Ockham Efficiency Theorem for Stochastic Empirical Methods

Kevin T. Kelly
Conor Mayo-Wilson

Abstract

Ockham's razor is the principle that, all other things being equal, scientists ought to prefer simpler theories. In recent years, philosophers have argued that simpler theories make better predictions, possess theoretical virtues like explanatory power, and have other pragmatic virtues like computational tractability. However, such arguments fail to explain how and why a preference for simplicity can help one find *true* theories in scientific inquiry, unless one already assumes that the truth is simple. One new solution to that problem is the Ockham efficiency theorem (Kelly 2002, 2004, 2007a-d, Kelly and Glymour 2004), which states that scientists who heed Ockham's razor retract their opinions less often and sooner than do their non-Ockham competitors. The theorem neglects, however, to consider competitors following random ("mixed") strategies and in many applications random strategies are known to achieve better worst-case loss than deterministic strategies. In this paper, we describe two ways to extend the result to a very general class of random, empirical strategies. The first extension concerns expected retractions, retraction times, and errors and the second extension concerns retractions in chance, times of retractions in chance, and chances of errors.

1 Introduction

When confronted by a multitude of competing theories, all of which are compatible with existing evidence, scientists prefer theories that minimize free parameters, causal factors, independent hypotheses, or theoretical entities. Today, that bias toward simpler theories—known popularly as "Ockham's razor"—is explicitly built into statistical software packages that have become everyday tools for working scientists. But how does Ockham's razor help one find true theories any better than competing strategies could?[1]

Some philosophers have argued that simpler theories are more *virtuous* than complex theories. Simpler theories, they claim, are more explanatory, more easily falsified or tested, more unified, or more syntactically concise.[2] However, the scientific theory that truly describes the world might, for all we know in advance, involve multiple, fundamental constants or independent postulates; it might be difficult to test and/or falsify, and it might be "dappled" or lacking in underlying unity (Cartwright 1999). Since

the virtuousness of scientific truth is an empirical question, simplicity should be the conclusion of scientific inquiry, rather than its underlying premise (Van Frassen 1980).

Recently, several philosophers have harnessed mathematical theorems from frequentist statistics and machine learning to argue that simpler theories make more accurate predictions.[3] There are three potential shortcomings with such arguments. First, simpler theories can improve predictive accuracy *even when it is known that the truth is complex* (Vapnik 1998). Thus, one is led to an anti-realist stance according to which the theories recommended by Ockham's razor should be used as predictive instruments rather than believed as true explanations (Hitchcock and Sober 2004). Second, the argument depends essentially on randomness in the underlying observations (Forster and Sober 1994), whereas Ockham's razor seems no less compelling in cases in which the data are discrete and deterministic. Third, the assumed notion of predictive accuracy does not extend to predictions of the effects of novel interventions on the system under study. For example, a regression equation may accurately predict cancer rates from the prevalence of ash-trays but might be extremely inaccurate at predicting the impact on cancer rates of a government ban on ash-trays.[4] Scientific realists are unlikely to agree that simplicity has nothing to do with finding true explanations and even the most ardent instrumentalist would be disappointed to learn that Ockham's razor is irrelevant to vital questions of policy. Hence, the question remains, "How can a systematic preference for simpler theories help one find potentially complex, true theories?"

Bayesians and confirmation theorists have argued that simpler theories merit stronger belief in light of simple data than do complex theories. Such arguments, however, assume either explicitly or implicitly that simpler possibilities are more probable *a priori*.[5] That argument is circular—a prior bias toward complex possibilities yields the opposite result. So it remains to explain, without begging the question, why a prior bias toward simplicity is better for finding true theories than is a prior bias toward complexity.

One potential connection between Ockham's razor and truth is that a systematic bias toward simple theories allows for convergence to the truth *in the long run* even if the truth is not simple (Sklar 1977, Friedman 1983, Rosenkrantz 1983). In particular, Bayesians argue that prior biases "wash out" in the limit (Savage 1972), so that one's degree of belief in a theory converges to the theory's truth value as the data accumulate. But prior biases toward complex theories also allow for eventual convergence to the truth (Reichenbach 1938, Hempel 1966, Salmon 1966), for one can dogmatically assert some complex theory until a specified time t_0, and then revise back to a simple theory after t_0 if the anticipated complexities have not yet been vindicated. One might even find the truth *immediately* that way, if the truth happens to be complex. Hence, mere convergence to the truth does not single out simplicity as the best prior bias in the short run. So the elusive, intuitive connection between simplicity and theoretical truth is not explained by standard appeals to theoretical virtue, predictive accuracy, confirmation, or convergence in the limit.

It is, nonetheless, possible to explain, without circularity, how Ockham's razor finds true theories better than competing methods can. *The Ockham efficiency theorems* (Kelly 2002, 2004, 2007a-e, Kelly 2010, Kelly and Glymour 2004) imply that scientists who systematically favor simpler hypotheses converge to the truth in the long run *more efficiently* than can scientists with alternative biases, where efficiency is a

matter of minimizing, in the worst case, such epistemic losses as the total number of errors committed prior to convergence, the total number of retractions performed prior to convergence, and the times at which the retractions occur. The efficiency theorems are sufficiently general to connect Ockham's razor with truth in paradigmatic scientific problems such as curve-fitting, causal inference, and discovering conservation laws in particle physics.

