup>

This ‘contingency hypothesis’ has been most influential in the philosophy of biology, after John Beatty persuasively argued that no known biological generalizations are physically necessary.<sup>8</sup> The key premise in Beatty’s argument is that all biological generalizations have been found to have exceptions. Take Mendel’s law of segregation: ‘In diploid organisms, the probability that each allele at a genetic locus is transmitted to a child is 50%’. A few species have been found to have genes that do not segregate in a 50:50 fashion, but exhibit a bias in favor of one of the alleles. Or consider generalizations about genetic coding, which specify pairings between nucleic acids and amino acids. These generalizations, once thought universal, turn out to have exceptions in certain eukaryotes.

Exceptions are significant because they reveal the dependence of biological generalizations on conditions that look contingent from the perspective of physics. For instance, exceptions to generalizations about the genetic code make salient alternative ways this code could have been set up. Investigating these alternatives led Francis Crick to formulate the ‘frozen accident’ hypothesis: the idea that the genetic code is chemically arbitrary, but stable once established.<sup>9</sup>

It may be true nonetheless that the above generalizations can be associated with physically necessary conditionals of the form: under conditions C, G (where C describes some background conditions that – together with the physical laws – entail the generalization G). But this alone cannot be what distinguishes them from accidental regularities. Take a true sentence of the form ‘In year  $x$ , it rained every other Tuesday’. If the laws are deterministic, there is some contingent condition C such

<sup>7</sup> See Gerhard Schurz, “What Is Normal? An Evolution-Theoretic Foundation for Normic Laws and Their Relation to Statistical Normality,” *Philosophy of Science*, LXVIII, 4 (December 2001): 476–497; Sandra D. Mitchell, “*Ceteris Paribus* – An Inadequate Representation For Biological Contingency,” *Erkenntnis*, LVII (November 2002): 329–50; and Michael Strevens, “Physically contingent laws and counterfactual support,” *Philosopher’s Imprint*, VIII, 8 (August 2008): 1–20.

<sup>8</sup> John Beatty, “The Evolutionary Contingency Thesis,” in Gereon Wolters and James G. Lennox, eds., *Concepts, Theories, and Rationality in the Biological Sciences* (Pittsburgh: University of Pittsburgh Press, 1995), pp. 45–81.

<sup>9</sup> Francis Crick, “The Origin of the Genetic Code,” *Journal of Molecular Biology*, XXXVIII, 3 (December 1968): 367–379.

that ‘if C, then in year  $x$  it rained every other ‘Tuesday’ is physically necessary. Yet this doesn’t show that the pattern of rain in year  $x$  was non-accidental. We need a principled distinction between physically necessary conditionals with robust consequents, and ones with accidental consequents.

We have seen that physical necessity is neither necessary nor sufficient for robustness. If robustness is inherited by the macro from the micro, the inheritance principle is not as simple as: ‘All and only physical necessities are robust’. We need to find inheritance principles which ensure that nomic force i) reaches only those generalizations with the right structure and level of generality, and ii) can reach physically contingent generalizations. If no such principles could be given, egalitarianism would gain a lot of plausibility. But if we can provide such principles, there is no pressure to think that higher-level nomic structure is metaphysically autonomous.

## II. CRYSTALLIZATION

The modal structure that our physical theories posit outstrips the notion of ‘physical necessity’. Physical theories come equipped with state-spaces, which tell us not only which configurations of the fundamental physical stuff are possible, but also which states are similar or ‘nearby’. If we take the current physical state of our world and change the location of three fundamental particles by 1nm in random directions, we get a state that is physically nearby. If, instead, we move around all of the particles that make up New York City, we will get a state that is further away.

In what follows, I will develop the idea that our scientific efforts are primarily aimed the discovery of patterns that hold across our ‘modal neighborhood’: a class of possible worlds that are ‘similar’ to ours (where world similarity is defined in terms of distance in a physical state-space).

A generalization’s holding at all nearby worlds cannot, however, suffice to make it robust. The property of holding at nearby worlds behaves like a necessity modal (with a more restrictive accessibility relation than physical necessity). But robustness cannot be equated with any necessity modal, since it is not preserved under entailment.

Taking inspiration from Lewis’s influential account of laws<sup>10</sup>, I suggest that what makes some regularities stand out as scientifically elite is their ability to efficiently summarize our modal neighborhood. I call such regularities ‘crystallized’. Crystallized regularities are the axioms of the best

<sup>10</sup> David Lewis, *Counterfactuals* (Malden: Blackwell, 1973).

system for our modal neighborhood, where this system spans all levels of description. The rest of this section develops and refines this definition of crystallization, by defining the notions ‘modal neighborhood’ and ‘best system’ in foundationalist-friendly terms.

*II.1 Modal Neighborhoods.* Imagine possibility-space laid out around you. You are standing at the actual world, whose current state is  $s$ . Assuming physicalism,  $s$  can be exhaustively characterized in terms of the fundamental properties of the most basic physical objects (particles/space-time points), and the fundamental relations they bear to each other. To simplify the exposition, let us pretend that we are living in a classical world. Then, an exhaustive description of the current physical state consists of a specification of the position and momenta of each fundamental particle. The space of all possible states so described is known as ‘phase space’.

From the perspective of a given point in phase space (in this case, the point corresponding to state  $s$ ), we can ask which possible states are *nearby*. Then a state  $s'$  is near our state  $s$  if almost all the particles have similar positions and velocities in  $s'$  as they do in  $s$ . Given a state-description for  $s$  (that is, a specification of positions and momenta for each particle), we can create descriptions for nearby states by minimally modifying  $s$  – for example, by changing the location of a few hundred particles by a couple of meters, giving some particles a small velocity boost... We can measure the distance of these modified states from the actual state  $s$  by appeal to a metric on the state-space: in this case, a Euclidean metric on phase space is the natural choice.<sup>11</sup>

Once we have a notion of nearness for pairs of physical states, we can define nearness for worlds (which correspond to trajectories through state-space): Two worlds  $w$  and  $v$  are near at time  $t$  if (i) they have the same laws, and (ii) their states at  $t$  are near each other. If you collect all the worlds that are (or will be) near ours at some time in your lifetime, you get your ‘modal neighborhood’.

<sup>11</sup> Although phase space does not have intrinsic metric structure –there is no built-in definition of ‘distance’ between points – its structure does favor some ways of measuring similarity over others. For instance, it would be natural to pick a metric that obeyed the following constraints: two states that differ only in the position of one particle (by 1cm) are more similar to each other than two states that differ in the position of more particles (each by 1cm); if two states differ only in the position of a single particle, their degree of similarity should be inversely proportional to the distance between the particle’s respective positions.

Various choices of units for the different physical quantities (position and momentum in our case) yield different Euclidean metrics. The incommensurability of these different quantities needn’t be a problem; the account succeeds as long as all reasonably natural metrics make similar predictions.

To appreciate why the idea of a ‘modal neighborhood’ might be of interest, take the robust regularity ‘All ravens are black’. Suppose we have granted that this regularity is physically contingent: non-black ravens populate many physically possible worlds. They can be found not only in exotic/very distant physically possible worlds, but also in worlds very much like ours at some point in history. Rewind to a time long before common ravens existed, when the first RNA populations were ready to undergo Darwinian evolution. Minor alterations to the state of the world back then would likely lead to dramatic differences in the course of evolutionary history. Worlds that were ‘nearby’ back then may now contain populations of brown ravens. Thus, the ravenhood-blackness connection was once fragile.

Fast-forward to the present moment. Take the world’s complete state at some time in our epoch: the current time, yesterday, tomorrow, or a year from now. Consider minor alterations to that state: give a few particles a velocity boost, cut the wings off a mosquito in your office, delete a word in this paper. Each one of those ‘physically small’ alterations determines a physically similar possible state of the world. Evolve each of those states by the physical laws, in both temporal directions. (Assume for now that the laws are deterministic, so each state determines a single world). The minor alterations you made at the chosen time result in divergences from the actual world, both in the future and the past. In the nearby worlds with the wingless mosquito, you will not be bitten 20 times. So, you will finish your work earlier, you will arrive home earlier... and so on.

Will these differences affect the color of ravens? Probably not.<sup>12</sup> Even if you were to fiddle with a few raven eggs, you will not easily re-direct the evolutionary path of the species. In general, small physical changes will not affect the color of ravens, perhaps with the exception of very special targeted manipulations of many raven reproductive cells. ‘Ravens are black’ was once fragile, but it has become crystallized.

