Simplicity, Truth, and Topology

Kevin T. Kelly
Konstantin Genin
Hanti Lin

October 16, 2014

Abstract

Key words: simplicity, Ockham's Razor, topology, Popper, falsification, learning, induction, structuralism, specialization order, stratification, Hilbert space.

1 Introduction

We assume that science seeks true theories of the world—at least sometimes. When multiple, alternative theories compatible with information are available, scientists routinely favor the simplest, a bias known as *Ockham's razor*. Ockham's razor raises two fundamental questions, neither of which is settled.

I. What is empirical simplicity?

II. How does Ockham's razor lead one to true theories better than competing strategies would?

Neither question is easy. Simplicity appears to be an equivocal grab-bag of virtues, ranging from testability to minimization of entities, to maximization of symmetry, to minimization of parameters, causes, and principles. Moreover, whatever simplicity turns out to be, how could a fixed bias toward simplicity *possibly* guarantee better truth-finding performance when the truth might be complex? While a bias toward simplicity would obviously help in simple worlds, it would hurt in the complex ones—a mere example of robbing Peter to pay Paul, unless one begs the question by assuming, at the outset, that the world is simple.

There is a burgeoning literature on question (I) and (II), spread across philosophy, statistics, and machine learning, but no convincing answers have been forthcoming. There are frequentist attempts to explain Ockham's razor as a way to reduce predictive noise

(Akaike, 1974; Forster and Sober, 1994; Vapnik, 1998), but that instrumentalistic approach falls short of justifying belief in the theories, themselves. At the opposite extreme, Bayesians can post arbitrarily high betting quotients on inductive conclusions, and can explain Ockham's razor in a rationalistic way, as the product of conditioning over a wide class of plausible, prior probabilities that impose flattish distributions over theoretical parameters (Jeffreys, 1961; Bandyopadhayay et al., 1996; Wasserman, 2000; Myrvold, 2003). But that approach, while non-skeptical, does not begin to explain how such prior probabilities lead one to true theories better than alternative biases would—unless one appeals to the prior probabilities, themselves, which is circular.

This paper answers both questions in a unified way. The answer to question (I) is that simplicity is a topological order that emerges when one coarse-grains an inductive inference problem. The answer to question (II) is that Ockham's razor, relative to the emergent simplicity order, keeps one on the most direct route to the true answer to the inductive question, even though that route cannot be entirely direct. The plan is as follows. Section 2 introduces empirical problems. An empirical problem consists of a set of possible worlds partitioned into a countable collection of alternative answers by a theoretical question, and covered by a countable collection of possible information states. The information space specifies the kind of information one might receive in the future and the alternative answers to the question are alternative hypotheses, models, or theories. Everything in the following development is defined relative to an empirical problem, just as the analysis of algorithms is problem-relative in computer science. Section 3 introduces learners and convergence to the truth.

The next several sections of the paper develop a very general, topological theory of empirical simplicity, general enough to embrace ill-founded and even dense simplicity orders. Section 4 introduces the information topology induced by the information space and explains how standard epistemological concepts like verifiability, and refutability, are most fundamentally topological, rather than logical or probabilistic. Chapter 5 introduces the specialization pre-order, a standard topological concept, and demonstrates that it is identical to Karl Popper's (1959) proposed definition of empirical simplicity. The topological theory of empirical simplicity is developed in section 6. A stratification of a topological space is a partition of the space into locally closed cells (called simplicity degrees) that are homogeneous, in the sense that all of the worlds in a cell bear the same specialization relations to all other cells. Simplicity, we propose, is just a stratification of the original problem that decides the original question, in the sense that each answer to the original question is verifiable, given a simplicity degree. The rich and attractive consequences of that brief definition are developed. Section 7 presents some examples, including discrete experimental outcomes through time, and continuous polynomial laws.

The discussion then turns from simplicity to Ockham's razor. In section 8, the focus shifts from solving the original problem, to identifying the true simplicity degree, which serves as one's reason for believing an answer to the original problem. Upper bounds on simplicity are refutable in general. Section 9 highlights simplicity concepts with the special feature that lower bounds are verifiable. The difference is reminiscent of Kuhn's celebrated distinction between normal and revolutionary science, except that it is topological rather than political. We construct canonical, Ockham solutions to problems of either type. In the general case, an arbitrary enumeration of simplicity degrees is employed to impart the requisite, additional bias necessary for convergence to the truth. Section 10 defines Ockham's razor relative to simplicity. Ockham's vertical razor requires that one infer a disjunction whose disjuncts are closed downward in the simplicity order, restricted to degrees compatible with current information. Ockham's horizontal razor, which is restricted to simplicity concepts with open upward sets, requires that every minimal degree compatible with current information be included as a disjunct. When there are infinite descending chains of simplicity degrees, that generalizes to the requirement that disjuncts be co-initial in the simplicity order.

With simplicity and Ockham's razor on the table, we turn to the definition of optimally direct convergence to the truth in section 11. Our proposal has an ancient precedent:

> Fools dwelling in darkness, wise in their own conceit, and puffed up with vain knowledge, go round and round, staggering to and fro, like blind men led by the blind.[1]

Going round and round implies a credal cycle (returning to a view you once held and then abandoned) and staggering to and fro implies a credal reversal (repudiating a view you once held). Both explicate aspects of indirectness of one's approach to the truth.[2] We compare methods in terms of their worst-case performance, in both senses. Worst-case optimality means doing as well as an arbitrary convergent solution to the problem in terms of worst-case performance in every simplicity degree. Worst-case sub-optimality means that some alternative method does at least as well in each simplicity degree and better in some simplicity degree. Worst-case admissibility means not being worst-case sub-optimal. In order to extend the argument to arbitrary simplicity orders, we compare cycle and reversal sequences directly by means of sub-sequence relations, rather than through intermediate, numerical losses associated with such sequences.

The development culminates, in section 12, with a series of results that single out Ockham's razors in terms of directness of approach to the truth, relative to an underlying simplicity concept. In the case of open upward sets, Ockham's vertical razor is necessary for worst-case cycle optimality and Ockham's horizontal razor is necessary for worst-case reversal optimality. In general, Ockham's vertical razor is necessary for worst-case cycle optimality. How does the argument escape from the apparently insuperable problem of

[1] Excerpt from the *Katha Upanishad* (Müller, 1900).

[2] In formal learning theory, reversals of opinion are called mind-changes (Case et al., 1995) and cycles are called U-shaped learning (Carlucci et al., 2005). However, it is usually assumed that learners are forced to choose a full answer to the question at each stage, whereas we allow learners to produce arbitrary disjunctions of answers. In that more general setting, changing one's mind could mean strengthening belief, dropping a belief, or both. Reversals occur only when belief is both strengthened and weakened.

robbing Peter to pay Paul? It turns out that favoring a complex possibility over a simpler one results in extra cycles and reversals of opinion, in some possibility in the favored, *complex* simplicity degree. So *both* Peter and Paul get robbed when complex simplicity degrees are favored over simple ones! That is the epistemological crux of the matter.

Finally, the theory of simplicity developed in section 6 is intended to be as weak as possible, given the goal of justifying Ockham's razor. However, that raises a question about the objectivity of simplicity, since the entire argument is simplicity-relative. Section 13 presents some plausible, further principles that narrow down the range of possible simplicity concepts for a given empirical problem considerably, particularly for problems in which simplicity intuitions are strong. However, there are also highly symmetrical problems in which simplicity breaks symmetry. We propose, as a meta-principle for all possible simplicity theories, that the set of all simplicity concepts for a given problem be closed under problem symmetries.

The proposed paper extends and improves upon the general approach of Kelly and Glymour (2004); Kelly (2007a, 2010); Luo and Schulte (2006); Martin and Osherson (1998) which was inspired by earlier work in formal learning theory (Putnam, 1965; Kugel, 1977; Case and Smith, 1983; Osherson et al., 1986; Freivalds and Smith, 1993; Kelly, 1996; Martin and Osherson, 1998) and in the philosophy of science, where simplicity and truth are major themes in the scientific realism debate (Popper, 1959; Glymour, 1980; Van Fraassen, 1980). The major advance in this study is the topological theory of simplicity relative to problems defined, abstractly, in terms of possible world semantics and the lifting of the justification of Ockham's razor to that setting.[3] The generality is consonant with an increasing awareness that topology is the right setting for learning theoretic analysis (Kelly, 1996; Vickers, 1996; Martin et al., 2006; Schulte et al., 2007; Yamamoto and de Brecht, 2010; Case and Kötzing, 2013; Baltag et al., 2014). Furthermore, earlier versions of the argument were based on retraction minimization. Shifting the focus to cycles and reversals results in a far more general and direct optimality argument for Ockham's razor and also reveals the important distinction between Ockham's vertical and horizontal razors.

2 Empirical Problems

Let W be a set of points called *possible worlds*. A *proposition* is a subset of W. Then logical operations correspond to set theoretic operations: $A \wedge B = A \cap B$, $A \vee B = A \cup B$, $\neg A = W \setminus A$, and A entails B iff $A \subseteq B$.

An *information basis* on W is a countable collection $\mathcal{I}$ of propositions such that $\mathcal{I}$ covers W and, for each world w and $E, F \in \mathcal{I}$ such that $w \in E \cap F$, there exists $G \in \mathcal{I}$

[3]The concept of stratification of a topological space is familiar in algebraic geometry (Stratification., http://www.encyclopediaofmath.org/index.php?title=Stratification), which studies the numerical stability of solutions to polynomial equations. We arrived at the concept independently, based entirely on its aptness as a theory of empirical simplicity.

such that $w \in G \subseteq E \cap F$. In other words, every world presents at least some information state (possibly vacuous), and for any true, finite conjunction of information states, there is a true information state that entails the conjunction. The assumption of countability follows Alan Turing's (1936) argument that infinite gradations of input information are indistinguishable.

It is now a familiar point in philosophy (Hintikka, 2007), logic (Groenendijk, 2009), and theoretical linguistics (Roberts, 2004) that questions guide inquiry and discourse. We propose that empirical simplicity emerges when a question is imposed on an information basis. A *question* $\mathcal{Q}$ on W is modeled as a countable partition of W into mutually exclusive and exhaustive, non-empty propositions called *potential answers* to $\mathcal{Q}$. Note that question $\mathcal{Q}$ satisfies the requirements for an information basis, and may be thought of as the *output* information basis. Let $\phi_{\mathcal{Q}}(w)$ denote the unique answer Q true in w, which is just the unique cell of $\mathcal{Q}$ that contains w. The function $\phi_{\mathcal{Q}}$ is called the *canonical map* for $\mathcal{Q}$. Question $\mathcal{Q}$ *refines* question $\mathcal{Q}'$ iff each answer to $\mathcal{Q}$ entails some answer to $\mathcal{Q}'$. The questions on W constitute a lattice under the refinement order, with the least refined question $\mathcal{Q}_{\top} = \{W\}$ on top and the most refined question $\mathcal{Q}_{\perp} = \{\{w\} : w \in W\}$ on the bottom.

Let $\mathcal{I}$ be an information basis on W and let $\mathcal{Q}$ be a question on W. Then $\mathfrak{I} = (W, \mathcal{I})$ is an *information space*, $\mathfrak{Q} = (W, \mathcal{Q})$ is a *question space* and $\mathfrak{P} = (W, \mathcal{I}, \mathcal{Q})$ is an *empirical problem.*[4] It is convenient to identify $\mathfrak{P}$ with the pair $(\mathfrak{I}, \mathcal{Q})$, so that it is easy to change the question. Henceforth, we assume, without further comment, that $\mathfrak{P} = (W, \mathcal{I}, \mathcal{Q}) = (\mathfrak{I}, \mathcal{Q})$.

Let $\mathcal{I}(w)$ denote the set of all propositions in $\mathcal{I}$ that contain w. Call $\mathcal{I}(w)$ the *local information basis* at w, which contains every information state that would ever be entered in w. Furthermore, define the *restriction* of collection $\mathcal{S}$ of propositions to $E \in \mathcal{I}$ as follows:

$$\mathcal{S} \restriction E \quad = \quad \{S \cap E : S \in \mathcal{S}\} \setminus \{\varnothing\}.$$

Define the restricted problem—the problem faced in light of new information E, to be $\mathfrak{P} \restriction E = (E,\ \mathcal{I} \restriction E,\ \mathcal{Q} \restriction E)$. Define restricted information and question spaces similarly.

3 Convergence to the Truth

We view inquiry as a process of processing information and producing answers relevant to the question, in a way that should arrive, eventually, at the right answer. Let $\mathcal{Q}^*$ denote the closure of $\mathcal{Q}$ under union, so $\mathcal{Q}^*$ contains arbitrary disjunctions of answers to $\mathcal{Q}$. A *learning method* for problem $\mathfrak{P}$ is a function $\lambda : \mathcal{I} \to \mathcal{Q}^*$ such that λ never produces the inconsistent proposition $\varnothing$.

[4]In section 13 we discuss some reasons to think that a problem involves a set of questions, nested by preuppositions, rather than just a single question.

Solution of $\mathfrak{P}$ in the limit demands that the learner stabilize, eventually, to the true answer to $\mathfrak{P}$ on increasing information. Stabilization to the truth has been a central theme in epistemology since Plato's discussion of knowledge as stabilized true belief in the *Meno*. Say that $E \in \mathcal{I}(w)$ *locks* λ onto the truth in w, $\mathfrak{P}$ iff $\lambda(F) = \phi_{\mathcal{Q}}(w)$, for all $F \in \mathcal{I}(w)$ such that $F \subseteq E$.[5] Let $\mathsf{Lock}_{\mathfrak{P}}(\lambda, w)$ denote the set of all $E \in \mathcal{I}(w)$ such that E locks λ onto the truth in w, $\mathfrak{P}$. Then say that λ *solves* $\mathfrak{P}$ in the limit iff $\mathsf{Lock}_{\mathfrak{P}}(\lambda, w)$ is non-empty, for each $w \in W$.

4 Information Topology

The logic of inquiry is grounded in the *information topology* of the problem faced, a view that is becoming increasingly accepted in the area of computational learning theory (Kelly, 1996; Vickers, 1996; Martin et al., 2006; Schulte et al., 2007; Yamamoto and de Brecht, 2010; Baltag et al., 2014). Let $\mathcal{I}^*$ denote the closure of $\mathcal{I}$ under arbitrary union and let $\mathfrak{I}^* = (W, \mathcal{I}^*)$ and $\mathfrak{P}^* = (\mathfrak{I}^*, \mathcal{Q})$. Then $\mathfrak{I}^*$ is the topological space induced by $\mathfrak{I}$, which is called the *information topology*. Many of the most fundamental epistemological concepts are invariant features of the information topology. The propositions in $\mathcal{I}^*$ are the *open propositions* (of $\mathfrak{I}^*$), and elements of $\mathcal{I}^*(w)$ are *neighborhoods* of w (in $\mathfrak{I}^*$). We henceforth omit reference to $\mathfrak{I}$ unless ambiguity would result.

The open propositions are exactly the *verifiable* propositions, where H is verifiable iff, for each $w \in H$ there exists $E \in \mathcal{I}(w)$ such that $E \subseteq H$. Closed propositions are complementary to open propositions and correspond to *refutable* propositions. Clopen (closed-open) propositions are, therefore, *decidable*.

World w is an *interior point* of H iff there exists $E \in \mathcal{I}(w)$ such that $E \subseteq H$—i.e., if w presents information that verifies H. The *interior* $\mathsf{int}(H)$ of H is the set of all interior points of H, so it is just the proposition that H will be verified. World w is a *boundary point* of H iff w is an interior point of neither H nor $\neg H$. Thus, boundary worlds are worlds in which the binary question $\{H, \neg H\}$ is never answered deductively by information. The *boundary* $\mathsf{bdry}(H)$ of H is the set of all boundary points of H.

Furthermore, $\mathsf{bdry}(\mathcal{Q})$ is the union of the boundaries of the answers to $\mathcal{Q}$ and $\mathsf{int}(\mathcal{Q})$ is the union of the interiors of the answers to $\mathcal{Q}$. In that case, we have the dual relation $\mathsf{bdry}(\mathcal{Q}) = \neg\mathsf{int}(\mathcal{Q})$. Then $\mathsf{bdry}(\mathcal{Q})$ is the proposition that no answer will ever be verified and $\mathsf{int}(\mathcal{Q})$ is the proposition that some answer will be verified (i.e., that information will decide the question deductively). A *purely inductive* question is a question whose boundary is W, so one necessarily faces the problem of induction with respect to it—there is no prospect of crucial information that would settle the question. A *decidable* question has interior W, so information is destined to settle it deductively. In general, both the interior and the boundary of a question may be non-empty.

[5] The locking terminology follows Jain et al. (1999).

World w is a *closure point* of H iff w presents no information incompatible with H. Equivalently, w is a closure point of H iff w is an interior point of H or a boundary point of H. The *closure* $\mathsf{cl}(H)$ of H is the set of all closure points of H. Thus, $\mathsf{cl}(H) = \mathsf{int}(H) \cup \mathsf{bdry}(H)$. It is easy to show that H is open iff $H = \mathsf{int}(H)$ and H is closed iff $H = \mathsf{cl}(H)$.

5 Falsifiability and the Specialization Pre-Order

Define the *specialization pre-order* over propositions in $\mathfrak{I}^*$ as follows:

$$G \preceq_{\mathfrak{I}^*} H \quad \text{iff} \quad G \subseteq \mathsf{cl}_{\mathfrak{I}^*}(H).$$

Then G entails that one faces the problem of induction with respect to H, if H is false. As usual, drop the $\mathfrak{I}^*$ subscript when no ambiguity arises. Define $\prec$ to hold when $\preceq$ holds one way and not the other and define $\cong$ to hold when $\preceq$ holds both ways.

Karl Popper proposed three distinct, but similar, concepts of empirical simplicity. According to Popper's second concept (Popper, 1959), G is *as falsifiable* as H iff every information state incompatible with H is also incompatible with G. Then it is easy to check that:

Proposition 1. *Popper's falsifiablity order is identical to the specialization pre-order.*

One important shortcoming of Popper's proposal is that $\preceq$ is only a pre-order, so "simplicity cycles" are possible. For an easy example, suppose that possible worlds are natural numbers, so $W = \mathbb{N}$. Information states are upward-closed subsets of $\mathbb{N}$. Suppose that the question is the maximally refined question $\mathcal{Q}_\perp$. Then it is easy to check that $\{i\} \preceq \{j\}$ iff $i \leq j$, as one would expect—if we think of the problem as counting entities or reaction types, then theories that posit more are usually considered to be more complex. But now coarsen the question to $\mathcal{Q}_{\mathsf{prty}}$, whose answers are "even" and "odd". In that case, we have a $\preceq$ cycle between "even" and "odd", regardless of current information. However, if the current information is lower bound 3, it seems that "odd" is simpler than "even" and if the information is lower bound 4, then "even" is simpler than "odd". The cyclic simplicity order does not recover that judgment. It would do so, however, if the coarse question $\mathcal{Q}_{\mathsf{prty}}$ is replaced with the maximally refined question $\mathcal{Q}_\perp$. That eliminates the cycle by breaking it into an infinite spiral that alternates in parity. Moreover, given lower bound 3, the simplest answer to $\mathcal{Q}_\perp$ is $\{3\}$, which entails "odd" and rules out "even".