One gap in the efficiency argument for Ockham's razor is that worst-case loss minimization is demonstrated only with respect to deterministic scientific methods. Among game theorists, it is a familiar fact that random strategies can achieve lower bounds on worst-case loss than deterministic strategies can, as in the game "rock-paper-scissors", in which playing each of the three actions with equal probability achieves better worst-case loss than playing any single option deterministically can. Thus, an important question is: "Do scientists who employ Ockham strategies find true theories more efficiently than do arbitrary, *randomized* scientific strategies?" In this paper, we present a new *stochastic* Ockham efficiency theorem that answers the question in the affirmative. The theorem implies that scientists who deterministically favor simpler hypotheses fare no worse, in terms of the losses considered, than those who employ randomizing devices to select theories from data. The argument is carried out in two distinct ways, for *expected* losses and for losses *in chance*. For example, expected retractions are the expected number of times an answer is dropped prior to convergence, whereas retractions in chance are the total drops in probability of producing some answer or another. A larger ambition for this project is to justify Ockham's razor as the optimal means for inferring true statistical theories, such as acyclic causal networks. It is expected that the techniques developed here will serve as a bridge to any such theory—especially those pertaining to losses in chance.

2 Empirical Questions

Scientific theory choice can depend crucially upon subtle or arcane *effects* that can be impossible to detect without sensitive instrumentation, large numbers of observations, or sufficient experimental ingenuity and perseverance. For example, in curve fitting with inexact data[6] (Kelly and Glymour 2004, Kelly 2007a-e, 2008), a quadratic or second-order effect occurs when the data rule out linear laws, and a cubic or third-order effect occurs when the data rule out quadratic laws, etc. (figure 62). Such effects are subtle in the above sense because, for example, a very flat parabola may generate data that appear linear even in fairly large samples. For a second example, when explaining particle reactions by means of conservation laws, an effect corresponds to a reaction violating some conservation law (Schulte 2001). When explaining patterns of correlation with a linear causal network, an effect corresponds to the discovery of new partial correlations that imply a new causal connection in the network (Spirtes et al. 2000, Schulte, Luo, and Greiner 2007, Kelly 2008, Kelly 2010). To model such cases, we assume that each potential theory is uniquely determined by the empirical effects it implies and we assume that empirical effects are phenomena that may take arbitrarily long to appear but that, once discovered, never disappear from scientific memory.

Formally, let E be a non-empty, countable (finite or countably infinite) set of *empir-*

ical effects.[7] Let K be the collection of possible effect sets, any one of which might be the set of all effects that will ever be observed. We assume in this paper that each effect set in K is finite. The true effect set is assumed to determine the correctness (truth or empirical adequacy) of a unique theory, but one theory may be correct of several, distinct effect sets. Therefore, let $\mathscr{T}$, the set of possible *theories*, be a partition of K. Say that a theory T is *correct* of effect set S in K just in case S is an element of T. If S is in K, let T_S denote the partition cell of $\mathscr{T}$ that contains S, so that T_S represents the unique theory in $\mathscr{T}$ that is correct if S is the set of effects that will ever be observed. Say that $Q = (K, \mathscr{T})$ is an *empirical question*, in which K is the *empirical presupposition* and $\mathscr{T}$ is the set of *informative answers*. Call K the *uninformative* answer to Q, as it represents the assertion that some effect set will be observed. Let $\mathscr{A}_Q$ be the set of all answers to Q, informative or uninformative.

An *empirical world* w is an infinite sequence of finite effect sets, so that the nth coordinate of w is the set of effects observed or detected at stage n of inquiry. Let S_w denote the union of all the effect sets occurring in w. An empirical world w is said to be *compatible with* K just in case S_w is a member of K. Let W_K be the set of all empirical worlds compatible with K. If w is in W_K, then let $T_w = T_{S_w}$, which is the unique theory correct in w. Let $w|n$ denote the finite initial segment of w received by stage n of inquiry. Let F_K denote the set of all finite, initial segments of worlds in W_K. If e is in F_K, say that e is a *finite input sequence* and let e_- denote the result of deleting the last entry in e when e is non-empty. The set of effects presented along e is denoted by S_e, and let K_e denote the restriction of K to finite sets of effects that include S_e. Similarly, let $\mathscr{T}_e$ be the set of theories $T \in \mathscr{T}$ such that there is some S in K_e such that $T_S = T$. The *restriction* Q_e of question Q to finite input sequence e is defined as $(K_e, \mathscr{T}_e)$.

3 Deterministic Methodology

A *deterministic method* or *pure strategy* for pursuing the truth in problem Q is a function M that maps each finite input sequence in F_K to some answer in $\mathscr{A}_Q$. Method M *converges to the truth* in Q (or *converges* in Q for short) if and only if $\lim_{i\to\infty} M(w|i) = T_w$, for each world w compatible with K. Our focus is on how best to find the truth, so we consider only deterministic methods that converge to the truth.

Methodological *principles* impose short-run restrictions on methods. For example, say that M is *logically consistent* in Q if and only if M never produces an answer refuted by experience, i.e., $M(e)$ is in $\mathscr{A}_{Q_e}$, for all $e \in F_K$.

The methodological principle of main concern in this paper is Ockham's razor. Consideration of the polynomial degree example suggests that more complex theories are theories that predict more *relevant* effects, where an effect is relevant only if it changes the correct answer to Q. To capture this intuition, define a *path* in K to be a nested, increasing sequence of effects sets in K. A path $(S_0, \ldots, S_n)$ is *skeptical* if and only if T_{S_i} is distinct from $T_{S_{i+1}}$, for each i less than n. Each step along a skeptical path poses the classical problem of induction to the scientist, since effects in the next effect set could be revealed at any time in the future.