This suggests that epochs differ with respect to their crystallized regularities. This should not come as a surprise. I have granted already that some crystallized generalizations depend on contingent background conditions. In the case of ravens, for instance, their color depends on conditions in the distant past (before raven evolution had taken its course), but also on conditions of the current environment: for example, the absence of established human practices of bleaching ravens. We do not normally expect these background conditions to change in the near future. But, given their

<sup>12</sup> Some readers may worry that evolving nearby states towards the past with the dynamical laws will often yield anti-entropic world histories. I address this concern at the end of this section.

contingency, they may well change some day. And if this contingency doesn't threaten the explanations and inferences we base on the generalization, the fact that it will one day cease to hold should not either.

Similar reasoning suggests that different spatial regions may support different robust regularities at a given time. Suppose you were to learn about a human population that traveled from Earth to another planet long ago, and has been established there ever since. Their visual and auditory systems may have subtly changed in response to their new environment, meaning that our best psychological theories do not describe them accurately. Arguably, this should not make any more of a difference to our Earth-bound psychological theorizing than what happens in other physically possible worlds.<sup>13</sup>

To allow for crystallization to vary across epochs and locations, I will relativize the notion of a modal neighborhood to a spatio-temporal region  $R$ . Suppose, for instance, that our region of interest is that occupied by Earth from 1000 to 3000. The modal neighborhood of  $R$  at our home world  $w$  is a set of physically possible ways  $R$  could have been which are nearby from the perspective of  $w$ . I'll use world-region pairs  $\langle R, v \rangle$  to represent ways region  $R$  might have been (the way  $R$  is at world  $v$  is a way  $R$  could have been at the home world  $w$ ). More precisely:

#### MODAL NEIGHBORHOOD

The modal neighborhood of  $\langle R, w \rangle$  is the set of all pairs  $\langle R, v \rangle$  such that:

- 1)  $v$  and  $w$  have sufficiently similar (complete) states at some time in  $R$ , and
- 2)  $v$  obeys  $w$ 's laws.<sup>14</sup>

<sup>13</sup> Someone might deny that this faraway civilization is *human*. However, there is surely some natural kind that encompasses both civilizations (plausibly, the biological notion 'human species'). Let 'pan-human' pick out that broader kind. In the imagined scenario, we would still like to predict and explain seemingly non-accidental patterns concerning the broader kind (for example, 'All pan-humans studied by X's lab in 2020 perceived stimulus  $S$  as a rigid object'). Now, if all robust psychological generalizations are about the narrow kind *human*, then explanations and predictions of such regularities (concerning pan-humans) would have to rely on a generalization linking humans and pan-humans (for instance, 'All members of the human species in earth are humans'). Thus, not all region-relativity would be eliminated. (Thanks to an anonymous referee for raising this issue.)

<sup>14</sup> This definition assumes primitive trans-world identity for spatio-temporal regions, as well as a privileged simultaneity structure. These assumptions simplify the presentation, but are not essential to the account. The definition could be rephrased in terms of Cauchy surfaces and counterpart relations on space-time regions.

For example, take  $\langle \text{Earth-today, actual world} \rangle$  as our domain of interest. The following recipe yields worlds in its neighborhood. Pick any time  $t$  in the relevant epoch: in this case, any time today. Identify the state of the world at  $t$  – call it  $s$ . Make a physically small change to  $s$  within the relevant spatial region; for instance, make the velocity of a bird in Antarctica slightly slower. This change takes us to a nearby state  $s'$ . The physically possible world(s) whose state at  $t$  is  $s'$  will be one of many worlds in the modal neighborhood of  $\langle \text{Earth-today, actual world} \rangle$ . Note that this world must differ from actuality at times before and after  $t$ . Shortly before  $t$ , the mental state of the bird would have been different. Shortly after  $t$ , the mental state of anything perceiving the bird would register a different velocity. And so on...<sup>15</sup>

We can repeat this procedure for any time  $t$  today; any world we reach is in the modal neighborhood. Thus, the modal neighborhood of  $\langle \text{Earth-today, actual-world} \rangle$  contains a continuous infinity of richly varying worlds. But, by construction, these worlds have at least two things in common: 1) they obey the actual laws of physics, and 2) they closely resemble the actual world at some time today (but different worlds resemble the actual world at different times). To be crystallized relative to *Earth-today*, a generalization  $G$  ought to hold locally at enough of these neighboring worlds.

Before moving on to discuss best systems, I have like to draw attention to an important issue concerning time's arrow. The crystallization account will need to be paired with some solution to the problem of time's arrow, which has been widely discussed in the foundations of statistical mechanics. The problem is this: many of the nearby trajectories which are compatible with the dynamical laws have very strange pasts. If we take arbitrary nearby states, and evolve backwards with the dynamical laws, we will end up with lots of world-histories that exhibit a strange entropy gradient: entropy increases towards the past as well as the future. This means that, in many of these worlds, we find all sorts of thermodynamic processes happening 'in reverse': ice unmelting, coffee spontaneously getting warm...

<sup>15</sup> This recipe for identifying a worlds' neighbors (with respect to region  $R$ ) is closely related to Tim Maudlin's recipe for evaluating counterfactuals, in *The Metaphysics within Physics* (Oxford: Oxford University Press, 2007). According to Maudlin, to evaluate 'P at  $t > Q$ ' we start with a description of the world at  $t$ , modify it in accordance with an instruction encoded in the antecedent of the counterfactual (make it such that P at  $t$  while changing as little as possible), and let the laws 'generate' a physical model (a world) by taking the modified state description and evolving it with the laws. An important difference between Maudlin's recipe for 'constructing' nearby worlds and my own is that I'm using the laws to evolve states in both temporal directions, while Maudlin only uses them to generate what is in the future of the state in question. I will explain why I favor the time-symmetric recipe later in this section.

I do not aspire to provide a novel solution to this problem here. It is enough for my purposes that my account fits well with existing proposals. For concreteness, I shall adopt a view defended primarily by David Albert and Barry Loewer, according to which the fundamental laws of physics include not only the dynamical laws, but also a law about the initial conditions of the universe. This additional law, known as the ‘Past Hypothesis’, specifies that the world began in a low-entropy state so (as a matter of nomological necessity);<sup>16</sup> this will guarantee that anti-entropic worlds are extremely rare in our modal neighborhoods.

Adopting the Past Hypothesis as a fundamental law is not our only option. Another option is to invoke a time-asymmetric notion of world-similarity in constructing the modal neighborhood. Note that my procedure for constructing the modal neighborhood of a region-world pair  $R, w$  was time-symmetric. I considered taking time-slices intersecting the region  $R$ , modifying them, and then evolving the resulting states with the laws, both towards the future and towards the past. But I could have opted instead for a time-asymmetric, non-backtracking distance metric. I could have said, for example, that a world is near ours relative to  $R$  if and only if it is slightly different at some time  $t$  in  $R$ , identical before  $t$ , and respects the laws of our world as much as possible. A region  $R$ ’s modal neighborhood would be constructed by taking some time slice  $t$  intersecting  $R$ , modifying it slightly, and evolving forwards with the dynamical laws, while leaving everything in its past fixed.<sup>17</sup> If we adopted this time-asymmetric recipe, the reversibility worries that afflict attempts to reduce thermodynamics to physics would not arise: almost all worlds in the neighborhood would have normal entropy gradients.<sup>18</sup>

Readers who are skeptical of the past hypothesis (conceived as a law) may prefer this alternative construction of the modal neighborhood. I chose not to rely on a time-asymmetric notion of world nearness/similarity because I did not want to assume that temporal asymmetries could be fully explained independently of robustness structure. It seems to me that relying on a time-asymmetric similarity metric in the account of higher-level laws amounts to putting an asymmetry into non-fundamental sciences by hand. But we might have hoped for something more: to see how the whole

<sup>16</sup> David Z. Albert, *Time and Chance* (Cambridge: Harvard University Press, 2000); Barry Loewer, "Two accounts of laws and time," *Philosophical Studies*, CLX, 1 (August 2012): 115–37.