6 Empirical Simplicity

Recall, from the preceding section, that $\preceq$ can have cycles over the original question $\mathcal{Q}$, so the task is to somehow sift an alternative question $\mathcal{S}$ from $\mathcal{Q}$ such that $\preceq$ is anti-symmetric (cycle-free) over $\mathcal{S}$ and still somehow reflects upon the original question. Then the problem $\mathfrak{S} = (W, \mathcal{I}, \mathcal{S})$ may be viewed as a *simplicity* concept for the original problem $\mathfrak{P} = (W, \mathcal{I}, \mathcal{Q})$, and the answers to $\mathfrak{S}$ can be called *simplicity degrees*. The resulting partially ordered set (poset) $(\mathcal{S}, \preceq)$ then corresponds to a simplicity structure for $\mathfrak{P}$, and the elements of $\mathcal{S}$ are *simplicity degrees*. The induced, simplicity pre-order on worlds is just:

$$w \preceq_{\mathcal{S}} v \quad \text{iff} \quad \phi_{\mathcal{S}}(w) \preceq \phi_{\mathcal{S}}(v).$$

We aim at a weak proposal whose conditions are plausible and that suffices to prove that Ockham's razor is the best strategy for solving inductive problems. Further conditions are discussed in section 13.

Here are some handy concepts definable from $(\mathcal{S}, \preceq)$, for arbitrary collection of propositions $\mathcal{X}$. Let $E \subseteq W$. Then the image $\phi_{\mathcal{S}}(E) = \{\phi_{\mathcal{S}}(w) : w \in E\}$ is just the set of all simplicity degrees $C \in \mathcal{S}$ that are logically compatible with information E. Define $\mathsf{Min}_{\mathcal{S}}(E)$ to be the set of all $\preceq$-minimal elements of $\phi_{\mathcal{S}}(E)$. For $H \in \mathcal{S}^*$, let $H_{\preceq}$ denote the set of all $C \in \mathcal{S}$ such that $C \preceq H$. Let $H^*_{\preceq}$ denote the set of all $G \in \mathcal{S}^*$ such that $G \preceq H$. The definitions for $\prec$, $\succeq$ and, $\succ$ are similar.

6.1 Local Closure

Perhaps the most paradigmatic simplicity concept is polynomial degree. Let worlds be polynomial functions $Y = w(X)$. An *information state* consists of all polynomial functions passing through some finite set of rectangles in the real plane.[6] Call that information space $\mathfrak{I}_{\mathsf{poly}}$. The question $\mathcal{Q}_{\mathsf{deg}}$ is to determine the degree of the true polynomial function. Let C_n assert that the true degree is exactly n. Then C_{n+1} has the property that, if $w \in C_{n+1}$, then one eventually receives information E_w in C_{n+1} (namely, rectangles that rule out all C_i for $i \leq n$), *given* which, if C_{n+1} is false, then more rectangles will be seen that rule out all polynomials in C_{n+1}. More generally, say that proposition C is *conditionally refutable* iff, for all $w \in C$, there exists $E \in \mathcal{I}^*(w)$ such that $C \cap E$ is closed in $\mathfrak{I}^* \restriction E$. Say that $\mathcal{S}$ is conditionally refutable in $\mathfrak{I}^*$ iff each $C \in \mathcal{S}$ is.

The shift to the restricted space $\mathfrak{I}^* \restriction E$ in the definition of conditional refutability can be eliminated as follows. In topology, C is said to be *locally closed* iff C is a difference of open propositions.[7] Then we have:

Proposition 2. *C is conditionally refutable iff C is locally closed.*

Say that $\mathcal{S}$ is locally closed iff each $C \in \mathcal{S}$ is locally closed (in $\mathfrak{I}^*$). Our first requirement on simplicity concepts is that they be locally closed. In terms of locally closed sets, conditional refutability may be thought of as follows.

[6]This data model, known as *uncertain but bounded error* is discussed in Glymour (2001). The example is developed rigorously in section 7.

[7]The moniker "locally closed" actually connotes conditional refutability, but we follow the standard definition.

Proposition 3. *Suppose that $\mathcal{S}$ is locally closed in $\mathfrak{I}^*$. Suppose, further, that $C \in \mathcal{S}$, that A, B are open, and that $C = A \setminus B$. Then:*

$$A \cap \bigcup C_{\prec} \;\; = \;\; \varnothing.$$

It is *not* true in general that $\bigcup C_{\succeq}$ is open. When $\mathcal{S}$ satisfies that, stronger property, say that $\mathcal{S}$ has *open upward sets*. We discuss that special case in section 9 below.

A crucial, further consequence of local closure is that it forces the simplicity order to be a partial order—simplicity cycles are ruled out:

Proposition 4. *If $\mathcal{S}$ is locally closed for $\mathfrak{I}^*$, then $\mathcal{S}$ is partially ordered by $\preceq$.*

Finally, local closure guarantees solvability in the limit. Let countable partition $\mathcal{S} = \{C_i : i \in \mathbb{N}\}$ be locally closed in $\mathfrak{I}$. Let $C_i = A_i \setminus B_i$, for B_i, $C_i \in \mathcal{I}^*$. Say that $E \in \mathcal{I}$ *activates* C_i iff $E \subseteq A_i$ and $E \not\subseteq B_i$. Let $\lambda_{\mathsf{enm}}(E) =$ the first C_i activated by E, if there is one, and $\bigcup \phi_{\mathcal{S}}(E)$ otherwise. Then λ_{enm} solves $\mathfrak{S}$.

Proposition 5. *Suppose that question $\mathcal{S}$ is locally closed in $\mathfrak{I}$. Then λ_{enm} solves $\mathfrak{S}$ in the limit.*

6.2 Homogeneity

The $\preceq$ relation is a partial order over locally closed partition $\mathcal{S}$, but there is still an important shortcoming in the idea. If degree $C = \{w, v\}$ is heterogeneous, in the sense that $\{w\} \prec D$ and $\{v\} \not\prec D$, then $C \not\prec D$, so the problem of induction from w to D is invisible in the simplicity order $(\mathcal{S}, \preceq)$. Implicit in the concept of a simplicity degree is that one need only know the degree of the world to know which problems of induction one faces in that world. Then the partial order $(\mathcal{S}, \preceq)$ can serve as a kind of "epistemic road map", in the sense that one's location in a city determines the set of all cities one can reach from that city, without further information about one's position in the city. Accordingly, say that proposition C is *homogeneous* for $\mathcal{S}$ in $\mathfrak{I}^*$ iff:

$$\{w\} \preceq D \quad \text{implies} \quad C \preceq D,$$

for all $w \in C$ and $D \in \mathcal{S}$. When each $C \in \mathcal{S}$ is homogeneous for $\mathcal{S}$ in $\mathfrak{I}^*$, say that $\mathcal{S}$ is homogeneous in $\mathfrak{I}^*$.

Homogeneity forces the information topology to align with the simplicity order on degrees.

Proposition 6. *Suppose that partition $\mathcal{S}$ is homogeneous in $\mathfrak{I}^*$ and that $C \in \mathcal{S}$. Then:*

$$\mathsf{cl}(C) = \bigcup C_{\preceq}.$$

Another important consequence of homogenity is that it holds the simplicity relation rigid in light of new information.

Proposition 7. *If $\mathcal{S}$ is homogeneous in $\mathfrak{I}^*$, then, for all $C, D \in \mathcal{S}$ such that $C \cap E \neq \varnothing$:*

$$C \preceq_{\mathfrak{I}^*} D \quad \textit{iff} \quad C \cap E \preceq_{\mathfrak{I}^* \restriction E} D \cap E.$$

6.3 Stratification

Say that $\mathcal{S}$ *stratifies* $\mathfrak{I}^*$ iff $\mathcal{S}$ is both homogeneous and locally closed in $\mathfrak{I}^*$.[8] A stratification ensures that the disjunction of all simplicity degrees simpler than a given degree is refutable, which is crucial if simplicity is to be a guide to inquiry.

Proposition 8. *Suppose that $\mathcal{S}$ stratifies $\mathfrak{I}^*$ and that $C \in \mathcal{S}$. Then $\bigcup C_{\prec}$ is closed.*

As a consequence, the open differences that witness local closure assume a normal form that lines up with the simplicity order:

Proposition 9. *If $\mathcal{S}$ stratifies $\mathfrak{I}^*$, then the local closure of $C \in \mathcal{S}$ is witnessed by:*

$$C = \bigcup C_{\not\prec} \setminus \bigcup C_{\not\preceq}.$$

If $\mathcal{S}$ has open upward sets, then there exists witness:

$$C = \bigcup C_{\succeq} \setminus \bigcup C_{\succ}.$$

6.4 σ-Homogeneity

It is natural to extend the definition of homogeneity from $\mathcal{S}$ to $\mathcal{S}^*$, as follows. Say that proposition C is *σ-homogeneous* iff

$$\{w\} \preceq H \quad \text{implies} \quad \phi_{\mathcal{S}}(w) \preceq H,$$

for all $w \in C$ and $H \in \mathcal{S}^*$. A *σ-stratification* is a σ-homogeneous stratification. It makes a difference:

Proposition 10. *Homogeneity is logically weaker than σ-homogeneity.*

[8] Stratifications have been a familiar topic in the area of algebraic geometry, since the 1950s (Stratification., http://www.encyclopediaofmath.org/index.php?title=Stratification). The terminology is not entirely standardized. Sometimes, what we call a stratification is referred to as a *good* stratification (Stacks Project, http://stacks.math.columbia.edu/tag/09XY), whereas a stratification is a weaker concept.

We do not insist upon σ-homogeneity, as it can be subtle to verify in paradigmatic applications, and our main results do not depend upon it. But the following considerations speak strongly in its favor.

The *relevant* responses to question $\mathcal{S}$ are disjunctions of answers to $\mathcal{S}$. The image $\phi_{\mathcal{S}}(H) = \{\phi_{\mathcal{S}}(w) : w \in H\}$ is exactly the set of all answers to $\mathcal{S}$ that are logically compatible with proposition H. Furthermore, $(\phi_{\mathcal{S}}^{-1} \circ \phi_{\mathcal{S}})(H) = \bigcup \phi_{\mathcal{S}}(H)$ is the disjunction of all such answers, which is the *strongest relevant consequence* of H in $\mathcal{S}$:

$$\mathsf{rel}_{\mathcal{S}}(H) \quad = \quad (\phi_{\mathcal{S}}^{-1} \circ \phi_{\mathcal{S}})(H) \quad = \quad \bigcup \{\phi_{\mathcal{S}}(w) : w \in H\}.$$

Thus, $\mathsf{rel}_{\mathcal{S}}$ is the "coarse-graining" of H by $\mathcal{S}$. It is immediate from the definition that:

Proposition 11. $\mathcal{S}^* \cap \mathcal{I}^* \;\subseteq\; \mathsf{rel}_{\mathcal{S}}(\mathcal{I}^*) \;\subseteq\; \mathcal{S}^*$.

The $\mathsf{rel}_{\mathcal{S}}$ operation commutes with union:

Proposition 12. *Let $\mathcal{H}$ be an arbitrary collection of propositions over W. Then:*

$$\mathsf{rel}_{\mathcal{S}}(\mathcal{H}^*) \quad = \quad \mathsf{rel}_{\mathcal{S}}(\mathcal{H})^*.$$

One may think of $\mathsf{rel}_{\mathcal{S}}(E)$, for information state E, as an "empirical effect", in the sense in which scientists speak of a "second-order effect" or "the photo-electric effect". One would surely expect empirical effects to be empirically verifiable (open). That natural condition is *equivalent* to σ-homogeneity.

Proposition 13. *Let $\mathcal{S}$ partition topological space $\mathfrak{I}^*$.*

$$\mathcal{S} \textit{ is } \sigma\textit{-homogeneous for } \mathfrak{I}^* \quad \textit{iff} \quad \mathsf{rel}_{\mathcal{S}}(\mathcal{I}) \subseteq \mathcal{I}^*.$$

The following corollary summarizes the preceding, three propositions:

Proposition 14. *Let $\mathcal{S}$ partition topological space $\mathfrak{I}^*$. Then:*

$$\mathcal{S} \textit{ is } \sigma\textit{-homogeneous for } \mathfrak{I}^* \quad \textit{iff} \quad \mathsf{rel}_{\mathcal{S}}(\mathcal{I})^* = \mathsf{rel}_{\mathcal{S}}(\mathcal{I}^*) = \mathcal{S}^* \cap \mathcal{I}^*.$$

Moreover, the set of all empirical effects is an information basis in its own right, so one may simply replace the original information space with the space of empirical effects, to arrive at an abstract version $\mathsf{rel}_{\mathcal{S}}(\mathfrak{P}) = (W, \mathsf{rel}_{\mathcal{S}}(\mathcal{I}), \mathcal{Q})$ of the original problem $\mathfrak{P}$:

Proposition 15. *Suppose that $\mathcal{S}$ is σ-homogeneous for $\mathfrak{I}$. Then $\mathsf{rel}_{\mathcal{S}}(\mathcal{I})$ is an information basis on W.*

Finally, when $\mathcal{S}$ is σ-homogeneous, the simplicity order is preserved under restriction, over all of $\mathcal{S}^*$:

Proposition 16. *If $\mathcal{S}$ is σ-homogeneous in $\mathfrak{I}^*$, then proposition 7 holds, with $\mathcal{S}^*$ in place of $\mathcal{S}$.*

We close with a pair of examples illustrating the pathologies of non-σ-homogeneous partitions. The first example possesses an information state whose strongest relevant consequence is unverifiable. Let $W = \mathbb{N}$. Let $\mathcal{I}$ consist of upward-closed subsets of $\mathbb{N}$, under the standard order, together with the special information state $\{0\}$. Let $\mathcal{S} = \{0, 1\} \cup \{\{n\} : n > 1\}$. It is intuitively odd to associate the isolated world 0 with the generic world 1. However, $\preceq$ is just the identity relation over $\mathcal{S}$, so homogeneity is satisfied trivially. Local closure is also satisfied, since $\{0, 1\} = W \setminus \{n \in \mathbb{N} : n > 1\}$, so $\mathcal{S}$ is a stratification. But $\{0\}$ is an information state and $\mathsf{rel}_{\mathcal{S}}(\{0\}) = \{0, 1\}$ is not open, so σ-homogeneity fails by proposition 13. To see how, note that $\{0\} \not\prec \{\{n\} : n > 1\}$ and $\{1\} \prec \{\{n\} : n > 1\}$. The natural remedy is to split the gerrymandered degree $\{0, 1\}$.

A similar example illustrates failure of $\mathsf{rel}_{\mathcal{S}}(\mathcal{I})$ to be an information basis and failure of $\mathsf{rel}_{\mathcal{S}}(\mathcal{I}^*)$ to be a topological space, for information basis $\mathcal{I}$. Let $\mathbb{N}_0$ and $\mathbb{N}_1$ be two disjoint copies of the natural numbers and let a_0, a_1 be two further worlds. Let $W = \mathbb{N}_0 \cup \mathbb{N}_1 \cup \{a_0, b_0\}$. Each information state is either a singleton $\{n_i\}$, for $i \leq 1$, or a set of the form $\{a_i\} \cup X$, where X is an upward-closed subset of $\mathbb{N}_i$, for $i \leq 1$. Let $\mathcal{S} = \{\{a_0, a_1\}\} \cup \{\{n_0\} : n_0 \in \mathbb{N}_0\} \cup \{\{n_1\} : n_1 \in \mathbb{N}_1\}$. Then $\mathcal{S}$ stratifies $\mathcal{I}^*$, but violates σ-homogeneity, in light of proposition 15, because $\mathsf{rel}_{\mathcal{S}}(\mathcal{I})$ is not an information basis and $\mathsf{rel}^*_{\mathcal{S}}(\mathcal{I})$ is not a topological space. The violation is witnessed by $\{a_0, a_1\} \in \mathcal{S}$, because $a_0 \prec \mathbb{N}_0$ and $a_1 \not\prec \mathbb{N}_0$. The natural remedy is to split the gerrymandered degree $\{a_0, a_1\}$.

6.5 Deciding the Original Problem

Stratification depends entirely on the information topology $\mathfrak{I}^*$. The third requirement relates $\mathfrak{I}^*$ to the question $\mathcal{Q}$. Say that $\mathcal{S}$ *decides* $\mathcal{Q}$ in $\mathfrak{I}^*$ iff $H \cap C$ is open in $\mathfrak{I} \restriction C$, for each $C \in \mathcal{S}$ and $H \in \mathcal{Q}$. Therefore, the problem of induction is always external to simplicity degrees, where it is represented explicitly by the simplicity relation, due to homogeneity. That is a natural condition, if the idea is for $\mathfrak{S}$ to make the epistemological structure $\mathfrak{P}$ explicit in terms of $\preceq$. Furthermore, when $\mathcal{S}$ decides $\mathcal{Q}$, one can think of answers to $\mathcal{S}$ as *reasons* for answers to $\mathcal{Q}$, in the sense that answering $\mathcal{S}$ suffices, in light of incoming information, for answering $\mathcal{Q}$. Let λ be a convergent solution to $\mathfrak{S}$. Turn λ into a method for the original problem $\mathfrak{P}$ as follows.

$$\lambda_{\mathcal{Q}}(E) = \bigcup \phi_{\mathcal{Q}}(\lambda(E) \cap E).$$

Then:

Proposition 17. *Suppose that $\mathcal{S}$* decides *$\mathcal{Q}$ in $\mathfrak{I}^*$ and that λ solves $\mathfrak{S}$ in the limit. Then $\lambda_{\mathcal{Q}}$ solves $\mathfrak{P}$ in the limit.*

6.6 Simplicity Defined

In light of the preceding discussion, we propose that $\mathcal{S}$ is a *simplicity concept* for $\mathfrak{P}$ iff $\mathcal{S}$ is a stratification of $\mathfrak{I}^*$ that decides $\mathcal{Q}$. In other words:

(s1). $\mathcal{S}$ is locally closed in $\mathfrak{I}^*$;

(s2). $\mathcal{S}$ is homogeneous in $\mathfrak{I}^*$;

(s3). $\mathcal{S}$ decides $\mathcal{Q}$ in $\mathfrak{I}^*$.

If $\mathcal{S}$ is also σ-homogeneous, say that $\mathcal{S}$ is a σ-simplicity concept for $\mathfrak{P}$.