Define the *empirical complexity* $c_{Q,e}(S)$ of effect set S in K to be the result of

subtracting 1 from the length of the longest skeptical path to S in K_e (we subtract 1 so that the complexity of the simplest effect sets in K is zero). Henceforth, the subscript Q will be dropped to reduce clutter when the question is clear from context. The complexity $c_e(T)$ of theory T in $\mathscr{T}$ is defined to be the *least* empirical complexity $c_e(S)$ such that S is in T. For example, it seems that the theory "either linear or cubic" is simpler, in light of linear data, than the hypothesis "quadratic" and that the theory "quadratic" is simpler in light of quadratic data than "linear or cubic". The complexity $c_e(w)$ of world w is just $c_e(S_w)$. The nth empirical *complexity cell* $C_e(n)$ in the *empirical complexity partition* of W_K is defined to be the set of all worlds w in K such that $c_e(w) = n$.

Answer A is *Ockham* in K at e if and only if $A = K$ or A is the unique theory T such that $c_e(T) = 0$. Method M satisfies *Ockham's razor* in K at e if and only if $M(e)$ is Ockham at e. Note that Ockham's razor entails logical consistency and does not condone choices between equally simple theories. A companion principle, called *stalwartness*, is satisfied at e if and only if $M(e) = M(e_-)$ when $M(e_-)$ is Ockham at e. Ockham's razor and stalwartness impose a plausible, diachronic pattern on inquiry. Together, they ensure that theories are visited in order of ascending complexity, and each time a theory is dropped, there may be a long run of uninformative answers until a new, uniquely simplest theory emerges and the method becomes confident enough in that theory to stop suspending judgment.

Say that a skeptical path in Q is *short* if and only if, first, it is not a proper subsequence of any skeptical path in Q and second, there exists at least one longer skeptical path in Q. Then Q *has no short skeptical paths* if and only if for each e in F_K, there exists no short skeptical path in Q_e. Commonly satisfied sufficient conditions for nonexistence of short skeptical paths are (i) that all skeptical paths in Q are extendable and (ii) that $(K, \subset)$ is a ranked lattice and each theory in T implies a unique effect set. The problem of finding polynomial laws of unbounded degree and the problem of finding the true causal network over an arbitrarily large number of variables both satisfy condition (i). The problem of finding polynomial laws and the problem of finding the true causal network over a fixed, finite set of variables both satisfy condition (ii) (Kelly and Mayo-Wilson 2010b).

4 Deterministic Inquiry

We consider only methods that converge to the truth, but justification requires more than that—a justified method should pursue the truth as directly as possible. Directness is a matter of reversing course no more than necessary. A fighter jet may have to zig-zag to pursue its quarry, but needless course reversals during the chase (e.g., performance of acrobatic loops) would likely invite disciplinary action. Similarly, empirical science may have to *retract* its earlier conclusions as a necessary consequence of seeking true theories, in the sense that a theory chosen later may fail to logically entail the theory chosen previously (Kuhn 1970, Gärdenfors 1988), but needless or gratuitous reversals en route to the truth should be avoided. We sometimes hear the view that minimizing retractions is a merely pragmatic rather than a properly epistemic consideration. We disagree. Epistemic justification is grounded primarily in a method's connection with

the truth. Methods that needlessly reverse course or that chase their own tails have a weaker connection with the truth than do methods guaranteed to follow the most direct pursuit curve to the truth.

Let M be a method and let w be a world compatible with K (or some finite initial segment of one). Let $\rho(M,w,i)$ be 1 if M retracts at stage i in w, and let the *total retraction loss* in world w be $\rho(M,w) = \sum_{i=0}^{\infty} \rho(M,w,i)$. If e is a finite input sequence, define the preference order $M \leq^{\rho}_{e,n} M'$ among convergent methods to hold if and only if for each world w in complexity set $C_e(n)$, there exists world w' in empirical complexity cell $C_e(n)$ such that $\rho(M,w) \leq \rho(M',w')$. That amounts to saying that M does as well as M' in terms of retractions, *in the worst case*, over worlds of complexity n that extend e. Now define:

$$
\begin{aligned}
M <^{\rho}_{e,n} M' \quad &\text{iff} \quad M \leq^{\rho}_{e,n} M' \text{ and } M' \not\leq^{\rho}_{e,n} M; \\
M \leq^{\rho}_{e} M' \quad &\text{iff} \quad M \leq^{\rho}_{e,n} M', \quad \text{for each } n; \\
M \ll^{\rho}_{e} M' \quad &\text{iff} \quad M \leq^{\rho}_{e,n} M', \quad \text{for each } n \text{ such that } C_e(n) \text{ is nonempty.}
\end{aligned}
$$

Consider the comparison of M with alternative methods one might adopt when the last entry of finite input sequence e has just been received (and no theory has yet been chosen in response thereto). There is no point comparing one's method M in light of e with methods that did something different from M in the past along e, since the past cannot be changed. Accordingly, say that M is *efficient* in terms of retractions given e if and only if M is convergent and for each convergent competitor M' that produces the same outputs as M along e_-, the relation $M \leq^{\rho}_{e} M'$ holds. In contrast, say that M is *beaten* in terms of retractions given e if and only if there exists convergent M' that agrees with M along e_- such that $M' \ll^{\rho}_{e} M$. The concepts of efficiency and being beaten are relative to e. When such a concept holds for every e in F_K, say that it holds *always* and when the concept holds at each e' in F_K that extends e, say that it holds *from e onward*.