<sup>17</sup> Michael Strevens, *Depth: An Account of Scientific Explanation* (Cambridge: Harvard University Press, 2008), at p. 290, gestures at the view that higher-level laws (robust regularities) are merely regularities that extend to nearby worlds, where ‘nearby’ is understood under the non-backtracking interpretation. While he ends up rejecting this sort of view (on the grounds that regularities don’t explain their instances), I expect that he would be sympathetic to the idea that robust regularities hold throughout modal neighborhoods (given the non-backtracking construction).

<sup>18</sup> Thanks to an anonymous referee for suggesting this alternative.

structure of non-fundamental science (including temporal asymmetries in non-fundamental laws) arises from physics.

There is, however, an interesting alternative approach which is not objectionably time-asymmetric. We would start by defining not one notion of crystallization but two: future-crystallization and past-crystallization. A regularity is future-crystallized if and only if it features in a good summary of the physically possible worlds that are similar in the past, and diverge in the future. A regularity is past-crystallized if and only if it features in a good summary of the physically possible worlds that are similar in the future, and diverge in the past. We could then try to explain why future-crystallized regularities are more significant to embedded agents like us than past-crystallized generalizations (why they are more capable of playing the robustness role). The challenge would be to have an explanation for this that didn't itself smuggle other higher-level temporal asymmetries (for example, the fact that we can influence the past but not the future). Pursuing this alternative strategy requires a more rigorous exploration of the foundations of statistical mechanics and its relation to the asymmetry of agency, which I will have to postpone for future work.

*II.2 Balancing Theoretical Virtues.* In its original formulation, the best systems account was meant to be an account of fundamental laws.<sup>19</sup> It has since been used to motivate an egalitarian picture where special sciences can have their own 'best systems', expressed in their own vocabularies.<sup>20</sup> However, the foundationalist tradition has not drawn on the insights from Lewis's account of laws to shed light on how statements about the macro-world inherit their elite nomic status from the underlying physics. This is what I propose to do in what follows. I will repurpose what I take to be the most promising aspects of Lewis's best systems view to give a foundationalist account of robust regularities at all levels of description.

Systems will be assessed along three dimensions: i) how faithfully they describe the modal neighborhood (*accuracy*), ii) how much they tell us about the modal neighborhood (*informativeness*), and iii) how easily they can be expressed in suitable vocabularies (*simplicity*). The best system is the set of statements that strikes the best balance of these virtues.

<sup>19</sup> David Lewis, *Counterfactuals*, *op. cit.*

<sup>20</sup> Callender & Cohen, "Special Sciences, Conspiracy and the Better Best System Account of Lawhood," *op. cit.*  
Cohen & Callender, "A Better Best System Account of Lawhood," *op. cit.*

*Simplicity.* A system is simple to the extent that it has a short formulation. Since the length of a statement is language-dependent, we need a privileged language. What could be a principled choice? Lewis suggests that all candidate systems be stated in a ‘fundamental language’, that is, a language whose predicates denote fundamental properties.<sup>21</sup> But this will not do in the present context, where the goal is to allow for systems that describe reality at many levels of description (that is, in the vocabulary of many special sciences).

Instead, I will draw on an idea due to Michael Hicks and Jonathan Schaffer<sup>22</sup>. They modify Lewis’s criterion to allow non-fundamental quantities (such as acceleration) into the lawbook. Their idea is to let candidate systems be written in any language whatsoever, and take a system’s simplicity to be a function of i) its length, and ii) the degrees of naturalness of its primitive predicates. On this approach, ‘Emeralds are grue’ will be complex despite superficially appearing simple, because its predicates denote highly unnatural properties. (A property’s degree of naturalness could be taken as primitive, it could be understood in terms of ease of definability in fundamental terms<sup>23</sup>, or it could be taken to depend on our interests and/or psychological makeup).<sup>24 25</sup>

*Informativeness.* Start from a familiar idea: a system is informative when it rules out many possibilities; it is uninformative when it rules out few possibilities. Although this cannot be understood straightforwardly in terms of the *number* of possibilities, there is something in this idea that is worth keeping.

Imagine you and I are trying to locate a spider in a large empty room. You have been told that the spider is in a particular corner; I have only been told which half of the room the spider is in. Clearly, you are more informed than me. But how much more? A natural answer suggests itself: it depends on

<sup>21</sup> David Lewis, "New Work for a Theory of Universals," *Australasian Journal of Philosophy*, LXI, 4 (December 1983): 343–77.

<sup>22</sup> Michael Townsend Hicks and Jonathan Schaffer, "Derivative Properties in Fundamental Laws", *The British Journal for the Philosophy of Science*, LXVIII, 2 (September 2015): 411–50.

<sup>23</sup> David Lewis, "New Work for a Theory of Universals," *Australasian Journal of Philosophy*, LXI, 4 (December 1983): 343–77.

<sup>24</sup> See Barry Loewer, "Laws and Natural Properties," *Philosophical Topics*, XXXV, 1/2 (Spring/Fall 2007): 313–28; Callender & Cohen, "Special Sciences, Conspiracy and the Better Best System Account of Lawhood," *op. cit.*; Cohen & Callender, "A Better Best System Account of Lawhood," *op. cit.*

<sup>25</sup> I argue elsewhere ("Naturalness by Law", manuscript) for a version of the second option. I suggest that we define the simplicity of a candidate system as a function of 1) syntactic simplicity in a higher-level language, and 2) ‘semantic simplicity’, which tracks the complexity of the real definitions of the terms involved. However, the crystallization account is compatible with any conception of degrees of naturalness, including ones in which naturalness is anthropocentric and/or context-dependent.

how much bigger my region of possible locations is than yours. Although both of our information-states are compatible with continuum many possible spider-locations, there is a sense in which you've ruled out more possibilities than I have. This is because, using a natural measure on the space of possible locations, we can say that the region of possibilities associated with my information-state is larger than the region of possibilities associated with yours.

We can generalize this idea by imposing a measure on sets of worlds: the informativeness of a system  $S$  relative to a region  $R$  is inversely proportional to the measure of the set of all worlds  $v$  such that  $S$  holds at  $\langle R, v \rangle$ .<sup>26</sup>

Does the dependence on a choice of measure suggest that informativeness is not an objective property of systems? Not if some measures are privileged from the perspective of physics. For instance, Classical Statistical Mechanics has an impressive track record of explanations that rely on a standard measure on phase space. Although we can define various measures on this space, there is one measure that respects the intrinsic structure of the space (giving a natural notion of 'volume'), and uniquely satisfies a natural dynamical constraint: a region of measure  $m$  always evolves, under Hamiltonian dynamics, into another region of measure  $m$ . This natural measure on phase space induces a natural measure on the space of physically possible worlds, which can then be plugged in to the above definition of informativeness.

Admittedly, more needs to be said to support the connection between a measure's naturalness and its appropriateness for defining theoretical virtues (such as informativeness). My conjecture is that, in worlds like ours, 'how many' possibilities an agent can rule out (by the lights of the natural measure) will correlate better with agents' abilities to achieve their goals – and, therefore, natural measures track quantities that we care about. A rigorous argument for this conjecture would require a much deeper understanding of agency (and its physical basis), and falls outside the scope of this paper.

*Accuracy.* Accuracy is not standardly considered a virtue to be balanced with simplicity and informativeness, except in the case of probabilistic generalizations, where a closely related notion ('fit') is introduced. For Lewis, a set of statements will not even enter the best systems competition if it

<sup>26</sup> We may need to refine this definition to favor systems whose information can be easily 'extracted'. Predicting what will happen to an ecosystem on the basis of the physical laws alone would be extremely hard, because ecosystems are highly complex when described in physical vocabulary. This same situation may be relatively easy to model at a higher level, if there are dynamical generalizations about ecosystems specifiable in macro-terms. These generalizations could be used to (tractably) derive predictions about the system without specifying its initial state in full detail. (Making this precise will have to be postponed for future work).

contains a single falsehood. This, however, seems unmotivated when we're dealing with special science generalizations. It seems odd to require generalizations to be exactly true to feature in the best system, when it is difficult to come up with a single example of a strictly true generalization that is considered elite in the special sciences. I suggest, instead, that we drop this requirement. We can take accuracy to be a third virtue, which can sometimes be sacrificed to some degree in favor of the other two virtues.

The best systems competition, as I envisage it, looks for systems that can function as summaries of any world-region pair in a modal neighborhood. This departure from standard best systems accounts will require a non-standard notion of accuracy: accuracy relative to a modal neighborhood. I will take the accuracy of a system, relative to a modal neighborhood, to be given by its average accuracy throughout that neighborhood. (If we're dealing with indeterministic physical laws that assign probabilities to world-histories, we should weigh accuracy-scores throughout the modal neighborhood accordingly, so that exceptions in improbable worlds are less costly).