Proposition 18. *Conditions (s1)-(s3) are logically independent. Furthermore, σ-homogeneity does not follow from (s1)-(s3).*

If $\mathcal{S}$ is a simplicity concept for $\mathfrak{P}$, then $(\mathcal{S}, \preceq)$ is a partially ordered set, by proposition 4. Let $\mathsf{Simp}_{\mathfrak{I}}(\mathcal{S}, \mathcal{Q})$ abbreviate that $\mathfrak{S} = (W, \mathcal{I}, \mathcal{S})$ is a simplicity concept for problem $\mathfrak{P} = (W, \mathcal{I}, \mathcal{Q})$. Define $\mathsf{Simp}_{\mathfrak{I}}(\mathcal{S})$ to hold iff there exists question $\mathcal{Q}$ such that $\mathsf{Simp}_{\mathfrak{I}}(\mathcal{S}, \mathcal{Q})$, in which case $\mathcal{S}$ is a *simplicity concept* for $\mathfrak{I}$ and $\mathfrak{S} = (\mathfrak{I}, \mathcal{S})$ is a *simplicity problem.* Define σ-Simp similarly. As usual, drop the subscripts when no ambiguity arises. Then:

Proposition 19. Simp *is transitive.*

Furthermore:

Proposition 20. $\mathsf{Simp}(\mathcal{S})$ *iff* $\mathsf{Simp}(\mathcal{S}, \mathcal{S})$.

Hence:

Proposition 21. Simp *pre-orders the set of all simplicity concepts for* $\mathfrak{I}$.

Examples presented below easily refute both symmetry and anti-symmetry. Crucially, simplicity concepts are preserved under restriction by new information, so one never has to seek a new simplicity concept in light of new information, as one would hope, if simplicity is to be a diachronic guide to inquiry.

Proposition 22. *If* $\mathsf{Simp}_{\mathfrak{I}}(\mathcal{S}, \mathcal{Q})$ *then* $\mathsf{Simp}_{\mathfrak{I}\restriction E}(\mathcal{S} \restriction E, \mathcal{Q} \restriction E)$, *for all* $E \in \mathcal{I}$.

Finally:

Proposition 23. *Propositions 19-22 hold, as well, for* σ-Simp.

7 Examples

We begin with highly abstract problems, whose simplicity concepts are fairly obvious. Consider a very elementary inductive problem in which there is a number n concealed in a box, and λ is informed of successively greater lower bounds on n, with the guarantee that λ is informed, eventually, that the lower bound is n. The problem is inductive, because there is no stage at which the information available rules out that the truth is $m > n$. More generally, replace $(\mathbb{N}, \leq)$ with an *arbitrary*, countable, partially ordered set $(W, \leq)$ and let the information space $\mathfrak{I}^{\uparrow}(W, \leq)$ be the set of all *upward sets* $w^{\uparrow} = \{v \in W : w \leq v\}$, for $w \in W$. Then $(W, \leq) = (\mathfrak{I}^{\uparrow}(W, \leq), \mathcal{Q}_{\perp})$ is the *Alexandrov* problem $\mathfrak{S}^{\uparrow}(W, \leq)$ on $(W, \leq)$, since $(\mathfrak{I}^{\uparrow}(W, \leq))^*$ is known as the Alexandrov topology generated by $(W, \leq)$.

Alexandrov problem $\mathfrak{S}^{\uparrow}(W, \leq)$ wears its simplicity order on its sleeve, since $\{w\} \preceq \{v\}$ iff $w \leq v$. Furthermore, $\mathfrak{S}^{\uparrow}(W, \leq)$ is a simplicity problem. Condition (s1) is easily met, because $\{w\} = w^{\uparrow} \setminus \bigcup_{v>w} v^{\uparrow}$. Local closure and σ-homogeneity are satisfied trivially by $\mathcal{Q}_{\perp}$. Some coarsenings of $\mathcal{Q}_{\perp}$ may also count as simplicity concepts. Condition (s3) rules out the union of ordered degrees, and (s2) imposes a strong, global requirement on the fusion of un-ordered degrees—e.g., no finite chain can be fused with an infinite chain.

Suppose that $W = \mathbb{N}$, that the information space is $\mathfrak{I}^{\uparrow}(\mathbb{N}, \leq)$, and the question $\mathcal{Q}_{\mathsf{parity}}$ is whether n is even or odd. That question is not a simplicity concept for itself, since there is a simplicity cycle between the two answers, in violation of proposition 4. But $\mathcal{Q}_{\perp}$ is a simplicity concept for $\mathcal{Q}_{\mathsf{parity}}$ and, moreover, is the unique simplicity concept for $\mathcal{Q}_{\mathsf{parity}}$, because any coarsening either combines successive degrees, violating (s3), or skips an intermediate degree, violating (s1). So when the question is over-coarsened, the proposed theory of simplicity can sometimes isolate the essential, underlying simplicity concept that refines it.

We now shift attention to a slightly different way of generating a problem from a given, partial order $(W, \leq)$. Given partial order $(W, \leq)$, let information basis $\mathcal{I}^{\mathsf{co}-\downarrow}(W, \leq)$ consist of all non-empty complements of downward closed sets in $(W, \leq)$ and let $\mathfrak{I}^{\mathsf{co}-\downarrow}(W, \leq)$ and $\mathfrak{S}^{\mathsf{co}-\downarrow}(W, \leq)$ be the associated information space and problem with question $\mathcal{Q}_{\perp}$. Call such problems T_D problems, since $\mathfrak{I}$ is the weakest T_D topology on W with specialization order $\preceq$ (Aull and Thron, 1962). Such problems are still simplicity problems, and the order $(W, \leq)$ is still isomorphic to the simplicity concept $(\mathcal{Q}_{\perp}, \preceq)$.

The preceding problems abstract from time and wear their simplicity concepts on their sleeves. Less abstractly, assume that λ receives sequential, discrete inputs, and the question pertains to the sequence of inputs λ will receive in the future. Let I denote the countable set of all possible such inputs, and let W be some collection of infinite sequences of inputs. Let $w \upharpoonright t$ denote $(w_0, \ldots, w_{t-1})$. The information imparted by observation of finite sequence e of inputs is just $[e]$ = the set of all $w \in W$ that extend e. Let $\mathcal{I}_{\mathsf{seq}} = \{[w \upharpoonright n] : w \in W \text{ and } n \in \mathbb{N}\}$. Then $\mathcal{I}_{\mathsf{seq}}$ is an information basis. One can formulate many inductive problems by varying W and the question asked. For example, suppose that $I = \mathbb{N} \cup \{*\}$, where asterisk is a non-numeric input. Let W_{fin} contain all infinite sequences of inputs that have finite range. For finite set S of natural numbers, let $w \in C_S$ iff the set of numbers occurring in w is exactly S, and let the range question $\mathcal{Q}_{\mathsf{rng}}$ denote the set of all C_S such that S is a finite subset of $\mathbb{N}$. Let $w \in C_n$ iff $w \in C_S$ and $|S| = n$ and let the counting question $\mathcal{Q}_{\mathsf{cnt}}$ denote the set of all C_n such that $n \in \mathbb{N}$. Then we have the intuitive simplicity relations:

Proposition 24.

$$\begin{aligned} C_n \preceq_{\mathfrak{I}_{\mathsf{seq}}} C_{n'} \quad &\textit{iff} \quad n \leq n'; \\ C_S \preceq_{\mathfrak{I}_{\mathsf{seq}}} C_{S'} \quad &\textit{iff} \quad S \subseteq S'. \end{aligned}$$

The proof is intuitive—no matter which inputs one has seen so far, one could always see a new input later. Furthermore:

Proposition 25. *The questions $\mathcal{Q}_{\mathsf{cnt}}$ and $\mathcal{Q}_{\mathsf{rng}}$ are σ-homogeneous simplicity concepts for themselves and for one another, over $\mathfrak{I}_{\mathsf{seq}} \upharpoonright W_{\mathsf{fin}}$.*

The proposed account of simplicity also applies to many standard kinds of scientific problems involving continuous quantities, without prior translation into logic or infinite temporal sequences. To see how it all works, consider the paradigmatic problems of inferring polynomial degrees. Let W_{cts} be the set of all continuous functions with support on $[a, b]$. Let $\mathbf{a} = (a_1, \ldots, a_n)$ be a finite vector of real numbers such that $a_n \neq 0$. A polynomial function is a function expressible as $f_{\mathbf{a}} = \sum_{i \leq n} a_i x^i$. Let $W_{\mathsf{poly}} \subset W_{\mathsf{cts}}$ denote the set of all polynomial functions with support on the interval $[a, b]$. Then n is the *degree* of $f_{\mathbf{a}}$, and the *form* of $f_{\mathbf{a}}$ is the set S of non-zero positions in $\mathbf{a}$. Let D_n denote all $f \in W_{\mathsf{poly}}$ of degree n and let F_S denote the set of all $f \in W_{\mathsf{poly}}$ with form S. Then question $\mathcal{Q}_{\mathsf{deg}} = \{D_n : n \in \mathbb{N}\}$ asks what the degree of the true law is, and $\mathcal{Q}_{\mathsf{form}} = \{F_S : S \text{ is a finite subset of } \mathbb{N}\}$ asks what its form is—the questions of theory choice and model selection, respectively. Data are finite sets of open coordinate rectangles in the real plane, which are just non-empty cross products $(x_1, x_2) \times (y_1, y_2)$. Each such rectangle corresponds to an inexactly observed data point. Given finite set R of such rectangles, the information state E_R is the set of all $f \in W_{\mathsf{cts}}$ that have non-empty intersection with each element of R. Let $\mathcal{I}_{\mathsf{rec}}$ denote the set of all such information states. Then $\mathcal{I}_{\mathsf{rec}}$ is an information basis, for given $E, E' \in \mathcal{I}_{\mathsf{rec}}$, the conjunction $E \cap E'$ is the set of all continuous functions that pass through $R \cup R'$, for corresponding, finite sets of rectangles R, R'. Thus, $\mathfrak{I}_{\mathsf{cts}} = (W_{\mathsf{cts}}, \mathcal{I}_{\mathsf{rec}})$ is an information space. Similarly, $\mathfrak{I}_{\mathsf{poly}} = \mathfrak{I}_{\mathsf{cts}} \upharpoonright W_{\mathsf{poly}}$ is an information space. As one would expect, polynomial degree and form are simplicity concepts for themselves. Furthermore, they are simplicity concepts for one another.

Proposition 26. *The questions $\mathcal{Q}_{\mathsf{deg}}, \mathcal{Q}_{\mathsf{form}}$ are simplicity concepts for themselves, and for one another, over $\mathfrak{I}_{\mathsf{poly}}$. Furthermore, $\mathcal{Q}_{\mathsf{deg}}$ is σ-homogeneous.*

8 Simplicity as a Reason

The main issue is to provide an epistemic justification for favoring simpler theories over complex ones. A theory is simple, if it is entailed by a simplicity degree that is minimal in the simplicity order, given current information. It therefore facilitates the development that follows to focus on the problem $\mathfrak{S}$ of inferring simplicity degrees, rather than answers to the original problem $\mathfrak{P}$, which may, in general, cut across the simplicity degrees. Proposition 17 ensures that the simplicity problem is solvable in the limit if and only if the original problem is. Furthermore, one may think of $\lambda_{\mathfrak{P}}(E)$ as *basing* its conclusions about $\mathfrak{P}$ on conclusions about $\mathfrak{S}$ and on current information E. It is possible that $\lambda_{\mathfrak{P}}$ stabilizes to a conclusion in $\mathfrak{P}$ *before* reversing opinion concerning a simplicity degree in $\mathcal{S}$, which amounts to an inductive example of a Gettier case (1963). Finally, basing conclusions on simplicity does not beg the fundamental question at issue, which remains why simpler simplicity degrees should be favored over complex ones. Therefore, we henceforth focus attention on solving the simplicity problem $\mathfrak{S}$, rather than the given problem $\mathfrak{P}$.

9 Simplicity with Open Upward Sets

Thomas Kuhn's (1962) celebrated distinction between "normal" and "revolutionary" science is highly suggestive, but notoriously vague (Masterson, 1970). A similar, but clearer, distinction is whether simplicity considerations alone suffice to solve the problem, or some further "control structure" is necessary to guide convergence to the truth. In the former case, all Ockham agents solving the same simplicity problem must agree. In the latter case, rational disagreement is possible. Consider the following Ockham strategy for simplicity problem $\mathfrak{S}$:

$$\lambda_{\mathsf{Min}}(E) \;\; = \;\; \bigcup \mathsf{Min}_{\mathfrak{S}}(E).$$

Recall that simplicity problem $\mathfrak{S}$ has *open upward sets* iff $\bigcup C_{\succeq}$ is open, for all $C \in \mathcal{S}$. Then σ-homogeneity comes for free:

Proposition 27. *If $\mathcal{S}$ is homogeneous and has open upward sets, then $\mathcal{S}$ is σ-homogeneous.*

As an immediate corollary, every Alexandrov problem $\mathfrak{S}^{\uparrow}(W, \leq)$ is σ-homogeneous.
The same is true of the polynomial degree problem:

Proposition 28. *The question $\mathcal{Q}_{\mathsf{deg}}$ has open upward sets over $\mathfrak{I}_{\mathsf{poly}}$.*

It may come as a surprise that the polynomial form problem does not have open upward sets, so λ_{Min} cannot solve it in the limit.

Proposition 29. *The question $\mathcal{Q}_{\mathsf{form}}$ does not have open upward sets over $\mathfrak{I}_{\mathsf{poly}}$.*

Here is a simplified example that illustrates how that can happen. Let $\mathbb{N}_1$ and $\mathbb{N}_2$ be disjoint copies of the natural numbers and let $\leq$ be the standard order on both copies, with no ordering between the copies. Then λ_{Min} solves $\mathfrak{S}^{\uparrow}(\mathbb{N}_1 \cup \mathbb{N}_2, \leq)$ in the limit, because one of the chains is refuted eventually, after which there is a unique, minimal degree compatible with the information at each stage. But in $\mathfrak{S}^{\mathsf{co}-\downarrow}(\mathbb{N}_1 \cup \mathbb{N}_2, \leq)$, neither infinite chain is ever ruled out. Think of each infinite chain as a "paradigm", each step along which is a sharply testable "articulation"—e.g., the first chain could correspond to polynomial degree and the second could correspond to trigonometric polynomial degree. Neither paradigm is ever refuted by information, but each articulation of each paradigm is refutable. Hence, there are always two minimal possibilities in light of the information, and λ_{Min} fails to converge to the truth. The problem is bi-laterally symmetric with respect to the two chains. Therefore, a bias or control structure extrinsic to the problem's structure is required to break the symmetry and select one minimal possibility over the other. Many such strategies are possible. A Popperian (1959) strategy is to favor the paradigm whose current articulation is simpler (in terms of rank in the order on $\mathbb{N}$). A more Lakatosian (1970) strategy is to favor the paradigm that was re-articulated least recently, even though its simplicity rank may be higher. Both converge to the truth in a way that respects the (partial) simplicity order.

In the preceding examples, there at least exists an alternative simplicity concept that does have open upward sets—just coarsen the partial order by rank, in which case Popper's rank method is mandated by Ockham's razor. But some simplicity problems have no simplicity concept with open upward sets, so the general case is not always avoidable. For a very elementary example of that kind, the information space $\mathfrak{S}^{\mathsf{co}-\downarrow}(\mathbb{N}, =)$ reflects the situation in which each singleton is refutable but not verifiable, and there is no simplicity order over the singletons.

Proposition 30. *Question $\mathcal{Q}_{\perp}$ is its own simplicity concept over $\mathfrak{S}^{\mathsf{co}-\downarrow}(\mathbb{N}, =)$, but has no simplicity concept with open upward sets over $\mathfrak{S}^{\mathsf{co}-\downarrow}(\mathbb{N}, =)$.*

Another example is like the preceding one, except that there are now infinitely many paradigms (disjoint copies of $\mathbb{N}$). Neither Popper's rule nor Lakatos' rule suffices to solve problems with infinitely many paradigms (disjoint copies of $\mathbb{N}$), since, at every stage, infinitely many of the paradigms may have escaped re-articulation. We now present a method that works for such problems and, furthermore, for *every* simplicity problem $\mathfrak{S}$, even when the order types of the paradigms are so structurally diverse that complexity is "incommensurable" (non-rankable) over all of them. Since $\mathcal{S}$ is countable, choose arbitrary enumeration $\mathcal{S} = \{C_i : i \in \mathbb{N}\}$. Think of it as a symmetry-breaking, epistemic preference over answers to $\mathcal{S}$ that may have nothing to do with the structure of $\mathfrak{S}$. The obvious idea is to let $\lambda(E) =$ the first $C_i \in \mathsf{Min}_{\mathfrak{S}}(E)$, if there is one, and the vacuous conclusion $\bigcup \phi_{\mathcal{S}}(E)$ otherwise. However, that procedure allows for the possibility that some C_j prior to C_j that is not simplest in light of E_0 *becomes* simplest in light of further information E_1, in which case the proposed method would drop the truth after having found it, only to return to it later, resulting in a cycle of conclusions. That is also a violation of *rational monotonicity*, a familiar requirement on rational belief revision (Alchourrón et al., 1985; Pearl, 1990; Lehmann and Magidor, 1992), which holds of λ in $\mathfrak{S}$ iff, $F \subset E$ and $\lambda(E) \cap F \neq \varnothing$ implies $\lambda(F) \subseteq \lambda(E)$, for all $E, F \in \mathcal{I}(w)$. The trick for learning in a rationally monotone way that respects simplicity is to withhold conclusion C_i until information rules out all prior answers that *might* become minimal in light of future information, because they are not more complex than C_i. That is possible whenever $\mathcal{S}$ stratifies $\mathfrak{I}$. Let $\mathsf{Ordmin}_{\mathfrak{S}}(E)$ denote the set of all $C_i \in \mathsf{Min}_{\mathfrak{S}}(E)$ such that $C_i \prec C_j$, for all $j < i$ such that $C_j \in \phi_{\mathcal{S}}(E)$. In other words, current experience rules out every competitor prior to C_i that is not more complex than C_i. Note that $\mathsf{Ordmin}_{\mathfrak{S}}(E)$ is either empty, or contains a unique element. So it makes sense to define:

$$\lambda_{\mathsf{Ordmin}}(E) \quad = \quad \begin{cases} \text{the unique } C_i \in \mathsf{Ordmin}_{\mathfrak{S}}(E) & \text{if } \mathsf{Ordmin}_{\mathfrak{S}}(E) \neq \varnothing; \\ \mathsf{rel}_{\mathcal{S}}(E) & \text{otherwise.} \end{cases}$$

Method $\lambda_{\mathsf{Ordmin}}$ implements a version of Ockham's razor relative to the given enumeration, but the enumeration amounts to an extraneous, supplementary bias subordinate to simplicity. So $\lambda_{\mathsf{Ordmin}}$ interweaves "revolutionary" and "normal" scientific decisions in a natural way. Method $\lambda_{\mathsf{Ordmin}}$ works even if the simplicity order is ill-founded, dense, or has chains of order type greater than ω.