5 Deterministic Ockham Efficiency Theorems

A stalwart, Ockham strategy M is guaranteed to converge to the truth as long as M does not return the uninformative answer K for eternity. But other strategies also converge to the truth, so it remains to explain why one should follow Ockham's razor *now*. The Ockham efficiency theorems answer that more difficult question.

Theorem 1 (deterministic Ockham efficiency theorem) *Let the loss be retractions. Assume that question $Q = (K, \mathscr{T})$ has no short skeptical paths, that each theory in $\mathscr{T}$ is correct for a unique effect set, and that method M converges to the truth and is logically consistent. Then the following are equivalent:*

1. *method M is always Ockham and stalwart;*
2. *method M is always efficient;*
3. *method M is always unbeaten.*

Proof: Consequence of theorem 4 below. ⊣

The above theorem asserts that Ockham's razor and stalwartness are not merely sufficient for efficiency; they are both *necessary*. Furthermore, any method that is ever inefficient is also beaten at some time. Thus, convergent methods are cleanly partitioned into two classes: those that are efficient, Ockham, and stalwart, and those that are either not Ockham or not stalwart and are, therefore, beaten.

The main idea behind the proof is that nature is in a position to *force* an arbitrary, convergent method to produce successive theories $(T_{S_0}, \ldots, T_{S_n})$, with arbitrary time delays between the successive retractions, if there exists a skeptical path $(S_0, \ldots, S_n)$ in Q.

Lemma 1 (forcing deterministic changes of opinion) *Let e be a finite input sequence of length l, and suppose that M converge to the truth in Q_e. Let $(S_0, \ldots, S_n)$ be a skeptical path in Q_e such that $c_e(S_n) = n$, let $\varepsilon > 0$ be arbitrarily small and let natural number m be arbitrarily large. Then there exists world w in $C_e(n)$ and stages of inquiry $l = s_0 < \ldots < s_{n+1}$ such that for each i from 0 to n, stage s_{i+1} occurs more than m stages after s_i and $M_{w|j} = T_{S_i}$, at each stage j such that $s_{i+1} - m \leq j \leq s_{i+1}$.*

Proof: To construct w, set $e_0 = e$ and $s_0 = l$. For each i from 0 to n, do the following. Extend e_i with world w_i such that $S_{w_i} = S_i$. Since M converges in probability to the truth, there exists a stage s such that for each stage $j \geq s$, $M_{w|j} = T_{S_i}$. Let s' be the least such s. Let $s_{i+1} = \max(s', s_i) + m$. Set $e_{i+1} = w_i | s_{i+1}$. The desired world is w_n, which is in $C_e(n)$, since $S_{w_n} = S_n$. ⊣

Any non-circular argument for the unique truth-conduciveness of Ockham's razor must address the awkward question of how one does worse at finding the truth by choosing a complex theory *even if that theory happens to be true*. The Ockham efficiency argument resolves the puzzle like this. Suppose that convergent M violates Ockham's razor at e by producing complex theory T_{S_n} of complexity n. Then there exists a skeptical path $(S_0, \ldots, S_n)$ in Q_e. Nature is then in a position to *force M* back to T_{S_0} and then up through $T_{S_1}, \ldots, T_{S_n}$, by the retraction forcing lemma, for a total of $n+1$ retractions. A stalwart, Ockham method, on the other hand, would have incurred only n retractions by choosing T_{S_0} through T_{S_n} in ascending order. Therefore, the Ockham violator is *beaten* by each convergent, stalwart Ockham competitor (figure 62.b). Incidentally, the Ockham violator also traverses a *needless*, epistemic *loop* $T_n, T_0, \ldots, T_n$, an embarrassment that cannot befall an Ockham method. A similar beating argument can be given for stalwartness. Non-stalwart methods are beaten, since they start out with an avoidable, extra retraction. Furthermore, the retraction-forcing lemma allows nature to force every convergent method through the ascending sequence $T_{S_0}, T_{S_1}, \ldots, T_{S_n}$, so normal Ockham methods are *efficient* (figure 62.a). Thus, normal Ockham strategies are efficient and all non-Ockham or non-stalwart strategies are not just inefficient, but beaten as well. This sketch is suggestive but ignores some crucial cases; the details are spelled out in the proof of the more general theorem 4, which is provided in full detail in the appendix.

Theorem 1 does not imply that stalwart Ockham methods *dominate* alternative methods, in the sense of doing better in every world or even as well in every world—a

violation of Ockham's razor can result in no retractions at all if nature is kind enough to refute all simpler theories immediately after the violation occurs. Nor are stalwart Ockham methods *minimax* solutions, in the usual sense that they achieve lower worst-case loss *simpliciter*—every method's overall worst-case loss is infinite if there are worlds of every empirical complexity, as in the case of discovering polynomial laws. The unique superiority of stalwart Ockham strategies emerges only when one considers a hybrid decision rule: dominance in terms of worst-case bounds over the cells of a complexity-based partition of possible worlds. The same idea is familiar in the theory of computational complexity (Garey and Johnson 1979). There, it is also the case that cumulative computational losses such as the total number of steps of computation are unbounded over all possible worlds (i.e., input strings). The idea in computational complexity theory is to partition input strings according to *length*, so that the worst-case computational time over each partition cell exists and is finite. That partition is not arbitrary, as it is expected that computational time rises, more or less, with input length. In the case of inquiry, inputs never cease, so we plausibly substitute empirical complexity for length. Again, it is expected that retractions rise with empirical complexity. Then we seek methods that do as well as an arbitrary, convergent method, in terms of worst-case bounds over every cell of the empirical complexity partition.