I have not yet said how we will quantify a system's accuracy with respect to each world-region pair. We will need a degreed notion of accuracy, which counts some falsehoods as being more accurate than others. For instance, 'All ravens are black' (although false) should count as more accurate than 'All ravens are white'.

A simple-minded metric may work for generalizations like 'All ravens are black'. We can just ask how numerous and widespread the exceptions are (that is, how many non-black ravens are there?). However, we need something subtler to deal with generalizations that depart from this simple form, such as generalizations connecting variables that take continuously many values. Take the ideal gas law, which mathematically relates various macro-variables of gases (pressure, volume, and temperature, number of molecules). No real gas satisfies this generalization exactly, but almost all gases will approximately satisfy it. To measure the generalization's accuracy, we take each of the instances in the relevant region (that is, each gas located within the region), measure the divergence between its macro-properties at various times (its pressure, volume, and temperature) and those of the most similar ideal gas, and then aggregate those divergences (across various macro-properties for each gas, and then across all gases in the region).

*II.3 Probabilistic Axioms.* I have been restricting my attention to non-probabilistic generalizations. It is a nice feature of my account that it can be straightforwardly extended to deal with probabilistic generalizations, regardless of whether the underlying physical laws are themselves deterministic or

stochastic. The strategy for this extension comes from Lewis<sup>27</sup>; I will merely explain how to adapt it to our context.

Consider a simple world that consists of a random sequence of coin tosses: HTTH HHTHTTHTHH... If someone asked you to tell them something concise and informative about this world, what would you say? By listing the whole sequence, you could be maximally informative, but you would not be concise. You could say something less specific, like ‘about half the tosses came up Heads’. But you could be even more helpful by giving your interlocutor the following set of instructions: for any sufficiently long sub-sequence, expect about half Heads and half Tails; for each particular outcome, be indifferent between Heads and Tails; expect many consecutive Tails (or Heads) to occur quite rarely... These instructions may not sound concise as they stand, but they can be packaged into a single ‘chance axiom’: an assignment of a number between 0 and 1 to the relevant pair of properties. In this case: <coming up Heads, being a coin toss> is assigned the number 1/2.

Unlike statements, chance axioms do not inform by telling us which propositions to put in our ‘belief-boxes’. Instead, they directly tell us how to adjust our credences in various propositions. In our simple coin-world, the statement  $Pr(\text{Heads, coin toss}) = 1/2$  tells us that, absent specific information about the outcome, we should be indifferent between Heads and Tails.<sup>28</sup> It also tells us that we should expect roughly equal numbers of Heads and Tails in the long run, and so on.

Chance axioms, so understood, are not simply true or false, and they do not rule out any possibilities.<sup>29</sup> Thus, the notions of informativeness and accuracy characterized above will not apply. However, these axioms can be judged for informativeness and accuracy on the basis of the credences that they recommend. There are standard ways of quantifying how ‘opinionated’ a distribution is, which we could use to quantify systems’ informativeness (assuming we’re dealing with systems that specify complete probability distributions over world-region pairs).<sup>30</sup> Moreover, standard notions of

<sup>27</sup> See David Lewis, "Humean Supervenience Debugged," *Mind*, CIII, 412 (October 1994): 473–90.

<sup>28</sup> I’m assuming there is some principle along the lines of Lewis’s Principal Principle connecting chance axioms and credences.

<sup>29</sup> I’m not suggesting that natural language statements about objective probabilities fail to express propositions. They are naturally interpreted as saying something like ‘the best system recommends such-and-such credences’.

<sup>30</sup> The technical notion is ‘entropy’. (Not to be confused with thermodynamic entropy, which is a property of physical systems). Roughly, the entropy of a probability function  $Pr$  specifies the amount of information you *expect* to get from a subsequent observation – under the idealization that you will observe event  $E$  with probability  $Pr(E)$ . A uniform distribution has maximum entropy, and a distribution that assigns probability one to a single outcome (in our case, a single world-region pair) has zero entropy.

‘distance from the truth’ (relative to each world in the modal neighborhood) yield a natural combined estimate of accuracy and informativeness.<sup>31</sup>

Not all candidate systems explicitly specify complete probability distributions. A candidate system may consist of a single probabilistic axiom:  $\forall x Pr(black(x) | raven(x)) = 0.99$ . We can think of such an axiom as stating a constraint on probability distributions.  $Pr$  satisfies this constraint if and only if for every pair of propositions  $p$  and  $q$ , if  $p$  says of some object that it is black and  $q$  says of that same object that it is a raven then  $Pr(p | q) = 0.99$ . To evaluate the accuracy and informativeness of a system like this one, we can use a simple trick: associate the system with the least informative distribution that satisfies the constraints imposed by the system. In a sense, this distribution incorporates only the information embodied by the system, so we can let the system inherit its accuracy/informativeness score(s) from the associated distribution.

The crystallization account is now complete: the crystallized regularities within a region are the axioms of the best system for the region’s modal neighborhood. In what follows I will argue that this account vindicates the intuitive distinctions that the foundationalist wanted to draw, and also promises to illuminate the importance of robust regularities for beings like us. But before doing this, I would like to draw attention to two ‘free parameters’ within the account, and say something about how I understand them.

How nearby does  $\langle R, v \rangle$  have to be in order to be in the modal neighborhood of  $\langle R, w \rangle$ ? I will leave this *proximity threshold* unspecified. The crystallization account succeeds as long as there is some (non-extreme) setting of this threshold that characterizes the notion of robustness that we care about.

The best systems account also requires a *balancing function* specifying how much each of the virtues matters to the system’s overall quality. Where we draw the boundary between crystallized and

<sup>31</sup> For distributions defined on a finite set of propositions, a natural choice is the Brier score: the average squared distance between the subject’s credence in atomic proposition  $P$  and the ‘true value’ of  $P$  (that is, 0 if  $P$  is false, and 1 if it is true). The ‘distance from the truth’ idea can be generalized to the continuous case, provided that the hypothesis-space has metric structure. See for instance the ‘continuous ranked probability score’ (CRPS), in Thomas A. Brown, "Admissible Scoring Systems for Continuous Distributions," *RAND Corporation* (August 1974). The Brier score and the CRPS in fact capture at the same time the uncertainty of the distribution, as well as its ‘reliability’ (See Hans Hersbach, "Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems," *Weather and Forecasting*, xv, 5 (October 2000): 559–70.) Thus, if accuracy is measured in accordance to one of these rules, we only need to consider two virtues: accuracy score and simplicity.

non-crystallized regularities depends on the choice of balancing function. For some choices, perhaps, only generalizations stated in physical vocabulary (for example, Schrodinger's equation) come out as crystallized. But the more weight we put on informativeness, the more generalizations will count as crystallized. This is the key to understanding why crystallized regularities are not just the equations that describe with perfect accuracy the dynamics of fundamental particles (or fields). We can get more informative systems about the modal neighborhood(s) we inhabit if we begin to tolerate some inaccuracy, as well as greater complexity. The most powerful summaries of our modal neighborhood(s) will include, besides physical generalizations, stable generalizations written in the vocabulary of biology, neuroscience, geology, and so on.

Thinking of the balancing function and the proximity threshold as free parameters allows us to draw fine distinctions between the statuses of different sciences and different generalizations within a science. We can make sense of the idea that some generalizations are 'more' crystallized than others, if they feature in the best system under strictly more parameter-values. For instance, chemical and thermodynamic generalizations could turn out to be more crystallized than generalizations in biology and economics. This fits nicely with a conception of robustness defended by Sandra Mitchell<sup>32</sup>, on which the nomic statuses of regularities vary along multiple dimensions such as: stability, tractability, strength, idealization, and so on.

We can ask how various kinds of human theorizing relate to the above parameters. I will not settle this issue here, but I expect that, for any given context, there is a range of parameter-values that characterize the distinctions we care about equally well. This is as it should be with reductions of human concepts: the candidate metaphysical analyses should be precise, and the mapping from our vague terms to those candidate analyses should be indeterminate.

### III. Robustness as Crystallization

According to the 'crystallization account', the notion of crystallization defined in §2 accounts for the notion of robustness characterized in §1.1 by its theoretical role. The goal of this section is to show why this claim should appeal to foundationalists.

