A Close Shave with Realism:

How Ockham's Razor Helps Us Find the Truth[1]

Kevin T. Kelly

Department of Philosophy

Carnegie Mellon University

kk3n@andrew.cmu.edu

December 6, 2002

[1]This is a circulation draft subject to revision without notice. Comments are most welcome. The author is indebted to Joseph Ramsey, Oliver Schulte, Clark Glymour, Mark Ravaglia, Peter Tino, Julia Pfeiffer, Marion Ledwig, and Noriko Nagata for helpful comments on drafts of this paper.

Abstract

Many distinct theories are compatible with current experience. Scientific realists recommend that we choose the simplest. Anti-realists object that such appeals to "Ockham's razor" cannot be truth-conducive, since they lead us astray in complex worlds. I argue, on behalf of the realist, that always preferring the simplest theory compatible with experience is necessary for efficient convergence to the truth in the long run, even though it may point in the wrong direction in the short run. Efficiency is a matter of minimizing errors or retractions prior to convergence to the truth.

0.1 Realism and Ockham's Razor

There are infinitely many, incompatible theories consistent with any finite amount of experience, so how can we choose among them? The answer is easy: apply "Ockham's razor".

> In choosing a scientific hypothesis in light of evidential data, we must frequently add to the data some methodological principle of simplicity in order to select out as "preferred' one of the many different possible hypotheses, all compatible with the specified data (Sklar 1977; p. 100).

Why be guided by simplicity? Scientific realists would like to say something like this.

> ... [A]mong the ... theories consistent with our observational data, some are better explanations than others in virtue of their greater simplicity or elegance or unifying power and ... these virtues are *indications that those theories are true* (Papineau 1997; p. 9, my emphasis).

Great! But there's a catch.

> ... [T]he connection between *simplicity* and *truth* seems so dubious. ...[I]f no argument can establish such a connection, what reasons do we have in the first place for invoking simplicity in our mechanism for choosing (Sklar 1977; p. 132)?

Indeed, how could there be such a connection? A *fixed* bias toward simplicity (or toward anything else) can no more indicate truth than a broken thermometer can indicate temperature. An indicator has to be *sensitive* to what it indicates, but Ockham's razor favors simplicity no matter what. At least the errors incurred by Ockham's razor are corrected, eventually, by future experience.

> It is this feature that makes the rule, "Adopt the simplest hypothesis compatible with the present data," seem more innocuous than might first appear. For even if we do make this choice, we are not stuck with it, in the sense that ongoing experimentation

> can "test" our choice and, conceivably, reject it in favor of some more complex hypothesis (Sklar 1977; pp. 132-33).

But that doesn't explain much, since a mistaken presumption that the world is complex would also be corrected, eventually, by future experience. Nor does it help to stack the deck in favor of simplicity in advance.

> All we have to say is that the simpler laws have the greater prior probabilities (Jeffreys 1985; p. 47).[1]

To preach beyond the converted, realists need to explain how simplicity leads to the truth better than other biases without presupposing, in a narrowly circular manner, the very bias whose efficacy is to be explained.[2] My purpose in this paper is to provide such an explanation, along the following lines. Yes, every prior bias is corrected by future experience in the long run. And yes, simplicity cannot indicate truth in the short run unless we presuppose that the truth is likely to be simple. Nonetheless, I will show that conformity with Ockham's razor is

[1] There can be a prior bias toward simple worlds even when there is no prior bias toward simple theories. Suppose there are two theories, a "Ptolemaic" one with lots of free parameters and a "Copernican" theory with none. We are urged to "keep the door open" to the simple theory by assigning it a real-valued probability greater than zero (perhaps far less than the prior probability of the complex theory). But that implies that each parameter setting of the complex theory is infinitely less probable than each parameter setting of the simple theory (i.e., there is an infinite, prior bias toward simple *worlds*). Suppose that the simple theory entails the data but the complex theory does so only when the parameter is set to a special value (e.g., Ptolemy had to assume that the epicycles of superior planets are synchronized with the deferent of the sun). Then the (subjective) likelihood of the data given the complex theory is infinitesimal, whereas that given the unified theory it is unity. By Bayes' theorem, the complex theory gets mauled (Rosenkrantz 1983), and we say that "it would be a *miracle* if it were true. But the miracle, as Hume would say, is in ourselves, rather than in the world, because we started out with an infinite personal bias in favor of simple worlds over complex ones; a bias packaged as "fairness" at the level of theories. When the data are random, the story is similar, if less extreme: the simpler theory is favored to the extent that prior probability is not concentrated over high-likelihood parameter values in the disunified theory.

[2] Absent such an explanation, it is tempting to diagnose realism as a case of wishful thinking, in which our desire for simple theories is confused with evidence that such theories are true (van Fraassen 1980).

necessary for arriving at the truth *as efficiently as possible* in the long run, where efficiency is a matter of minimizing worst-case, cumulative *epistemic* costs, such as errors or retractions of earlier theories prior to convergence.[3] So to reject Ockham's razor as a principle of theory choice is to reject the most efficient possible means for finding the truth.

0.2 The Main Results

An **empirical problem** is assumed to consist of a **question** together with a **presupposition**, which specifies the range of possible worlds over which success "matters", for whatever reason (cf. Lewis 1996 for a list of reasons). To eliminate nuisance cases, it is assumed, throughout, that the presupposition is consistent, in the sense that it does not rule out all worlds. Each world determines a unique, **correct answer** to the question and a potentially infinite **input stream** that is fed, bit by bit, to the scientist or learner. I assume nothing at all about the set of possible inputs except that they are presented in discrete stages of inquiry. In particular, they need not be linguistic, symbolic, or even consciously accessible by the agent. A **method** or learning disposition maps each finite sequence of inputs compatible with the problem's presupposition to potential answers to the question or to the uninformative output '?' that indicates unwillingness to choose an answer. A method **solves** a problem just in case it **converges** (stabilizes eventually) to the correct answer to the question in each relevant possibility.

A method solves a problem under a finite **resource bound** if the total cognitive costs incurred in each possibility never exceed the bound. This idea is extended to infinite resource bounds in sections 4, 5, and 6. An **efficient solution** to a problem solves the problem under the least achievable resource bound.[4] The cognitive costs entertained in this study are retractions

[3] The approach builds upon a transfinite generalization due to R. Freivalds and C. Smith (1993) of the concept of "trial-and-error predicates" introduced by H. Putnam (1965). Putnam's idea is explored topologically in (Kelly 1996) and is applied to issues in the philosophy of science in (Schulte 1999a, 1999b, 2000, 20001).

[4] This gloss is oversimplified: an efficient solution must also solve each subproblem of the original problem under the least achievable bound (sections 5, and 6).

("taking back" an earlier answer) and errors (producing a mistaken answer). Thus, one may speak of **retraction-efficiency** and of **error-efficiency**. A problem is **efficiently solvable** just in case it has an efficient solution in either sense.

Ockham's razor says something like: never output an answer in a problem unless it is the simplest answer compatible with current experience.[5] An answer is as simple as its simplest worlds. So what makes a world simple? Notoriously, there is nothing intrinsic to an isolated world that makes it simple. Consider a world in which every input is green and another world in which the inputs turn blue at some stage i. For if "grue" means "green up to i and blue thereafter" and "bleen" means "blue up to i and green thereafter", a world's description is "bent" in the grue/bleen language just in case it is "straight" in the blue/green language and conversely (Goodman 1983). Hence, simplicity is often thought to be a mere matter of taste or of description. If the realist is to *explain* how simplicity helps us find the truth, however, simplicity must somehow be anchored in the structure of empirical problems, themselves. That is just the approach I will follow. Consider, for example, the question whether the color of the inputs will ever change. There is a structural sense in which constantly colored worlds are simpler than worlds of changing color in this problem: the experience presented by a constantly colored world always agrees with the experience presented by some changing color world, but each changing color world eventually stops agreeing with all constantly colored worlds. One might describe the situation like this: in constantly colored worlds, Nature eternally "reserves the right" to present an "anomaly" later by exhibiting a color change, but in changing color worlds she eventually exercises her right to change the color and never gets an opportunity to make the world appear constantly colored thereafter.

The example suggests a problem-relative notion of **degrees of simplicity**. A world has simplicity degree zero if it presents experience that verifies an answer to the question (i.e., it ends up exercising all of Nature's options to present anomalies) so it is as "misleading" as the

[5]This statement is right in spirit, but requires some subtle correction when the problem requires an infinite retraction bound (cf. section 8).

problem allows a world to be.) Inductively, world w has simplicity degree $n + 1$ if there are abribrarily long periods of time during which worlds of simplicity n that fail to satisfy the same answer as w present the same experience as w. Hence, simpler worlds are more constant, not in themselves, but in the problem-relative, methodological sense that they constantly refrain from exercising Nature's options to present anomalies. The approach is extended to infinite simplicity degrees in sections 7 and 8, which involves some subtle modifications of the preceding sketch. The following propositions all hold for the fully general versions of the concepts presented in detail there.

Here is the main result of the paper, which establishes the intended connection between simplicity and truth.

Proposition 1 *If a method solves a problem retraction-efficiently or error-efficiently, then the method complies with Ockham's razor in the problem.*

The proof is presented in Appendix I.[6] The approach is Leibnizian. Yes, Ockham's razor leads to disaster in complex worlds and has led us into painful retractions in the past (Laudan 1981), so we imagine we can improve upon it by sometimes preferring complex hypotheses over simple ones. However, that leads to still more retractions in other complex worlds, so Ockham's razor is, after all, the most efficient possible strategy for finding the right answer (cf. section 3 for some simple examples).

Proposition 1 vindicates both sides of the realism debate. The anti-realist is correct that Ockham's razor cannot be shown to indicate truth in the short run without begging the question and that any initial bias is compatible with long-run convergence. But the realist can respond, without begging the question or appealing to other primitive principles of "rationality" or "confirmation", that *efficient* convergence in the long run singles out Ockham's razor uniquely.

Proposition 1 cannot be strengthened in the case of error-efficiency, for the error-efficient solutions to a problem are *precisely* the solutions that comply with Ockham's razor. Hence,

[6] As are the proofs of all propositions occurring in the text.

Ockham's razor *exhausts* the short-run, methodological consequences of error-efficiency.

Proposition 2 *If a method solves an efficiently solvable problem, then the method is an error-efficient solution if and only if the method complies with Ockham's razor in the problem.*

Propositon 1 can be strengthened in the case of retraction-efficiency, however, and the result points, again, toward realism. While anti-realists may question successful theories on general, skeptical grounds, realists see no virtue in retracting the simplest theory unless nature provides a concrete sign (in the form of an anomaly) that it is time to do so. The **anomaly complexity** of a problem is the least upper bound on the simplicity degrees of the worlds that satisfy its presupposition (simple problems don't have hard anomalies that simple worlds can refuse to present.) **Anomalies** occur when the anomaly complexity of the problem drops, for that is when Nature exercises an opportunity to "veer" into an easier problem.[7] The **retention principle** states that one may retract in a problem only when an anomaly occurs (cf. section 7). The retention principle cannot be deduced from error-efficiency, because skeptical retreat to '?' never counts as an error. On the other hand, the retention principle completely characterizes retraction-efficiency.

Proposition 3 *If a method solves an efficiently solvable problem, then the method is a retraction-efficient solution if and only if the method complies with the retention principle in the problem.*

So in light of proposition 3, retraction-efficiency implies two interlocking features of scientific practice: choosing the simplest hypothesis if you choose one at all *and* hanging on to it until an anomaly indicates trouble. This rebuts the received view among philosophers of science that a short-run theory of "confirmation" is required to explain concrete evidential judgments because long-run reliability considerations are too weak to do so (e.g., Earman 1992, pp. 218-219). Indeed, the only freedom that remains for a retraction-efficient method is the amount of experience

[7] Cf. section 7 for a slight refinement of this definition in case of infinite simplicity. Nature may "veer" without taking advantage of us when she does so, just as a chess player can move her bishop without useful effect. These events are counted as "null" anomalies.

it demands before "leaping" to the (unique) Ockham answer and even confirmation theorists tend to leave this issue open. Bayesians locate learning speed in arbitrary, prior probablities and even Rudolph Carnap (1950) associated it with a tweakable parameter on his inductive logic.

By propositions 1 and 3, *every problem solution that satisfies the retention principle also satisfies Ockham's razor*.[8] That is somewhat surprising, for the retention principle says nothing about the material character of the answers we choose, as long as we hang on to them in the right circumstances; whereas Ockham's razor restricts the material character of our answers and says nothing about when we should retain them. The principles are linked by the method's guaranteed success as follows: if we must hang on to an answer until after an anomaly, then the answer had better be right if no anomalies occur. The Ockham answer turns out to have just this property, since simple answers are true in simple worlds and simple worlds present fewer anomalies (all of which is explained in detail below).

An important question remains. Efficiency implies conformity with Ockham's razor, but when is efficiency, itself, possible? Here is a natural, sufficient condition. When the presupposition of a problem is false, the learning method is free to say anything (it's operating outside of its intended range). But wouldn't it be nice if the method were to converge, at least in the limit, to '?' (i.e., "something's wrong") when the presupposition is false, so that we would eventually know that something is wrong? Such a method may be said to be **self-disqualifying**.[9] Now

[8]The converse fails: retracting once to '?' doesn't alter the facts about convergence and doesn't count as a violation of Ockham's razor.

[9]The expression "whenever the presupposition is false" is trickier than it might first appear, for although it makes sense to speak of the possibilities in P that make A false, it is unclear what the set of worlds in which P is supposed to be. Fortunately, something more concrete suffices for our purposes. Consider the set of all possible inputs that might eventually be seen given that the presupposition of the problem is true. Construct all possible ω-sequences of these inputs. For each such input sequence that is incompatible with the presupposition, construct a unique, **virtual** world that presents it and that makes the presupposition false. Call the result of adding these worlds to the original problem the **completion** of the problem. Now it makes sense to require that a self-disqualifying method converge to the truth when the presupposition is true and converge to '?' otherwise in the *completed* problem.

we have:

Proposition 4 *if a problem has a self-disqualifying solution, then it is efficiently solvable, so the preceding propositions hold non-trivially.*

In other words, the deductions of Ockham's razor from efficiency and of the retention principle from retraction-efficiency stand unless it is impossible to determine the falsity of our empirical presuppositions even in the ideal limit of inquiry.[10]

0.3 Efficiency Arguments for Ockham's Razor

Before proceeding to general definitions and proofs, it is helpful to consider some simple examples that illustrate the basic ideas.