Theorem 1 provides a motive for staying on the stalwart, Ockham path, but does not motivate returning to the path after having once deviated from it. In other words, theorem 1 provides an *unstable* justification for Ockham's razor. For example, suppose that method M selects T_1 twice in a row before any effects are observed, and suppose that method O reverts to a stalwart, Ockham strategy at the second stage of inquiry. Then nature can still force M to retract in the future to T_0, but O has already performed *that* retraction, so reversion to Ockham's razor does not result in fewer retractions. However, the inveterate Ockham violator retracts *later* than necessary, and efficient convergence to the truth also demands that one retract as soon as possible, if one is going to retract at all. It is common in economic analysis to discount losses incurred later, which may suggest the opposite view that retractions should be delayed as long as possible. Epistemology suggests otherwise. If nature is in a position to force one to retract T in the future by presenting only true information, then one's belief that T does not constitute knowledge, even if T is true.[8] By a natural extension of that insight, more retractions prior to arriving at the truth imply greater distance from knowledge, so getting one's retractions over with earlier brings one closer to knowledge and reduces epistemic loss.

To make this idea precise, let $\gamma(M,w,i)$ be a *local loss function*, which is a function that assigns some nonnegative quantity to M in w at stage i (e.g., $\rho(M,w,i)$ is a local loss function). Define the *delay* to accumulate quantity u of loss γ, where u is a nonnegative real number, as:

$$(\mathsf{D}i)\ (\gamma(M,w,i) \geq u) = \text{ the least stage } j \text{ such that } \sum_{i=0}^{j} \gamma(M,w,i) \geq u,$$

with the important proviso that the expression denotes 0 if there is no such stage j. In the deterministic case, $\rho(M,w)$ is always a natural number. The time delay to the kth

retraction is just:

$$\tau(M,w,k) = (\mathsf{D}i)\ (\rho(M,w,i) \geq k).$$

It remains to compare methods in terms of worst-case retraction times. It is not quite right to compare each method's delay to each retraction; for consider the output sequences $\sigma = (T_0, T_1, T_2)$ and $\sigma' = (T_0, T_0, T_2)$. Sequence σ has an earlier elapsed time to the first retraction, but it still seems strictly worse than σ'; for the retraction delays in σ are at least as late as those in σ' if one views the first retraction in σ as an "extra" retraction and ignores it. Ignoring extra retractions amounts to considering a local loss function γ such that $\gamma(M,w,i) \leq \rho(M,w,i)$, for each M,w,i. In that case, say that $\gamma \leq \rho$. Accordingly, define $M \leq^{\tau}_{e,n} M'$ to hold if and only if there exists local loss function $\gamma \leq \rho$ such that for each w in $C_e(n)$ there exists w' in $C_e(n)$ such that:

$$\tau(M,w,k) \leq (\mathsf{D}i)\ (\gamma(M,w',i) \geq k).$$

Define $<^{\tau}_{e,n}$, $\leq^{\tau}_{e}$ and $\ll^{\tau}_{e}$ as was done for ρ. Now define efficiency and beating from e onward in terms of retraction times by substituting τ for ρ in the corresponding definitions provided in the preceding section.

Theorem 2 (deterministic, stable Ockham efficiency theorem) *Let the loss be retraction times. Assume that question Q_e has no short skeptical paths and that method M converges to the truth. Then the following are equivalent:*

1. *method M is Ockham and stalwart from e onward;*
2. *method M is efficient from e onward;*
3. *method M is unbeaten from e onward.*

Proof: Consequence of theorem 4 below. $\dashv$

Retraction may be viewed as a strategy for eliminating error, so it is of interest to check whether theorem 2 can be strengthened to include the total number of errors committed as a loss. Let $\varepsilon(M,w,i)$ assume value 1 if M produces a theory incorrect of S_w at stage i and value 0 otherwise. Define the *cumulative errors* of M in w as $\varepsilon(M,w) = \sum_{i=0}^{\infty} \varepsilon(M,w,i)$. Violating Ockham's razor at e also increases the worst-case error bound over complexity cell $C_e(0)$. Why? We claim that any method that is Ockham from e onward never errs after e in any world in $C_e(0)$, whereas any method that violates Ockham's razor at e errs at least once in some world in $C_e(0)$. In every world w in $C_e(0)$, there is some stage n_w at which T_w becomes the uniquely simplest theory compatible with experience, and moreover, there is no stage between e and n_w such that some other theory $T \neq T_w$ is uniquely simplest. Because every Ockham method refuses to answer anything other than the unique simplest theory (when it exists) after e, it follows such methods commit no errors in any world in $C_e(0)$. In contrast, if M violates Ockham's razor at e, then M returns some theory T that is not uniquely simplest at e. Hence, there is some theory $T' \neq T$ such that $c_e(T') = 0$, and it follows that M commits at least one error in every world in which T' is true.

We focus on retractions and their times primarily because violating Ockham's razor at e yields more retractions in *every* non-empty complexity cell $C_e(n)$, whereas the

Ockham violator does worse in terms of errors only in $C_e(0)$. The reason for the weaker result in the error case is, in a sense, trivial—the worst-case bound on total errors is infinite in every non-empty complexity cell $C_e(n)$ other than $C_e(0)$ for *all* convergent methods, including the stalwart, Ockham methods. To see why, recall that nature can force an arbitrary, convergent method M to converge to some theory T of complexity n and and to produce it arbitrarily often before refuting T (by lemma 1). Thereafter, nature can extend the data to a world w of complexity $n+1$ in which T is false, so M incurs arbitrarily many errors, in the worst case, in $C_e(n+1)$. Retractions and retraction times are not more important than errors; they are simply more *sensitive* than errors at exposing the untoward epistemic consequences of violating Ockham's razor.