<sup>32</sup> This fits nicely with a conception of robustness defended by Sandra D. Mitchell, "Dimensions of Scientific Law," *Philosophy of Science*, LXVII, 2 (June 2000): 242–65, on which the nomic statuses of regularities vary along multiple dimensions such as: stability, tractability, strength, idealization, and so on.

The claim that crystallization accounts for robustness can be understood in two ways. On one reading, the thesis says that the definition of crystallization stated before is itself a definition of robustness in more fundamental terms. On a weaker reading, the thesis states that, in worlds like ours, every robustness fact holds in virtue of a corresponding crystallization fact. This weaker thesis allows robustness facts that are grounded in some other way in distant worlds. The choice between these two theses will depend on subtle metaphysical and meta-semantic issues that I would like to bracket here; thus, I will only be arguing for their disjunction.

In what follows I will argue, first, that the account explains the intuitive contrast between ‘local necessities’ (such as ‘All ravens are black’) and mere accidents (such as ‘All the pennies in my pocket are silver’), and in doing so vindicates the core intuitions motivating foundationalism (§3.1). Then I will show that adopting the crystallization account allows us to draw a further nomic distinction that is crucial for a theory of explanation (§3.2). Finally, I will suggest that thinking of robustness in terms of crystallization takes us a step towards a more ambitious foundationalist goal: understanding – from the perspective of physics – the importance that robust regularities have for physical agents like us. §3.3 sketches the beginnings of such a story.

*III.1 Local Necessities.* I started out drawing a contrast between ‘All the pennies in my pocket are silver’ and ‘All ravens are black’. While the former seems accidental, the latter does not. The challenge was to identify the physical basis for this modal difference, given that both regularities depend on physically contingent factors. The crystallization account has the resources to explain this contrast.

Let me start by introducing the notion of a ‘local necessity’:

#### LOCAL NECESSITY

To be ‘locally necessary’ (relative to  $R,w$ ) is to be metaphysically necessitated by the robust regularities at  $\langle R,w \rangle$ .

If we accept that crystallization accounts for robustness, we can conclude that a regularity is ‘locally necessary’ (relative to  $R,w$ ) if and only if it is metaphysically necessitated by the axioms of the best system of the modal neighborhood for  $\langle R,w \rangle$ . A regularity is accidental if it not locally necessary.

I cannot offer decisive proof that generalizations like ‘All ravens are black’ will be necessitated by the axioms of the best system. Similarly, proponents of the best systems account of fundamental laws cannot prove that Schrodinger’s equation (or whatever equation will come to replace it) will feature

in the best system. All I will offer is a plausibility argument that statements like these are either in the best system, or entailed by it.

As far as we know, ‘All ravens are black’ is highly accurate in our world, at least if we restrict ourselves to a certain epoch on Earth – say, the 21<sup>st</sup> Century. It is plausible, moreover, that this regularity holds in the corresponding modal neighborhood. Choose any time in this century, and find physically possible worlds that are very similar to ours at that time. Surely you will not see pink ravens populating Earth in this world a decade later or a decade before.

Besides being accurate throughout nearby worlds, ‘All ravens are black’ appears to be informative, simple, and mentions no gerrymandered properties. Given this, it is plausibly an axiom of the best system, unless it already follows from other axioms. In either case, it will count as a ‘local necessity’ and therefore as non-accidental. The same can be said for many other biological generalizations, which may be contingent but have a genetic basis that is widespread and resistant to small physical changes. For example: ‘Tigers have stripes’, ‘Olives are high in sodium’, or ‘Humans are subject to confirmation bias’.

This plausibility argument extends beyond generalizations with a genetic basis. Some crystallized regularities may be sustained by powerful and widespread cultural mechanisms: for example, English speakers say ‘hello’ when they encounter a friend... The psychological features underlying these dispositions are spread across many minds, and sustained by a complex dynamic of social incentives. This makes it plausible that small physical changes will not disrupt them. If so, these generalizations are highly accurate throughout the kinds of modal neighborhoods we care about. Since they are also plausibly simple and informative, we can expect them to be axioms of the best system (if not already entailed by other axioms).<sup>33</sup>

You may wonder whether the account’s yielding the robustness of such mundane generalizations as ‘English speakers say ‘hello’...’ is not a symptom of excessive permissiveness. I think not: many true generalizations will be too modally fragile from the perspective of physics to make it to the best system of the modal neighborhood for a region of interest.

To see how the account rules out accidental generalizations, let us consider a couple of cases. Start with our simple example: ‘All the coins in my pocket are silver’. This generalization will be highly inaccurate throughout any sufficiently expansive modal neighborhood. It fails in all the worlds where

<sup>33</sup> As this example illustrates, crystallized regularities can depend on arbitrary conventions. Such regularities may have been fragile when the convention was just starting, but they are now crystallized.

I find a lucky penny on my way to work, or where I get five pennies instead of a nickel at the local café.

Now take a more interesting case, which has been used to illustrate the difference between foundationalism and egalitarianism.<sup>34</sup> Imagine that, in some spatio-temporal region of our world, coins happen to land Heads 90% of the time, despite the physical characteristics of coins and coin-tossers being just as they are around here. The foundationalist regards these patterns as flukes, not autonomous robust regularities. After all, they do not seem like the kinds of regularities that our physical laws are capable of reliably sustaining.

The crystallization account can accommodate this foundationalist intuition. Take an arbitrary time-slice intersecting the spatio-temporal region R where coins landed heads 90% of the time. As before, make small changes to the state of R at the relevant time: change the trajectory of a bird, the position of a few specks of dust, or the speed of a few air molecules. Many such changes will make a difference to the exact velocity and height with which coins are tossed after  $t$ . Given the symmetry of the coin (and the underlying physics) the frequency of heads in almost all of these worlds will end up being much lower than 0.9. If so, the generalization ‘Coins land heads with probability 0.9’ will be highly inaccurate in nearby worlds, and any system is plausibly better off without it.

The moral is that we can allow for contingent regularities to figure in the best system for non-fundamental phenomena, without having to give up foundationalism altogether. The crystallization account shows that robust regularities can be dependent on, and sufficiently constrained by, the fundamental physical structure, even if they are not determined by the physical laws alone. This, I think, is the first step to vindicating the foundationalist worldview.

*III.2 Axioms and Theorems.* As we saw already, differences in modal force are not the only differences that we deem scientifically relevant. ‘All ravens are black-or-green’ does not share the explanatory status of ‘All ravens are black’, but neither is accidental. The present account has the resources to draw these fine-grained distinctions. Crystallized regularities are the axioms of the best system. They are to be distinguished from mere local necessities (the statements they necessitate). Unlike the axioms, mere local necessities do not have a distinguished explanatory status.

Let me now illustrate the work that the axioms/local necessities distinction can do for us, by discussing a few cases mentioned before. Suppose ‘All ravens are black’ is crystallized relative to *earth-*

<sup>34</sup> This example was suggested to me by David Albert.

2019. Then ‘All ravens are black or green’ will be deductively entailed by the best system for this region, and will thereby count as a ‘local necessity’ relative to *earth-2019*. Note, however, that the simpler generalization ‘All ravens are black’ will exclude ‘All ravens are black or green’ from the axioms of the best system. Adding the disjunctive statement as an axiom would yield a more complex system, with no gain in accuracy or informativeness. Thus, the system that didn’t explicitly include the disjunctive statement was better.

Now consider ‘All gremeralds are grue’. By my account of simplicity, generalizations that have short formulations in highly natural vocabulary are simpler than generalizations that only have comparably short generalizations in unnatural vocabulary. Keeping this assumption in mind, compare the pair of gruesome generalizations (‘Gremeralds are grue’, ‘Graphires are bleen’) with the pair (‘Emeralds are green’, ‘Sapphires are blue’). Given the connection between simplicity and naturalness, the latter pair is simpler. And given that the pairs are equally informative and accurate, it’ll be the latter that gets into the best system’s axioms – if either does.

Finally, recall the variant of Kyburg’s example: the generalization ‘Hexed caffeinated beverages keep humans awake’ (where  $x$  is ‘hexed’ iff certain ‘hexing’ gestures and utterances were made over it). I can explain why the coffee I hexed this morning kept me up, by saying that caffeine keeps humans awake, and that hexed coffee was caffeinated. My explanation would be worse had I cited instead the generalization ‘Hexed caffeinated beverages keep humans awake’. The distinction between axioms and theorems helps explain this asymmetry: ‘Hexed caffeinated beverages keep humans awake’ follows from the simpler and more informative claim, namely: ‘Caffeine keeps humans awake’. Given this, the hexed caffeine generalization is excluded from the axioms of the best system.