Uniformity of nature. Recall the question whether the color of the observations will ever change. The simplest answer, both intuitively and according to the concept of simplicity sketched above, is that the color will remain constant. Here is a Leibnizian argument for complying with Ockham's razor (Schulte 1999a, b). Suppose you can answer the question in the limit, but you violate Ockham's razor at some stage. The only way to do so is to guess that the color will change prior to seeing it change. Nature is free to present the same color for as long as it takes to get you to conclude that it will never change, on pain of converging to the wrong answer if it never changes. That counts as one retraction. Nature is now free to change the color, forcing you to revise to the non-uniform answer (on pain of converging to the wrong answer) for a total of two retractions.[11] Had you followed Ockham's advice, however, you would have retracted at most once in the worst case (i.e., when the first color change occurs). Hence, any deviation from

[10]The condition is not necessary for efficiency. For example, let A_0 = "the input stream will converge to 0" and let A_0 be the only answer to problem P, so that every possible input stream converges to 0. Then P is trivially solvable with 0 retractions, but P has no self-disqualifying solution (cf. Kelly 96, chapter 4).

[11]Note the tension between minimizing retractions and converging to the right answer.

Ockham's razor results in an extra retraction in the worst case.[12]

The argument derives strict adherence to Ockham's razor entirely from the slender premise that we want to converge to the truth as efficiently as possible. Moreover, it does not beg the question by presuming that the world is simple or that it is probably simple— the reasoning is entirely worst-case, so each world counts as much as any other. In this respect, the argument recalls Popperian (1968) themes. Boldly conjecturing the simplest hypothesis compatible with experience is the most efficient possible means for converging to the truth, but there is no sense in which the simplest answer is better supported, better confirmed, more probable, more important, or otherwise more "weighty" than any other answer in the short run.

The argument's loyalty to uniform color evaporates if the empirical problem is altered. For example, if one asks whether the world is eternally grue, then the same argument recommends that we output the grue answer first. But according to the preceding definition of simplicity, the grue answer is the methodologically simplest answer in this problem (the grue world withholds anomalies forever but the constantly green world doesn't), so the argument defends Ockham's razor after all. That is as it should be, for if the realist's explanation of the efficacy of Ockham's razor is to be robust over problem variations, simplicity must adapt itself to the underlying contours of a wide range of potential problems.[13]

The efficiency argument generalizes to problems of higher complexity if a slight complication is attended to. Suppose that the color begins with green and alternates between green and blue

[12]The idea of counting retractions prior to convergence was employed for purely logical ends by Hilary Putnam in (1965). Counting retractions as a definition of the intrinsic difficulty or complexity of an empirical problem has seen a great deal of study in computational learning theory. A good reference and bibliography may be found in (Jain et al. 1999).

[13]It has been objected that anybody who speaks grue-ese would ask a grue question, so that simplicity depends on vocabulary after all (cf. Chart 2000 and Schulte 2000 for a response). If the thesis is that grue speakers couldn't even understand a question about color, then it is plainly false, since the terms are interdefinable. If it is, rather, that grue speakers have a psychological propensity to ask questions about uniform grueness rather than uniform greenness, then it is irrelevant to the thesis that simplicity is conceptually anchored in the branching structure of the scientific problem one faces.

at most twice. Ask what the color will be after the last change. The simplest answer is that the current color will be the last, since the simplest worlds in light of current experience are those in which the color never changes again. Hence, the obvious Ockham method solves the problem with at most two retractions (each color change is an anomaly). But suppose you deviate from Ockham's advice by persisting with "green" upon receiving the successive inputs (green, blue) and that you comply with Ockham's razor thereafter. In the worst case, the observations are (green, blue, blue, ..., green) and you say (green, green, blue, ..., green, ...), so you retract at most twice. That matches Ockham's performance, so it seems that Ockham's razor is not deducible from retraction efficiency even in problems requiring just two retractions.

The trouble is that efficiency is not simply a matter of overall resource consumption. It is a matter of always minimizing resource consumption "from now on", for otherwise "slush funds" accumulated from past shenanigans can conceal future inefficiency. Consider the **subproblem** in which the input sequence (green, blue) has just been received. This subproblem is solved by Ockham's razor with just one retraction (when blue flips back to green). But since you say "green" upon seeing (green, blue), a sufficiently long run of blue experience will force you to revise to "blue" (on pain of converging to the wrong answer), after which Nature is free to flip the color back to green, eliciting two retractions. So the efficiency argument works again if we require that the method minimize worst-case retractions in each subproblem, where the subproblem **rooted** at finite input sequence σ is just like the original problem except that the method starts at the end of σ (so no retractions occur until σ is extended) and all possible worlds incompatible with σ are deleted.[14] The efficiency argument now generalizes to n color changes. If we know in advance that the color may shift at most n times, then retraction-efficiency (in each subproblem) requires that we never guess more shifts than we have seen and that the

[14] This requirement is analogous to **subgame perfection** in game theory. Schulte (1999a, 1999b) derives Ockham's razor at each stage by imposing time admissibility in conjunction with minimaxing errors, but time admissibility forces the agent to output a possible answer to the question at each stage of inquiry, which seems too strict when outputs are viewed as full beliefs.

currently observed color will be the last.

Polynomial degree. In the context of curve fitting, simplicity is often identified with low polynomial degree. Suppose we know that the degree of the true curve is at most n and that we can specify an arbitrary value of the independent variable and receive a measurement of the dependent variable whose error is less than a fixed, nonzero bound ϵ. The problem is to find the true degree of the curve. If we guess the lowest degree compatible with experience, at most n retractions are required, for each curve of order $k+1$ eventually presents inputs incompatible (up to ϵ error) with all curves of order k. Now suppose that an order higher than 0 (constant functions) is guessed first. Nature can present inputs drawn from a constant function until the method retracts to the answer "degree $= 0$". As long as the data currently presented do not touch the border of the ϵ envelope around a curve of order k, there exists a curve of order $k+1$ from which the same data could have been presented up to error ϵ. Using this fact, Nature can force the method to rectract n more times for a total of $n+1$ retractions. So one is obligated, on grounds of efficiency, to prefer the simplest degree compatible with experience. The simplest worlds in this problem are worlds in which the true law is constant. Successive anomalies occur when the observed data points fall outside of the ϵ envelope of every curve of lower order.[15]

Unity. Copernican astronomy, Newtonian physics, the wave theory of optics, evolutionary theory, and chaos theory all won their respective revolutions by providing unified, low-parameter explanations of phenomena for which their competitors required many. I cannot, of course, show that this kind of simplicity leads to the right answer when two answers are compatible with the same experience for eternity, for then no method driven entirely by experience could be guaranteed to arrive at the right answer. So assume a principle of "plenitude" stating that

[15]Popper (1968) observed that lower polynomial degrees are more "falisifiable" than higher degrees, but he never explained how preferring the most falsifiable hypothesis helps us find the truth. The preceding argument provides such an explanation: it minimizes retractions prior to convergence. For a different explanation of the merits of simplicity in curve fitting, cf. (Forster and Sober 2001). Their proposal is based on the Akaike criterion for statistical model selection. As my proposal is not explicitly statistical, providing a systematic comparison of the two ideas is not trivial.

if the world is disunified, Nature is obligated to show us some (possibly obscure) violations of the misleading regularities eventually. This dodges the question of how to choose among theories that are indistinguishable in principle, but even if there are such theories, the more pervasive and interesting question is why we should prefer the unified theory when disunified alternatives are still compatible with experience received *so far* (Churchland 1981). If there are but n such regularities under consideration, then by an argument similar to those already given, one must choose the most unified theory compatible with experience (i.e, the theory that implies that all of the regularities compatible with the current inputs will remain unviolated forever). For otherwise, Nature can maintain all of the regularities until we retract and adopt the most unified theory. Thereafter, Nature can elicit one more retraction for each of the n regularities (by making it appear to be true until we "take the bait") for a total of $n+1$ retractions; but if we are always guided by a taste for unity, we succeed with at most n retractions in each possibility.

Conservation laws in particle physics (Schulte 2001). It is both intuitive and a matter of standard practice (Ford 1963) to propose the most *restrictive* conservation laws consistent with the currently observed reactions. But what does such practice have to do with finding the right answer as efficiently as possible? Suppose there are n types of particles. Then each possible reaction may be represented by an n-vector of positive integers (each component corresponds to the total number of particles of a given kind coming out of the reaction minus the total number going in). A conservation law assigns some conserved quantity to each particle type and specifies that the total quantity going into each reaction equals the total quantity coming out (Valdez-Perez and Erdmann 1994). Hence, the conservation law determined by a given assignment is consistent with all and only those possible reactions that are orthogonal to the assignment. Accordingly, each nontrivial conservation law determines an $n-1$ dimensional subspace of the original vector space. Positing more conserved quantities determines lower-dimensional subspaces. The most restrictive conservation theory compatible with the observations is then the least-dimensional subspace containing the observed reactions (i.e., the **span** of the observations restricted to vectors with integral coordinates). Always guessing the integral span of the observed reactions is guaranteed to succeed with at most n retractions (one retraction each time the dimension of this subspace increases), but choosing a less restrictive law before it is necessary to do so risks more retractions (Nature is free to exhibit only reactions compatible with the current span until the method retracts to a smaller subspace). Worlds in which the tightest such theory survives forever are simplest and anomalies occur when the dimensionality of the observed reactions increases.

0.4 A Transfinite Efficiency Argument

The preceding problems are all solvable under finite retraction bounds, but the efficiency argument can be extended to a wide range of problems for which no such bound exists. To see how, suppose that there is a curtain behind which there may be at most one box that may contain some particles. If the box exists, it is revealed, eventually, at a time of Nature's choosing (prior to the removal of any eggs) and after the box is revealed, Nature must reveal each egg in the box, without replacement, at times of her own choosing. The size of the revealed box provides a visual upper bound on how many eggs it can contain. The question is simply whether there is a box behind the curtain and, if so, how many eggs it contains (a more interesting version with the same structure: "will we ever discover an upper bound on the number of kinds of subatomic particles and if so, what are the finitely many conservation laws governing their interactions?"). The Ockham hypothesis prior to seeing the box is that no box exists.

There is no finite, a priori bound on the number of retractions required to solve the whole problem (a box of any size might appear), so the finitely bounded efficiency arguments presented above do not apply. But the least, infinite retraction bound, ω, is achieved by *every* solution (by definition, convergence implies at most finitely many retractions). So it seems that one cannot derive Ockham's razor from efficient convergence using infinite bounds, either.

R. Freivalds and C. Smith (1993) have devised an ingenious solution to this dilemma. Counting retractions *up* to an infinite, ordinal number is indeed trivial, but counting *down* from such a number is not— because there are no infinite, descending chains of ordinal numbers. Assume that there is an **accountant** who manages the learner's resources. The accountant begins with some initial, possibly infinite, ordinal **allotment** of funds. Occasionally, she makes unit **withdrawals** to cover the retractions performed by the method (assume that each retraction costs one unit) and specifies the current account balance at each stage. The accountant is required to be **empirical**, in the sense that her actions depend only on the inputs currently available to the learner.[16] If the total sum withdrawn in each world matches the total number of retractions performed by the learner in that world, say that the accountant **covers** the learner's retractions. When a unit is withdrawn from limit ordinal funds λ, the balance must somehow be reduced to an ordinal less than λ. Since $\lambda - 1$ is undefined in standard ordinal arithmetic, it is up to the accountant to decide what the new balance will be. But she does so at her peril, for if she selects a value that is too low, the learner may needlessly run out of funds prior to convergence, in which case she and the learner fail as a team. Tentatively, say that a method **solves** a problem with **α-bounded retractions** just in case the method solves the problem and there exists an accountant whose allotment is initialized to α whose withdrawals cover the method's retractions. A problem is **solvable** with α-bounded retractions just in case there exists a method that solves it with α-bounded retractions. The **retraction complexity** of a problem is the least ordinal bound under which the problem is solvable.

Recall the example of the curtain that may conceal an egg-box. Suppose we follow the obvious Ockham method, guessing that there is no box until it is seen and guessing the observed number of eggs thereafter. An accountant starting with allotment ω can withdraw a unit and decrement to n when a box (of size n) is observed and can withdraw one more unit for each egg observed thereafter. Since we already know that no finite allotment suffices, the problem's retraction complexity is exactly ω. The idea generalizes in a natural way to higher ordinals. For example, putting the box-and-curtain problem behind another curtain gives rise to a problem

[16]Given the input stream, the accountant can reconstruct the behavior of the learner by simulating the learner on the inputs, so it is not necessary to pass the learner's acts to the accountant directly.

of retraction complexity $\omega + 1$, and so forth.

Now it is possible to derive Ockham's razor from efficiency in the box-and-curtain problem. A method is **retraction-efficient** just in case it solves the problem under the least feasible retraction bound in each subproblem. Suppose that a method believes in the box before it is seen and that the accountant's initial allotment is the least feasible value ω (by the preceding argument). Nature is free to withhold the box until the method retracts to the view that there is no box (on pain of failing to converge to the right answer). Nature can continue to withhold the box until the empirical accountant withdraws a unit to cover the retraction and decrements the allotment to some finite n to balance the infinite books. Now Nature is free to exhibit a box of size $n + 1$ and to withhold eggs until the method converges to "empty box". Nature can exact another retraction for each of the $n + 1$ eggs that might fit in the box, for a total of $n + 2$ retractions, but the total withdrawal is at most $n + 1$. Since allotment ω suffices to solve this problem, the method that violates Ockham's razor is not retraction-efficient.

0.5 Retraction Efficiency Defined

The preceding concept of retraction-efficiency is too lenient to yield proposition 1 in its full generality, for consider the following variant of the box-and-curtain problem. Keeping everything else the same, add the presupposition that the box will eventually appear. This problem is easier than the original version, because one can solve it without having to "leap" to the risky conclusion that there is no box after a sufficiently long run of box-less experience. Now consider a method that violates Ockham's razor by guessing "two eggs" a priori and that follows the obvious Ockham strategy thereafter. The initial retraction can be covered by a unit withdrawal when the accountant decrements the initial allotment of ω to the observed box size. Thereafter, everything works as usual. Since no finite retraction bound suffices to solve this problem, the method is retraction-efficient in each subproblem in spite of its a priori violation of Ockham's principle.

The trouble is that both versions of the problem have retraction complexity ω, but the "box for sure" version is easier than the "possibly no box" version because the appearance of the box is certain in the former, but constitutes an anomaly with respect to uniformly box-less experience in the latter. A natural remedy is to introduce a new quantity $\overleftarrow{\omega}$ less than ω and greater than each finite n, from which no withdrawal may be taken.[17] The easy version of the problem can be solved under allotment $\overleftarrow{\omega}$, because the answer "no eggs" can be retained when the box is observed, so when the box is observed, the accountant can decrement $\overleftarrow{\omega}$ to the observed box size without making a withdrawal. More generally, let $\overleftarrow{\lambda}$ be a new number strictly between limit ordinal λ and each ordinal $\beta < \lambda$ and let the **extended ordinals** be the result of adding these new numbers to the ordinals.[18] For accounting purposes, it suffices to specify how to subtract finite quantities from extended ordinal allotments.[19] Let $\overleftarrow{\lambda} - k$ be undefined (we don't want to

[17] A more direct solution is to allow no withdrawals from limit ordinals, so that the easy version has retraction complexity ω and the hard version has retraction complexity $\omega + 1$. That works fine for retraction complexity, but it makes a mess of the deduction Ockham's razor from error efficiency (propositions 1 and 3) and destroys the natural analogy between error and retraction complexity (proposition 5).