Proposition 31. *If $\mathcal{S}$ stratifies $\mathfrak{I}$, then (i) $\lambda_{\mathsf{Ordmin}}$ is rationally monotonic in $\mathfrak{S}$ and (ii) $\lambda_{\mathsf{Ordmin}}$ solves $\mathfrak{S}$ in the limit, no matter how $\mathcal{S}$ is enumerated.*

10 Ockham's Razors

We begin with some natural conditions on λ that are more basic than Ockham's razor. First, we assume that λ is *deductively cogent* in the sense that $\lambda(E)$ is consistent with and entails $\mathsf{rel}_{\mathcal{S}}(E)$:

$$\varnothing \;\neq\; (\phi_{\mathcal{S}} \circ \lambda)(E) \;\subseteq\; \phi_{\mathcal{S}}(E).$$

Furthermore, say that $H \in \mathcal{S}^*$ is σ-homogeneous *given* $E \in \mathcal{I}$ iff $\{w\} \preceq \lambda(E)$ implies $\phi_{\mathcal{S}}(w) \preceq H$, for all $w \in \mathsf{rel}_{\mathcal{S}}(E)$. Then λ is σ-homogeneous iff $\lambda(E)$ is σ-homogeneous given E, for each $E \in \mathcal{I}$. Every method for $\mathfrak{S}$ is σ-homogeneous, if $\mathfrak{S}$ is, but sufficient care guarantees σ-homogeneity over arbitrary simplicity concepts. For example:

Proposition 32. *If $\mathfrak{S}$ is a simplicity problem, then $\lambda_{\mathsf{Ordmin}}$ is deductively cogent and σ-homogeneous in $\mathfrak{I}^*$.*

Henceforth, we consider only methods that are deductively cogent and σ-homogeneous.

In light of the preceding restrictions, Ockham's razor serves to constrain the choice among non-empty, σ-homogeneous subsets of $\phi_{\mathcal{S}}(E)$. Ockham's razor is a preference for simpler simplicity degrees. Preference—epistemic or otherwise—should have consequences for choices. If $C \in \phi_{\mathcal{S}}(E)$ is logically compatible with $\lambda(E)$, that is an epistemic compliment by λ to C. If $D \in \phi_{\mathcal{S}}(E)$ and $D \prec C$, then λ should pay the same compliment to D as well, if λ has an epistemic preference based on simplicity. Say that output disjunction $H \in \mathcal{S}^*$ satisfies Ockham's *weak, vertical* razor in response to E iff:

$$\phi_{\mathcal{S}}(E) \cap \{C_{\preceq} : C \in \phi_{\mathcal{S}}(H)\} \quad \subseteq \quad \phi_{\mathcal{S}}(H).$$

Note, however, that a compliment is also paid to $H = \lambda(E)$, itself, which may be disjunctive, so the compliment should be extended to each C as simple as H that is compatible with E:

$$\phi_{\mathcal{S}}(E) \cap H_{\preceq} \quad \subseteq \quad \phi_{\mathcal{S}}(H).$$

That is the full version of Ockham's vertical razor. Say that λ satisfies the vertical razor iff $\lambda(E)$ does, with respect to each $E \in \mathcal{I}$. Since $\phi_{\mathcal{S}}(H) \subseteq H_{\preceq}$, by the definition of $H_{\preceq}$, the vertical razor is equivalent to:

$$\phi_{\mathcal{S}}(E) \cap H_{\preceq} \quad = \quad \phi_{\mathcal{S}}(H).$$

For the same reason, the weak, vertical razor follows from the full one. The converse implication fails—for example, in problem $(\mathcal{I}_{\mathsf{seq}}, \mathcal{Q}_{\perp}) \upharpoonright W_{\mathsf{fin}}$, the weak version countenances the apparently non-Ockham conclusion $H_{100} =$"there are exactly 100 effects" when no effects have been seen, but the full version plausibly rules it out, because the singleton containing the effect-free world is simpler than H_{100}.

Karl Popper (1959) linked simplicity to falsifiability. In fact, Ockham's vertical razor is equivalent to the requirement that one's conclusions be falsifiable, given the σ-homogeneity requirement on learning methods:

Proposition 33. *Let $H \in \mathcal{S}^*$ be σ-homogeneous in $\mathfrak{I}^*$. Then H is closed (refutable) in $\mathfrak{I}^* \upharpoonright E$ iff H is vertical Ockham in response to E.*

Ockham's vertical razor is satisfied by plumping for $\bigcup \mathcal{H}$ such that $\mathcal{H} \subseteq \mathsf{Min}_{\mathfrak{S}}(E)$ is finite, without providing any *horizontal* advice constraining which disjuncts to include in $\mathcal{H}$. Simplicity is not the only bias in science—e.g., reliance only on known causal mechanisms—and those other biases may help to select a unique, simplest answer. But, in other cases, such as data mining or causal discovery over large variable sets, plausibility considerations can be lacking, and then it sounds questionable to choose among the simplest possibilities. Then, it is natural to insist that every element of $\mathsf{Min}_{\mathfrak{S}}(E)$ be taken seriously:

$$\mathsf{Min}_{\mathfrak{S}}(E) \quad \subseteq \quad (\phi_{\mathcal{S}} \circ \lambda)(E).$$

That principle is vacuous when $\mathsf{Min}_{\mathfrak{S}}(E)$ is empty, due to infinite, descending chains. Therefore, Ockham's *horizontal* razor demands, more generally, that $(\phi_{\mathcal{S}} \circ \lambda)(E)$ be *co-initial* in $(\phi_{\mathcal{S}}(E), \preceq)$, in the sense that, for each $C \in \phi_{\mathcal{S}}(E)$, there exists $D \in \lambda(E)$ such that $D \preceq C$.

Neither razor implies the other in general, and they work together as a natural team—the horizontal razor ensures breadth of coverage over simplicity anti-chains and the vertical razor fills in the gaps along chains. The vertical razor is, at least, no impediment to solving arbitrary simplicity problem $\mathfrak{S}$:

Proposition 34. *If $\mathcal{S}$ stratifies $\mathfrak{I}$, then $\lambda_{\mathsf{Ordmin}}$ is a vertical Ockham solution to $\mathfrak{S}$.*

We now present a strategy that satisfies both razors and that solves every simplicity problem with open upward sets. A very conservative approach is to return C if E verifies open set $\bigcup C_{\succeq}$, for $C \in \mathcal{S}$, and to return the vacuous, relevant proposition $\mathsf{rel}_{\mathcal{S}}(E)$ otherwise. We present a more interesting and aggressive method, that agrees with λ_{Min} when $\mathcal{S}$ has a well-founded simplicity order. Let $\mathsf{Nmb}_{\mathfrak{S}}(E)$ (for "no minimum below") denote the set of all $C \in \phi_{\mathcal{S}}(E)$ such that no $D \in \mathsf{Min}_{\mathfrak{S}}(E)$ satisfies $D \prec C$. Define the Nmb method:

$$\lambda_{\mathsf{Nmb}}(E) \quad = \quad \bigcup \mathsf{Nmb}_{\mathfrak{S}}(E).$$

Proposition 35. *Suppose that $\mathfrak{S}$ is a simplicity problem with open upward sets. Then:*

1. *λ_{Nmb} is deductively cogent and σ-homogeneous;*
2. *λ_{Nmb} solves $\mathfrak{S}$ in the limit;*
3. *$\lambda_{\mathsf{Nmb}}(E)$ satisfies Ockham's horizontal and vertical razors in $\mathfrak{S}$;*
4. *$\lambda_{\mathsf{Nmb}}(E)$ is closed in $\mathfrak{I}^* \restriction E$.*
5. *$\lambda_{\mathsf{Nmb}} = \lambda_{\mathsf{Min}}$, if $(\phi_{\mathcal{S}}(E), \preceq)$ is well-founded.*

11 Optimal Inductive Strategies

The aim is to justify Ockham's razor as the best strategy for answering inductive questions. Inductive inquiry is non-monotonic, so course reversals are inevitable along the way. But one can, at least, insist upon the most direct route possible, where directness is explicated in terms of epistemic reversals and cycles. Sequence $a = (A_0, \ldots, A_n)$ of propositions in $\mathcal{S}^*$ is a *reversal sequence* iff $A_i \cap A_{i+1} = \varnothing$, for each i from 1 to $n-1$. A *cyclic* sequence is a reversal sequence with the additional property that the terminal entry A_n logically entails the initial entry A_0.

Our argument is based on worst-case considerations. The standard objection to such reasoning is that it is too pessimistic, but that gets the situation backwards in this case, because we obtain a strong, non-circular vindication of Ockham's razor, in contrast to the circular arguments that inevitably result from expected-case reasoning based on simplicity-biased prior probabilities.

The usual approach to worst-case reasoning is to tally worst-case losses incurred by each method in each world, to compute the supremum over all worlds in a given simplicity degree and then to compare those worst-case bounds for alternative methods. That works for very elementary examples like $\mathcal{Q}_{\mathsf{cnt}}$, in which all maximal simplicity chains are ordered as the natural numbers. But when there are infinite descending chains, that approach would result in infinite worst-case loss in each degree occurring in the chain, trivializing all method comparisons. Therefore, we adopt the more refined approach of comparing output sequences directly, without assigning intervening, numerical losses.[9] Define $(A_0, \ldots, A_n) \leq (B_0, \ldots, B_m)$ to hold iff $m = n$ and $\varnothing \neq A_i \subseteq B_i$, for each $i \leq n$. Then, if b is a reversal sequence, a is also a reversal sequence that reverses at least as sharply. Moreover, if b is a cycle sequence, then a carries out the cycle at least as sharply.

An *information history* in $\mathfrak{I}$ is a finite, non-empty, downward-nested sequence $e = (E_0 \supset \ldots \supset E_n)$ of elements of $\mathcal{I}$. Let $\lambda(e) = (\lambda(E_0), \ldots, \lambda(E_n))$. Let $E \in \mathcal{I}$ and $C \in \mathcal{S}$. Let $\mathsf{Hst}_{\mathfrak{I}}$ denote the set of all information histories $(E_0 \supset \ldots \supset E_n)$, and let $\mathsf{Hst}_{\mathfrak{I}}(C, E)$ denote the set of all information histories starting with an information state included in E and ending with an information state compatible with C.

Let λ, λ' be methods for $\mathfrak{S}$ and let $C \in \mathcal{S}$. Define the the worst-case comparison:

> $\lambda' \leq^{\mathsf{rev}}_{(\mathfrak{S},C,E)} \lambda$ iff for every $e \in \mathsf{Hst}_{\mathfrak{I}}(C, E)$ such that $\lambda(e)$ is a reversal sequence, there exists $e' \in \mathsf{Hst}_{\mathfrak{I}}(C, E)$ such that $\lambda'(e') \leq \lambda(e)$.

Then define $\lambda' \leq^{\mathsf{rev}}_{\mathfrak{S},E} \lambda$ to hold iff $\lambda' \leq^{\mathsf{rev}}_{(\mathfrak{S},C,E)} \lambda$ holds in each $C \in \mathcal{S}$. The relation $\leq^{\mathsf{rev}}_{(\mathfrak{S},C,E)}$ is a pre-order. The corresponding, strict order holds if the pre-order holds one way and not the other and the corresponding equivalence relation holds if the pre-order holds both ways.

Say that λ is *reversal optimal* given $E \in \mathcal{I}$ in $\mathfrak{S}$ iff (i) λ solves $\mathfrak{S}$ in the limit and (ii) $\lambda' \leq^{\mathsf{rev}}_{\mathfrak{S},E} \lambda$, for every λ' such that λ' solves $\mathfrak{S}$ in the limit. Say that λ is *reversal sub-optimal* given E in $\mathfrak{S}$ iff (i) λ fails to solve $\mathfrak{S}$ in the limit, or (ii) $\lambda <^{\mathsf{rev}}_{\mathfrak{S},E} \lambda'$, for some λ' such that λ' solves $\mathfrak{S}$ in the limit. Finally, λ is reversal optimal in $\mathfrak{S}$ iff λ is reversal optimal in $\mathfrak{S}$ in each $E \in \mathcal{I}$, and λ is reversal sub-optimal in $\mathcal{S}$ iff λ is reversal sub-optimal in some $E \in \mathcal{I}$. Repeat the entire chain of definitions, substituting cycles for reversals.

Note that $\lambda' <^{\mathsf{rev}}_{(\mathfrak{S},C,E)} \lambda$ can be described by saying that λ weakly dominates λ' in terms of worst-case bounds over the simplicity degres in $\mathcal{S}$. Thus, sub-optimality ranges from weak dominance ($\mathcal{S} = \mathcal{Q}_{\perp}$) to being non-minimax ($\mathcal{S} = \mathcal{Q}_{\top}$). Therefore, one advantage of a more refined simplicity concept $\mathcal{S}$ (e.g., a partial order rather than a total order) is that the comparison more closely approximates weak dominance in the finer concept. But

[9]The same idea was applied to retractions and retraction times in (Kelly, 2007b).

there is a limit to refinement, due to the assumption that questions are countable, and to the preference for eventual, unique minima, when possible.

12 A Strategic Justification of Ockham's Razor

The aim is to provide a strategic justification of Ockham's razor, in terms of directness of convergence to the truth, as reflected by course-reversals and cycles. When $\mathcal{S}$ has open upward sets, the vertical and horizontal Ockham method λ_{Nmb} is both cycle-optimal and reversal-optimal:

Proposition 36. *Let $\mathfrak{S}$ be a simplicity problem with open upward sets. Then λ_{Nmb} is a reversal-optimal and cycle-optimal solution to $\mathfrak{S}$.*

Therefore, cycle and reversal optimality are achievable, for simplicity problems with open upward sets. Moreover, conformity with Ockham's horizontal razor is *equivalent* to reversal optimality:

Proposition 37. *Let $\mathfrak{S}$ be a simplicity problem with open upward sets that is solved in the limit by λ. Then the following are equivalent:*

1. *λ is reversal optimal in $\mathfrak{S}$;*
2. *λ is not reversal sub-optimal in $\mathfrak{S}$;*
3. *λ satisfies Ockham's horizontal razor in $\mathfrak{S}$.*

In the general case, the method $\lambda_{\mathsf{Ordmin}}$ is cycle-free and, hence, cycle-optimal, in arbitrary simplicity problem $\mathfrak{S}$.

Proposition 38. *Let $\mathfrak{S}$ be a simplicity problem. Then $\lambda_{\mathsf{Ordmin}}$ is cycle-optimal for $\mathfrak{S}$.*

Thus, cycle-optimality is achievable. Furthermore, every method that violates Ockham's vertical razor is sub-optimal with respect to cycles, so one ought not to violate it.

Proposition 39. *Let $\mathfrak{S}$ be a simplicity problem. Suppose that method λ for $\mathfrak{S}$ violates Ockham's vertical razor for $\mathfrak{S}$ at E. Then λ is cycle-sub-optimal in $\mathfrak{S}$, as witnessed by $\lambda_{\mathsf{Ordmin}}$.*

One would like to say, in addition, that Ockham's vertical razor is also sufficient, in some sense, for cycle optimality. However, nothing prevents the learner from gratuitously cycling between vertical Ockham outputs when $\mathfrak{S}$ lacks open upward sets—some further, diachronic restriction is required, such as rational monotonicity. But less than that is required to vindicate Ockham's razor—it is enough that Ockham's razor is both necessary and sufficient for being *able* to avoid cycles in the future, from that point onward. And that is true—it is always possible to continue by means of $\lambda_{\mathsf{Ordmin}}$, after H is refuted.

Proposition 40. *Suppose $\mathfrak{S}$ is a simplicity problem and λ solves $\mathfrak{S}$ in the limit. Let $E \in \mathcal{I}$, and suppose that $\lambda(E)$ is σ-homogeneous for E. Then the following are equivalent:*

1. *$\lambda(E)$ satisfies Ockham's vertical razor with respect to E.*
2. *There exists a method λ' for $\mathfrak{S} \restriction E$ such that:*
   (a) *$\lambda'(E) = \lambda(E)$;*
   (b) *λ' solves $\mathfrak{S} \restriction E$ in the limit;*
   (c) *λ' is cycle-optimal in $\mathfrak{S} \restriction E$.*

Regarding Peter and Paul, violations of Ockham's horizontal razor that are also violations of the vertical razor incur extra reversals at each degree *higher* than the omitted simple disjunct. Violations of the horizontal razor that are not also violations of the vertical razor can result in extra reversals only in the simple, omitted degree. But worst-case reversals are not improved for the favored, simple degree, since zero reversals are already achieved there by λ_{Nmb}. Of course, there has to be some advantage for the favored, simple degree, and that is an improvement in time to convergence to the true answer, at the expense of alternative, simple degrees.

We close with a back-and-forth theorem analogous to proposition 40.

Proposition 41. *Suppose $\mathfrak{S}$ is a simplicity problem with open upward sets, and λ solves $\mathfrak{S}$ in the limit. Let $E \in \mathcal{I}$. Then the following are equivalent:*

1. *$\lambda(E)$ satisfies Ockham's vertical and horizontal razors with respect to E.*
2. *There exists a method λ' for $\mathfrak{S} \restriction E$ such that:*
   (a) *$\lambda'(E) = \lambda(E)$;*
   (b) *λ' solves $\mathfrak{S} \restriction E$ in the limit;*
   (c) *λ' is both cycle and reversal optimal in $\mathfrak{S} \restriction E$.*

That concludes our argument for the uniquely optimal truth-conduciveness of Ockham's razor.

13 Simplicity and Objectivity

The justification of Ockham's razor is one traditional puzzle concerning simplicity. Another is whether simplicity is a mere, subjective consideration. Principles (s1-s3) are weak, by design, since our focus was on justifying Ockham's razor relative to a simplicity concept, for which those principles suffice. But, of course, it is interesting to consider further principles that plausibly narrow down the range of possible simplicity concepts for a given problem. We present the following principles more as rules of thumb, to be applied with judgment, rather than as hard, necessary conditions.

To begin with, the many virtues of σ-homogeneity have already been discussed:

s4. It is preferable that $\mathcal{S}$ be σ-homogeneous for $\mathfrak{I}^*$.

Furthermore, it is very natural to prefer simplicity concepts with open upward sets, when the choice arises—e.g., polynomial degree over polynomial form.

s5. It is preferable that $\mathcal{S}$ have open upward sets in $\mathfrak{I}^*$.