Perhaps ‘Caffeine keeps humans awake’ is entailed by other generalizations which are themselves axioms – in which case it will not be an axiom either. Even so, the distinction between axioms and local necessities may help us account for the fact that ‘Hexed caffeinated beverages keep humans awake’ doesn’t explain its instances, while ‘Caffeine keeps humans awake’ does. Here is a sketch of how this could go. Suppose that a locally necessary generalization explains one of its instances if and only if the generalization figures in every ‘suitable’ argument from the axioms to that instance (where ‘suitability’ requires an optimal balance of simplicity/naturalness, generalizability, and other such features). An account along these lines may give us the resources to distinguish between local necessities that can be cited to explain their instances (such as ‘Caffeine keeps humans awake’) and

those that are too specific (such as ‘Hexed caffeinated beverages keep humans awake’), or too disjunctive (such as ‘Caffeine-or-sugar keeps humans awake’) to explain their instances.<sup>35</sup>

*III.3 Why Care about Robustness?* The foundationalist vision will be vindicated if all nomic facts can be understood in terms of physics. But we could hope for more: we could hope to understand, from the perspective of physics, why higher-level nomic facts are of interest to intelligent beings. Why should physical agents care about the distinction between robust and accidental regularities? In this section I will outline a strategy for answering this question that draws on the crystallization account.

Essential to our survival is the ability to select the best action available to us, given the state of our environment. Knowledge about the actual world does not suffice for these purposes if we cannot draw on it to infer what *would* happen if we were to act in one way or another. Even if we could predict the choices we will actually make, this would not tell us what we need to know: which of the available choices would be best for us, given our goals. Knowledge of robust regularities is crucial to us because we can rely upon it when reasoning hypothetically about possible ways to intervene on our environment. As I will now explain, this knowledge enables us to trace connections between merely possible actions, and the states of affairs that would result from these actions.

When reasoning hypothetically, we distinguish between certain facts that are settled independently of us (e.g. the laws of physics), and facts that are contingent on our decisions. This determines our ‘agential possibilities’. Roughly, a state of affairs P is agentially possible from an agent’s perspective at a time *t* if, for some possible decision D which is available to her at *t*, she treats  $D \rightarrow P$  (or probably P) as fixed.

Which regularities should an agent hold fixed by default in hypothetical reasoning? One might have thought it is only physical necessities (that is, entailments of the physical laws). On this picture, what is possible for an agent is given by what the agent’s decisions (together with the agent’s specific knowledge about the present) can physically necessitate.<sup>36</sup> For example, the light’s turning on is

<sup>35</sup> This is merely a promissory note. There have been more rigorous attempts to characterize what makes a derivation ‘suitable’. See, for instance, Philip Kitcher, “Explanatory Unification,” *Philosophy of Science*, XLVIII, 4 (December 1981): 507–31. A parallel issue has been discussed in the literature on mathematical explanation: what makes a mathematical proof explanatory? See Mark Steiner, “Mathematical Explanation,” *Philosophical Studies*, XXXIV, 2 (August 1978): 135–51; and Marc Lange, “Explanatory Proofs and Beautiful Proofs,” *Journal of Humanistic Mathematics*, VI, 1 (January 2016): 8–51.

<sup>36</sup> Since what an agent knows about the present includes memories/records, we expect such an agent ends up holding fixed lots of past events as well.

agentially possible for you because, holding fixed what you know about the present, your decision to flip a switch physically necessitates the light's turning on.

The crystallization account suggests an interesting alternative: treating physically contingent regularities as fixed may be advantageous, if these regularities are crystallized with respect to an agent's modal neighborhood. This idea can be motivated with a few thought-experiments.

Suppose that 'Caffeine keeps humans awake' holds in almost all worlds in the modal neighborhood of Earth-today. Consider two worlds in this modal neighborhood. In the first world, you are sleepy but desire to stay awake at  $t$ . The second world is almost exactly alike at  $t$ , except for your wanting to fall asleep. Now consider what happens if you hold fixed the caffeine regularity in hypothetical reasoning. In both situations you will conclude that you would stay awake if you were to consume caffeine, and you would fall asleep otherwise. As a result, you will choose to consume caffeine in the first situation, and not in the second. Since the caffeine regularity likely holds in any arbitrarily chosen worlds in the modal neighborhood, it is plausible that the following is true of the modal neighborhood: 'If I hold fixed the caffeine regularity, then my goals to stay awake/sleep will correlate with my staying awake/sleeping'.

Now consider an agent that is otherwise similar to you, but that only holds fixed physically necessary regularities, together with whatever she knows about the present state of the world. Could such an agent exploit the caffeine-awake connection on the basis of physical necessities? It depends on what kind of agent we have in mind. Suppose that, at the time of her decision ( $t$ ), the agent knows all the true statements of the form 'It is nomologically necessary that if condition X obtains at  $t$ , consuming caffeine shortly after  $t$  will keep me awake'. Let C be the class of all conditions that satisfy the above statement. If the agent knows enough about the present condition of her environment to determine whether a member of C obtains in her current situation, then she will be in a position to infer that caffeine would keep her awake. Assuming our agent is logically omniscient, this will lead her to drink caffeine whenever she wants to stay awake. But we – and all other physical agents – are importantly different from this agent. Firstly, it may be difficult for us to find out whether the enabling conditions C obtain in a given circumstance. Secondly, it might be difficult for us to draw the required inferences (if the set of enabling conditions is complex enough). In light of this, we're probably better off holding fixed 'Caffeine keeps people awake'.

These observations may be taken to suggest that the more true generalizations we hold fixed in hypothetical reasoning, the better off we are – in which case there is nothing special about crystallized regularities. But this is not right. Consider what happens to an agent who holds fixed

regularities that are too modally fragile. Imagine I get sleepy at 10pm every day, and conclude this regularity is robust. The night before an important deadline I would like to stay awake past 10pm, but I hold fixed the regularity that I take to be robust. This leads me to conclude that, regardless of what I do (e.g. how much caffeine I consume), I will be sleepy by 10pm. If I do this repeatedly throughout my life, ‘I get sleepy at 10pm’ may in fact turn out to be true – but only because, in mistaking it for a robust regularity, I missed the opportunity to break it. There is a clear sense in which my deliberation is sub-optimal here: another physically possible agent which is otherwise similar, but doesn’t treat ‘I will be sleepy by 10 pm’ as fixed, will likely do better than me in achieving the same goals.

These thought-experiments motivate the following hypothesis: physical agents who hold fixed crystallized regularities in reasoning are (on average) better adapted to their modal neighborhoods than agents that hold fixed fewer or more regularities; that is, they tend to fulfill their goals in more nearby possible worlds. Why should an agent care about the extent to which a strategy succeeds in nearby worlds? Because, plausibly, the extent to which it succeeds in nearby worlds typically coincides with the extent to which it succeeds in the actual world. A strategy is unlikely to systematically fulfill one’s goals in one’s home world without doing so throughout one’s modal neighborhood.<sup>37</sup>

It is thus possible that agents in general need to accomplish the following feat: develop cognitive procedures that incrementally uncover the most efficient summaries of the modal neighborhoods that they inhabit. If so, then to better understand the connections between robustness and other central aspects of our theorizing (explanation, prediction, hypothetical/counterfactual reasoning, controlled interventions...), we should be trying to understand how each of these cognitive activities fits within an overall strategy for solving the above problem, given the constraints that physical systems are subject to.

I have offered foundationalists three reasons to think that robustness coincides with crystallization. Firstly, this view identifies a physical basis for the accidental/non-accidental distinction that we find in the special sciences, where many regularities are treated as having a kind of natural necessity, despite their physical contingency (§3.1). Secondly, the crystallization account vindicates distinctions in

<sup>37</sup> One might think that the crystallization account needs a claim of this sort anyway: if actual frequencies didn’t typically align with frequencies throughout the modal neighborhood, it is hard to see how we could get evidence for crystallization statements on the basis of actual frequencies. By ‘typical’ I have in mind something like high physical probability: P is typical if and only if it is probable according to any probability distribution that is reasonable by the lights of physics.

explanatory status between regularities that are alike in their modal strength, by giving simplicity in higher-level vocabularies a central role (§3.2). Thirdly, the account promises to shed light on the question: why do we bother categorizing some regularities as ‘robust’? One reason is that these regularities form a system that we can rely upon to carry out tractable inferences in deliberative contexts (§3.3).