[18] I picture $\overleftarrow{\lambda}$ as follows. Surely, there is enough in the infinite tail of λ to extract the tiny quantity n. However, removing n inflicts "indefinite, finite damage" on the tail, resulting in $\overleftarrow{\lambda}$ (think of the damaged tail as being foggy or indeterminate, like a quantum state). Until this foggy "damage" is "amputated" to leave a clean, ordinal "stump", no further withdrawals from the tail are possible.The foggy injury to $\overleftarrow{\lambda}$ doesn't infect any ordinal initial segment $\beta < \lambda$ of $\overleftarrow{\lambda}$, so after the removal of n, it is possible to decrement $\overleftarrow{\lambda}$ to any such β. Since β has an undamaged tail, it is now possible to withdraw another finite quantity from its tail.

[19] Although it is not necessary for the following results, it is interesting and natural to extend ordinal arithmetic to the extended ordinals by extended ordinal recursion as follows, where $\mathrm{osup}(S)$ denotes the *ordinal* supremum of the set S.

$$
\begin{aligned}
\overrightarrow{\alpha} &= \mathrm{osup}\{\alpha\};\\
s(\alpha) &= \alpha + 1 \text{ if } \alpha \text{ is ordinal and } = \alpha \text{ otherwise};\\
\alpha + (\beta + 1) &= s(\alpha + \beta);\\
\alpha + \lambda &= \mathrm{osup}\{\overrightarrow{\alpha} + \beta : \beta < \lambda \text{ and } \beta \text{ is ordinal}\};\\
\alpha + \overleftarrow{\lambda} &= \overleftarrow{\alpha + \lambda};\\
\alpha \cdot (\beta + 1) &= (\alpha \cdot \beta) + \beta;
\end{aligned}
$$

withdraw funds from non-ordinal allotments anyway). To compute $\alpha - k$, for arbitrary ordinal α and natural number k, first write α as the unique sum $\lambda + n$, where λ is a limit ordinal, and then return $\overleftarrow{\lambda}$ if $n > k$ and return the ordinal $\lambda + (n - k)$ otherwise. The only additional assumption regarding the extended ordinals is the distribution law: $\omega \cdot \overleftarrow{\lambda} = \overleftarrow{\omega \cdot \lambda}$.[20] The new numbers suggest revised accounting rules:

1. after a unit is withdrawn from ordinal α, the new balance may not exceed $\alpha - 1$;
2. balance $\overleftarrow{\lambda}$ may be decremented to any extended ordinal $< \overleftarrow{\lambda}$;
3. no withdrawal may be taken from balance $\overleftarrow{\lambda}$.

One more refinement is required if we are to deduce Ockham's razor from efficiency, for suppose that a method starts out with the non-Ockham answer "two eggs" and the initial allotment is $\overleftarrow{\omega}$. The method can retract to "no eggs" when the box is seen and now the accountant cannot make an *immediate* withdrawal. But she can, nevertheless, decrement $\overleftarrow{\omega}$ to

$$\begin{aligned} \alpha \cdot \lambda &= \text{osup}\{\overrightarrow{\alpha} \cdot \beta : \beta < \lambda \text{ and } \beta \text{ is ordinal}\}; \\ \alpha \cdot \overleftarrow{\lambda} &= \overleftarrow{\alpha \cdot \lambda}. \end{aligned}$$

Extended ordinal exponentiation is defined analogously. Here are some examples: $\overset{\rightleftarrows}{\lambda} = \lambda$ (arrows cancel); $k + \overleftarrow{\lambda} = \overleftarrow{\lambda}$ (any finite quantity can be extracted for free from the head of damaged ordinal); $\overleftarrow{\lambda} + k = \overleftarrow{\lambda}$ (a fixed, finite quantity can't repair the indefinite, finite damage inflicted when a finite quantity is removed from the tail of an infinite ordinal); $\overleftarrow{\lambda} + \lambda' = \lambda + \lambda'$ (since any finite quantity can be extracted from the head of λ', it suffices to repair the indefinite, but finite damage in $\overleftarrow{\lambda}$); $\overleftarrow{\lambda} + \overleftarrow{\lambda}' = \lambda + \overleftarrow{\lambda}' = \overleftarrow{\lambda + \lambda'}$ (there is enough in the head $\overleftarrow{\lambda}'$ to repair the tail of λ, but the tail of $\overleftarrow{\lambda}'$ is still damaged); $\overleftarrow{\lambda} \cdot \lambda' = \lambda \cdot \lambda'$ (there is no last copy of $\overleftarrow{\lambda}$ in the product so each tail is fixed by the head of its successor); $\overleftarrow{\lambda} \cdot (\beta + 1) = (\lambda \cdot \beta) + \overleftarrow{\lambda} = \overleftarrow{\lambda + (\beta + 1)}$ (each copy of $\overleftarrow{\lambda}$ is repaired by its successor, except for the last); $\lambda \cdot \overleftarrow{\lambda}' = \overleftarrow{\lambda} \cdot \overleftarrow{\lambda}' = \overleftarrow{\lambda \cdot \lambda'}$ (if $\lambda \cdot \lambda'$ has its tail damaged, the damage is not in any of the copies of λ since each is repaired by its successor, so the damage must be in the manner in which the undamaged copies of λ are ordered, namely λ').

[20] The distribution law is used below only in the theory of error complexity. Notice that it is one of the clauses in the recursive definition of multiplication presented in footnote 19. Its intuitive motivation in terms of damage is given in the gloss of the last example in footnote 19.

$n+1$ when a box of size n is observed, using the surplus unit later as a "slush fund" to cover the needless retraction of "two eggs".

Once again, the culprit is undue tolerance for slush funds in the proposed definition of efficiency. A natural remedy is to require that that the accountant be **stingy**, in the sense that she never allocates more funds than necessary in subproblems of lower retraction complexity.[21] Now, the original version of the problem has retraction complexity ω, whereas the easy version has complexity $\varpi < \omega$, so the concept of retraction complexity is sensitive enough to distinguish the two problems.[22] Furthermore, retraction-efficiency implies Ockham's razor in the revised problem, for suppose that the method starts out with "one egg" under allotment ϖ. Nature is free to exhibit a box of size 0. The stingy accountant must decrement the allotment immediately to 0 without withdrawing a unit from ϖ, so no withdrawals can ever be made to cover the eventual retraction of "one egg". Finally, Ockham's razor is still deducible in the original version of the problem. For suppose that the method outputs an answer entailing that there is a box prior to seeing the box. Nature can withhold the box until the method retracts to "no box" and the accountant decrements ω. If the accountant decrements ω to some finite n, then the earlier argument works. If she decrements ω to ϖ, Nature presents a box of size 0 and the stingy accountant must decrement ϖ immediately to 0 without withdrawing a unit, so the second retraction (of "no box") is never covered by a withdrawal.

I close the section with a precise definition of the concept of retraction complexity that has just been outlined. Accountant C **covers** the retractions of method M in world w **starting**

[21] Another obvious remedy is to require that retractions be covered by *immediate* withdrawals. But that strict requirement is not feasible when we turn to the deduction of Ockham's principle from error-efficiency since errors, unlike retractions, cannot be noticed immediately by an empirical accountant.

[22] We already know that ω suffices for the hard problem and that ϖ suffices for the easy one. We also know that no finite n suffices for either. The argument that the retraction complexity of the original problem exceeds ϖ is similar to the deduction of Ockham's razor. Nature can withhold the box until the method converges to "no box". Now a box of size 0 can be exhibited. The stingy accountant must immediately decrement ϖ to 0, without making a withdrawal from ϖ, so the inevitable retraction of "no box" is never covered by a withdrawal.

from finite input sequence σ just in case the total withdrawal by C in w after σ is at least as great as the total number of retractions by M in w after σ. The **retraction complexity** $r(P;\sigma)$ of problem P given σ is the least extended ordinal α such that some method solves the subproblem of P rooted at σ under retraction bound α (solution under bound α remains to be defined). If there is no such α, then let $r(P;\sigma) = \infty$. Say that C is α-**retraction-stingy in** P **at** σ just in case for each extension τ of σ such that $r(P;\tau) < \alpha$, C allots $\leq r(P;\tau)$ on τ. Now define: M **solves the subproblem** of P **rooted at** σ **under retraction bound** α just in case M solves the subproblem of P rooted at σ and there is an accountant C such that

1. for each world w in P that presents σ, C covers the retractions of M in w starting from σ,
2. C allots α at σ, and
3. C is α-retraction-stingy in P at σ.[23]

Say that M **is retraction-efficient** in P just in case $r(P;\sigma) < \infty$ and for each σ compatible with P, M solves the subproblem of P rooted at σ under retraction bound $r(P;\sigma)$.[24] No method is retraction-efficient if it is impossible to solve P under an extended ordinal bound, so say that P is **efficiently solvable** given σ just in case $r(P;\sigma) < \infty$.

[23]The definition is an extended ordinal recursion (the extended ordinals are well-ordered). When $\alpha = 0$, the third condition becomes vacuous, so the base case is well-defined.

[24]One might expect that stinginess does the work of requiring retraction-efficiency in each subproblem, but that is not so: another way to hide violations of Ockham's razor is to make unnecessary withdrawals to cover future violations of Ockham's razor. For suppose that there is a box of size three on the table and that prior to seeing any eggs God happens to tell us that the box contains at most one egg. The retraction-stingy accountant decrements the allotment immediately to unity and withdraws one unit. The method now leaps to the conclusion that the box has one egg, violating Ockham's razor. Nature withholds eggs until the method retracts to "no eggs" and then shows the egg, eliciting two retractions in the subproblem rooted at God's announcement, but the accountant can withdraw another unit when the egg appears, covering all the retractions, so the method succeeds with the fewest possible retractions in the overall problem. It does not do so in the subproblem rooted at God's announcement, however, since at most one unit can be withdrawn in that subproblem (recall that the withdrawal at the announcement does not count in the corresponding subproblem).

0.6 Error-Efficiency Defined

Solution under error bound α is defined just like its contraction counterpart, except for the following, two amendments.

1. the accountant's total withdrawal in a world after σ must be as great as the total number of errors committed from the end of σ (inclusive) onward;[25]

2. the accountant may withdraw any finite number of available units at one time.

The first condition is obvious. The second arises from an essential difference between retractions and errors: retractions are directly observable by the (empirical) accountant, whereas errors are not.[26] Like the learner, the accountant must wait for a concrete empirical anomaly indicating potential trouble with the learner's current output.[27] Since Nature is free to withhold the sign as long as she pleases, the learner may produce an answer arbitrarily often before the occurrence of the sign indicates trouble. It makes sense, therefore, to allow the accountant to withdraw a lump sum covering all prior stages at which the current theory was output when the sign finally appears. Indeed, it is necessary to do so, for consider the simple question whether there is a ball behind the curtain. If one insists upon unit withdrawals, as in the retraction case, then

[25]Recall that we start counting retractions only after the end of σ. That is because errors occur "at an instant", whereas retractions and withdrawals are essentially extended across successive stages of inquiry.

[26]Some authors have excluded error avoidance as a constraint on scientific method because it is no easier for a third party to tally the scientist's errors than it is for the scientist to avoid making them (cf. Donovan et al. 1992). But that conclusion is premature, as is clear from proposition 2.

[27]It should not be supposed that errors are always verifiable, as in the egg-box problem. If the question is whether the inputs will converge to green or to blue and it is known a priori that they will alternate between green and blue at most n times, then the accountant will *provisionally* conclude that "convergence to green" was an error when a blue input is received, even though "convergence to green" is not conclusively refuted by the blue input. Hence, the accountant may have to over-withdraw. For suppose the method says "convergence to green" for a long time and then says "convergence to blue" once before converging to "convergence to green". Then when the blue observation is seen, the accountant may agree with the learner that "convergence to green" is an error and cover all the putative errors with a huge lump withdrawal, when in fact only one error is committed.

this problem is not solvable under any extended ordinal error bound α; an unattractive result.[28] Given lump sum withdrawals, initial allotment ω suffices, for if the ball appears at stage n, it suffices to withdraw the lump sum n from ω when the ball appears. Error bound $\overleftarrow{\omega}$ does not suffice, however,[29] so the error complexity of the problem, under the proposed definition, is exactly ω.

0.7 Simplicity, Complexity, and Anomalies Defined

Relative to problem P, each world w will be assigned an extended ordinal **simplicity** degree $s(w; P)$. In accordance with the approach sketched earlier, this will be done so that worlds that present fewer potential anomalies (relative to P) will be assigned lower simplicity degrees.

It helps to introduce a few auxiliary concepts that can be defined in terms of such simplicity degrees. The **simplicity** degree $s(A; \sigma, P)$ of answer A in problem P is defined as the least upper (extended ordinal) bound on the simplicity degrees of the worlds presenting inputs σ in which A is correct in w. The **anomaly complexity** $s(P; \sigma)$ of problem P given finite input sequence σ is defined as the least upper bound on the simplicity degrees of the worlds in P that present σ.[30] World simplicity will be defined so that anomaly complexity determines retraction

[28] For suppose there is no ball. Then since there is no infinite descending chain of ordinal allotments and at most one unit is withdrawn at each decrement, the accountant withdraws at most j units in the limit. Also, since the learner succeeds, she converges eventually to "no ball". Nature is now free to withhold the ball until the learner outputs "no ball" at least $j + 2$ times before presenting the ball. When the ball is presented, stinginess demands that the allotment drop immediately to zero, with a withdrawal of at most one unit. Hence, the total withdrawal is at most $j + 1$, but at least $j + 2$ errors were committed by the learner.

[29] For suppose the accountant does not decrement $\overleftarrow{\omega}$ prior to seeing the ball. Nature is free to withhold the ball until the learner converges to "no ball". Then when the ball is revealed, stinginess forces the accountant to decrement $\overleftarrow{\omega}$ immediately to zero. Since no withdrawal is possible from $\overleftarrow{\omega}$, the total withdrawal is zero, so the errors are never covered. Now suppose that the accountant does decrement $\overleftarrow{\omega}$ prior to seeing the ball, say to k. No withdrawal is allowed from $\overleftarrow{\omega}$. Nature is free to withhold the ball until the learner converges to "no ball" and produces that answer at least $k + 1$ times.

[30] Anomaly complexity generalizes two standard, topological complexity concepts (cf. Kechris 1991). Cantor-

and error complexity as follows.