Nonetheless, one may worry that retractions and errors trade off in an awkward manner, since avoiding retractions seems to promote dogmatism, whereas avoiding errors seems to motivate skeptical suspension of belief. Such tradeoffs are inevitable in some cases, but not in the worst cases that matter for the Ockham efficiency theorems. Consider, again, just the easy (Pareto) comparisons in which one method does as well as another with respect to every loss under consideration. Let $\mathscr{L}$ be some subset of the loss functions $\{\rho, \varepsilon, \tau\}$. Then the *worst-case Pareto* order and *worst-case Pareto dominance* relations in $\mathscr{L}$ are defined as:

$$\begin{array}{lll} M \leq_e^{\mathscr{L}} M' & \text{iff} & M \leq_e^{\gamma} M', \quad \text{for all } \gamma \in \mathscr{L}; \\ M \ll_e^{\mathscr{L}} M' & \text{iff} & M \leq_e^{\mathscr{L}} M' \text{ and } M \ll_e^{\gamma} M', \quad \text{for some } \gamma \in \mathscr{L}. \end{array}$$

Efficiency and beating may now be defined in terms of $\leq_e^{\mathscr{L}}$ and $\ll_e^{\mathscr{L}}$, just as in the case of ρ. The following theorem says that the Ockham efficiency theorems are driven primarily by retractions or retraction times, but errors can go along peacefully for the ride as long as only easy loss comparisons are made.

Theorem 3 (Ockham efficiency with errors) *Assume that $\mathscr{L} \subseteq \{\rho, \varepsilon, \tau\}$ and that the loss concept is $\leq_{\mathscr{L}}$. Then:*

1. *theorem 1 continues to hold if $\rho \in \mathscr{L}$ or $\tau \in \mathscr{L}$;*
2. *theorem 2 continues to hold if $\tau \in \mathscr{L}$.*

Proof: Consequence of theorem 4 below.[9] $\dashv$

6 Stochastic Inquiry

The aim of the paper is to extend the preceding theorems to mixed strategies. As discussed above, the extension is of interest since the Ockham efficiency theorems are based on worst-case loss with respect to the cells of an empirical complexity partition and, in some games, stochastic (mixed) strategies can achieve better worst case loss than can deterministic (pure) strategies. We begin by introducing a very general collection of stochastic strategies.

Recall that a deterministic method M returns an answer A when finite input sequence e is provided, so that $p(M(e) = A) = 1$. Now conceive of a method more

generally as a random process that produces answers with various probabilities in response to e. Then one may think of M_e as a random variable, defined on a probability space $(\Omega, \mathscr{F}, p)$, that assumes values in $\mathscr{A}$. A random variable is a function defined on Ω, so that $M_e(\omega)$ denotes a particular answer in $\mathscr{A}$. A *method* is then a collection $\{M_e : e \text{ is in } F_K\}$ of random variables assuming values in $\mathscr{A}$ that are all defined on an underlying probability space $(\Omega, \mathscr{F}, p)$.[10] In the special case in which $p(M_e = A)$ is 0 or 1 for each e and answer A, say that M is a *deterministic* method or a *pure* strategy.

Let M be a method and let e in F_K have length l. Then the *random output sequence* of M in response to e with respect to ω is the random sequence $M_{[e]}(\omega) = (M_{e|0}(\omega), \ldots, M_{e|l}(\omega))$. Note that the length of $M_{[e]}(\omega)$ is $l+1$, so the length of $M_{[e_-]}(\omega)$ is l. In particular, $M_{[()]}(\omega) = ()$, so $M_{[()]} = ()$ expresses the vacuous event Ω. If $\mathscr{S}$ is an arbitrary collection of random output sequences of M along e and D is an event in $\mathscr{F}$ of nonzero probability, then the conditional probability $p(M_{[e]} \in \mathscr{S} \mid D)$ is defined.

Consider the situation of a scientist who is deciding whether to keep method M or to switch to some alternative method M' after e has been received. In the deterministic case, it doesn't really matter whether the decision is undertaken before M produces its deterministic response to e or after, since the scientist can predict perfectly from the deterministic laws governing M how M will respond to e. That is no longer the case for methods in general—the probability that $M_e = A$ may be fractional prior to the production of A but becomes 1 thereafter. However, the case of deciding after the production of A reduces to the problem of deciding before because we can model the former case by replacing M_e with a method that produces A in response to e deterministically. Therefore, without loss of generality, we consider only the former case.

The methodological principles of interest must be generalized to apply to stochastic methods. Let e be in F_K and let D be an event of nonzero probability. Say that M is *logically consistent* at e given D if and only if:

$$p(M_e \in \mathscr{A}_{Q_e} \mid D) = 1.$$

Say that M is *Ockham* at e given D if and only if:

$$p(M_e \text{ is Ockham at } e \mid D) = 1.$$

Finally, say that M is *stalwart* at e given D if and only if:

$$p(M_e = T \mid M_{e_-} = T \wedge D) = 1,$$

when T is Ockham at e and $p(M_{e_-} = T \wedge D) > 0$. This plausibly generalizes the deterministic version of stalwartness—*given* that you produced an answer before and it is still Ockham, keep it for sure.