Pathological simplicity concepts remain. For a spectacular example, $\mathcal{Q}_\perp$, whose simplicity order is trivially flat, is a simplicity concept for $\mathcal{Q}_{\mathsf{cnt}}$ in $\mathfrak{I}_{\mathsf{seq}} \upharpoonright W_{\mathsf{fin}}$, in spite of (s1-s5).[10]

Clearly, $\mathcal{Q}_{\mathsf{cnt}}$ is a better simplicity concept for $\mathcal{Q}_{\mathsf{cnt}}$ than $\mathcal{Q}_\perp$ is. One diagnosis is that the simplicity order over $\mathcal{Q}_\perp$ is trivially flat—it is plainly more apt to represent, explicitly, all of the simplicity relations that already hold among answers to the original question. That idea is captured in an elegant and general way by the relation of simulation, a central concept in modal logic (van Benthem, 1983) and automata theory (Milner, 1971). Let $\mathcal{Q}$, $\mathcal{Q}'$ be questions over $\mathfrak{I}^*$. Say that $\mathcal{Q}'$ is an *upward simulation* of $\mathcal{Q}$ iff for all $C \in \mathcal{S}$, if $\phi_{\mathcal{S}}(w) \preceq C$, then there exists $D \in \mathcal{S}'$ such that $D \cap C \neq \varnothing$ and $\phi_{\mathcal{S}'}(w) \preceq D$. The corresponding preference principle, which easily rules out $\mathcal{Q}_\perp$ as a simplicity concept for $\mathcal{Q}_{\mathsf{cnt}}$, is:

s6. It is preferable that $\mathcal{S}$ be an upward simulation of $\mathcal{Q}$ in $\mathfrak{I}^*$;

Principles (s1-s6) rule out trivial simplicity concepts like $\mathcal{Q}_\perp$ and $\mathcal{Q}_\top$, but they allow for a range of disagreeing, non-trivial variants. Consider the problem $\mathfrak{P}_{\mathsf{cnt}} = (\mathfrak{I}_{\mathsf{seq}} \upharpoonright W_{\mathsf{fin}}, \mathcal{Q}_{\mathsf{cnt}})$, for which $\mathcal{Q}_{\mathsf{cnt}}$ is a natural simplicity concept. Now define the "grue-like" (Goodman, 1983) bijective map $\psi_{k,k'} : W_{\mathsf{fin}} \to W_{\mathsf{fin}}$ as follows. Inspect the first $k + k'$ entries in w. If they are all *, then pass over the first k' asterisks and replace the subsequent k' asterisks with $0, 1, \ldots, k-1$. If w starts with k' asterisks followed by $0, 1, \ldots, k-1$, then replace the first $k + k'$ entries with *. Otherwise, leave w unaltered. Then $\psi_{k,k'}(\mathcal{Q}_{\mathsf{cnt}})$ satisfies (s1-s6) for $\mathcal{Q}_{\mathsf{cnt}}$ in $\mathfrak{I}_{\mathsf{seq}} \upharpoonright W_{\mathsf{fin}}$. However, $\mathcal{Q}_{\mathsf{cnt}}$ and $\psi_{k,k'}(\mathcal{Q}_{\mathsf{cnt}})$ are not the same question described differently—they are distinct sets of propositions. Since the original question is part of the problem, and simplicity is problem-relative, there exist structural conditions that rule out the gerrymandering. One plausible and direct prohibition against the simplicity concept $\mathcal{S}$ gerrymandering the given question $\mathcal{Q}$ is as follows, where $\mathcal{Q} \wedge \mathcal{Q}'$ denotes the greatest common refinement of $\mathcal{Q}$ and $\mathcal{Q}'$. Then, due to its cross-cutting character, the gerrymandered simplicity concept is ruled out by:

s7. It is preferable that $\mathcal{S} \wedge \mathcal{Q} \subseteq \mathcal{S} \cup \mathcal{Q}$.

[10]The dual concept $\mathcal{Q}_\top$ is already ruled out by s3.

A stronger principle is to leave well enough alone:

s8. It is preferable that the simplicity concept modify the original question as little as possible.

Principles (s3, s6, s7, s8) rule out potential simplicity concepts based on asymmetries in the original question. But the question can also also be symmetrical, as in the problem $(\mathfrak{I}_{\mathsf{seq}} \upharpoonright W_{\mathsf{fin}}, \mathcal{Q}_{\perp})$. In that case, gerrymandered variants are no longer ruled out. One might hope that $\mathcal{Q}_{\mathsf{cnt}}$ is right, and that its gerrymandered variants are still wrong, for reasons missed by (s1-s7). However, that judgment cannot be grounded in the empirical structure of the problem as presented and, hence, could not have anything to do with arriving at the truth as directly as possible by purely empirical means. A *symmetry* of problem $\mathfrak{P}$ is a self-homemorphism of $\mathfrak{I}^*$ that is also an automorphism of $\mathcal{Q}$. Each symmetry of a problem preserves the structure of the problem perfectly. Since every bijection preserves $\mathcal{Q}_{\perp}$, each homeomorphism of $\mathfrak{I}^*$ is a symmetry of $(\mathfrak{I}^*, \mathcal{Q}_{\perp})$, so there is no structural criterion for ruling out $\psi_{k,k'}(\mathcal{Q}_{\mathsf{cnt}})$ as a simplicity concept for $\mathcal{Q}_{\perp}$ in $\mathfrak{I}_{\mathsf{seq}} \upharpoonright W_{\mathsf{fin}}$. It is still open to the metaphysically inclined to add hidden structure to empirical problems—e.g., there might be some sort of brute, metaphysical dependence or correlation between our biases and the nature of the world we live in (e.g., providence before Darwin and evolution thereafter). By ad hoc adjustment of the hidden dependence, any feature of empirical hypotheses could be a guide to truth, but the irony of justifying Ockham's razor with hidden, untestable forces or tendencies is palpable.

Although the intuition that $\mathcal{Q}_{\mathsf{cnt}}$ is the "right" concept cannot be grounded in problem $(\mathfrak{I}_{\mathsf{seq}} \upharpoonright W_{\mathsf{fin}},\ \mathcal{Q}_{\perp})$, it is plausibly explained as an ambiguity in the problem to be solved. A natural way to pose the question $\mathcal{Q}_{\perp}$ over $\mathfrak{I}_{\mathsf{seq}} \upharpoonright W_{\mathsf{fin}}$ is "how many effects will occur, and when does each one occur?". That could be viewed as an expression of the question $\mathcal{Q}_{\perp}$, but it could also be viewed as a sequence of two questions, ordered by presupposition (an occurrence time presupposes an occurence). The first question corresponds to $\mathcal{Q}_{\mathsf{cnt}}$, which breaks symmetry in $\mathcal{Q}_{\perp}$. Then, the following preference is very plausible:

s9. In light of a set of questions ordered by presupposition, it is preferable to satisfy (s4-s8) with respect to more deeply presupposed questions in the sequence.

By that principle, (s6) continues to rule out $\mathcal{Q}_{\perp}$ as a simplicity concept for $\mathcal{Q}_{\mathsf{cnt}}$ and (s7) continues to rule out gerrymandered simplicity concepts for $\mathcal{Q}_{\mathsf{cnt}}$. Of course, the presupposed question could be a gerrymandered question, in which case that would be the appropriate simplicity concept.

We have considered problems in which both the information space and the question are symmetrical, so one's choice among gerrymandered variants is arbitrary, as well as problems in which the question breaks symmetries in the information space. It is also possible that the information space, itself, is asymmetrical. For example, let the information space be $C_{\leq 1} = C_0 \cup C_1$, and let the question be $\mathcal{Q}_{\perp}$, which is perfectly symmetrical.

Note, however, that each $w \in C_1$ is an interior point of $\phi_{\mathcal{Q}_\perp}(w) = \{w\}$, whereas the world $v \in C_0$ is not interior to its answer $\phi_{\mathcal{Q}_\perp}(w) = \{w\}$. Here is one way to extract simplicity from that asymmetry, even when the question is symmetrical. Let $\mathfrak{P}$ be an arbitrary, empirical problem. Let the boundary of $\mathcal{Q}$ be the set of all boundary points of answers to $\mathcal{Q}$. Define the *Cantor-Bendixson* problem[11] $\mathfrak{S}_{\mathfrak{P}}$ generated by $\mathfrak{P}$ as follows. First, define downward-nested subsets of W by transfinite recursion:

$$\begin{aligned} X_0 &= W; \\ X_{\alpha+1} &= \mathsf{bdry}_{\mathfrak{I}^* \upharpoonright X_\alpha}(\mathcal{Q}); \\ X_\gamma &= \bigcap_{\alpha<\gamma} X_\alpha, \text{ for limit ordinal } \gamma. \end{aligned}$$

Since X_α is nested downward, there exists a least ordinal α^* such that $X_\beta = X_{\beta+1}$. Define:

$$\begin{aligned} C_\alpha &= X_\alpha \setminus X_{\alpha+1}, \text{ for } \alpha < \alpha^*; \\ \mathcal{S}_{\mathfrak{P}} &= \{C_\alpha : \alpha \leq \alpha^*\}; \\ W_{\mathfrak{P}} &= W \setminus X_{\alpha^*}. \end{aligned}$$

Let $\mathfrak{I}_{\mathfrak{P}} = \mathfrak{I} \upharpoonright W_{\mathfrak{P}}$, and $\mathfrak{S}_{\mathfrak{P}} = (\mathfrak{I}_{\mathfrak{P}}, \mathcal{S}_{\mathfrak{P}})$. There is no guarantee that $W_{\mathfrak{P}} = W$. When that happens to be the case, then say that problem $\mathfrak{P}$ is *scattered*. Then:

Proposition 42. *If $W_{\mathfrak{P}} \neq \varnothing$, then $\mathcal{S}_{\mathfrak{P}}$ satisfies (s1-s6) for $\mathfrak{P} \upharpoonright W_{\mathfrak{P}}$, such that the order type of $(\mathcal{S}_{\mathfrak{P}}, \succeq)$ is α^*. Thus, $\mathfrak{S}_{\mathfrak{P}}$ is a simplicity concept for $\mathfrak{P}$, if $\mathfrak{P}$ is scattered.*

Cantor-Bendixson rank was explored by (Freivalds and Smith, 1993) (Martin et al., 2006) as a measure of empirical problem complexity, rather than as a theory of empirical simplicity. In light of the proposition, that work can also be viewed as studying a special case of empirical simplicity. One considerable advantage of the approach is that it constructs a unique, total simplicity ranking if the problem is scattered:

s10. The simplicity concept $\mathcal{S}_{\mathfrak{P}}$ is preferable, if $\mathfrak{P}$ is scattered.

Alas, the paradigmatic examples $\mathcal{Q}_{\mathsf{cnt}}$ and $\mathcal{Q}_{\mathsf{deg}}$ are not scattered, so (s10) does not apply—the proposed, non-constructive axioms for simplicity are far more generally applicable. Moreover, the Cantor-Bendixson solution to $(C_0 \cup C_1,\ \mathcal{Q}_\perp)$ is not always intuitively correct. Suppose that question is expressed as: "will an effect be observed at stage 1 and, if not, will one be observed later, and if so, when?" Now, the first question is empirically decidable, by waiting until stage 1, so neither answer is more complex than the other. After that is settled, an effect in the future is more complex than no effect, for the usual reasons. Principle (s9) gives the right result, and the Cantor-Bendixson procedure applied directly to question $\mathcal{Q}_\perp$ does not. Principle (s9) also gives the right answer—in agreement with the Cantor-Bendixson procedure—when the question is posed as "will there be an effect and, if so, when?".

[11] The standard Cantor-Bendixson construction is the special case for problems with the trivial question $\mathcal{Q}_\top = \{W\}$.

14 Conclusion and Projects

We define empirical problems very generally, in terms of an information basis and a question. We provide three axiomatic conditions (s1)-(s3) and a collection of further principles that recover simplicity concepts from empirical problem structure. The theory is general enough to allow for problems with ill-founded and dense simplicity orders. Some simplicity concepts have open upward sets. In the general case, additional, subjective biases are required to solve the problem. The distinction is reminiscent of Kuhn's celebrated distinction between "normal" and "revolutionary" science, but it is grounded firmly in problem structure. Relative to a simplicity concept, we identify two, independent components of Ockham's razor, the vertical razor and the horizontal razor. We define optimally direct convergence to the truth, in terms of minimization of cycles and reversals of opinion, prior to convergence to the true answer to the question, from information provided by the information basis. We establish that Ockham's vertical razor is necessary for cycle-optimalitiy and Ockham's horizontal razor is necessary for reversal-optimality, if it is compatible with convergence to the truth at all. Cycle-optimal performance is feasible in all problems that are their own simplicity concepts, and reversal-optimal performance is feasible in all such problems.

We close, customarily, with some open questions and projects. We did not settle whether every problem solvable in the limit has a simplicity concept in the sense of (s1-s3).

It would be interesting to compare the proposed approach systematically to competing proposals, such as Kolmogorov complexity, VC dimension, and Bayes factors.

Furthermore, the examples presented in this paper just scratch the surface. The ideas developed in our handling of the polynomial case can be extended naturally to other classes of functions, to functions of arbitrarily high arity, and to functionals (i.e., to systems of differential equations).

It remains to extend the entire development to stochastic models and theories, but the preceding results come close. The simplicity theory already applies to statistical models (understood in the usual way as sets of sampling distributions over a fixed collection of random variables)—choose the information basis $\mathcal{I}$ to be metric balls with respect to a standard metric on probability measures, such as total variation metric. It is conjectured that the cycle-minimization results will also carry over, with convergence in probability replacing convergence and cycles in probability replacing cycles (i.e., probably producing an answer, then probably producing a mutually incompatible answer, and finally, probably producing the first answer again).[12] It is impossible to eliminate cycles in chance entirely, so the results will be indexed by a parameter $\alpha > 0$ that resembles significance, but that really reflects cycle-tolerance. The aim is to arrive at a comprehensive, frequentist

[12] Reversals pose a greater challenge for the optimality argument, because the robbing Peter to pay Paul problem re-emerges to a degree, due to overlapping tails of probability distributions. That issue is discussed with respect to retractions in Kelly and Glymour (2004).

framework for inductive inference of stochastic theories and models in which the usual practice of favoring the sharp null hypothesis over the complex alternative is literally an application of Ockham's vertical statistical razor to a binary question. Unlike the standard approach to frequentist model selection, which views models as instruments for prediction relative to the sampling distribution, the projected justification of Ockham's statistical razor would apply to counterfactual inductive inferences concerning modifications to the sampling distribution, such as causal discovery from non-experimental data (Kelly and Mayo-Wilson, 2010).

15 Proofs

Proof of proposition 2. Suppose that $C = E \setminus F$, for open E, F. Let $w \in C$. Then $w \in E$. Then $C \cap E = E \setminus F$, so C is closed in $\mathfrak{I}^* \upharpoonright E$.

For the converse, suppose that C is conditionally refutable. So for each $w \in C$, let C be closed in $\mathfrak{I}^* \upharpoonright E_w$. So there exists open F_w such that $C \cap E_w = E_w \setminus F_w$. Furthermore, by construction, $F_w \cap C = \varnothing$. Let $E = \bigcup_{w \in C} E_w$ and let $F = \bigcup_{w \in C} F_w$. Then $F \cap C = \varnothing$. Let $w \in C$. Then $w \in E_w \subseteq E$. But since $w \in C$, it follows that $w \notin F$. So $w \in E \setminus F$. Now, suppose that $w \notin C$. Suppose that $w \in E$. Then there exists $v \in W$ such that $w \in E_v$. Since $w \notin C$, it follows that $w \in F_v \subseteq F$. So $w \notin E \setminus F$. Thus, $C = E \setminus F$. $\square$

Proof of proposition 3. Suppose that $w \in A \cap \bigcup C_{\prec}$. Then there exists $D \in \mathcal{S}$ such that $w \in D \prec C$. Since $C \cap D = \varnothing$, and $w \in D$, it follows that $w \in B$. So B is an open neighborhood of w disjoint from C. So $\{w\} \npreceq C$. So $D \npreceq C$. Contradiction. $\square$

Proof of proposition 4. It is a standard fact that $\preceq$ is a pre-order over arbitrary subsets of W, so it suffices to show that $\preceq$ is anti-symmetric over $C, D \in \mathcal{S}$. Suppose $C \preceq D$ and $C \neq D$. Since $\mathcal{S}$ is locally closed, $C = A \setminus B$ for A, B open. Since $C \preceq D$, there is $w \in A \cap D$. But since $w \in D \neq C$, $w \in B$. Since B is open and disjoint from C, $\{w\} \npreceq C$ and $D \npreceq C$. $\square$

Proof of proposition 5. Let $w \in W$. Let i be least such that $w \in C_i$. So $w \in A_i$. Let X denote the set of all $j < i$ such that $w \in A_j$. Then $w \in B_j$, for all $j \in X$. Let $D = A_i \cap \bigcap_{j \in X} B_j$, which is open. So there exists $E \in \mathcal{I}(w)$ such that $E \subseteq D$. Let $F \in \mathcal{I}(w)$ such that $F \subseteq E$. Then $F \subseteq A_i$ and $F \not\subseteq B_i$, so F activates C_i. Consider $j < i$. If $j \notin X$, then $F \not\subseteq A_j$. If $j \in X$, then $F \subseteq B_j$. So F does not activate C_j. Thus, $\lambda_{\mathsf{enm}}(F) = C_i$. So $E \in \mathsf{Lock}_{\mathfrak{S}}(\lambda_{\mathsf{enm}}, w)$. $\square$

Proof of proposition 6. Homogeneity is used in the second step of the following argument:

$$\begin{aligned} w \text{ is a closure point of } C \quad &\text{iff} \quad \{w\} \preceq C \\ &\text{iff} \quad \phi_{\mathcal{S}}(w) \preceq C \\ &\text{iff} \quad \phi_{\mathcal{S}}(w) \in C_{\preceq} \\ &\text{iff} \quad w \in \bigcup C_{\preceq}. \end{aligned}$$

□

Proof of proposition 7. The left-to-right side is immediate. For the converse, suppose that $C \cap E \preceq_{\mathfrak{I}^*\lceil E} D \cap E$. Then $\{w\} \preceq_{\mathfrak{I}^*\lceil E} D \cap E$, for all $w \in C \cap E$. Since $C \cap E$ is non-empty, there exists w such that preceding statement holds. Thus, $w \in \mathsf{cl}_{\mathfrak{I}^*\lceil E}(D \cap E)$. So $w \in \mathsf{cl}_{\mathfrak{I}^*} D$. So $\{w\} \preceq_{\mathfrak{I}^*} D$. So $C \preceq_{\mathfrak{I}^*} D$, by homogeneity of $\mathcal{S}$. □

Proof of proposition 8. Recall that $\bigcup C_{\preceq}$ is closed, by homogeneity and proposition 6. Again, by homogeneity, $W \setminus C_{\preceq} = C_{\not\preceq}$. So $C_{\not\preceq}$ is open. Let $C = A \setminus B$, for A, B open, by local closure. Then $A \subseteq \bigcup C_{\not\prec}$, by local closure and proposition 3. So $\bigcup C_{\not\prec} = A \cup \bigcup C_{\not\preceq}$, which is open. □

Proof of proposition 9. For the first claim, recall that $\preceq$ is a partial order, due to local closure and proposition 4. So it is immediate that $C = \bigcup C_{\not\prec} \setminus \bigcup C_{\not\preceq}$. Furthermore, $\bigcup C_{\not\prec}$ is open, by homogeneity and proposition 8 and $\bigcup C_{\not\preceq}$ is open, by local closure, homogeneity, and proposition 6.