Proposition 5 *Let $s(P;\sigma) < \infty$. Then*

1. $r(P;\sigma) = s(P;\sigma)$.

2. $e(P;\sigma) = \omega \cdot s(P;\sigma)$.

3. *Hence, P is efficiently solvable just in case $s(P) < \infty$.*

An intriguing corollary of proposition 5 is that each retraction is worth ω errors: $e(P;\sigma) = r(P;\sigma) \cdot \omega$. That is because Nature can elicit arbitrarily many errors from us by coaxing us to converge to the wrong answer before revealing the crucial experience that leads us to retract it.

The definition of world simplicity is by recursion on the extended ordinals. Here is the base case.

> A is 0-**verified** by σ in P just in case each world w in P that presents σ satisfies A.
>
> $s(w;P) = 0$ just in case w presents some finite sequence σ that 0-verifies some answer to P.

The first clause is straightforward: 0-verification is just verification (relative to the presupposition of P). The second clause says that worlds of simplicity 0 (i.e., worlds of maximum complexity in P) are just the worlds that eventually verify some answer. That is intuitive in light of the examples, for recall the problem of a visible egg-box of size n. The least simple worlds in this problem are those in which the box is full, at which time the answer "n eggs" is verified.

Now for the inductive case. We are entitled to assume that β-verification and $s(w;P) = \beta$ are already defined for all extended ordinals $\beta < \alpha$. It is also defined whether $s(w;P) < \alpha$

Bendixson rank corresponds to ordinal anomaly complexity when the problem is maximally refined (each answer is compatible with exactly one input stream). Kuratowski's difference hierarchy corresponds to ordinal anomaly complexity when the question has two answers (yes or no). Anomaly complexity refines both by adding non-ordinal complexity classes and allows for questions that are neither binary nor maximally refined.

or whether $s(w; P) \geq \alpha$, since both depend only on values of $s(w; P)$ less than α. Begin by generalizing the concept of verification as follows:

> A is **α-verified** by σ in P just in case each world w in P that presents σ such that $s(w; P) \geq \alpha$ satisfies A.

Hence, α-verification is just verification relative to the additional presupposition that the actual world is at least α-simple. The idea is of interest in its own right. Think of "Normal" science (Kuhn 1996) as waiting for verification relative to a tacit simplicity presupposition. Scientific revolutions occur when this simplicity presupposition is undermined, after which we must wait for another answer to be verified relative to a weaker simplicity presupposition.

The inductive case of the definition of simplicity can be understood by "working backward" from proposition 5.1. Suppose that some answer A is α-verified, where α is an ordinal. If we output A under allotment α, we are safe from bankruptcy, so far as retractions are concerned, for if A is false, a future drop in ordinal complexity provides both a signal to retract A and an opportunity for the stingy accountant to withdraw a unit from α to cover the retraction.

When $\alpha = \overleftarrow{\lambda}$, however, it is not necessarily safe to output α-verified answer A, because the drop in anomaly complexity that signals potential trouble for A does not afford a chance to make a withdrawal. But it is safe to output A if we already know that it *will* be safe, in the sense of the preceding paragraph, to output A when the problem's complexity finally drops to an ordinal value. Say that a possible extension τ of σ is α-**dry** given σ just in case for each τ' extending σ and properly extended by τ, if $s(P; \tau') < \alpha$ then $s(P; \tau')$ is non-ordinal. So if α is non-ordinal, one never gets a chance to make a withdrawal along τ after σ has been received. Hence, it is only safe to output an answer under a non-ordinal allotment $\overleftarrow{\lambda}$ at σ if the answer is $s(P; \tau)$-verified in each $\overleftarrow{\lambda}$-dry extension of σ. Now define:

> $s(w; P) = \alpha$ just in case $s(w; P) \geq \alpha$ and w eventually presents σ such that:
>
> s1. A_w is α-verified by σ in P and

s2. if α is not an ordinal, then for each τ that is α-dry in P given σ, A_w is $s(P;\tau)$-verified by τ in P.

If there is no such extended ordinal α, let $s(w;P) = \infty$.

An **anomaly** of severity α occurs when the anomaly complexity of P drops from ordinal value α to some lower value. This implies that the simplest worlds have all been eliminated, so the reason for thinking that one's previous answer was simplest has been undercut. There is a sense in which reductions in non-ordinal complexity may also be viewed as anomalies, but only in a methodologically "benign" sense, for the simplest answer can never change when non-ordinal complexity drops. Since the anomaly complexity of a problem never drops in its simplest worlds, these are the worlds that are "uniformly" anomaly-free, which explains why they are the simplest. Also, the anomaly complexity of a problem is the least upper bound on the severity degrees of potential anomalies, as its name suggests. The **retention principle**, which characterizes retraction-efficiency according to proposition 3, requires that retractions occur only when anomalies occur.

To illustrate these concepts, consider the problem in which we are required to determine whether there is a box behind the curtain and, if so, how many eggs it contains. Worlds in which there is a box of size n containing k eggs have simplicity degree $n - k$. When a box of size n has been seen and k eggs have been observed, the anomaly complexity of the problem is $n - k$. Hence, the complexity of the problem eventually drops to the actual world's simplicity, and each drop from an ordinal value is an anomaly. The most severe such anomaly is when the box is first observed, which is an anomaly of order ω. When the kth egg is seen in a box of size n, an anomaly of order $n - k$ occurs. The world in which there is no box has simplicity ω, making it the simplest world, so the whole problem has complexity ω. At each anomaly a new answer emerges as the simplest answer compatible with experience.

Next, consider the variant of the preceding problem in which we only have to count the number of eggs behind the curtain, without saying whether or not there is a box. Consider the world in which no box exists. This world $\overleftarrow{\omega}$-verifies "no eggs". Moreover, each $\overleftarrow{\omega}$-dry τ is a finite input sequence along which the box appears at the last stage if it appears at all. Prior to seeing the box along such a sequence, the answer "no eggs" is $\overleftarrow{\omega}$-verified. Since no eggs are revealed until after the box appears, "no eggs" is n-verified when a box of size n appears. Hence, clause (s2) is satisfied, so the simplicity of the no box world is $\overleftarrow{\omega}$ rather than ω. Since this is the simplest world in the problem, the complexity of the problem is also $\overleftarrow{\omega}$. The appearance of a box of size n reduces the problem's anomaly complexity from $\overleftarrow{\omega}$ to n, but does not count as an anomaly since $\overleftarrow{\omega}$ is not an ordinal. Intuitively, the appearance of the box is not disturbing in this problem, since such an appearance is already built into the problem's presuppositions a priori.

0.8 Ockham's Razor Defined

The most natural statement of Ockham's razor is to output an answer in response to finite input sequence σ only if it is among the simplest answers compatible with σ. This principle is necessary for both error and retraction-efficiency.[31] On the other hand, it is not sufficient for either kind of efficiency, even in problems in which the simplest answer in light of current experience is always unique. For example, suppose that we will eventually either see a box or hear a bell, but not both. If we see a box, we must report the number of eggs in the box. If we hear a bell, the right answer is to report the time at which the bell rings. The answer "empty box" is the unique, simplest answer a priori[32] but no method that outputs this answer is either error-efficient or retraction-efficient.[33]

The preceding problem has no simplest worlds a priori. This situation is epistemically

[31] Necessity is an immediate corollary of propositions 1 and 7.

[32] Its simplicity degree is $\overleftarrow{\omega}$.

[33] For suppose we output "empty box" a priori. Then Nature is free to ring the bell right away. The stingy accountant must decrement immediately to zero without making a withdrawal, so our retraction/error is never covered.

"unstable", since a simplest world is destined to emerge eventually[34], at which time the overall complexity of the problem will drop. It is, therefore, natural to require that we output an answer only if it is (the unique) **simplest-world** answer, where an answer is a simplest-world answer given σ just in case the problem has simplest worlds given σ and the answer is correct in all such worlds.[35] This principle is sufficient[36] for both error and retraction-efficiency, but it is not necessary. To refute necessity, modify the box and curtain problem so that it is guaranteed that the box will occur and the question is whether it contains an even or odd number of eggs. An efficient solution may output "even" a priori, (since the answer "even" needn't be retracted after the box appears), but "even" is not a simplest-world answer, since there are no simplest worlds in the problem prior to seeing the box.

Another idea is to output an answer only if it is the uniquely simplest answer in light of experience. Unfortunately, this principle is neither necessary nor sufficint for efficiency.[37] The right principle lies somewhere between choosing arbitrarily among simplest answers and waiting until some simplest worlds compatible with experience emerge and experience singles out one of them. It allows one to output an answer A that is not yet a simplest world answer if it has already been verified that A *will* be the uniquely simplest answer just prior to the next anomaly. Accordingly, say that A is the **Ockham answer** in problem P given finite input sequence σ

[34]I.e., when the finite σ witnessing the world's maximum simplicity is experienced.

[35]Equivalently, the simplest-world answer given σ is the (unique) answer that is $s(P;\sigma)$-verified given σ.

[36]Sufficiency is an immediate corollary of propositions 1 and 7.

[37]The preceding, even/odd example shows that it is not necessary. To show that it is not sufficient requires a rather tortured example, which indicates how finely the logical hairs must be split to obtain proposition 2. Suppose that there is a box behind the curtain and that a bell may ring at stage 2. The question is whether the bell rings at stage 2 and if not, how many eggs the box contains, and if it contains at least one egg, what the size of the box is. The answer "0 eggs" has simplicity ϖ a priori. The answer "box of size n with k eggs" has simplicity $n - k$ a priori. Finally, the answer "bell at stage 2" has simplicity 0 a priori. Hence, "0 eggs" is the unique, simplest answer a priori (by a long shot). Suppose we output "0 eggs" a priori. Then Nature is free to ring the bell at stage 2. The stingy accountant must decrement the allotment of ϖ units immediately to 0 so we go bankrupt when revise "0 eggs" to "bell at stage 2". Had we stalled with '?' until after stage 2, however, we could have succeeded under retraction bound ϖ.

just in case

o1. A is $s(P;\sigma)$-verified by σ in P and

o2. if $s(P;\sigma)$ is non-ordinal, then for each τ that is $s(P;\sigma)$-dry in P given σ, A is $s(P;\tau)$-verified by τ in P.

The preceding definition is closely related to the definition of world simplicity: one need only replace α with $s(P;\sigma)$ and A_w with A. The latter substitution is necessary to yield a methodological principle, since the learner doesn't know a priori what the actual world w is.

The Ockham answer is unique, if it exists, so there is no need for extra principles that select among several Ockham answers.

Proposition 6 *If $s(P;\sigma) < \infty$, then there is at most one Ockham answer in P given σ.*

The version of **Ockham's razor** that is assumed in propositions 1 and 2 is: in problem P, output answer A in response to finite input sequence σ only if A is the (unique) Ockham answer in A given σ. The reader may verify that the intuitive instances of Ockham's razor illustrated in section 3 are all instances of this general version. Also, the Ockham answer concept is logically sandwiched between the two, plausible, alternative formulations just considered.

Proposition 7 *Let $s(P;\sigma) < \infty$. Then each of the following statements implies its successor, but not conversely.*

1. *A is the simplest world answer in P given σ;*
2. *A is the Ockham answer in P given σ;*
3. *A is one of the simplest answers in P given σ.*[38]

Furthermore, the fine distinctions matter only in subproblems of non-ordinal complexity. So in most cases, "choose the simplest hypothesis" suffices.

Proposition 8 *If $s(P;\sigma)$ is an ordinal, then the following statements are equivalent:*

[38] The implications are proved in Appendix I. The even/odd example witnesses the failure of both converses.

1. *A is the simplest world answer in P given* σ*;*

2. *A is the Ockham answer in P given* σ*;*

3. *A is the uniquely simplest answer in P given* σ*.*

0.9 Piece-meal Efficiency Defined

Although proposition 4 provides an appealing, sufficient condition for deriving Ockham's razor from efficiency, it is counterbalanced by the following, rather strong, necessary condition:

Proposition 9 *If a problem is efficiently solvable, then there is a world compatible with the problem in which some answer to the problem is eventually verified.*

It follows, for example, that the problem of counting the total number of color changes is not efficiently solvable under the presupposition that the color changes at most finitely often, for no count is ever completely verified. Hence, the presupposition is not decidable in the limit.[39]

In some such cases, Ockham's razor can still be derived from **piece-meal** efficiency (Schulte 2001). A solution to a problem is piece-meal efficient just in case it minimizes retractions in each decision problem determined by an answer to the question. The **decision problem** determined by answer A in P is the problem whose answers are A and $\neg A$, where $\neg A$ denotes the complement of A with respect to the presupposition of P. The definition of efficiency must also be adjusted so that producing any answer in P incompatible with A is understood as an output of $\neg A$, so that changing one's mind immediately from one such answer to another does not count as a retraction. Similarly, producing some answer A' distinct from A counts as an

[39]To see why, Nature is always free to stop showing us new color changes until the learner takes the bait (on pain of converging to the wrong answer) and concludes that the color will change only finitely often. Then Nature can display another color change. In the limit, the presupposition that the color changes at most finitely often is false but the learner doesn't converge to "false". As Kant might have put it, the presupposition's truth value lies "beyond all possible experience". Indeed, the finite divisibility of matter was one of Kant's antinomies of pure reason (cf. Kelly 1996, chapter 3).

error only if answer A is correct. Now suppose that the learner violates Ockham's razor in the decision problem determined by the answer "n color changes" (under the preceding conventions for interpreting other answers). One may argue, in the usual way, that the learner is not efficient in this decision problem, so the learner is not piece-meal efficient in the full problem. The same argument applies to the unrestricted versions of the curve fitting and particle conservation law problems.

In light of proposition 9, piece-meal efficiency is achievable only if each answer or its complement is verified in some world. That is not always the case. For example, suppose it is known in advance that the observations will change color only finitely often and the question is what the final color will be. The simplest answer seems to be that the convergent color will be the current color (i.e. that the color will be uniform from now on), but piece-meal efficiency does not require us to produce it .[40] There is a natural refinement of the problem that is piece-meal solvable: just ask how often the color changes and what the initial color was. Piece-meal efficiency dictates choosing an answer in the refined problem that entails that the current color will be the last. That approach, however, depends upon the selection of a refined subproblem, and the "naturalness" of the choice may depend, tacitly, upon the very simplicity intuitions we would like efficiency to explain.

0.10 References

Carnap, R. (1950) *Logical Foundations of Probability*, Chicago: University of Chicago Press.

Chart, D. (2000) "Schulte and Goodman's Riddle", *The British Journal for the Philosophy of Science*, 51: 147-149.

Churchland, P. (1981) "The Anti-Realist Epistemology of Van Fraassen's *The Scientific Image*, *Pacific Philosophical Quarterly*, 63.

[40]This problem is analogous in structure to the problem of inferring limiting relative frequencies from sequential test outcomes (cf. Salmon 1967).