The concepts pertaining to inquiry and efficiency must also be generalized. Say that M *converges to the truth* over K_e given event D if and only if:

$$\lim_{i \to \infty} p(M_{w|i} = T_w \mid D) = 1,$$

for each world world w in W_{K_e}.

Each of the above methodological properties is a relation of form $\Phi(M, e \mid D)$. In particular, one can consider $\Phi(M, e \mid M_{[e_-]} = \sigma)$, for some random output sequence

σ of M along e_- such that $p(M_{[e_-]} = \sigma) > 0$, in which case Φ is said to hold of M *at* (e, σ). When Φ holds of M at each pair (e', σ') such that e' is in $F_{K,e}$ and σ' is a random output sequence of M along e'_- such that $p(M_{[e'_-]} = \sigma') > 0$, then say that Φ holds *from* (e, σ) *onward*. When Φ holds from $((), ())$ onward, say that Φ holds *always*. For example, one can speak of M always being stalwart or of M converging to the truth from (e, σ) onward.

Turn next to epistemic losses. There are two ways to think about the loss of a stochastic method: as *loss in chance* or as *expected loss*. For example, T is retracted *in chance* at e if the probability that the method produces T drops at e. Define, respectively, the total *errors in chance* and *retractions in chance* at i in w given D such that $p(D) > 0$ to be:

$$\begin{aligned} \widehat{\varepsilon}(M, w, i \mid D) &= \sum_{T \neq T_w} p(M_{w|i} = T \mid D); \\ \widehat{\rho}(M, w, i \mid D) &= \sum_{T \in \mathscr{T}} p(M_{w|(i-1)} = T \mid D) \ominus p(M_{w|i} = T \mid D), \end{aligned}$$

where $x \ominus y = \max(x - y, 0)$. For $\hat{\gamma}$ ranging over $\hat{\rho}$, $\hat{\varepsilon}$, define the *total* loss in chance to be: $\widehat{\gamma}(M, w \mid D) = \sum_{i=0}^{\infty} \widehat{\gamma}(M, w, i \mid D)$. Retractions in chance can be fractional. Define the delay to accumulate u retractions in chance as $\widehat{\tau}(M, w, u \mid D) = (\mathsf{D}i)\ (\widehat{\gamma}(M, w, i) \geq u)$.

Now consider expected losses. Think of losses as random variables. A *random* local loss function is a nonnegative function of form $\gamma(M, w, i, \omega)$, where ω ranges over the samples space Ω. For example, define $\rho(M, w, i, \omega)$ to have value 1 if $M_{[w]}(\omega)$ exhibits a retraction at stage i and to have value 0 otherwise. For fixed M, w, i, let $\gamma_{M,w,i}(\omega) = \gamma(M, w, i, \omega)$. Then $\rho_{M,w,i}$ and $\varepsilon_{M,w,i}$ are random variables. If $\gamma_{M,w,i}(\omega)$ is a random variable, then the delay time $(\mathsf{D}i)\ (\gamma_{M,w,i}(\omega) \geq k)$ is a random variable and the sum $\sum_{i=0}^{\infty} \gamma_{M,w,i}(\omega)$ is a random variable on the extended real numbers; so $\rho_{M,w}(\omega)$, $\varepsilon_{M,w}(\omega)$, and $\tau_{M,w,k}(\omega)$ are random variables on the extended real line.

The next problem is to compare two methods M, M' in terms of worst-case loss in chance or expected loss at e of length l. Each stochastic method has its own probability space $(\Omega, \mathscr{F}, p)$ and $(\Omega', \mathscr{F}', p')$, respectively. Recall that M and M' are being compared when the last entry of e has been presented and M, M' have yet to randomly produce corresponding outputs. Suppose that, as a matter of fact, both M and M' responded to e_- by producing, with chances greater than zero, the same random trajectory σ of length l. Let $\widehat{\gamma}$ be $\widehat{\rho}$ or $\widehat{\varepsilon}$, and let γ be ρ or ε. Then, as in the deterministic case, define $M \leq^{\widehat{\gamma}}_{e,\sigma,n} M'$ (respectively $M \leq^{\gamma}_{e,\sigma,n} M'$) to hold if and only if for each w in $C_e(n)$, there exists w' in $C_e(n)$ such that:

$$\begin{aligned} \widehat{\gamma}(M, w \mid M_{[e_-]} = \sigma) &\leq \widehat{\gamma}(M', w' \mid M'_{[e_-]} = \sigma); \\ \mathrm{Exp}_p(\gamma_{M,w} \mid M_{[e_-]} = \sigma) &\leq \mathrm{Exp}_{p'}(\gamma_{M',w'} \mid M'_{[e_-]} = \sigma). \end{aligned}$$

Methods can be compared in terms of expected retraction times just as in the deterministic case. Define the comparison $M \leq^{\tau}_{e,\sigma,n} M'$ to hold if and only if there exists random local loss function $\gamma \leq \rho$ such that for every world w in $C_e(n)$, there exists world w' in $C_e(n)$ such that for each k:

$$\mathrm{Exp}_p(\tau_{M,w,k} \mid M_{[e_-]} = \sigma) \leq \mathrm{Exp}_{p'}((\mathsf{D}i)\ (\gamma_{M',w',i}(\omega) \geq k) \mid M'_{[e_-]} = \sigma).$$