For the second claim, $C = \bigcup C_{\succeq} \setminus \bigcup C_{\succ}$, since $\preceq$ is a partial order. Since $\mathcal{S}$ has open upward sets, $C_{\succeq}$ is open. Furthermore, for each $D \succ C$, we also have that $D_{\succeq}$ is open, so $C_{\succ} = \bigcup_{D \succ C} D_{\succeq}$ is open. □

Proof of proposition 10. The former clearly implies the latter. To refute the converse implication, consider the problem in which $W = \mathbb{N} \cup \{a, b\}$, for $a, b \notin \mathbb{N}$. The information states in $\mathcal{I}$ are $\{a\}$, together with $\{b\} \cup \{n, n+1, \ldots\}$, for each n. Let the degrees in $\mathcal{S}$ be $\{a, b\}$ together with $\{n\}$, for each $n \in \mathbb{N}$. Then $\mathcal{S}$ is homogeneous, but not σ-homogeneous for $\mathfrak{I}^*$. □

Proof of proposition 11. Suppose that $H \in \mathcal{I}^* \cap \mathcal{S}^*$. Since $H \in \mathcal{I}^*$, we have that $\mathsf{rel}_{\mathcal{S}}(H) \in \mathsf{rel}_{cs}(\mathcal{I}^*)$. Since $H \in \mathcal{S}^*$, it follows that $H = \mathsf{rel}_{\mathcal{S}}(H)$. So $H \in \mathsf{rel}(\mathcal{I}^*)$. The second inclusion is immediate. □

Proof of proposition 12. Let $v \in (\phi_{\mathcal{S}}^{-1} \circ \phi_{\mathcal{S}})(\bigcup \mathcal{H}) = \bigcup \{\phi_{\mathcal{S}}(w) : w \in \bigcup \mathcal{H}\}$. So there exists $H \in \mathcal{H}$ and $w \in H$ such that $\phi_{\mathcal{S}}(w) = \phi_{\mathcal{S}}(v)$. So $w \in (\phi_{\mathcal{S}}^{-1} \circ \phi_{\mathcal{S}})(H)$ and therefore $w \in \bigcup (\phi_{\mathcal{S}}^{-1} \circ \phi_{\mathcal{S}})(\mathcal{H})$. For the converse inclusion, let $v \in \bigcup (\phi_{\mathcal{S}}^{-1} \circ \phi_{\mathcal{S}})(\mathcal{H}) = \bigcup \{(\phi_{\mathcal{S}}^{-1} \circ \phi_{\mathcal{S}})(H) : H \in \mathcal{H}\}$. So there exists $H \in \mathcal{H}$ such that $v \in (\phi_{\mathcal{S}}^{-1} \circ \phi_{\mathcal{S}})(H) = \bigcup \{\phi_{\mathcal{S}}(w) : w \in H\} \subseteq \bigcup \{\phi_{\mathcal{S}}(w) : w \in \bigcup \mathcal{H}\} = (\phi_{\mathcal{S}}^{-1} \circ \phi_{\mathcal{S}})(\bigcup \mathcal{H})$. □

Proof of proposition 13. Suppose that $\mathcal{S}$ is σ-homogeneous for $\mathfrak{I}^*$. Let E be open. Let $\mathcal{X} = \phi_{\mathcal{S}}(E)$ and let $\mathcal{Y} = \mathcal{S} \setminus \mathcal{X}$. Let $w \in \bigcup \mathcal{X}$. Then there exists $v \in \phi(w) \cap E$. So E witnesses that $\{v\} \not\preceq \bigcup \mathcal{Y}$. By σ-homogeneity, $\phi_{\mathcal{S}}(v) = \phi_{\mathcal{S}}(w) \not\preceq \bigcup \mathcal{Y}$. So there exists open F_w such that $\phi_{\mathcal{S}}(w) \subseteq F_w$ and $F_w \cap \bigcup \mathcal{Y} = \varnothing$, so $F_w \subseteq \bigcup \mathcal{X}$. So $\bigcup \mathcal{X} = \bigcup_{w \in E} F_w$, which is open.

For the converse, suppose that $\mathcal{S}$ is not σ-homogeneous for $\mathfrak{I}^*$. So there exist $w, v \in W$ and $\mathcal{Y} \subseteq \mathcal{S}$ such that $\phi_{\mathcal{S}}(w) = \phi_{\mathcal{S}}(v)$, and $\{w\} \not\preceq \bigcup \mathcal{Y}$, and $\{v\} \preceq \bigcup \mathcal{Y}$. Since $\{w\} \not\preceq \bigcup \mathcal{Y}$ there exists $E \in \mathcal{I}^*(w)$ such that $E \cap \bigcup \mathcal{Y} = \varnothing$. Then $X = (\phi^{-1} \circ \phi)(E) = \bigcup \phi_{\mathcal{S}}(E)$ is disjoint from $\bigcup \mathcal{Y}$. Suppose, for contradiction, that X is open. Note that $\phi_{\mathcal{S}}(w) = \phi_{\mathcal{S}}(v) \in \phi_{\mathcal{S}}(E)$. So X witnesses that $v \not\preceq \bigcup \mathcal{Y}$. Contradiction. $\square$

Proof of proposition 15. For (1), let $w \in W$. Then there exists $E \in \mathcal{I}$ such that $w \in E$. So $w \in \mathsf{rel}_{\mathcal{S}}(E) \in \mathsf{rel}_{\mathcal{S}}(\mathcal{I})$. So $\mathsf{rel}_{\mathcal{S}}(\mathcal{I})$ covers W. Next, let $w \in C, C' \in \mathsf{rel}_{\mathcal{S}}(\mathcal{I})$. Then $C, C' \in \mathcal{I}^*(w)$, by proposition 13. So $C \cap C' \in \mathcal{I}^*(w)$. Furthermore $\mathsf{rel}_{\mathcal{S}}(C \cap C') = C \cap C'$. So, letting $E = C \cap C'$, $E \in \mathsf{rel}_{\mathcal{S}}(w)$ and $E \subseteq C \cap C'$.

For (2), $\mathsf{rel}_{\mathcal{S}}(\mathfrak{I})^*$ is an information topology by (1). The identity follows by proposition 12. $\square$

Proof of proposition 16. Proceed as in the proof of proposition 7 $\square$

Proof of proposition 18. First, let $\mathfrak{P} = (W, \mathcal{I}, \mathcal{Q})$, where $W = \{0, 1\}$, $\mathcal{I}$ denotes the upward sets of W, and $\mathcal{Q} = \mathcal{Q}_\perp$. Then $\mathcal{S} = \{W\}$ satisfies (s1), (s2) but not (s3). Next, let $W = \{0, 1, 2\}$, let $\mathcal{I} = \{\{0, 1\}, \{1\}, \{2\}\}$, and let $\mathcal{S} = \{\{0, 2\}, \{1\}\}$. Then (s1), (s3) are satisfied, but not (s2). Let $\mathfrak{P} = (W, \mathcal{I}, \mathcal{Q})$, where $W = \mathbb{N}$, $\mathcal{I}$ denotes the upward sets of $\mathbb{N}$, and $\mathcal{Q}$ is the question "n is even" vs. "n is odd". Then $\mathcal{S} = \mathcal{Q}$ satisfies (s2)-(s3), but not (s1). For the second claim, recall the first example at the end of section 6.4. $\square$

Proof of proposition 19. Conditions (s1) and (s2) don't depend on the question. For (s3), suppose that $\mathcal{Z}$ decides $\mathcal{Y}$ decides $\mathcal{X}$. Let $X \in \mathcal{X}$ and $Z \in \mathcal{Z}$. We need to show that X is open in $\mathfrak{I} \restriction Z$. Since $\mathcal{Y}$ decides $\mathcal{X}$, we have that X is open in $\mathfrak{I} \restriction Y$, for each $Y \in \mathcal{Y}$. So for each $Y \in \mathcal{Y}$, there exists open A_Y such that $A_Y \cap Y = X \cap Y$. Similarly, since $\mathcal{Z}$ decides

$\mathcal{Y}$, we have that for each $Y \in \mathcal{Y}$, there exists open B_Y such that $B_Y \cap Z = Y \cap Z$. Thus:

$$\begin{aligned} X \cap Z &= \left(\bigcup_{Y \in \mathcal{Y}} (X \cap Y)\right) \cap Z \\ &= \bigcup_{Y \in \mathcal{Y}} X \cap Y \cap Z \\ &= \bigcup_{Y \in \mathcal{Y}} A_Y \cap Y \cap Z \\ &= \bigcup_{Y \in \mathcal{Y}} A_Y \cap B_Y \cap Z \\ &= \left(\bigcup_{Y \in \mathcal{Y}} (A_Y \cap B_Y)\right) \cap Z. \end{aligned}$$

□

Proof of proposition 20. Suppose that $\mathsf{Simp}(\mathcal{S}, \mathcal{Q})$. Then $\mathcal{S}$ satisfies (s1) and (s2) with respect to $\mathfrak{I}$. For (s3), $\mathcal{S}$ trivially decides itself. The converse is immediate. □

Proof of proposition 22. For (s1), let $C \in \mathcal{S} \restriction E$. So there exists $D \in \mathcal{S}$ such that $D \cap E = C$. By (s1) for $\mathcal{S}$, we have that D is locally closed, so let $C = A \setminus B$, for open A, B. So $D = C \cap E = (A \setminus B) \cap E = (A \cap E) \setminus (B \cap E)$. Both $(A \cap E)$ and $(B \cap E)$ are open in $\mathcal{I} \restriction E$. So C is locally closed in $\mathcal{I} \restriction E$.

For (s2), suppose that $w \in E$ and $C \in \mathcal{S} \restriction E$ and $\{w\} \preceq_{\mathcal{I} \restriction E} C$. There exists $D \in \mathcal{S}$ such that $C = D \cap E$. By proposition 7, $\{w\} \preceq D$. By (s2) for $\mathcal{S}$, we have $\phi_{\mathcal{S}}(w) \preceq D$. So by proposition 7, again, $\phi_{\mathcal{S}}(w) \preceq_{\mathcal{I} \restriction E} D \cap E = C$.

For (s3), let $H \in \mathcal{Q} \restriction E$, and let $C \in \mathcal{S} \restriction E$. There exist $H' \in \mathcal{Q}$ and $C' \in \mathcal{S}$ such that $H = H' \cap E$ and $C = C' \cap E$. By (s3) for $\mathcal{S}$, we have that H' is open in $\mathfrak{I} \restriction C'$. So there exists open X such that $H' = X \cap C'$. So $H = H' \cap E = X \cap C' \cap E = X \cap C \cap E$. So H is open in $\mathfrak{I} \restriction (C \cap E)$. So $\mathcal{S} \restriction E$ decides $\mathcal{Q} \restriction E$ in $\mathfrak{S} \restriction E$. □

Proof of proposition 23. Substitute an appeal to proposition 16 for the the appeal to 7 in the proof of proposition 22. □

Proof of proposition 25. Consider the claim that $\mathcal{Q}_{\mathsf{rng}}$ is a simplicity concept for itself. For (s1), note that $C_S = A_S \setminus B_S$, where A_S is the disjunction of all $[e]$ such that the numbers occurring in e are exactly those in S and B_S is the disjunction of all $[e]$ such that some number missing from S occurs in e. For (s2), suppose that $w, v \in C_S$ and that $\mathcal{H} \subseteq \mathcal{Q}_{\mathsf{rng}}$. Then $w \preceq \bigcup \mathcal{H}$ iff there exists $C_S \in \mathcal{H}$ such that $\mathsf{rng}(w) \setminus \{*\} \subseteq S$ iff there exists $C_S \in \mathcal{H}$ such that $\mathsf{rng}(v) \setminus \{*\} \subseteq S$ iff $v \preceq \bigcup \mathcal{H}$. Condition (s3) is trivially satisfied. The proof that $\mathcal{Q}_{\mathsf{cnt}}$ is a simplicity concept for itself is similar. To see that $\mathcal{Q}_{\mathsf{cnt}}$ is a simplicity concept for $\mathcal{Q}_{\mathsf{rng}}$, one must only re-check (s3). Note that $\bigcup (C_S)_{\succeq}$ is the union of all $[e]$ such that $S \subseteq \mathsf{rng}(e)$, so $\bigcup(C_S)_{\succeq}$ is open. But $C_S \cap C_n = \bigcup(C_S)_{\succeq} \cap C_n$, so C_S is open in $\mathfrak{I}_{\mathsf{seq}} \upharpoonright E$. Condition (s3) is immediate in the case of $\mathcal{Q}_{\mathsf{rng}}$ being a simplicity concept for $\mathcal{Q}_{\mathsf{cnt}}$, since $C_n \cap C_S$ is either C_n or $\varnothing$. $\square$

Let $\mathcal{I}_{\mathsf{L}^2}$ be the information basis induced by the L^2 norm on W_{cts}. Then $\mathfrak{I}_{\mathsf{L}^2} = (W_{\mathsf{cts}}, \mathcal{I}_{\mathsf{L}^2})$ is a Hilbert space.

Lemma 1. *If S is a finite subset of $\mathbb{N}$, $F_{<S}, F_{\leq S}$ are closed in $\mathfrak{I}_{\mathsf{L}^2}$.*

Proof of lemma 1. Every finite dimensional subspace of a Hilbert space is closed. Finite unions of closed sets are closed. $\square$

Lemma 2. *If $D \subseteq W_{\mathsf{cts}}$ is closed in $\mathfrak{I}_{\mathsf{L}^2}$, it is closed in $\mathfrak{I}_{\mathsf{cts}}$.*

Proof of lemma 2. Suppose $D \subseteq W_{\mathsf{cts}}$ is not closed in $\mathfrak{I}_{\mathsf{rec}}$. Then there is $f \in \mathsf{cl}_{\mathfrak{I}_{\mathsf{rec}}}(D) \setminus D$. Exploiting continuity and the compactness of $[a, b]$, we can construct a Cauchy sequence $\{f_i\} \subset D$ converging to f in the L^2 norm. Since every closed subset of a complete metric space is complete, it must be that D is not closed in $\mathfrak{I}_{\mathsf{L}^2}$. $\square$

Lemma 3. *If S, S' are finite subsets of $\mathbb{N}$ and $f \in F_S$ then $\{f\} \preceq_{\mathfrak{I}_{\mathsf{cts}}} S'$ iff $S \subseteq S'$.*

Proof of lemma 3. $\Leftarrow$: It suffices to show that for all $f \in F_S$ and $\epsilon > 0$, there exists $g \in F_{S'}$ such that $\sup |f - g| < \epsilon$. Since the x^i are continuous, $|\sum_{i \in S' \setminus S} x^i|$ is continuous as well. Since $[a, b]$ is closed and bounded, $|\sum_{i \in S' \setminus S} x^i|$ attains a maximum M on $[a, b]$ by the extreme value theorem. So letting $a = \epsilon/M$, and $g = f + \sum_{i \in S' \setminus S} a x^i$, $\sup |f - g| < \epsilon$.
$\Rightarrow$: Immediate from lemmas 1 and 2. $\square$

Proof of proposition 26. By lemmas 1 and 2, since $F_{<S}, F_{\leq S}$ are closed in $\mathfrak{I}_{\mathsf{L}^2}$, they are closed in $\mathfrak{I}_{\mathsf{cts}}$ and therefore closed in the restriction $\mathfrak{I}_{\mathsf{poly}}$. Therefore $F_S = F_{\not< S} \setminus F_{\not\leq S}$ is locally closed in $\mathfrak{I}_{\mathsf{poly}}$. Since $D_n = F_{\not\leq \{1,\ldots,n-1\}} \cap F_{\leq \{1,\ldots,n\}}$ it is locally closed by lemmas 1 and 2. Homogeneity of both questions follows from lemma 3. Obviously, both questions decide themselves and polynomial form decides polynomial degree. The fact that polynomial degree decides polynomial form follows from lemma 2 and the fact that every finite-dimensional subspace of a Hilbert space is closed. Finally, we establish σ-homogeneity. Let $H = \bigcup_{i \in S} D_i$, for some $S \subseteq \mathbb{N}$. Suppose that $\{w\} \preceq D_j$, for some $j \in S$. By homogeneity, $D_n = \phi_{\mathcal{Q}_{\mathsf{deg}}}(w) \preceq D_j$, so $\phi_{\mathcal{Q}_{\mathsf{deg}}}(w) \preceq H$. Alternatively, $\{w\} \not\preceq D_j$, for all $j \in S$. So $k = \mathsf{max}(S)$ is finite and $n > k$. By homogeneity, $D_k \prec D_n$. Choose $E \in \mathcal{I}$ such that $D_n \subseteq E$ and $E \cap D_k = \varnothing$. Hence, $E \cap D_j = \varnothing$, for each $j \leq k$. So $D_n \cap E \cap H = \varnothing$, which suffices, since $D_n \subseteq E$. $\square$

Proof of proposition 27. Suppose that $\mathcal{S}$ has open upward sets. Let $\mathcal{H} \subseteq \mathcal{S}$ and $w \in W$. Suppose that $\{w\} \preceq \bigcup \mathcal{H}$. Then, since $\bigcup(\phi_{\mathcal{S}}(w)_{\succeq})$ is open, it follows that some $D \in \mathcal{H}$ is in $\phi_{\mathcal{S}}(w)_{\succeq}$. So $\phi_{\mathcal{S}}(w) \preceq D \subseteq \bigcup \mathcal{H}$, by homogeneity. Hence, $\phi_{\mathcal{S}}(w) \preceq \bigcup \mathcal{H}$. $\square$

Proof of proposition 28. By lemmas 1 and 2, $D_{<n} = F_{\{1,\ldots,n-1\}}$ is closed, so $D_{\geq n}$ is open. Let $w \in D$. Then, there exists $E \in \mathcal{I}(w)$ such that $E \subseteq D_{\geq n}$. So $F \subseteq D_{\geq n}$, for all $F \in \mathcal{I}(w)$ such that $F \subseteq E$. Then $D_n \in \mathsf{Min}_{\mathcal{S}}(F)$. So E is locking for λ_{Min}. $\square$

Lemma 4. *If $D \subseteq W_{\mathsf{cts}}$ is open in $\mathfrak{I}_{\mathsf{cts}}$, it is open in $\mathfrak{I}_{\mathsf{unf}}$.*