Donovan, A., Laudan, L., and Laudan, R. (1992) *Scrutinizing Science*, Baltimore: Johns Hopkins University Press.

Earman, J. (1992), *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory, Cambridge: MIT Press.*, M.I.T. Press.

Ford, K. (1963) *The World of Elementary Particles*, New York: Blaisdell.

Ellis, B. (1985) "What Science Aims to Do", in *Images of Science*, P. Churchland and C. Hooker eds., Chicago: University of Chicago Press, pp. 48-74.

Forster, M. and E. Sober (2001) "Why Likelihood", forthcoming in Mark Taper and Subhash Lele eds., *Likelihood and Evidence*, Chicago and London: University of Chicago Press.

Freivalds, R. and C. Smith (1993) "On the Role of Procrastination in Machine Learning", *Information and Computation* 107: pp. 237-271.

Friedman, M. (1983) *Foundations of Space-Time Theories*, Princeton: Princeton University Press.

Glymour, C. (1980) *Theory and Evidence*, Princetion: Princeton University Press.

Goodman, N. (1983) *Fact, Fiction, and Forecast*, fourth edition, Cambridge: Harvard University Press.

Jain, S., Osherson, D., Royer, J. and Sharma A. (1999) *Systems that Learn* 2nd ed., Cambridge: M.I.T. Press.

Jeffreys, H. (1985) *Theory of Probability*, Third edition, Oxford: Clarendon Press.

Kechris, A. (1991) *Classical Descriptive Set Theory*, New York: Springer.

Kelly, K. (1996) *The Logic of Reliable Inquiry*, New York: Oxford.

Kelly, K (2000) "The Logic of Success", *British Journal for the Philosophy of Science*, 51:639-666.

Laudan, L. (1981) "A Confutation of Convergent Realism", *Philosophy of Science* 48, pp. 19-48.

Lewis, D. (1996) "Elusive Knowledge", *Australasion Journal of Philosophy* 74: 549-67.

Papineau, (1997) *The Philosophy of Science*, New York: Oxford University Press.

Popper, K. (1968) *The Logic of Scientific Discovery*, New York: Harper.

Putnam, H. (1965) "Trial and Error Predicates and a Solution to a Problem of Mostowski", *Journal of Symbolic Logic* 30: 49-57.

Salmon, W. (1967) *The Logic of Scientific Inference*, Pittsburgh: University of Pittsburgh Press.

Schulte, O. (1999a) "The Logic of Reliable and Efficient Inquiry", *The Journal of Philosophical Logic*, 28:399-438.

Schulte, O. (1999b), "Means-Ends Epistemology", *The British Journal for the Philosophy of Science*, 50: 1-31.

Schulte, O. (2000),"What to Believe and What to Take Seriously: A Reply to David Chart concerning the Riddle of Induction", *The British Journal for the Philosophy of Science*, 51: 151-153.

Schulte, O. (2001) "Inferring Conservation Laws in Particle Physics: A Case Study in the Problem of Induction", *The British Journal for the Philosophy of Science*, Forthcoming.

Sklar, L. (1977) *Space, Time, and Spacetime*, Berkeley: University of California Press.

Martin, E. and D. Osherson (1998). *Elements of Scientific Inquiry*, Cambridge: M.I.T. Press.

Rosenkrantz, R. (1983) "Why Glymour is a Bayesian", in *Testing Scientific Theories*, J. Earman ed., Minneapolis: Universitiy of Minnesota Press.

Valdes-Perez and Erdmann (1994) "Systematic Induction and Parsimony of Phenomenological Conservation Laws", *Computer Physics Communications* 83: 171-180.

van Fraassen, B. (1980) *The Scientific Image*, Oxford: Oxford University Press.

0.11 Appendix I: Proofs of Main Propositions

Finite input sequence τ **extends** finite input sequence σ just in case σ is no longer than τ and each entry in σ agrees with the corresponding entry in τ. Also, τ **properly extends** σ just in case τ extends σ but is not identical to σ. Let $P|\sigma$ denote the restriction of P to worlds in P that present σ. If α is an extended ordinal, then let $\overrightarrow{\alpha}$ denote the least ordinal $\geq \alpha$. Hence, $\overset{\rightleftarrows}{\lambda} = \lambda$. Let $V_\alpha(A; \sigma, P)$ abbreviate "A is α-verified by σ in P". Let $e(w, M; \sigma)$ denote the total number of errors committed by M in w from the end of σ onward, let $r(w, M; \sigma, P)$ be the total number of retractions performed by M in w after the end of σ and let $u(w, C; \sigma, P)$ be the total units withdrawn by C after the end of σ. Let $\sigma * x$ denote the concatenation of input x onto the end of finite input sequence σ.

Nature can exact a retraction shortfall of k from C, M in the subproblem of P rooted at σ just in case there exists a w in $P|\sigma$ such that $r(w, M; \sigma, P) - u(w, C; \sigma, P) > k$. Say that C is **committed to** covering the retractions of M in P at σ just in case C is determined never to leave a shortfall unless she runs out of funds; i.e., for each w in $P|\sigma$, if $r(w, M; \sigma, P) > u(w, C; \sigma, P)$ then for some τ extending σ in w, $C(\tau)$ allots 0. Commitment to covering errors or time can be defined similarly.

The following methods and accountants will be mentioned below.

$M_0(\sigma)$ outputs the (unique) Ockham answer A given σ, if there is one, and outputs '?' otherwise.

$C_0(\sigma)$ allots $s(P;\sigma)$ and withdraws exactly one unit at σ if an anomaly occurs at σ.

$C_1(\sigma)$ allots $\omega \cdot s(P;\sigma)$ and withdraws $n =$ the length of σ if an anomaly occurs at σ.[41]

All lemmas cited in the following proofs may be found in appendix II.

Proof of proposition 1, retraction case. Suppose that M violates Ockham's razor on finite input sequence σ. Suppose for reductio that M is retraction-efficient in P, so $s(P;\sigma) = \alpha < \infty$ and M solves the subproblem rooted at σ under retraction bound α, by proposition 5. Hence, there is an accountant C such that in each world w in $P|\sigma$,

a. M converges to the correct answer for w and

b. C covers the retractions of M in w starting from σ and

c. $C(\sigma)$ allots α and

d. C is α-retraction-stingy in P at σ.

This implies statements (a, b, c) in the antecedent of proposition 10.

Case A: α is an ordinal. Then $M(\sigma) = A$ but $\neg V_\alpha(A;\sigma,P)$, so let w in $P|\sigma$ be such that w does not satisfy A and $s(w;P) = \alpha$. Then by (a), M eventually retracts A on some τ properly extending σ in w. Hence, by (b), C must withdraw at least one unit in w after receipt of σ. Let δ be the first moment properly extending σ in w such that all withdrawals by C in w precede δ and A is retracted at least once along δ after σ. Let u be the total withdrawal in w after σ, so u is also the total withdrawal along δ after σ. Hence, $C(\delta)$ allots $< \alpha$. By proposition 18 there are three cases.

Case A.1: $C(\delta)$ allots $\overleftarrow{\alpha}$. Then $u \leq 1$ since at most one unit may be withdrawn at a stage. Since $s(w;P) = \alpha$, proposition 10 implies that Nature can exact at least a unit shortfall from C, M in the subproblem rooted at δ. Since the unit withdrawal along δ after σ is matched by a retraction in the same interval, Nature can exact a unit shortfall from C, M in the subproblem rooted at σ. Contradiction.

[41] The lump withdrawals are possible because the only ordinal allotments are limit ordinal allotments.

Case A.2: $C(\delta)$ allots β such that there exists $k > 0$, $\overrightarrow{\beta} + k = \alpha$. Then by proposition 10, Nature can exact a shortfall of at least k units from C, M in the subproblem rooted at δ. Also, $u \leq k$, just balancing the shortfall in the subproblem rooted at δ. But a retraction also occurs after σ along δ, so Nature can exact a unit shortfall from C, M in the subproblem rooted at σ.

Case A.3: $C(\delta)$ allots β such that for each $k > 0$, $\overrightarrow{\beta} + k < \alpha$. Then by (d) and proposition 10, Nature can exact an arbitrarily large shortfall from C, M in the subproblem rooted at δ. Choose this shortfall to be $u + 1$, so Nature can exact a unit shortfall from C, M in the subproblem rooted at σ.

Case B: α is not an ordinal. Let $M(\sigma) = A$. If $\neg V_\alpha(A; \sigma, P)$, revert to case A. Else, there exists an α-dry θ given σ in P such that $\neg V_{s(P;\theta)}(A; \theta, P)$. So we may choose w' in $P|\theta$ such that $s(w'; P) = s(P; \theta)$ and w' does not satisfy A. Nature is free to continue presenting inputs from w', so eventually M converges to $A_{w'}$ (by a) and hence retracts A, say at $\theta' * x$ properly extending σ. Since efficiency requires the minimization of retractions in each subproblem, the accountant must eventually make a withdrawal in the subproblem rooted at θ' to account for this retraction, say at δ extending $\theta' * x$. By (c, d) and proposition 5, C always allots the current anomaly complexity on inputs extending σ. So since θ is α-dry given σ and α is non-ordinal, C makes no withdrawal along θ after the end of σ. Hence, δ properly extends θ. So $C(\theta)$ allots $< s(w'; P) = s(P; \delta)$. By proposition 10, Nature can exact a unit shortfall in the subproblem rooted at θ and, hence, in the subproblem rooted at σ, since no withdrawals occur along θ after σ. $\dashv$.

Proof of proposition 1, error case. Suppose that M fails to be patiently Ockham, say on finite input sequence σ. Then $M(\sigma) = A$, where A is not Ockham in P given σ. Suppose for reductio that M is error-efficient in P, so M solves the subproblem rooted at σ under retraction bound α, where $\omega \cdot s(P; \sigma) = \alpha < \infty$ by proposition 5. Hence, there is an accountant C such that in each world w in $P|\sigma$,

a. M converges to the correct answer for w and

b. C covers the errors of M in w starting from σ and

c. $C(\sigma)$ allots α and

d. C is α-error-stingy in P at σ.

This implies statements (a, b, c) in the antecedent of proposition 11.

Case A: α is an ordinal. Then $M(\sigma) = A$ but $\neg V_\alpha(A; \sigma, P)$. Let w in $P|\sigma$ be such that w does not satisfy A and $s(w; P) = \alpha$. So by (a), M commits an error in w at σ, in the subproblem of P rooted at σ. Hence, by (b), C must withdraw at least one unit in w after σ (recall that errors at the root of a subproblem must be accounted for). Let δ be the first moment in w after σ at which a withdrawal is taken. Hence $C(\delta)$ allots $< \alpha$, by (c). By (b, d) and proposition 11, Nature can exact an arbitrarily large shortfall from C, M in the subproblem rooted at δ, so the shortfall can be chosen to exceed the amount withdrawn along δ after σ, contradicting (b).

Case B: α is not an ordinal. Let $M(\sigma) = A$. If $\neg V_\alpha(A; \sigma, P)$, revert to case A. Else, there exists an α-dry θ given σ in P such that $\neg A_{s(P;\theta)}(A; \theta, P)$. So we may choose w' in $P|\theta$ such that $s(w'; P) = s(P; \theta) = \gamma$ and w' does not satisfy A, so an error is committed by M in w' at σ. The accountant C must eventually make a withdrawal in the subproblem rooted at σ to account for this error, say at δ properly extending σ. By (b, d) and proposition 5, C allots $s(P; \sigma')$, for each σ' compatible with P. So since θ is α-dry given σ and α is non-ordinal and the backarrow operation distributes over products, C cannot make a withdrawal along θ after the end of σ. Hence, δ properly extends θ. So $C(\delta)$ allots $< \gamma = s(w'; P) = s(P; \delta)$. So by (d) we may apply proposition 11 to exact from C, M a shortfall greater than the total amount withdrawn between the end of σ and the end of δ. $\dashv$

Proof of proposition 2. Given proposition 1, it suffices to show that each solution to efficiently solvable P that satisfies Ockham's razor is error-efficient. Let M satisfy the principle and solve P. Method M never outputs an answer unless M_0 does. Hence, M produces no more errors than M_0. So by proposition 14, the errors of M are covered by C_1. By proposition 11, C_1 is $s(P)$-stingy, so M is error efficient. $\dashv$

Proof of proposition 3. For the "only if" side, suppose that M solves P. Suppose that M violates the retention principle on finite input sequence $\sigma * x$ compatible with P. Then $M(\sigma) = A$, $M(\sigma * x) \neq A$. For reductio, suppose that M is retraction efficient and let C witness this fact. By proposition 5 and the reductio hypothesis, (*) $C(\delta)$ allots $s(P;\delta) < \infty$ for each δ compatible with P.

Case A: $s(P;\sigma)$ is not an ordinal. Since $\sigma * x$ is compatible with P, let w be in $P|\sigma * x$.

Case A.1: suppose that for all τ extending $\sigma * x$ in w, $s(P;\tau)$ is not an ordinal. Then by (*), C never gets to make a withdrawal in w after $\sigma * x$, so C does not cover the retractions of M in the subproblem rooted at $\sigma * x$. Contradiction.

Case A.2: suppose that there is a τ extending $\sigma * x$ in w such that $s(P;\tau)$ is an ordinal. Choose τ' to be the least such. Then by (*), C cannot make a withdrawal to cover the retraction at $\sigma * x$ until after τ'. Since $s(P;\tau')$ is an ordinal, we may choose w' in $P|\tau'$ such that $s(w';P) = s(P;\tau')$. The retraction occurs in w', so Nature is free to present inputs from w' after $\sigma * x$ until C makes a withdrawal, say at δ extending $\sigma * x$. By choice of τ' and (*), δ properly extends τ'. Then $C(\delta)$ allots $< s(w';P) = s(P;\delta)$. So (by the reduction hypothesis and proposition 10) it is possible for Nature to exact at least a unit shortfall from C, M in the subproblem rooted at δ. Since no withdrawals are made before δ in the subproblem rooted at σ (by *), there is a shortfall of at least one unit in the subproblem rooted at σ in w'. Contradiction.

Case B: $s(P;\sigma) = s(P;\sigma * x)$. By case A, we may assume that $s(P;\sigma)$ is an ordinal. So since $\sigma * x$ is compatible with P, we may choose w in $P|\sigma * x$ such that $s(w;P) = s(P;\sigma * x)$. By the reductio hypothesis, C must withdraw a unit in w after $\sigma * x$ in order to account for the retraction at $\sigma * x$ in the subproblem rooted at σ, say at δ extending $\sigma * x$. Hence, $C(\delta)$ allots $< s(w;P) = s(P;\delta)$. So (by the reductio hypothesis and proposition 10) Nature can exact a shortfall from C, M of at least one unit in w after δ. Since the first withdrawal occurs after σ at δ, this yields a unit shortfall overall in the subproblem rooted at σ. Contradiction.