## IV. COMPARISONS

In the rest of the paper, I will discuss how the crystallization account relates to two approaches to robustness that appear in the literature. The first is a contemporary version of the reductionist tradition, which goes back to Putnam and Oppenheim (1948); the second takes robustness to be a form of counterfactual stability. I will argue that both are compatible with, and usefully supplemented by, the notion of crystallization.

*IV.1 The Mentaculus.* As I explained in §2, my preferred version of the crystallization account assumes that the physical lawbook includes, alongside the dynamical laws, a law stating that the universe began in a very low entropy state (the ‘Past Hypothesis’). Without it, I said, the modal neighborhood would be plagued by anti-entropic worlds, and many familiar time-asymmetric generalizations would not turn out to be crystallized.

Albert and Loewer have argued that the Past Hypothesis, together with a uniform probability distribution over all initial micro-conditions that realize it, gives us the foundation for all of science.<sup>38</sup> Those who are sympathetic to their view may wonder why, having taken the Past Hypothesis on board, I need the notion of crystallization at all. I will argue, however, that Albert and Loewer should welcome the notion of crystallization I’m providing, since it complements their account of the physical basis for special science theorizing.

Let us grant that the Past Hypothesis, together with a probability distribution over initial conditions, is in fact sufficient to derive probabilistic versions of all thermodynamic generalizations (as Albert has argued).<sup>39</sup> Even granting this, we’ll probably still need to appeal to theoretical virtues to prevent disjunctive/gruesome and uninformative statements counting as robust. But do we need the

<sup>38</sup> See Albert, *op. cit.* and Loewer, *op. cit.*

<sup>39</sup> *Ibid.*

notion of the modal neighborhood at all? Why not say that the robust regularities are those that feature in the best system of all physically possible worlds?

The importance of the notion of the modal neighborhood can be made vivid with an example. Consider some hypothesis about brain lateralization in chimpanzees: ‘the left hemisphere of a chimpanzee’s brain is more active than the right hemisphere during task *T*. Neuroscience has proceeded under the assumption that such regularities can be robust. But presumably, those left/right asymmetries would not have their basis in the laws of physics, whether these laws include the Past Hypothesis or not. Such asymmetries depend on the chimpanzee’s evolutionary history, which could have just as easily gone a different way. Given this, brain lateralization hypotheses could not feature in the best system of all physically possible worlds.

Albert would say that the left-hemisphere statement is highly likely (if true), but only once we conditionalize on some historical event which led to the left-hemisphere regularity. I agree, but I require an account of the rules for conditionalizing in this way. I can cook up all sorts of propositions on which to conditionalize. If the laws are deterministic, then for any event *e* there will be true propositions *P* about the past such that  $Pr(e|P) = 1$  according to Albert’s lawbook. But this doesn’t show that the event was non-accidental, explanatory, inductively learnable...

What I’m providing is a principled way to meet the above requirement. The interesting background conditions (the ones that we tend to implicitly ‘conditionalize on’) are not simply all those that figure as antecedents of nomologically necessary statements. They are ‘crystallization conditions’: stable background conditions which – together with the laws – sustain the regularities that are crystallized within the region in question.

*IV.2 Counterfactual Stability Accounts.* The idea that robustness is a form of counterfactual stability features prominently in the literature, and it was an important source of inspiration for my account.<sup>40</sup> I would like to end by explaining how my account relates to this proposal.

<sup>40</sup> Sandra D. Mitchell, "Dimensions of Scientific Law," *op. cit.*, argues that laws are less contingent than mere accidents, because they would hold under a wider range of background conditions. James Woodward and Christopher Hitchcock, "Explanatory Generalizations, Part I: A Counterfactual Account," *Nous*, XXXVII, 1 (March 2003): 1–24, defend the idea that regularities are robust if and only if they would remain invariant under certain kinds of interventions. In a similar vein, Michael Strevens, "The Explanatory Role of Irreducible Properties." *Nous*, XLVI, 4 (December 2010): 754–80), argues that ‘All ravens are black’ has a special nomic status despite its contingency, due to the ‘physical inertia’ of the underlying coloration mechanism (which he cashes out in terms of counterfactuals). See also: Michael Strevens, "Physically contingent laws and counterfactual support," *op. cit.*

Counterfactual accounts of robustness share the following schematic form:

G is robust only if for all/many/most relevant counterfactual antecedents  $q$ ,  $q > G$  is accurate.

For reasons I have already discussed, satisfying a counterfactual stability condition of this sort cannot suffice for robustness: gruesome regularities can be as stable as physical laws, without thereby attaining the status of robust regularities. Nonetheless, I think counterfactual stability accounts identify an important necessary condition for robustness, which sets them apart from mere accidents. In what follows, I explain how my account vindicates this idea.

Accuracy throughout the modal neighborhood is closely related to a notion of counterfactual stability, which I will call ‘physical stability’. To be physically stable relative to R just is to be stable under all ‘relevant<sub>R</sub>’ physical counterfactuals – where a counterfactual is relevant<sub>R</sub> iff its antecedent stipulates a small enough change to the state of the world (at some time in R). More precisely:

A generalization G is *physically stable* relative to R,  $w \equiv_{\text{def}}$

For any time  $t$  in R, and any (total) physical state  $s$  that is similar enough to  $w$ 's (total) physical state at  $t$ :

It is true at  $w$  that ( $s$  obtains at  $t > G$  is true in R).

In the above condition, ‘similar’ should be understood in terms of a distance metric on the fundamental state-space (as explained in §2), and the conditional ‘ $>$ ’ should be understood in a technical sense:

$s$  obtains at  $t > P \equiv_{\text{def}}$

$P$  holds in all the physically possible world(s) whose total physical state at  $t$  is  $s$ .

Call these conditionals ‘physical counterfactuals’. The antecedent of a physical counterfactual takes us to a class of physically possible worlds: worlds whose state at some time  $t$  is as that antecedent stipulates. If the consequent holds where that antecedent takes us, the counterfactual is true.<sup>41</sup> (When dealing with indeterministic laws, we could weaken the requirement that *all* the antecedent worlds to

<sup>41</sup> See Maudlin, *op cit.*, for an account of counterfactuals along these lines.

be consequent worlds. It may be enough that the set of antecedent-worlds in which the consequent doesn't hold have measure zero, or close to zero).

Since no relevant<sub>R</sub> physical counterfactual takes us outside of R's modal neighborhood, perfect accuracy throughout R's neighborhood suffices for physical stability relative to R. Conversely, all the worlds in R's neighborhood are reached by some relevant<sub>R</sub> physical counterfactual. This means that stability under all relevant<sub>R</sub> physical counterfactuals suffices for perfect accuracy throughout this neighborhood. We can conclude that a regularity is physically stable relative to R if and only if it is perfectly accurate throughout R's neighborhood.

The connection between physical stability and accuracy is less straightforward once we consider generalizations that are less than perfectly accurate. We might expect that high accuracy almost everywhere in the modal neighborhood would correspond to a slightly weakened physical stability condition, such as:

A generalization  $G$  is *highly physically stable* relative to  $R$ ,  $w \stackrel{\text{def}}{=} \text{For almost every antecedent of the form 's is the state at } t$ ,

where  $t$  is some time in  $R$ , and  $s$  is a total physical state similar enough to  $w$ 's physical state at  $t$ , then:

It is true at  $w$  that ( $s$  obtains at  $t \rightarrow G$  is highly accurate in  $R$ ).<sup>42</sup>

However, there is no logical entailment from high accuracy in the modal neighborhood to high physical stability (or vice versa). This is because there needn't be an even measure-theoretic correspondence between sets of relevant<sub>R</sub> counterfactual antecedents, and the sets of worlds they reach. A generalization  $G$  may be physically unstable, despite being highly accurate in almost all worlds in the modal neighborhood, if there are small islands of  $G$ -offending worlds are reached by sufficiently many counterfactual antecedents. Nonetheless, I see no reason to think that our modal neighborhoods will contain many such 'hyper-accessible' (or 'hyper-inaccessible') islands. If so, it is safe to assume that – in worlds like ours – only the highly physically stable regularities relative to  $R$  are highly accurate throughout  $R$ 's modal neighborhood. Thus, any generalization that is low in physical stability is likely not an axiom (or theorem) of the relevant modal neighborhood's best summary. This vindicates the intuitive evidential connection between physical instability and accidentality.