Proof of lemma 4. Let R be a rectangle and let $f \in E_R \in \mathfrak{I}_{\mathsf{cts}}$. Let $(c, f(c)) \in R$. By continuity of f there is a rectangle $R_f = G \times N$ centered at c such that $graph(f, G) \subset R_f$ and $R_f \subseteq R$. Then $f \in E_{R_f} \in \mathfrak{I}_{\mathsf{unf}}$. Furthermore, $E_{R_f} \subseteq E_R$. Finally, $E_R = \cup_{f \in E_R} E_{R_f} \in \mathfrak{I}_{\mathsf{unf}}$ as required. $\square$

Lemma 5. *Suppose A is a subalgebra of W_{cts} which contains a non-zero constant function. Then A is dense in $\mathfrak{I}_{\mathsf{unf}}$ if and only if A separates points.*

Proof of lemma 5. Immediate consequence of the Stone-Weierstrass theorem. $\square$

Proof of proposition 29. Define the *graph* of f on $A \subset [a, b]$ as follows: $graph(f, A) = \{(x, f(x)) : x \in A\}$. Each rectangle $R = G \times N$ defines an information state $E_R = \{f : graph(f, G) \subset R\}$. Let $\mathcal{I}_{\mathsf{unf}}$ be the set of all such information states. $\mathfrak{I}_{\mathsf{unf}} = (W_{\mathsf{cts}}, \mathcal{I}_{\mathsf{unf}})$ is known as *the topology of uniform convergence* on W_{cts}. Say that a set $A \subseteq W_{\mathsf{cts}}$ *separates points* iff, for every two distinct points $x, y \in [a, b]$, there exists a function $f \in A$ with $f(x) \neq f(y)$. Let A be the algebra generated by x^3 and some non-zero constant function. Since x^3 is increasing, it separates points. By lemma 5, A is dense in $\mathfrak{I}_{\mathsf{unf}}$. By lemma 4, A is dense in $\mathfrak{I}_{\mathsf{cts}}$ and $\mathfrak{I}_{\mathsf{poly}}$ as well. Let $F_2 = \{ax^2 : a \in \mathbb{R}\} \in \mathcal{Q}_{\mathsf{form}}$. Then $F_2 \preceq A$. But, by lemma 3, for all $F \in \mathcal{Q}_{\mathsf{form}}$ such that $F \cap A \neq \varnothing$, $F_2 \npreceq F$. So $\mathcal{Q}_{\mathsf{form}}$ does not have open upward sets over $\mathfrak{I}_{\mathsf{poly}}$. $\square$

Proof of proposition 30. For (s1), each singleton is closed and, hence, locally closed in $\mathfrak{S}^{\mathsf{co}-\downarrow}(\mathbb{N}, =)$. Conditions (s2, s3) are trivial for $\mathcal{Q}_\perp$. Suppose that $\mathcal{S}$ contains only finite subsets of $\mathbb{N}$. Then $C \not\prec D$, for each $C, D \in \mathcal{S}$, so $\bigcup C_{\succeq} = C$, which is closed but not open ($C \prec W \setminus C$). Alternatively, suppose that $\mathcal{S}$ contains an infinite cell. Then (s3) is violated, since no singleton compatible with infinite C is open in the information topology restricted to C. In neither case is $\mathcal{S}$ a simplicity concept for $\mathcal{Q}_\perp$ that has open upward sets in $\mathfrak{S}^{\mathsf{co}-\downarrow}(\mathbb{N}, =)$. $\square$

Proof of proposition 31. For (ii), let $w \in C_i$. Let $J = \{j < i : C_i \not\prec C_j \text{ and } C_j \not\prec C_i\}$. Define:

$$G_i \;=\; \bigcup (C_i)_{\not\prec}; \quad G_j \;=\; \bigcup (C_j)_{\npreceq}; \quad G \;=\; G_i \cap \bigcap_{j \in J} G_j.$$

Then $w \in G_i$, because $C_i \not\prec C_i$ and, for each $j \in J$, we have that $w \in G_j$, since $C_i \npreceq C_j$. So $G \in \mathcal{I}^*(w)$, by stratification and propositions 6, 8. Let $E \in \mathcal{I}(w)$ such that $E \subseteq G$. Let $F \in \mathcal{I}(w)$ such that $F \subseteq E$. Let $j < i$. Suppose that $C_i \not\prec C_j$. Case 1: $C_j \prec C_i$.

Then $C_j \notin \phi_{\mathcal{S}}(F)$, since $F \subseteq (C_i)_{\not\prec}$. Case 2: $C_j \not\prec C_i$. Then $j \in J$, so $C_j \notin \phi_{\mathcal{S}}(F)$, since $F \subseteq \bigcup (C_j)_{\not\preceq}$. So, by contraposition, we have that $C_i \prec C_j$, for each $j < i$ such that $C_j \in \phi_{\mathcal{S}}(F)$. Furthermore, $C_i \in \phi_{\mathcal{S}}(F)$, by case 1. So $C_i \in \mathsf{Ordmin}_{\mathfrak{S}}(F)$. So $\lambda_{\mathsf{Ordmin}}(F) = C_i = \phi_{\mathcal{S}}(w)$.

For (i), suppose that $E, F \in \mathcal{I}$ such that $F \subset E$ and $\lambda(E) \cap F \neq \varnothing$. If $\lambda_{\mathsf{Ordmin}}(E) = \bigcup \phi_{\mathcal{S}}(E)$, we are done, so suppose that $\lambda_{\mathsf{Ordmin}}(E) = C_i \in \mathsf{Ordmin}_{\mathfrak{S}}(E)$. Then $C_i \in \mathsf{Ordmin}_{\mathfrak{S}}(F)$, by proposition 7 and the stratification property. □

Proof of proposition 32. Consider the case in which $\lambda_{\mathsf{Ordmin}}(E) = C \in \mathsf{Min}_{\mathcal{S}}(E)$. Then $\varnothing \neq C \subseteq \phi_{\mathcal{S}}(E)$. Also, since $C \in \mathcal{S}$, the homogeneity of $\mathfrak{S}$ ensures σ-homogeneity. Consider the alternative case in which $\lambda_{\mathsf{Ordmin}}(E) = \mathsf{rel}_{\mathcal{S}}(E)$. Then cogency is immediate and $\phi_{\mathcal{S}}(w) \subseteq \mathsf{rel}_{\mathcal{S}}(E)$, so σ-homogeneity holds as well. □

Proof of proposition 33. Suppose that H is σ-homogeneous in response to E. Suppose that $\phi_{\mathcal{S}}(E) \cap H_{\preceq} \subseteq \phi_{\mathcal{S}}(H)$. Let $w \in E$, and suppose that $w \in \mathsf{cl}(H)$. Then $\{w\} \preceq H$. So, by σ-homogeneity, $\phi_{\mathcal{S}}(w) \preceq H$. So $\phi_{\mathcal{S}}(w) \subseteq H$ and, hence, $w \in H$. Hence, H is closed in $\mathfrak{I}^* \restriction E$. For the converse, suppose that $\phi_{\mathcal{S}}(E) \cap H_{\preceq} \not\subseteq \phi_{\mathcal{S}}(H)$. Then there exists $C \in \phi_{\mathcal{S}}(E)$ such that $C \preceq H$ and $C \not\subseteq H$. Since $C \in \phi_{\mathcal{S}}(E)$, there exists $w \in E \cap C$ such that $w \in \mathsf{cl}(H)$ and $w \notin H$. So H is not closed in $\mathfrak{I}^* \restriction E$. □

Proof of proposition 34. Method $\lambda_{\mathsf{Ordmin}}$ satisfies the vertical razor, because either $\lambda_{\mathsf{Ordmin}}(E) =$ some $C \in \mathsf{Min}_{\mathfrak{S}}(E)$, or $\lambda_{\mathsf{Ordmin}}(E) = \bigcup \phi_{\mathcal{S}}(E)$, both of which are closed downward in $\phi_{\mathcal{S}}(E)$. Apply proposition 31. □

Proof of proposition 35. For (1), σ-homogeneity follows from the fact that the range of λ is $\mathcal{S}^*$, and $\mathcal{S}$ is σ-homogeneous, by proposition 27. Also, $\lambda_{\mathsf{Nmb}}(E) \subseteq \phi_{\mathcal{S}}(E)$, by definition. Furthermore, if $\mathsf{Min}_{\mathfrak{S}} \neq \varnothing$, then each $C \in \mathsf{Min}_{\mathfrak{S}}$ is a subset of $\lambda_{\mathsf{Nmb}}(E)$. If $\mathsf{Min}_{\mathfrak{S}} = \varnothing$, then $\lambda_{\mathsf{Nmb}}(E) = \phi(E)$. Either way, $\lambda_{\mathsf{Nmb}}(E) \neq \varnothing$. So λ_{Nmb} is deductively cogent.

For (2), let $w \in W$. Let $C = \phi_{\mathcal{S}}(w)$. Since $\mathcal{S}$ has open upward sets, we have that $C = \bigcup C_{\succeq} \setminus \bigcup C_{\succ}$, by proposition 13. Since $C_{\succeq}$ is open, by proposition 6, there exists $E \in \mathcal{I}(w)$ such that $E \subseteq C_{\succeq}$. Let $F \in \mathcal{I}(w)$ such that $F \subseteq E$. Then $C_{\succeq} = \phi_{\mathcal{S}}(F)$, by the definition of $\succeq$. Thus, $C \in \mathsf{Min}_{\mathfrak{S}}(F)$, so $\mathsf{Nmb}_{\mathfrak{S}}(F) = \{C\}$. Therefore, $\lambda_{\mathsf{Nmb}}(E) = C$, as required.

For (3), let $C \in \phi_{\mathcal{S}}(E)$. If there exists no $D \in \mathsf{Min}_{\mathfrak{S}}(E)$ such that $D \prec C$, then $C \in \mathsf{Nmb}_{\mathfrak{S}}(E)$. Alternatively, there exists $D \in \mathsf{Min}_{\mathfrak{S}}(E)$ such that $D \prec C$, so $D \in \mathsf{Nmb}_{\mathfrak{S}}(F)$. Therefore, λ_{Nmb} satisfies the horizontal razor. For the vertical razor, suppose that $C \in \mathsf{Nmb}_{\mathfrak{S}}(E)$ and that some $D \in \phi_{\mathcal{S}}(E)$ is simpler than C. Then no $G \in \mathsf{Min}_{\mathfrak{S}}(E)$ is simpler than D, else G is also simpler than C, contradicting the case hypothesis. So $D \in \mathsf{Nmb}_{\mathfrak{S}}(F)$.

For (4), notice that $\phi_{\mathcal{S}}(E) \setminus \lambda_{\mathsf{Nmb}}(E) = \bigcup\limits_{D \in \mathsf{Min}_{\mathfrak{S}}(E)} D_{\succ}$, which is open by assumption.

For (5), the inclusion $\mathsf{Min}_{\mathfrak{S}}(E) \subseteq \mathsf{Nmb}_{\mathfrak{S}}(E)$ holds generally. When $(\phi_{\mathcal{S}}(E), \preceq)$ is well-founded, the converse inclusion $\mathsf{Nmb}_{\mathfrak{S}}(E) \subseteq \mathsf{Min}_{\mathfrak{S}}(E)$ is immediate. □

Proof of proposition 36. Method λ_{Nmb} satisfies Ockham's horizontal and vertical razors in $\mathfrak{S}$ and solves $\mathfrak{S}$ in the limit, by proposition 35. For optimality, let $e = (E_1, \ldots, E_n) \in \mathsf{Hst}_{\mathfrak{S}}(C, E)$ and suppose that λ_{Nmb} solves $\mathfrak{S}$ in the limit.

Cycle case. By lemma 6.2, $\lambda_{\mathsf{Nmb}}(e)$ is not a cycle sequence. So, trivially, $\lambda_{\mathsf{Nmb}} \leq^{\mathsf{cyc}}_{(\mathfrak{S},C,E)}$ λ in E, for all methods λ.

Reversal case. Suppose that $\lambda_{\mathsf{Nmb}}(e)$ is a reversal sequence, for $e \in \mathsf{Hst}_{\mathfrak{S}}(C, E)$ By lemma 6.1, there exists chain $C_1 \prec \ldots \prec C_n$ in $(\phi_{\mathcal{S}}(E), \preceq)$ such that $C_i \in \mathsf{Nmb}_{\mathfrak{S}}(E_i)$, for $1 \leq i \leq n$. Construct $d = (D_0, \ldots, D_n) \in \mathsf{Hst}_{\mathfrak{I}}(C, E)$, as follows. Since $C_1 \in \mathsf{Nmb}_{\mathfrak{S}}(E_1) \subseteq \phi_{\mathfrak{S}}(E_1)$, there exists $w \in E \cap C_1$. Since λ solves $\mathfrak{S}$ in the limit, let $D_1 \in \mathcal{I}(w)$ such that $D_1 \subseteq E$ be locking for λ, w. Suppose that $n > 1$. Then, since $w \in C_1 \prec C_2$, it follows that $D_1 \cap C_2 \neq \varnothing$.

Now, suppose that we are given downward-nested $d_k = (D_1, \ldots, D_k)$, for $1 \leq k < n$, such that $C_{k+1} \cap D_k \neq \varnothing$. Choose $v \in D_k \cap C_{k+1}$. Since λ solves $\mathfrak{S}$ in the limit, let $D_{k+1} \in \mathcal{I}(w)$ such that $D_{k+1} \subseteq D_k$ be locking for λ, v. Suppose that $n > k + 1$. Then $w \in C_{k+1} \prec C_{k+2}$, so $D_{k+1} \cap C_{k+2} \neq \varnothing$.

By construction, d_n is downward-nested, and $D_1 \subseteq E$. Furthermore, since $C_n \preceq C$, it follows, from the choice of D_n, that $D_n \cap C \neq \varnothing$. So $d_n \in \mathsf{Hst}_{\mathfrak{S}}(C, E)$. Furthermore, we have that $\lambda(D_i) = C_i \subseteq \bigcup \mathsf{Nmb}_{\mathfrak{S}}(E_i) = \lambda_{\mathsf{Nmb}}(E_i)$, for $1 \leq i \leq n$. So $\lambda_{\mathsf{Nmb}} \leq^{\mathsf{rev}}_{(\mathfrak{S},C,E)} \lambda$. $\square$

Proof of proposition 37. Suppose that $\mathfrak{S}$ has open upward sets and that λ solves $\mathfrak{S}$ in the limit. It is immediate from the definitions that (1) implies (2). Next, argue by contraposition that (2) implies (3). Suppose that λ violates Ockham's horizontal razor at $E_1 \in \mathcal{I}$. Then $(\phi_{\mathcal{S}} \circ \lambda)(E_1)$ is not co-initial in $(\phi_{\mathcal{S}}(E_1), \preceq)$. So there exists $C_2 \in \phi_{\mathcal{S}}(E_1)$ such that $D \notin (\phi_{\mathcal{S}} \circ \lambda)(E_1)$, for all $D \in \phi_{\mathfrak{S}}(E_1)$ such that (*) $D \preceq C_2$. Let $w_2 \in E_1 \cap C_2$. Let $E_2 \in \mathcal{I}(w_1)$, such that $E_2 \subseteq E_1$, be locking for λ, w_2. Let $e = (E_1, E_2)$. Then $\lambda(e)$ is a reversal sequence in $\mathsf{Hst}_{\mathfrak{S}}(C_2, E_1)$. Let $d = (D_1, D_2) \in \mathsf{Hst}_{\mathfrak{S}}(C_2, E_1)$. Suppose that $\lambda_{\mathsf{Nmb}}(D_1) \subseteq \lambda(E_1)$, and $\lambda_{\mathsf{Nmb}}(D_2) \subseteq \lambda(E_1) = C_2$. Then $\mathsf{Nmb}_{\mathfrak{S}}(D_1) \subseteq (\phi_{\mathcal{S}} \circ \lambda)(D_1)$ and $\mathsf{Nmb}_{\mathfrak{S}}(D_2) = \{C_2\}$. So, by (*), we have that $\mathsf{Nmb}_{\mathfrak{S}}(D_1) \cap \mathsf{Nmb}_{\mathfrak{S}}(D_2) = \varnothing$. Since $C_2 \in \mathsf{Nmb}_{\mathfrak{S}}(D_2)$, we have that $C_2 \in \phi_{\mathcal{S}}(D_2) \subseteq \phi_{\mathcal{S}}(D_1)$. So $C_2 \in \phi_{\mathcal{S}}(D_1) \setminus \mathsf{Nmb}_{\mathfrak{S}}(D_1)$. Hence, there exists $D \in \mathsf{Min}_{\mathcal{S}}(D_1)$ such that $D \prec C_1$, by the definition of $\mathsf{Nmb}_{\mathfrak{S}}$, contradicting (*). Therefore, $\lambda \not\leq^{\mathsf{rev}}_{(\mathfrak{S},C,E)} \lambda_{\mathsf{Nmb}}$. But $\lambda_{\mathsf{Nmb}} \leq^{\mathsf{rev}}_{\mathfrak{S}} \lambda$, by proposition 36. So $\lambda_{\mathsf{Nmb}} <^{\mathsf{rev}}_{\mathfrak{S}} \lambda$.

To show that (3) implies (1), assume that λ, λ' solve $\mathfrak{S}$ in the limit and that λ also satisfies Ockham's horiziontal razor in $\mathfrak{S}$. Let $C \in \mathcal{S}$, $E \in \mathcal{I}$, and $e = (E_0 \supset \ldots \supset E_n) \in \mathsf{Hst}_{\mathfrak{I}}(C, E)$. Set $o = (\lambda(E_0), \ldots, \lambda(E_n))$. Suppose that o is a reversal sequence. Construct $c = (C_0 \prec \ldots \prec C_n)$ such that $\varnothing \neq C_i \subseteq \lambda(E_i)$, for $i \leq n$ and $C_i \cap C_{i+1} = \varnothing$, for $i < n$, as follows. In the base case, let C_n be an arbitrary element of $(\phi_{\mathcal{S}} \circ \lambda)(E_n)$, which exists because λ is a learning method for $\mathfrak{S}$. Let $0 < i < n$, and assume, inductively, that $C_{i+1} \in (\phi_{\mathcal{S}} \circ \lambda)(E_{i+1})$. Since λ satisfies Ockham's horizontal razor, we have that $(\phi_{\mathcal{S}} \circ \lambda)(E_i)$ is co-initial in $(\phi_{\mathcal{S}}(E), \preceq)$. So there exists $C_i \in (\phi_{\mathcal{S}} \circ \lambda)(E_i)$ such that $C_i \prec C_{i+1}$. Moreover, since o is a reversal sequence, we have that $C_{i+1} \cap C_i = \varnothing$.