For the "if" side, suppose that M solves efficiently solvable P and satisfies the retention principle. Recall that $C_0(())$ allots $s(P)$, and by proposition 10, C_0 is $s(P)$-stingy in P. By the retention principle, C_0 covers the retractions of M, so (by proposition 5) M is a retraction-efficient solution to P. $\dashv$

Proof of proposition 4. Suppose that M is a self-disqualifying solution to P and P has presupposition Q. Let $\neg Q$ denote the complement of the presupposition in the completed problem P'. Now suppose for reductio that P is not solvable under any extended ordinal retraction bound. Let the **unverifiable kernel** K denote the set of all w in Q such that $s(w; P) = \infty$ and let $P|K$ be the restriction of P to worlds in K. K is nonempty, else $s(w; P) \leq \alpha$, for some ordinal α. Let A_i denote the proposition "M converges to an answer after at most i retractions" and let B_i denote "M converges to '?' after at most i retractions", both taken over the completed problem P'.

Say that a finite input sequence σ is an n-**squeezer** just in case

1. σ is compatible with K;

2. σ verifies no A_i given K;

3. σ verifies no B_i in the completed problem P';

4. σ refutes $A_0, \ldots, A_{n-1}$ and $B_0, \ldots, B_{n-1}$.

Construct an infinite, nested sequence of n-squeezers as follows. Base: the empty sequence () is a 0-squeezer. Condition 1 follows from the reductio hypothesis and proposition 5. To obtain condition 2, suppose for reductio that () verifies some A_i. Then since M solves P and converges to an answer in i retractions in each world in K, we have by proposition 5 that $s(w; P|K) \leq i$, for each w in K. So by proposition 17, K is empty. Contradiction. For condition 3, if () verifies B_i, then Q is empty so P is trivially solvable, contradicting the reductio hypothesis. Condition 4 is trivial when $n = 0$.

Induction: suppose that σ is an n-squeezer. By condition 2, there exists a world w in K that is compatible with σ and that makes A_n false. Since w is in K, and hence is in P, B_n is also false, since B_n is true only in virtual worlds outside of P. Present inputs from w, continuing from where σ left off until both B_n and A_n are refuted, which must happen because both are refutable and false in w. Add one more input, if necessary, to ensure that the τ so presented properly extends σ. Then τ satisfies conditions 1 and 4 immediately, by the choice of w. To obtain condition 2, suppose for reductio that τ verifies some A_i. Then since M solves P, $s(P|K;\tau) \leq i$, by proposition 5. Hence, $s(w; P|K) \leq i$, since w presents τ. So by proposition 17, w is not in K. Contradiction. Condition 3 holds because w is in K and hence is in P and M succeeds in P so M does not converge to '?' in w. So τ is an $n+1$-squeezer.

Let e be the unique input stream that extends the sequence of nested n-squeezers so constructed. By the supervenience assumption, there is a unique world w that presents e in the completed problem P'. By property 4 at each stage along e, M is not a self-disqualifying solution to P. Contradiction.

To refute the converse, let $A_0 =$ "the input stream will converge to 0" and let A_0 be the only answer to problem P. P is solvable with 0 retractions, but P has no self-disqualifying solution (cf. Kelly 96, chapter 4). $\dashv$

Proof of proposition 5. Let $s(P;\sigma) < \infty$.

Part 1: M_0 solves the subproblem of P rooted at σ by proposition 12. Moreover, C_0 covers the retractions of M_0, by proposition 13. For each τ, $C_0(\tau)$ allots $s(P;\tau)$, so by the corollary to proposition 10, C_0 is $s(P;\sigma)$-retraction stingy given σ in P. Hence, $s(P;\sigma) \geq r(P;\sigma)$. Also, $s(P;\sigma) \leq r(P;\sigma)$ is an immediate consequence of the corollary to proposition 10.

Part 2: M_0 solves the subproblem of P rooted at σ by proposition 12. Moreover, C_1 covers the errors of M_0, by proposition 14. For each τ, $C_1(\tau)$ allots $\omega \cdot s(P;\sigma)$, so by the corollary to proposition 11, C_1 is $\omega \cdot s(P;\sigma)$-retraction stingy given σ in P. Hence, $\omega \cdot s(P;\sigma) \geq r(P;\sigma)$. Also, $\omega \cdot s(P;\sigma) \leq r(P;\sigma)$ is an immediate consequence of the corollary to proposition 11. $\dashv$

Proof of proposition 6. Suppose that A, A' are both Ockham in P given σ and $s(P;\sigma) < \infty$.

Case A: there exists a w presenting σ such that $s(w;P) = s(P;\sigma)$. By condition (o1), A, A' are both $s(P;\sigma)$-verified by σ, so $A = A'$, since answers are mutually exclusive.

Case B: there is no w presenting σ such that $s(w; P) = s(P; \sigma)$. Hence, $s(P; \sigma)$ is non-ordinal. By (o2), we have (†) for each $s(P; \sigma)$-dry θ extending σ, $V_{s(P;\theta)}(A; \sigma, P)$ and similarly for A'. Suppose for reductio that for each $s(P; \sigma)$-dry θ extending σ, each w' in $P|\theta$ satisfies $s(w'; P) < s(P; \theta)$. Then $s(P; \theta)$ is also a non-ordinal, so each unit extension of θ is also $s(P; \sigma)$-dry extending σ. Since σ is a $s(P; \sigma)$-dry extension of itself, it follows that there is no finite extension τ of σ compatible with P such that $s(P; \tau)$ is an ordinal. Hence, there is no finite extension τ of σ compatible with P such that $s(P; \tau) = 0$. Hence, there is no world w' in P in the subproblem rooted at σ such that $s(w'; P) = 0$. By proposition 15, $s(P; \sigma) = \infty$. Contradiction. Hence, there exists a $s(P; \sigma)$-dry θ extending σ such that some w' in $P|\theta$ satisfies $s(w'; P) = s(P; \theta)$. By (†), w' satisfies both A and A', so $A = A'$, since answers are mutually exclusive. $\dashv$

Proof of proposition 7. Suppose that $s(P; \sigma) < \infty$. $1 \Rightarrow 2$: suppose that A is the simplest-world answer given σ in P. Then there exists w in $P|\sigma$ such that $s(w; P) = s(P; \sigma) < \infty$ and each such world satisfies A, so $V_{s(P;\sigma)}(A; \sigma, P)$, which is condition (o1). Now suppose $s(P; \sigma) = s(w; P)$ is non-ordinal. Then for each w' such that $s(w'; P) = s(P; \sigma)$, for each $s(P; \sigma)$-dry θ, $V_{s(P;\theta)}(A_{w'}; \theta, P)$. Since A is correct in each such world w', we have that $A_{w'} = A$. Hence, condition (o2) holds.

$2 \Rightarrow 3$: suppose condition o1 holds for A, so that $V_{s(P;\sigma)}(A; \sigma, P)$. Hence, there is no w such that A is incorrect in w but $s(w; P) = s(P; \sigma)$. Hence, there is no A' such that A' is simpler than A. $\dashv$

Proof of proposition 8. Suppose $s(P; \sigma)$ is an ordinal. So we may choose $s(w; P) = s(P; \sigma)$. We already have $1 \Rightarrow 2 \Rightarrow$, by proposition 7.

$2 \Rightarrow 3$. Suppose A is Ockham. Then $V_{s(P;\sigma)}(A; \sigma, P)$. Hence, for each w' such that $s(w; P) = s(P; \sigma)$, A is correct in w'. Hence, A is correct in w. So A is a simplest answer in P given σ. Let A' be another such answer. Then since the simplicity of A' is an ordinal, A' is correct in some world w'' such that $s(w; P) = s(P; \sigma)$. But since A is correct in w'', $A = A'$, since answers are mutually exclusive.

$3 \Rightarrow 2$. Suppose that A is the uniquely simplest answer given σ. Since some answer is correct in w, some answer has simplicity $s(P;\sigma)$, so A has simplicity $s(P;\sigma)$. Since no other answer has simplicity $s(P;\sigma)$ and answers partition the presupposition, we have $V_{s(P;\sigma)}(A;\sigma,P)$. Since $s(P;\sigma)$ is ordinal, A is Ockham given σ.

$2 \Rightarrow 1$. Suppose that A is Ockham. So $V_{s(P;\sigma)}(A;\sigma,P)$. So A is correct in w and in all w' such that $s(w';P) = s(P;\sigma)$. So A is a simplest world answer given σ. $\dashv$

Proof of proposition 9. Immediate consequence of propositions 5 and 15. $\dashv$

0.12 Appendix II: Lemmas

Proposition 10 (fundamental retraction lemma)

Let $a(\sigma)$ denote the allotment of C at σ and suppose the following:

a. M is a method that succeeds in the subproblem of P rooted at σ,

b. C is $a(\sigma)$-retraction-stingy for the subproblem of P rooted at σ,

c. C is committed to covering the retractions of M in P at σ.

d. $\overrightarrow{a(\sigma)} + k \leq s(P;\sigma)$.

Then Nature can exact the specified shortfall from M, C in the subproblem rooted at σ, under the corresponding conditions.

	$a(\sigma)$	$k \geq$	shortfall
1.	*ordinal*	0	k
2.	*infinite ordinal*	2	*arbitrary* k'
3.	*non-ordinal*	-1	$k+1$
4.		1	*arbitrary* k'

Corollary: $r(P;\sigma) \geq s(P;\sigma)$.

Proof of proposition: by extended ordinal induction on $a(\sigma)$. Suppose (a-d). The following constructions will be used repeatedly, as they embody the two basic strategies Nature can use to exact retractions from M.

Nature's ordinal trick: suppose that $s(P;\sigma) > \beta$, where β is an ordinal. Then $s(P;\sigma) \geq \beta + 1$. By proposition 15, there exists w in $P|\sigma$ such that $s(w;P) = \beta + 1$. Nature is free to present inputs from w until some τ extending σ is reached such that $M(\tau) = A_w$ (by a). Since β is an ordinal $< \beta + 1$, $\neg V_\beta(A_w;\tau, P)$. So there exists a w' in $P|\tau$ such that $A_{w'} \neq A_w$ and $s(w';P) \geq \beta$. Let δ be the first stage in w' after which no further retractions by M or withdrawals by C occur, so δ properly extends τ. Let u denote the total withdrawal after σ along δ and let r be the total number of retractions after σ along δ. Then (by a), $r > 0$, since $M(\tau) = A_w \neq A_w$. So if $a(\sigma) > 0$ then $u > 0$ (by c), so $a(\delta) < a(\sigma)$. Then say that w', δ, r, s are obtained by Nature's β trick (from M, C in P at σ).

Nature's non-ordinal trick: suppose that $s(P;\sigma) > \overleftarrow{\lambda}$, where λ is a limit ordinal. Then $s(P;\sigma) \geq \lambda$. By proposition 15, there exists w in $P|\sigma$ such that $s(w;P) = \lambda$. Nature is free to present inputs from w until some τ extending σ is reached such that $M(\tau) = A_w$ (by a). Since $s(w;P) = \lambda$, w presents τ' extending τ such that $V_\lambda(A_w;\tau', P)$. So since $s(w,P) = \lambda$, there exists a θ extending τ' such that θ is $\overleftarrow{\lambda}$-dry in P given σ and $\neg V_{s(P;\theta)}(A_w;\theta, P)$. Since θ extends τ, $V_\lambda(A_w;\theta, P)$. So since $\neg V_{s(P;\theta)}(A_w;\theta, P)$, $s(P;\theta) < \lambda \leq s(P;\sigma)$ and there exists w' in $P|\theta$ such that $s(w';P) = s(P;\theta)$ and $A_{w'} \neq A_w$. Let δ be the first moment in w' such that from δ onward in w', M never retracts and C never withdraws. Let u denote the total withdrawal after σ in w' and let r be the total number of retractions after σ in w'. Then (by a), $r > 0$, since $M(\tau) = A_w \neq A_w$. So if $a(\sigma) > 0$ then $u > 0$ (by c). Then say that w', θ, δ, r, s are obtained by Nature's $\overleftarrow{\lambda}$ trick (from M, C in P at σ).

Base case: $a(\sigma) = 0$.

Part 1 base: $k = 0$. Nature exacts a 0 shortfall because no withdrawals are possible from $a(\sigma) = 0$.

Part 1 induction: suppose $k > 0$, so $s(P;\sigma) \geq k$. Obtain $w', \delta, r > 0, u$ by Nature's k trick.

Then $k-1 \leq s(w'; P) \leq s(P; \delta)$. By the induction hypothesis, Nature can exact a shortfall of $(k-1)$ from M, C in the subproblem rooted at δ. Let w'' witness that fact. Since $a(\sigma) = 0$, $u = 0$. So we have

$$\begin{aligned} r(w'', M; \sigma, P) - u(w'', C; \sigma, P) &\geq (r + r(w'', M; \delta, P)) - (u + u(w'', C; \delta, P)) \\ &= (r - u) + (r(w'', M; \delta, P)) - u(w'', C; \delta, P)) \\ &\geq (1 - 0) + (k - 1) = k. \end{aligned}$$

Parts 2, 3, 4 are trivial when $a(\sigma) = 0$.

Induction: $a(\sigma) > 0$. Assume the theorem for all σ, C', M' and for all extended ordinals $< a(\sigma)$.

Lemma I: Let τ extend σ and suppose C', M' satisfy (a, b, c) with respect to τ. If $C'(\tau)$ allots $< s(P; \tau) \leq a(\sigma)$ then for all M', Nature can exact a unit shortfall from C', M' in the subproblem rooted at τ.

Argue as in the proof of the corollary, calling the induction hypothesis instead of the proposition itself since $C'(\tau)$ allots $< a(\sigma)$. $\dashv$

Lemma II: Let τ extend σ. If $s(P; \tau) < a(\sigma)$ then $a(\tau) \leq s(P; \tau)$.

Suppose that τ extends σ and $s(P; \tau) < a(\sigma)$. Since C is $a(\sigma)$-retraction-stingy in the subproblem rooted at σ, it suffices to show that (*) $r(P; \tau) \leq s(P; \tau)$. Let $a_0(\zeta) = s(P; \zeta)$ denote the allotment of C_0 on ζ. By proposition 13, C_0 covers the retractions of M_0, which succeeds in the subproblem rooted at τ. It suffices for (*) to show that C_0 is $a_0(\tau)$-retraction-stingy in the subproblem rooted at τ. Let ζ extend τ and suppose that $r(P; \zeta) < a_0(\tau) = s(P; \tau)$. Suppose for reductio that $s(P; \zeta) > r(P; \zeta)$. So there exist M', C' witnessing this fact. Hence, (d) C' allots $< s(P; \zeta)$ on ζ. Also, (a) M' succeeds in the subproblem rooted at ζ, (b) C' is $a_0(\zeta)$-retraction-stingy in the subproblem rooted at ζ and (c) C' is committed to covering the retractions of M' in the subproblem rooted at ζ. Apply lemma I to exact a unit shortfall from M', C' in the subproblem rooted at ζ, which contradicts the choice of M', C'. $\dashv$

Part 1: suppose that $a(\sigma) > 0$ is an ordinal, $k \geq 0$ and $a(\sigma) + k \leq s(P; \sigma)$.