<sup>42</sup> 'Almost every time' will be cashed out in terms of a standard measure on temporal intervals.

I defined the notion of physical stability in terms of ‘micro-antecedents’ – that is, antecedents specifying total micro-conditions for the world at some time. This is one important respect in which it differs from stability conditions on robustness proposed in the literature, which take into consideration antecedents formulated in the macro-language of the generalization or scientific field of interest. For example, Marc Lange, Christopher Hitchcock, and James Woodward have all suggested that special sciences treat different classes of antecedents as relevant, depending on the ‘level’ at which they carve reality.<sup>43</sup> In what follows, I will explain how this kind of ‘macro-stability’ relates to accuracy (and physical stability) and I will argue that the crystallization account vindicates the claim that failures of macro-stability are indicative of accidentality.

For concreteness, I will focus on one particular conception of macro-stability, proposed by Hitchcock and Woodward.<sup>44</sup> They start from the idea that all special science generalizations are mathematical expressions that relate non-fundamental ‘variables’. A variable is a syntactic entity that represents a determinable property such as temperature or pressure; its values are numerical entities that represent the corresponding determinates. Hitchcock and Woodward propose that a generalization  $G$  is robust only if it is stable under counterfactual antecedents that set the variables in  $G$  to alternative values. For example, the ideal gas law should be stable under counterfactual antecedents that specify alternative distributions of pressures, temperatures, and volumes; economic generalizations should be stable under antecedents that change prices, supply, and demand, and so on.

For this kind of proposal to work, we must restrict our attention to ‘nearby’ values of the macro-variables in question. As Hitchcock and Woodward acknowledge, all generalizations break under ‘extreme’ settings of the variables in question, and extreme values are likely just those that are only realized in physically distant worlds.<sup>45</sup> Thus, their notion of stability can be characterized as follows:

A generalization  $G$  is *macro-stable* relative to  $R$ ,  $w \stackrel{\text{def}}{=} \text{def}$

For every variable  $X$  in  $G$ , every macro-object  $a$  which instantiates some value of  $X$  (in  $R$ ), and (almost) every number  $\epsilon$  sufficiently close to 0:

<sup>43</sup> See Marc Lange, "The Autonomy of Functional Biology: A Reply to Rosenberg," *Biology & Philosophy*, XIX, 1 (January 2004): 93–109; and James Woodward and Christopher Hitchcock, *op cit*.

<sup>44</sup> *Ibid*.

<sup>45</sup> *Ibid*.

It is true at  $w$  that ( $a$ 's  $X$ -value is greater/smaller by  $\epsilon$  units  $\rightarrow$   $G$  would be highly accurate).

In order to understand the connections between macro-stability, physical stability and crystallization, we need a more general proposal for evaluating counterfactuals than the one outlined before. I will use a Lewis/Stalnaker semantics, with some minor variations: roughly,  $A \rightarrow C$  is true relative to  $R$  iff  $C$  holds in almost all the nearest worlds where  $A$  obtains.<sup>46</sup> I will make one crucial assumption about the nearness relation, namely that all the worlds in the modal neighborhood of  $R$  are nearer than any world outside of it. I do not assume, however, that these assumptions will carry over to a semantic account of natural language counterfactuals. It is fine for our purposes if ' $\rightarrow$ ' does not capture all speaker intuitions about counterfactuals, so long as it identifies an important relation between propositions that our counterfactual talk tracks (in certain contexts).

Several proponents of macro-stability would reject the idea that counterfactuals reduce to notions like similarity and laws (as I will be supposing).<sup>47</sup> Lange and Woodward are happy to take counterfactual truths as primitive (that is, not in need of further analysis). But my goal (unlike theirs) is to understand robustness from a foundationalist perspective. In this context, primitivism about counterfactuals is unattractive.<sup>48</sup> Despite these differences between my project and theirs, I think that it will be instructive to compare the two ways of understanding stability, bracketing any differences in how we're understanding the counterfactual conditional ' $\rightarrow$ '.

Suppose we know that a generalization  $G$  is accurate throughout some modal neighborhood (of Earth throughout 2020, say). Can we conclude that it is also macro-stable relative to this region? I think we can. Think of how vast the corresponding modal neighborhood will be. Consider making small changes to the actual world on January 1<sup>st</sup>. You will find worlds where the weather on March 2<sup>nd</sup> is different, worlds where my coffee spills on October 30<sup>th</sup>, worlds where ten more ravens mate on Halloween, and worlds where the supply of shoes on Christmas day is 10% higher. Differences amplify as time goes by, yielding great variety in the modal neighborhood. Thus, we should expect the modal neighborhood to include worlds that differ from ours with respect to every macro-variable. We should

<sup>46</sup> Requiring that the consequent hold in *all* nearby worlds would simplify the account, but the corresponding stability conditions would end up being too demanding (especially given that our modal neighborhoods contains lots of tiny islands of anti-thermodynamic worlds).

<sup>47</sup> A notable exception is Michael Strevens, who employs a similarity-based non-backtracking reduction of counterfactuals. See Strevens, "Physically contingent laws and counterfactual support," *op. cit.*

<sup>48</sup> This is not, of course, a criticism of authors that take counterfactuals as primitive; these authors are not engaged in the same foundationalist project that I am.

expect, then, that almost all the relevant macro-antecedents will be true somewhere in the modal neighborhood. Since our robust generalization  $G$  approximately holds in almost all worlds of the modal neighborhood, relevant macro-antecedents will likely take us to worlds where  $G$  approximately holds.<sup>49</sup>

The above line of reasoning is far from a rigorous proof. However, I think it makes it credible that generalizations which are accurate across some region's modal neighborhood are also macro-stable within that region. This in turn means that that if a generalization fails to be macro-stable in some region, then it is unlikely to feature in the best system for that region (either as axiom or theorem).<sup>50</sup> If this is correct, then the crystallization account vindicates, from a foundationalist perspective, the guiding idea behind counterfactual stability accounts of robustness: counterfactual fragility is a clear sign of accidentality.

## CONCLUSION

The special sciences treat certain higher-level generalizations as nomically elite or 'robust'. To a first approximation, those generalizations do not seem privileged from the perspective of physics: they are physically contingent, and look highly complex when translated into physical vocabulary. Some philosophers take this to indicate autonomy from physics: they conclude that higher-level regularities do not inherit their robustness from the laws of physics. However, the crystallization account demonstrates how these marks of autonomy may be reconciled with foundationalism. On this account, robust regularities have a privileged status in virtue of efficiently summarizing patterns throughout our 'modal neighborhoods'.

The idea that much of our scientific theorizing aims at summarizing our modal neighborhoods provides a new framework for conceptualizing the connections between robustness, counterfactuals, explanation, and induction. This paper has only begun to explore these connections, conjecturing that

<sup>49</sup> Without a more precise proposal for counting macro-variable antecedents, it will not be possible to verify that this last step is licensed. But I hope that the prima facie plausibility of this reasoning suffices to illuminate the connection between the notions in question.

<sup>50</sup> Macro-stability does not suffice for accuracy throughout the modal neighborhood. Take a generalization  $G$ , which is macro-stable relative to  $R$ . For each variable  $X$  in  $G$ , mark the worlds in  $R$ 's modal neighborhood that are the closest  $X$ -variants (for each object  $a$ ). Given  $G$ 's macro-stability, it will be accurate at all the marked worlds. But it needn't be accurate at unmarked worlds. And unless our set of variables happens to be very comprehensive, these unmarked worlds will occupy a significant proportion of the modal neighborhood.

we rely on these summaries to reason counterfactually/hypothetically about nearby possibilities (for example, in deliberation). Future work should evaluate this conjecture more systematically, and also explore how modal neighborhoods connect to explanation and prediction. Why is our knowledge of the future so deeply intertwined with our knowledge of the modal neighborhood? How do our attempts to explain observable patterns in the actual world contribute to the construction of accurate summaries for modal neighborhoods of various sizes? Are there explanatory connections between summaries for various neighborhoods? If so, how do we discover them?