Next, construct $f = (F_0 \subset \ldots \subset F_n) \in \mathsf{Hst}_{\mathfrak{I}}(C, E)$ such that $C_{i+1} = (\phi_{\mathcal{S}} \circ \lambda')(F_i)$, for each $i \leq n$. Let $w_0 \in E \cap C_0$. Since λ' solves $\mathfrak{S}$ in the limit, there exists $F_0' \in \mathcal{I}(w)$ such that $\lambda(F) = C_0$, for all $F \subseteq F_0'$. So, since $\mathcal{I}$ is a topological basis, there exists $F_0 \in \mathcal{I}(w_0)$ such that $F_0 \subseteq F_0' \cap E_0$ and $\lambda'(F_0) = C_0$. Suppose, inductively, that $(F_0 \subset \ldots \subset F_i)$ and $(w_0, \ldots, w_i)$ are given, such that $F_j \in \mathcal{I}(w_j)$, and $\lambda'(F_j) = C_j$, for $j \leq i$. Since $C_i \prec C_{i+1}$, and $F_i \in \mathcal{I}(w_i)$, there exists $w_{i+1} \in F_i \cap C_{i+1}$. Since λ' solves $\mathfrak{S}$ in the limit, there exists $F_{i+1}' \in \mathcal{I}(w_{i+1})$ such that $\lambda(F) = C_{i+1}$, for all $F \subseteq F_{i+1}'$. So, since $\mathcal{I}$ is a topological basis, there exists $F_{i+1} \in \mathcal{I}(w_{i+1})$ such that $F_{i+1} \subseteq F_{i+1}' \cap F_i$ and $\lambda'(F_{i+1}) = C_{i+1}$. Since $C_i \neq C_{i+1}$, it follows, moreover, that $F_{i+1} \subset F_i$.

By construction, $f \in \mathsf{Hst}_{\mathfrak{I}}(C, E)$. So, since $o \leq c$ and $\lambda'(f) = c$, where $f \in \mathsf{Hst}_{\mathfrak{I}}(C, E)$, we have that $\lambda \leq^{\mathsf{rev}}_{\mathfrak{S},C,E} \lambda'$. □

Proof of proposition 38. Immediate consequence of proposition 31. □

Proof of proposition 39. Suppose that λ solves simplicity problem $\mathfrak{S}$ in the limit. Suppose that λ violates the vertical razor at E_1, which means that there exists $C_2 \notin (\phi_{\mathcal{S}} \circ \lambda)(E_1)$, such that $C_2 \in \phi_{\mathcal{S}}(E_1) \cap \lambda(E_1)_{\preceq}$. Since $C_2 \in \phi_{\mathcal{S}}(E_1)$, there exists $w_1 \in E_1 \cap C_2$. Let $E_2 \subseteq E_1$ be locking for λ, w_2, so $\lambda(E_2) = C_2$. Since $C_2 \in \lambda(E_1)_{\preceq}$, there exists $C_3 \in (\phi_{\mathcal{S}} \circ \lambda)(E_1)$ such that $C_2 \prec C_3$. Hence, there exists $w_3 \in E_2 \cap C_3$. Let $E_3 \subseteq E_2$ be locking for λ, w_3, so $\lambda(E_3) = C_3$. Then $e = (E_1, E_2, E_3) \in \mathsf{Hst}_{\mathfrak{S}}$, since $E_3 \subset E_2 \subset E_1$. Furthermore, $\lambda(e)$ is a reversal sequence. Finally, $C_3 \subseteq \lambda(E_1)$, so $\lambda(e)$ is cyclic. But $\lambda_{\mathsf{Ordmin}}(d)$ solves $\mathfrak{S}$, and is non-cyclic, for all $d \in \mathsf{Hst}_{\mathfrak{S}}$, by proposition 38. So $\lambda_{\mathsf{Ordmin}} <^{\mathsf{cyc}}_{\mathfrak{S}} \lambda$. □

Lemma 6. *Suppose that $\mathfrak{S}$ is a simplicity problem. Let $e = (E_1, \ldots, E_n) \in \mathsf{Hst}_{\mathfrak{I}}(C, E)$, for $C \in \mathcal{S}$ and $E \in \mathcal{I}$.*

1. *If $\lambda_{\mathsf{Nmb}}(e)$ is a reversal sequence, then there exists chain $C_1 \prec \ldots \prec C_n$ in $(\mathcal{S}, \preceq)$ such that $C_n \preceq C$ and $C_i \in \mathsf{Nmb}_{\mathfrak{S}}(E_i)$, for $i \leq n$;*

2. *$\lambda_{\mathsf{Nmb}}(e)$ is not a cycle sequence.*

Proof of lemma 6. For claim (1), define $C_1 \prec \ldots \prec C_n$ inductively, starting with C_n. Suppose that $C \in \mathsf{Nmb}_{\mathfrak{S}}(E_n)$. Then set $C_n = C$. Alternatively, suppose that $C \notin \mathsf{Nmb}_{\mathfrak{S}}(E_n)$. Note that $C \in \phi_{\mathcal{S}}(E_n)$, since $(E_1, \ldots, E_n) \in \mathsf{Hst}_{\mathfrak{I}}(C, E)$. Therefore, by the definition of $\mathsf{Nmb}_{\mathfrak{S}}$, there exists $D \in \mathsf{Min}_{\mathfrak{S}}(E_n) \subseteq \mathsf{Nmb}_{\mathfrak{S}}(E_n)$ such that $D \prec C$. Set $C_n = D$.

For the inductive step, let $C_{i+1} \prec \ldots \prec C_n$ be given, such that $C_{i+1} \in \mathsf{Nmb}_{\mathfrak{S}}(E_{i+1})$. Then $C_{i+1} \notin \mathsf{Nmb}_{\mathfrak{S}}(E_i)$, since $\lambda_{\mathsf{Nmb}}(e)$ is a reversal sequence. But $C_{i+1} \in \phi_{\mathcal{S}}(E_i)$, since $\mathsf{Nmb}_{\mathfrak{S}}(E_{i+1}) \subseteq \phi_{\mathcal{S}}(E_{i+1}) \subseteq \phi_{\mathcal{S}}(E_i)$. Therefore, by the definition of $\mathsf{Nmb}_{\mathfrak{S}}$, there exists $D \in \mathsf{Min}_{\mathfrak{S}}(E_i) \subseteq \mathsf{Nmb}_{\mathfrak{S}}(E_i)$, such that $D \prec C_{i+1}$. Set $C_i = D$.

For claim (2), suppose that $\lambda_{\mathsf{Nmb}}(e)$ is a cycle sequence. Then $C_n \in \mathsf{Nmb}_{\mathfrak{S}}(E_1)$, and $n > 1$. Then $C_1 \prec C_n$. Recall that $C_1 \in \mathsf{Min}_{\mathfrak{S}}(E_1)$, if $n \neq 1$. Thus, $C_n \in \mathsf{Nmb}_{\mathfrak{S}}(E_1)$. Contradiction. □

Proof of proposition 40. Suppose that (1) is false. Follow the argument for proposition 39 to force an arbitrary method satisfying (a) and (b) to perform a cycle starting at E, contrary to (c). Next, suppose that (1) is true. Then H is closed in $\mathfrak{I} \upharpoonright E$, by hypothesis and proposition 33. Therefore, H is refuted if false. Enumerate the elements of $\phi_{\mathcal{S}}(E)$. Let $\lambda'(E) = \lambda(E)$. For open $F \subset E$, let $\lambda'(F) = \lambda_{\mathsf{Ordmin}}(F \cap H)$ so long as $F \cap H \neq \varnothing$. Let $\lambda'(F) = \lambda_{\mathsf{Ordmin}}(F)$ otherwise. By the argument for proposition 38, $\lambda_{\mathsf{Ordmin}}$ avoids all cycles. $\square$

Proof of proposition 41. Let $\lambda'(E) = \lambda(E)$ and for all $F \in \mathcal{I}$, such that $F \subset E$, $\lambda'(F) = \lambda_{\mathsf{Nmb}}(F)$. λ' solves $\mathfrak{G} \upharpoonright E$ in the limit by proposition 35. First we show that (*) if $\lambda_{\mathsf{Nmb}}(F) \cap \lambda'(E) = \varnothing$ then $F \cap \lambda(E) = \varnothing$. Suppose that $\lambda_{\mathsf{Nmb}}(F) \cap \lambda'(E) = \varnothing$, then by definition of λ_{Nmb}, for $C \in \lambda'(E)$ either $C \notin \phi_{\mathcal{S}}(F)$ or there is $D \in \mathsf{Min}(F)$ such that $D \prec C$. Suppose that there is $D \in \mathsf{Min}(F)$ such that $D \prec C$. Since $\lambda(E)$ is vertical Ockham, $D \in \lambda(E)$. But then since $\mathsf{Min}(F) \subseteq \lambda_{\mathsf{Nmb}}(F)$, $\lambda'(E) \cap \lambda_{\mathsf{Nmb}}(F) \neq \varnothing$. Contradiction. So it must be that $F \cap \lambda(E) = \varnothing$. From (*) and lemma 6 it follows that λ' is cycle-free. It remains to show that λ' is reversal optimal. Since $\lambda'(F) = \lambda_{\mathsf{Nmb}}(F)$ for $F \subset E$ and λ_{Nmb} is reversal optimal we have that $\lambda \leq^{\mathsf{rev}}_{\mathfrak{G},C,F} \lambda'$ for every $C \subseteq F$ and λ that solves $\mathfrak{G} \upharpoonright E$ in the limit. It remains to show that for every (E, F) such that $\lambda'(e)$ is a reversal sequence, there is e' such that $\lambda(e') \leq \lambda'(e)$. By (*), if $(\lambda'(E), \lambda'(F))$ is a reversal sequence, then $F \cap \lambda'(E) = \varnothing$. But since $\lambda'(E)$ is horizontal Ockham, for every $C \subseteq \lambda'(F)$, there is $D \subseteq \lambda'(E)$ such that $D \prec C$. Let E_1 be locking for D, λ and E_2 be locking for C, λ. Then $\lambda(d) \leq \lambda'(e)$ for $d = (E_1, E_2)$ as required. To show that (2) entails (1), reproduce the arguments of propositions 39 and 37. $\square$

Proof of proposition 42. Condition (s1) holds because X_α is closed. In the base case, $X_0 = W$ is closed. For induction, the limit ordinal case is immediate, since closed sets are closed under intersection. For the successor ordinal case, suppose that X_α is closed, so $W \setminus X_\alpha$ is open. Suppose that $w \in X_\alpha \setminus X_{\alpha+1}$. Then w is an interior point of $\phi_{\mathcal{Q}}(w)$ in $\mathfrak{I}^* \upharpoonright X_\alpha$. So there exists open A_w such that $w \in A_w \cap W \setminus X_\alpha \subseteq \phi_{\mathcal{S}}(w)$. Let:

$$Z \;=\; W \setminus X_\alpha \cup \bigcup_{w \in X_\alpha \setminus X_{\alpha+1}} A_w.$$

Since Z is open, it suffices to show that $W \setminus X_{\alpha+1} = Z$. Suppose that $v \in W \setminus X_{\alpha+1}$. If $v \in W \setminus X_\alpha$, then $v \in Z$, so suppose that $v \in (W \setminus X_{\alpha+1}) \setminus (W \setminus X_\alpha) = X_\alpha \setminus X_{\alpha+1}$. Then $v \in A_v \subseteq Z$. Conversely, suppose that $v \in X_{\alpha+1}$. Then $v \notin W \setminus X_\alpha$, since $X_{\alpha+1} \subseteq X_\alpha$. Furthermore, v is in the boundary of $\phi_{\mathcal{Q}}(w)$ in $\mathfrak{I}^* \upharpoonright X_\alpha$, so v is in no A_w. So $v \notin Z$.

For conditions (s2, s4), we have that $\{w\} \not\preceq C_\alpha$ for each $\{w\} \in C_\beta$ such that $\beta < \alpha$, because $W \setminus X_\beta$ is open, by the argument for (s1). Next, suppose that $w \in C_\alpha$. Then $\{w\} \preceq C_{\alpha+1}$, by the construction of $X_{\beta+1}$.

For condition (s3), suppose that $\mathcal{Q}$ is not open in $\mathcal{I}^* \upharpoonright C_\alpha$. Then there exists a boundary point w of some $H \in \mathcal{Q}$ in $\mathcal{I}^* \upharpoonright C_\alpha \subseteq \mathcal{I}^* \upharpoonright X_\alpha$. But then $w \in X_{\alpha+1}$, which is disjoint from C_α.

For conditions (s5, s6), it suffices that the simplicity order over $\mathcal{S}_{\mathfrak{P}}$ is total, by the argument for (s2). The order type of the order also follows from the argument for (s2). $\square$

16 Acknowledgments

The authors are indebted to the John Templeton Foundation for its generous support under grant RQ-8236. We are also indebted to Thomas Icard for discussions concerning T_D spaces, and to Aeyaz Kayani for detailed corrections to the manuscript.

References

Hirotugu Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723, 1974.

Carlos E Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *The journal of symbolic logic*, 50(02):510–530, 1985.

CE Aull and WJ Thron. Separation axioms between t0 and t1. *Indag. Math*, 24:26–37, 1962.

Alexandru Baltag, Nina Gierasimczuk, and Sonja Smets. Epistemic topology (to appear). *ILLC Technical Report*, 2014.

Prasanta S Bandyopadhayay, Robert J Boik, and Prasun Basu. The curve fitting problem: A bayesian approach. *Philosophy of Science*, pages S264–S272, 1996.

Lorenzo Carlucci, John Case, Sanjay Jain, and Frank Stephan. Non u-shaped vacillatory and team learning. In *Algorithmic Learning Theory*, pages 241–255. Springer Berlin Heidelberg, 2005.

John Case and Timo Kötzing. Topological separations in inductive inference. In *Algorithmic Learning Theory*, pages 128–142. Springer, 2013.

John Case and Carl Smith. Comparison of identification criteria for machine inductive inference. *Theoretical Computer Science*, 25(2):193–220, 1983.

John Case, Sanjay Jain, and Arun Sharma. Complexity issues for vacillatory function identification. *Information and Computation*, 116(2):174–192, 1995.

Malcolm Forster and Elliott Sober. How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *The British Journal for the Philosophy of Science*, 45(1):1–35, 1994.

Rusins Freivalds and Carl H Smith. On the role of procrastination in machine learning. *Information and Computation*, 107(2):237–271, 1993.

Edmund Gettier. Is knowledge justified true belief? *Analysis*, 23(6):121–123, 1963.

Clark Glymour. *Theory and Evidence*. Princeton University Press, 1980.

Clark Glymour. Instrumental probability. *The Monist*, pages 284–300, 2001.

Nelson Goodman. *Fact, fiction, and forecast*. Harvard University Press, 1983.

Jeroen Groenendijk. Inquisitive semantics and pragmatics. In *Meaning, Content, and Argument: Proceedings of the ILCLI International Workshop on Semantics, Pragmatics, and Rhetoric (SPR-90)*, pages 41–72. Universidad del Pais Vasco, Servicio Editorial, 2009.

Jaakko Hintikka. *Socratic Epistemology: Explorations of Knowledge-Seeking by Questioning*. Cambridge University Press, 2007.

Sanjay Jain, Daniel N. Osherson, James S. Royer, and Arun Sharma. *Systems that Learn: an Introduction to Learning Theory*. MIT press, 1999.

Harold Jeffreys. Theory of probability, 1961.

Kevin Kelly. *The Logic of Reliable Inquiry*. 1996.

Kevin Kelly and Clark Glymour. Why probability does not capture the logic of scientific justification. In C. Hitchcock, editor, *Debates in the Philosophy of Science*, pages 94–114. Blackwell, 2004.

Kevin Kelly and Conor Mayo-Wilson. Causal conclusions that flip repeatedly and their justification. *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, pages 277–286, 2010.

Kevin T Kelly. Ockham's razor, empirical complexity, and truth-finding efficiency. *Theoretical Computer Science*, 383(2):270–289, 2007a.

Kevin T Kelly. Simplicity, truth, and the unending game of science. 2007b.

Kevin T Kelly. Simplicity, truth, and probability. 2010.

Peter Kugel. Induction, pure and simple. *Information and Control*, 35(4):276–336, 1977.

Thomas S. Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, 1962.

Imre Lakatos. Falsification and the methodology of scientific research programmes. *Criticism and the Growth of Knowledge*, 4:91–196, 1970.

Daniel Lehmann and Menachem Magidor. What does a conditional knowledge base entail? *Artificial intelligence*, 55(1):1–60, 1992.

Wei Luo and Oliver Schulte. Mind change efficient learning. *Information and Computation*, 204(6):989–1011, 2006.

Eric Martin and Daniel N Osherson. *Elements of scientific inquiry*. MIT Press, 1998.

Eric Martin, Arun Sharma, and Frank Stephan. Unifying logic, topology and learning in parametric logic. *Theoretical Computer Science*, 350(1):103–124, 2006.

Margaret Masterson. The nature of a paradigm. In Imre Lakatos and Alan Musgrave, editors, *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press, 1970.

Robin Milner. An algebraic definition of simulation between programs. In *Proceeding of the 2nd International Joint Conferences on Artificial Intelligence*. British Computer Science, 1971.

Friedrich Max Müller. *The Upanishads*. Oxford University Press, 1900.

Wayne C Myrvold. A bayesian account of the virtue of unification. *Philosophy of Science*, 70(2):399–423, 2003.

Daniel N Osherson, Michael Stob, and Scott Weinstein. *Systems that learn: An introduction to learning theory for cognitive and computer scientists.* The MIT Press, 1986.

Judea Pearl. System z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In *Proceedings of the 3rd Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 121–135. Morgan Kaufmann Publishers Inc., 1990.

Karl R. Popper. *The Logic of Scientific Discovery*. London: Hutchinson, 1959.

Hilary Putnam. Trial and error predicates and the solution to a problem of mostowski. *Journal of Symbolic logic*, pages 49–57, 1965.

Craige Roberts. Context in dynamic interpretation. In Laurence Horn and Gregory Ward, editors, *Handbook of Contmeporary Pragmatic Theory*, pages 197–220. Blackwell, 2004.

Oliver Schulte, Wei Luo, and Russell Greiner. Mind change optimal learning of bayes net structure. In *20th Annual Conference on Learning Theory (COLT), San Diego, CA, June 12-15*, 2007.

Stacks Project. Partitions and stratifications, http://stacks.math.columbia.edu/tag/09XY.

Stratification. *The Encyclopedia of Mathematics*. http://www.encyclopediaofmath.org/index.php?title=

Alan Mathison Turing. On computable numbers, with an application to the entscheidungsproblem. *J. of Math*, 58:345–363, 1936.

Johan van Benthem. Modal logic and classical logic. 1983.

Bas C Van Fraassen. The scientific image. 1980.

Vladimir Vapnik. *Statistical learning theory*, volume 2. Wiley New York, 1998.

Steven Vickers. *Topology Via Logic*. Cambridge University Press, 1996.

Larry Wasserman. Bayesian model selection and model averaging. *Journal of mathematical psychology*, 44(1):92–107, 2000.

Akihiro Yamamoto and Matthew de Brecht. Topological properties of concept spaces (full version). *Information and Computation*, 208(4):327–340, 2010.