Case A: $a(\sigma) + k$ is a successor ordinal. Obtain w', δ, r, u by Nature's $a(\sigma) + k - 1$ trick. Then $u, r > 0$ (by a, c) because $a(\sigma) > 0$. So (by c) $a(\delta) < a(\sigma)$. Then, by proposition 16, we have three possibilities.

Case A.1: $a(\delta)$ is ordinal and $a(\delta) + u \leq a(\sigma)$. Since $s(P;\delta) = s(w';P) \geq a(\sigma) + k - 1$, we have $a(\delta) + u + k - 1 \leq s(P;\delta)$. So (a, b, c) and the induction hypothesis yield that Nature can exact a shortfall of $u + k - 1$ from M, C in the subproblem rooted at δ. Let w'' witness this fact. So we have:

$$\begin{aligned} r(w'', M; \sigma, P) - u(w'', C; \sigma, P) &\geq (r + r(w'', M; \delta, P)) - (u + u(w'', C; \delta, P)) \\ &= (r - u) + (r(w'', M; \delta, P)) - u(w'', C; \delta, P)) \\ &\geq (1 - u) + (u + k - 1) = k. \end{aligned}$$

Case A.2: $a(\delta)$ is non-ordinal and $\overrightarrow{a(\delta)} + u - 1 \leq a(\sigma)$. Since $s(P;\delta) \geq s(w';P) = a(\sigma) + k - 1$, we have $\overrightarrow{a(\delta)} + k + u - 2 \leq s(P;\delta)$. Since $u, k > 0$, $u + k - 2 \geq 0$. So (a, b, c) and the induction hypothesis, Nature can exact a shortfall of $u + k - 1$ from M, C in the subproblem rooted at δ. Continue exactly as in case A.1

Case B: $a(\sigma) + k$ is a limit ordinal λ, so $k = 0$. Obtain w', θ, δ, r, u by Nature's $\overleftarrow{\lambda}$ trick. Suppose that there exists a moment ζ extending σ along δ such that $a(\zeta) < s(P;\zeta) \leq a(\sigma)$. Let ζ be the first such. Then by Lemma I, Nature can exact a unit shortfall from M, C in the subproblem rooted at ζ. Since the only withdrawal after σ along ζ occurs possibly at ζ, Nature exacts a zero shortfall in the subproblem rooted at σ, and we are done. So we may assume w.l.o.g. that for each ζ extending σ along δ, $a(\zeta) = a(\sigma)$ or $a(\zeta) \geq s(P;\zeta)$. By lemma II, for each ζ extending σ along δ, if $a(\zeta) < a(\sigma)$ then $a(\zeta) \leq s(P;\zeta)$. Hence, for each moment ζ extending σ along δ, either $a(\zeta) = a(\sigma)$ or $a(\zeta) = s(P;\zeta)$. Since $u > 0$, there is some ζ after σ along δ such that $a(\zeta) < a(\sigma)$. Let ζ be the least such. Let ζ' extend ζ along δ. If θ properly extends ζ', then since θ is $a(\sigma)$-dry given σ, $s(P;\zeta') = a(\zeta')$ is non-ordinal, so no withdrawal is allowed at the next stage. Now suppose that ζ' extends θ. Then $s(P;\zeta') = s(w';P) = s(P;\theta)$ so $a(\zeta') = a(\theta)$. So no further withdrawals occur, and hence $u \leq 1$. Since $r \geq 1$, w' witnesses that Nature exacts a 0 shortfall in the subproblem rooted at σ.

Part 2: suppose that $a(\sigma)$ is an infinite ordinal $\lambda + n$ and $k \geq 2$ and that $a(\sigma) + k = \lambda + n + k \leq s(P;\sigma)$. Perform Nature's (ordinal) $\lambda + n + k - 1$ trick to obtain w', δ, r, u. Then since $a(\sigma)$ is infinite, $r, u > 0$. So $a(\delta) < a(\sigma)$.

Case A: $n = 0$, so $a(\delta) < \lambda$. Since $k \geq 2$, $s(P;\delta) \geq s(w';P) \geq \lambda + 1$. Hence, for all $k', \overrightarrow{a(\delta)} + k \leq s(P;\delta)$. So by (a, b, c) and the induction hypothesis, Nature can exact an arbitrary, finite shortfall in the subproblem rooted at δ, and hence at σ.

Case B: $n > 0$, so $\overrightarrow{a(\delta)} \leq \lambda + n - 1$. Since $k \geq 2$, $s(P;\delta) \geq s(w';P) \geq \lambda + n + 1$. Hence, $s(P;\delta) \geq \overrightarrow{a(\delta)}$. So by (a, b, c) and the induction hypothesis (parts 2 and 4) Nature can exact an arbitrary, finite shortfall in the subproblem rooted at δ, and hence at σ.

Part 3: Suppose $a(\sigma) = \overleftarrow{\lambda}$ and $k \geq -1$.

Case A: $k = -1$. Then $\overset{\rightleftarrows}{\lambda} - 1 = \lambda - 1 = \overleftarrow{\lambda} \leq s(P;\sigma)$. By proposition 15, there is a w in $P|\sigma$ such that $s(w;P)$ is an ordinal $< \overleftarrow{\lambda}$. Let θ be the least moment in w such that $s(P;\theta)$ is ordinal. Suppose that there exists a moment ζ extending σ and properly extended by θ such that $a(\zeta) < s(P;\zeta) \leq a(\sigma)$. Then since $a(P;\zeta)$ is non-ordinal, for each k', $a(\zeta) + k' \leq s(P;\zeta)$. So by (a, b, c) and and the induction hypothesis, Nature can exact an arbitrary, finite shortfall in the subproblem rooted at ζ, and hence in the subproblem rooted at σ. So we may assume, w.l.o.g., that for each ζ extending σ and properly extended by θ, $a(\zeta) = a(\sigma)$ or $a(\zeta) \geq s(P;\zeta)$. By lemma II, for each ζ extending σ and properly extended by θ, if $a(\zeta) < a(\sigma)$ then $a(\zeta) \leq s(P;\zeta)$. Hence, for each moment ζ extending σ along prior to θ, either $a(\zeta) = a(\sigma) = \overleftarrow{\lambda}$ or $a(\zeta) = s(P;\zeta)$, which is also non-ordinal. So no withdrawals occur along θ after σ. Since $s(P;\theta) < s(P;\sigma)$ and $s(P;\theta)$ is non-ordinal and $a(\theta) \leq s(P;\theta)$ (by lemma II), we have by (a, b, c) and the induction hypothesis that Nature can exact a zero shortfall in the subproblem rooted at θ. Since no withdrawals occur after σ along θ, w witnesses that Nature can exact a 0 shortfall in the subproblem rooted at σ.

Case B: $k = 0$. Then $\overset{\rightleftarrows}{\lambda} = \lambda \leq s(P;\sigma)$. Obtain $w', \theta, \delta, r > 0, u > 0$ by Nature's $\overleftarrow{\lambda}$ trick. Suppose that there exists a moment ζ extending σ and properly extended by θ such that

$a(\zeta) < s(P;\zeta) \leq a(\sigma)$. Then since θ is $a(\sigma)$-dry given σ, $s(P;\zeta)$ is non-ordinal, so we have that for each k', $a(\zeta) + k' \leq s(P;\zeta)$. So by (a, b, c) and and the induction hypothesis, Nature can exact an arbitrary, finite shortfall in the subproblem rooted at ζ, and hence in the subproblem rooted at σ. So we may assume, w.l.o.g., that for each ζ extending σ and properly extended by θ, $a(\zeta) = a(\sigma)$ or $a(\zeta) \geq s(P;\zeta)$. By lemma II, for each ζ extending σ along δ, if $a(\zeta) < a(\sigma)$ then $a(\zeta) \leq s(P;\zeta)$. Hence, for each moment ζ extending σ and properly extended by θ, either $a(\zeta) = a(\sigma) = \overleftarrow{\lambda}$ or $a(\zeta) = s(P;\zeta)$, which is also non-ordinal. So no withdrawals occur along θ after σ. So all the $u > 0$ units withdrawn in w' are withdrawn after θ along δ. We have two cases to consider, by proposition 16.

Case B.1: $a(\delta)$ is ordinal and $a(\delta) + u \leq a(\theta)$. Then since $a(\delta) < a(\sigma)$ (because $u > 0$) and $a(\delta) + u \leq a(\theta) \leq s(P;\theta)$, we have by (a, b, c) and the induction hypothesis that Nature can exact a shortfall of u from M, C in the subproblem rooted at δ. Let w'' witness that fact. Calculate

$$\begin{aligned} r(w'', M; \sigma, P) - u(w'', C; \sigma, P) &\geq (r + r(w'', M; \zeta, P)) - (u + u(w'', C; \zeta, P)) \\ &= (r - u) + (r(w'', M; \zeta, P)) - u(w'', C; \zeta, P)) \\ &\geq (1 - u) + u \geq 1. \end{aligned}$$

Case B.2: $a(\delta)$ is non-ordinal and $\overleftarrow{a(\delta)} + u - 1 \leq a(\theta) \leq s(P;\theta)$. Then since $a(\delta) < a(\sigma)$ (because $u > 0$) and $a(\delta) + u \leq a(\theta) \leq s(P;\theta)$, we have by (a, b, c) and the induction hypothesis that Nature can exact a shortfall of u from M, C in the subproblem rooted at δ. Continue as in the preceding subcase.

Case C: $k > 0$. Covered under part 4.

Part 4: Suppose $a(\sigma)$ is a non-ordinal $\overleftarrow{\lambda}$ and $k \geq 1$. Then $s(P;\sigma) \geq \lambda + 1$. Obtain $w', \delta, r > 0, u > 0$ by Nature's $\lambda + 1$ trick. Then $s(P;\delta) \geq s(w';P) \geq \lambda$ but $a(\delta) < a(\sigma) = \overleftarrow{\lambda}$. Hence, for all k', $\overrightarrow{a(\delta)} + k' \leq s(P;\delta)$, so by (a, b, c) and the induction hypothesis, Nature can exact an arbitrary, finite shortfall from M, C in the subproblem rooted at δ and, hence, in the subproblem rooted at σ.

Proof of corollary: Suppose for reductio that $s(P;\sigma) > r(P;\sigma)$. So let M, C solve the subproblem rooted at σ under retraction bound $r(P;\sigma)$. Proposition 18 yields three cases.

Case: $a(\sigma) = \overleftarrow{s(P;\sigma)}$. Then by part 3 of proposition 10, Nature can exact a shortfall of $0+1$ from M, C in the subproblem rooted at σ. Contradiction.

Case: $\overrightarrow{a(\sigma)} + k = s(P;\sigma)$, for some finite $k > 0$. Then $\overrightarrow{a(\sigma)} + 1 \leq s(P;\sigma)$, so by parts 1 and 3 of proposition 10, Nature can exact a unit shortfall from M, C in the subproblem rooted at σ. Contradiction.

Case: $\overrightarrow{a(\sigma)} + k \leq s(P;\sigma)$, for each finite k. Then again parts 1 and 3 (or 2 and 4) of the proposition suffice for a nonzero shortfall. Contradiction. $\dashv$

Proposition 11 (fundamental error lemma)

Let $a(\sigma)$ denote the allotment of C at σ and suppose the following:

a. *M is a method that succeeds in the subproblem of P rooted at σ,*

b. *C is $a(\sigma)$-error-stingy for the subproblem of P rooted at σ,*

c. *C is committed to covering the errors of M in P at σ.*

d. *$a(\sigma) < \omega \cdot s(P;\sigma)$.*

Then Nature can exact an arbitrary, finite, error shortfall from M, C in the subproblem rooted at σ.

Corollary: $e(P;\sigma) \geq \omega \cdot s(P;\sigma)$.

The corollary is immediate from the proposition. The proposition is shown by extended ordinal induction on $a(\sigma)$. Suppose (a-d). Let n be the arbitrary error shortfall we would like Nature to exact from M, C in the subproblem of P rooted at σ.

Base case: $a(\sigma) = 0$. Then $s(P;\sigma) > 0$, by (d). By proposition 15, let w in $P|\sigma$ satisfy $s(w;P) = 1$. Nature is free to present w until some moment τ by which M outputs A_w at least $k + n$ times, by (a). Since $s(w;P) > 0$, we have $\neg V_0(A_w;\tau, P)$. So there exists w' in $P|\tau$ such that $A_{w'} \neq A_w$. Then M commits $k+n+1$ errors in w' but the withdrawal is at most $a(\sigma) = 0$, so Nature exacts an error shortfall of n from M, C in the subproblem of P rooted at σ.

Induction: $a(\sigma) > 0$. By (d) and proposition 19, we have two cases.

Case A: There exists $k > 0$ and ordinal $\beta < s(P;\sigma)$ such that $a(\sigma) \leq (\omega \cdot \beta) + k$. Since β is an ordinal, $s(P;\sigma) \geq \beta + 1$. By proposition 15, there exists w in $P|\sigma$ such that $s(w;P) = \beta + 1$. Nature is free to present inputs from w until some τ extending σ is reached such that M outputs A_w at least $n+k+1$ times (by a). Since β is an ordinal $< \beta + 1$, $\neg V_\beta(A_w; \tau, P)$. So there exists a w' in $P|\tau$ such that $A_{w'} \neq A_w$ and $s(w';P) = \beta$. Let δ be the first stage in w' after which no further errors by M or withdrawals by C occur, so δ properly extends τ. Let $e > k+n$ be the total number of errors committed along δ from σ onward.

Case A.1: $\beta = 0$. Then since $a(\sigma) \leq k$ and $e > k+n$, w' witnesses that Nature exacts an error shortfall of n from M, C in the subproblem rooted at σ.

Case A.2: $\beta > 0$. Then since $e > k+n$, we have (by c) that $\overrightarrow{a(\delta)} + k + n < a(\sigma) \leq (\omega \cdot \beta) + k$, so $\overrightarrow{a(\delta)} + n < \omega \cdot \beta$. Also, $s(P;\delta) \geq s(w';P) \geq \beta$, so $\overrightarrow{a(\delta)} + n < \omega \cdot s(P;\delta)$. So by the induction hypothesis, Nature can exact an arbitrary error shortfall in the the subproblem rooted at δ, and hence in the subproblem rooted at σ.

Case B: $s(P;\sigma)$ is a limit ordinal λ and $a(\sigma) = \overleftarrow{\omega \cdot \lambda} = \omega \cdot \overleftarrow{\lambda}$. Obtain w', θ, δ, r, s by Nature's $\overleftarrow{\lambda}$ trick from M, C in P at σ (cf. proof of proposition 10). Suppose that there exists a moment ζ extending σ and properly extended by θ such that $a(\zeta) < \omega \cdot s(P;\zeta) \leq a(\sigma)$. Then by (a, b, c) and and the induction hypothesis, Nature can exact an arbitrary, finite error shortfall in the subproblem rooted at ζ, and hence in the subproblem rooted at σ. So we may assume, w.l.o.g., that for each ζ extending σ and properly extended by θ, $a(\zeta) = a(\sigma)$ or $a(\zeta) \geq \omega \cdot s(P;\zeta)$.

By the following lemma, for each ζ extending σ along δ, if $a(\zeta) < a(\sigma)$ then $a(\zeta) \leq \omega \cdot s(P;\zeta)$. Hence, for each moment ζ extending σ prior to θ, either $a(\zeta) = a(\sigma) = \overleftarrow{\omega \cdot \lambda}$ or $a(\zeta) = \omega \cdot s(P;\zeta) = \overleftarrow{\omega \cdot \overrightarrow{s(P;\zeta)}}$ (by the arrow distribution rule), which is also non-ordinal. So no withdrawals occur along θ after σ. So the $u > 0$ units withdrawn in w' are withdrawn after θ along δ. Since $a(\theta) \leq \omega \cdot s(P;\theta)$, we have that $a(\delta) < a(\theta) \leq \omega \cdot s(P;\theta) = \omega \cdot s(w';P) = \omega \cdot s(P;\delta) \leq a(\sigma)$. So by (a, b, c) and the induction hypothesis, Nature can exact an arbitrary, finite error shortfall from M, C in the subproblem rooted at δ, and hence in the subproblem rooted at σ.

Lemma I: Let τ extend σ. If $\omega \cdot s(P;\tau) < a(\sigma)$ then $a(\tau) \leq \omega \cdot s(P;\tau)$. For suppose that τ extends σ and $\omega \cdot s(P;\tau) < a(\sigma)$. Since C is $a(\sigma)$-error-stingy in the subproblem rooted at σ, it suffices to show that (*) $e(P;\tau) \leq \omega \cdot s(P;\tau)$. Let $a_1(\zeta) = \omega \cdot s(P;\zeta)$ denote the allotment of C_1 on ζ. By proposition 14, C_1 covers the errors of M_0, which succeeds in the subproblem rooted at τ. Since $a_1(\tau) = \omega \cdot s(P;\tau)$, it suffices for (*) to show that C_1 is $\omega \cdot s(P;\tau)$-error-stingy in the subproblem rooted at τ. Let ζ extend τ and suppose that $e(P;\zeta) < a_1(\tau) = \omega \cdot s(P;\tau)$. Suppose for reductio that $\omega \cdot s(P;\zeta) > e(P;\zeta)$. So there exist M', C' witnessing this fact. Hence (d) C' allots $< \omega \cdot s(P;\zeta) \leq a(\sigma)$ on ζ. Also, (a) M' succeeds in the subproblem rooted at ζ, (b) C' is $a_1(\zeta)$-error-stingy in the subproblem rooted at ζ, and (c) C' is committed to covering the retractions of M' in the subproblem rooted at ζ. Apply the induction hypothesis to exact an arbitrary, finite error shortfall from M', C' in the subproblem rooted at ζ, contradicting the choice of M', C'. $\dashv$

Proposition 12 (success lemma) *If $s(w;P) < \infty$, then M_0 converges to A_w in w.*

Case A: $s(w;P)$ is an ordinal β. Then w presents τ such that $V_\beta(A_w;\tau,P)$. This situation persists for each τ' extending τ in w, so M_0 converges to A_w. Case B: $s(w;P) = \overleftarrow{\lambda}$. Then w presents τ such that (i) $V_{\overleftarrow{\lambda}}(A_w;\tau,P)$ and (ii) for all $\overleftarrow{\lambda}$-dry θ given τ, $V_{s(P;\theta)}(A_w;\theta,P)$. Hence, $s(P;\tau) = \overleftarrow{\lambda}$. As before, (i) persists for each τ' extending τ in w, so it suffices that (ii) also persists at each τ' extending τ in w. So let τ' extend τ in w.

Claim: if θ is $s(P;\tau')$-dry given τ', then θ is $\overleftarrow{\lambda}$-dry given τ. For suppose that θ is $s(P;\tau')$-dry given τ'. Then for each θ' extending τ' and properly extended by θ, if $s(P;\theta') < \overleftarrow{\lambda}$ then $s(P;\theta')$ is non-ordinal. Also, for each θ' extending τ and properly extended by θ, $s(P;\theta') = \overleftarrow{\lambda} = s(w;P) = s(P;\tau)$, so for each θ' extending τ and properly extended by θ, if $s(P;\theta') < \overleftarrow{\lambda}$ then $s(P;\theta')$ is non-ordinal. So θ is $s(P;\tau)$-dry given τ.

By the claim and by the choice of τ, each $s(P;\tau')$-dry θ given τ' satisfies $V_{s(P;\theta)}(A_w;\theta,P)$.

So (ii) also persists after τ in w. So M_0 converges to A_w in w. $\dashv$

Proposition 13 (retraction covering lemma) *If $s(P;\sigma) < \infty$, then C_0 covers the retractions of M_0 starting from σ in each world in $P|\sigma$.*

Let w be in $P|\sigma$, so $s(w;P) < \infty$. Since C_0 makes a withdrawal every time an ordinal allotment is dropped, it suffices to show the following claim: if $\tau * x$ is presented by w, τ extends σ, and $M_0(\tau) \neq M_0(\tau * x)$ then (a) $s(P;\tau * x) < s(P;\tau)$ and (b) $s(P;\tau)$ is an ordinal. Suppose that $M_0(\tau) = A$. Then (i) $V_{s(P;\tau)}(A;\tau,P)$ and (ii) if $s(P;\tau)$ is non-ordinal then for each $s(P;\tau)$-dry θ extending τ, $V_{s(P;\theta')}(A;\theta',P)$. Now suppose that $M_0(\tau * x) \neq A$. Then we have two cases.

Case A: $\neg V_{s(P;\tau*x)}(A;\tau * x,P)$. Then since anomaly complexity never rises, (a) follows from (i). Next, suppose for reductio that (b) fails, so that $s(P;\tau)$ is a non-ordinal. Then $\tau * x$ is $s(P;\tau)$-dry given τ. So since $M_0(\tau) = A$, the definition of M_0 implies that $V_{s(P;\tau*x)}(A;\tau*x,P)$, which contradicts the case hypothesis.

Case B: $s(P;\tau * x)$ is a non-ordinal and there exists an $s(P;\tau * x)$-dry θ' given $\tau * x$ such that $\neg V_{s(P;\theta')}(A;\theta',P)$. Suppose for reductio that (b) fails, so $s(P;\tau)$ is a non-ordinal. Then by the case hypothesis, $\tau * x$ is $s(P;\tau)$-dry given $s(P;\tau)$. Hence, $V_{s(P;\tau*x)}(A;\tau * x,P)$. Since $M_0(\tau * x) \neq A$, the definition of M_0 implies that there is some $s(P;\tau * x)$-dry θ' given $\tau * x$ such that $\neg V_{s(P;\theta')}(A;\theta',P)$. But since $s(P;\tau * x)$ is non-ordinal, θ' is also $s(P;\tau)$-dry given τ, contradicting (ii). Hence, (b) $s(P;\tau)$ is an ordinal. So since $s(P;\tau * x)$ is not an ordinal and complexity never rises, we have (a). $\dashv$

Proposition 14 (error covering lemma) *If $s(P;\sigma) < \infty$, then C_1 covers the errors of M_0 starting from σ in each world in $P|\sigma$.*

Suppose $s(w;\sigma,P) = \alpha$. By the claim in the proof of proposition 13, C_0 makes a unit withdrawal on $\tau * x$ if M_0 retracts at $\tau * x$. By definition, C_1 withdraws the length of $\tau * x$ if C_0 makes a withdrawal on $\tau * x$. But since M_0 converges to the right answer by proposition 12, errors cease after the last retraction, so the final withdrawal by C_1 covers all of them. $\dashv$

Proposition 15 (no ordinal complexity gaps) *If there exists w in $P|\sigma$ such that $s(w;P) < \infty$, then for each ordinal $\beta \leq s(w;\sigma)$ there exists w' in $P|\sigma$ such that $s(w';P) = \beta$.*

Corollary: *If $s(P;\sigma) < \infty$, then for each ordinal $\beta \leq s(P;\sigma)$, there is a w in $P|\sigma$ such that $s(w;P) = \beta$.*

For the corollary, suppose ordinal $\beta < s(P;\sigma) < \infty$. Since $\beta < s(P;\sigma)$, there is a w in $P|\sigma$ such that $s(w';P) > \beta$. Apply the proposition.

For the proposition, suppose there exists w in $P|\sigma$ such that $s(w;P) < \infty$. For reductio, let β be an ordinal $\leq s(w;P)$ and suppose that there is no w' in $P|\sigma$ such that $s(w';P) = \beta$. Now argue by extended ordinal induction that (*) for each extended ordinal $\gamma > \beta$, there is no w' in $P|\sigma$ such that $s(w';P) = \gamma$, which (together with the reductio hypothesis) contradicts $\beta \leq s(w;P)$.

When $\gamma = 0$, we have $\gamma = 0 = \beta$, so (*) is trivially true. Now suppose that $\gamma > \beta$ and that (*) holds for each extended ordinal $\gamma' < \gamma$. Suppose for reductio that $s(w';P) = \gamma$. Since $s(w';P) \leq \gamma$, there exists τ extending σ such that (a) $V_\gamma(A_{w'};\tau,P)$. Since $s(w';P) \geq \gamma$, we have that for each $\gamma' < \gamma$, for each τ' extending σ along w', (b.i) $\neg V_{\gamma'}(A_{w'};\tau,P)$ or (b.ii) γ' is non-ordinal. We are free to let $\gamma' = \beta$ and $\tau' = \tau$. Since β is an ordinal, case (b.i) obtains, so there is a world w'' in $P|\tau$ such that $s(w'';P) \geq \beta$ and $A_{w''} \neq A_{w'}$. By (a) and the fact that $A_{w''} \neq A_{w'}$, we also have $s(w'';P) < \gamma$. Hence, $\beta \leq s(w'';P) < \gamma$ and w'' is in $P|\sigma$, contradicting the induction hypothesis. $\dashv$

Proposition 16 (decrement lemma) *If C withdraws u units after σ along δ and allots $a(\sigma)$ at σ and $a(\delta)$ at δ, then*

1. *$a(\delta) + u \leq a(\sigma)$ if $a(\delta)$ is ordinal and*
2. *$\overrightarrow{a(\delta)} + u - 1 \leq a(\sigma)$ otherwise.*

Base case: $u = 0$. If $a(\delta)$ is ordinal, then $a(\delta) + 0 \leq a(\sigma)$ since the allotment never increases. If $a(\delta) = \overleftarrow{\lambda}$ then $\overrightarrow{a(\delta)} + 0 - 1 = \overset{\rightleftarrows}{\lambda} - 1 = \lambda - 1 = \overleftarrow{\lambda} = a(\delta) \leq a(\sigma)$, since the allotment never increases.

Induction: $u > 0$. Let $\tau * x$ be the stage be the stage at which the last withdrawal along δ occurs. Then $a(\tau) > a(\tau * x) \geq a(\delta)$.

Case: $a(\tau)$ is an ordinal and $a(\tau * x)$ is an ordinal. Then $a(\tau * x) + 1 \leq a(\tau) + 1$. By the induction hypothesis, $a(\tau) + (u - 1) \leq a(\sigma)$. Hence, $a(\tau * x) + u \leq a(\sigma)$. So if $a(\delta)$ is an ordinal, then $a(\delta) + u \leq a(\sigma)$. If $a(\delta)$ is not an ordinal, however, then $a(\delta) < a(\tau * x)$, so $\overrightarrow{a(\delta)} + u \leq a(\sigma)$.

Case: $a(\tau)$ is an ordinal and $a(\tau * x)$ is not. Hence, $\overrightarrow{a(\tau * x)} \leq a(\tau)$. By the induction hypothesis, $a(\tau) + u - 1 \leq a(\sigma)$, so $\overrightarrow{a(\delta)} + u - 1 \leq \overrightarrow{a(\tau * x)} + u - 1 \leq a(\tau) + u - 1 \leq a(\sigma)$, which suffices if $a(\delta)$ is non-ordinal. If $a(\delta)$ is an ordinal, then for each k, $a(\delta) + k < a(\tau * x)$, so $a(\delta) + u \leq a(\tau * x) \leq a(\sigma)$.

Case: $a(\tau)$ is not an ordinal. Hence, for each k, $\overrightarrow{a(\tau * x)} + k < a(\tau)$. Hence, $\overrightarrow{a(\delta)} + u \leq \overrightarrow{a(\tau * x)} + u \leq a(\sigma)$, which suffices whether or not $a(\delta)$ is an ordinal. $\dashv$

Proposition 17 (infinity given infinity is infinity) *Let K denote the set of all worlds in P such that $s(w; P) = \infty$. Let $P|K$ be the restriction of problem P to worlds in K. Then for each world w in K, $s(w; P|K) = \infty$.*

Suppose w is in K. Suppose for reductio that w presents some σ such that for each w' in $P|\sigma$, if $A_{w'} = A_w$ then $s(w'; P) < \infty$. Then let α denote the ordinal supremum of the set $\{s(w'; P) : w' \text{ presents } \sigma \text{ and } A_{w'} \neq A_w\}$. Hence, $s(w; P) \leq \alpha + 1 < \infty$, contradiction. So for each w in K, for each σ presented by w, there exists a w' in $(P|\sigma)|K = (P|K)|\sigma$ such that $A_{w'} \neq A_w$. So for each w in K, $s(w; P|K) > 0$. So by proposition 15, $s(w; P|K) = \infty$, for each $w \in K$. $\dashv$

Proposition 18 (trichotomy) *Let α, β be extended ordinals. Then $\beta < \alpha$ just in case*

1. *$\beta = \overleftarrow{\alpha}$ or*
2. *$\overrightarrow{\beta} + k = \alpha$, for some finite $k > 0$ or*
3. *$\overrightarrow{\beta} + k < \alpha$, for each finite k.*

Proposition 19 (dichotomy) *Suppose $\alpha < \omega \cdot \beta$. Then either*

1. $\alpha = \overleftarrow{\omega \cdot \beta}$ *or*

2. *there exists ordinal $\gamma < \beta$ and finite k such that $\alpha \leq (\omega \cdot \gamma) + k$.*