Comparing retraction times in chance is similar to comparing expected retraction times. Let $\widehat{\gamma}, \widehat{\delta}$ map methods, worlds, stages of inquiry, and measurable events to real numbers. A *local loss in chance* is a mapping $\widehat{\gamma}(M, w, i \mid D)$ that assumes nonnegative real values, where D is a measurable event of nonzero probability. Define $\widehat{\gamma} \leq \widehat{\delta}$ to hold if and only if $\widehat{\gamma}(M, w, i \mid D) \leq \widehat{\delta}(M, w, i \mid D)$, for each method M, world w, and measurable event D of nonzero probability. Define the comparison $M \leq^{\widehat{\tau}}_{e,\sigma,n} M'$ to hold if and only if there exists local loss in chance $\widehat{\gamma} \leq \widehat{\rho}$ such that for all w in $C_e(n)$ and for all $\varepsilon > 0$ there exists w' in $C_e(n)$ and there exists open interval I of length $\leq \varepsilon$ such that for all real numbers $u \geq 0$ such that u is not in I,

$$\widehat{\tau}(M, w, u' \mid M_{[e_-]} = \sigma) \leq (\mathsf{D}i)\ (\widehat{\gamma}(M', w, i \mid M'_{[e_-]} = \sigma) \geq u').$$

The only obvious difference from the definition for expected retraction times is the exemption of an arbitrarily small interval I of possible values for cumulative retractions in chance. The reason for the exemption is that stalwart, Ockham strategies can be forced by nature to retract fully at each step down a skeptical path, whereas some convergent methods can only be forced to perform $1 - \varepsilon$ retractions in chance at each step, for arbitrarily small ε. Since the time of non-occurring retractions in chance is 0, the retraction times in chance of an Ockham method would be incomparable with those of some convergent methods, undermining the efficiency argument. Allowing an arbitrarily small open interval of exceptions introduces no bias into the argument, since non-Ockham methods equally benefit from the exceptions. Still, they do worse.

Now define the obvious analogues of all the order relations in the deterministic case to arrive at the worst-case Pareto relations $\leq^{\mathscr{L}}_{e,\sigma}$ and $\ll^{\mathscr{L}}_{e,\sigma}$, where $\mathscr{L}$ is a set of losses γ or of losses in chance $\widehat{\gamma}$.

It remains to define efficiency and beating in terms of $\mathscr{L}$. The scientist cannot change the past, so if the scientist elects at e to follow a different method M' than her old method M, she is stuck with the theory choices σ made by M along e_-. So it is as if she always followed a method that produces σ deterministically in response to e_- and that acts like M thereafter. Accordingly, if e, σ have the same length l, define $M'[\sigma/e_-]$ to be just like M' except that $M'[\sigma/e_-]_{[e_-]}(\omega) = \sigma$, for each ω in Ω. Let $p(M_{[e_-]} = \sigma) > 0$. Say that method M is *efficient* in Q at (e, σ) with respect to the losses in $\mathscr{L}$ if and only if:

1. M converges to the truth given $M_{[e_-]} = \sigma$;
2. $M \leq^{\mathscr{L}}_{e,\sigma} M'[\sigma/e_-]$, for each alternative method M' that converges to the truth in Q_e.

Say that method M is *beaten* in Q at (e, σ) with respect to losses in $\mathscr{L}$ if and only if the second condition above holds with $\ll^{\mathscr{L}}_{e,\sigma}$ in place of $\leq^{\mathscr{L}}_{e,\sigma}$. Efficiency and being unbeaten are again relations of form $\Phi(M, e \mid D)$, so one can speak of them as holding always or from (e, σ) onward.

7 Stochastic Ockham Efficiency Theorem

Here is the main result.

Theorem 4 (stochastic Ockham efficiency theorem) *Theorem 3 extends to stochastic methods and losses in chance when "from e onward" is replaced with "from* (e,σ) *onward", for all* (e,σ) *such that* $p(M_{[e_-]} = \sigma) > 0$*. The same is true for expected losses.*

The proof of the theorem is presented in its entirety in the appendix. The basic idea is that nature can still force a random method to produce the successive theories along a skeptical path with arbitrarily high chance, if the method converges in probability to the truth. The following result entails lemma 1 as a special case and is nearly identical in phrasing and proof.

Lemma 2 (forcing changes of opinion in chance) *Let e be a finite input sequence of length l, and suppose that M converge to the truth in* Q_e*. Let* $p(D) > 0$*. Let* $(S_0, \ldots, S_n)$ *be a skeptical path in* Q_e *such that* $c_e(S_n) = n$ *and let* $\varepsilon > 0$ *be arbitrarily small and let natural number m be arbitrarily large. Then there exists world w in* $C_e(n)$ *and stages of inquiry* $l = s_0 < \ldots < s_{n+1}$ *such that for each i from* 0 *to n, stage* s_{i+1} *occurs more than m stages after* s_i *and* $p(M_{w|j} = T_{S_i} \mid D) > 1 - \varepsilon$*, at each stage j such that* $s_{i+1} - m \leq j \leq s_{i+1}$*.*

Proof: To construct w, set $e_0 = e$ and $s_0 = l$. For each i from 0 to n, do the following. Extend e_i with world w_i such that $S_{w_i} = S_i$. Since M converges in probability to the truth, there exists a stage s such that for each stage $j \geq s$, $p(M_{w|j} = T_{S_i} \mid D) > 1 - \varepsilon$. Let s' be the least such s. Let $s_{i+1} = \max(s', s_i) + m$. Set $e_{i+1} = w_i | s_{i+1}$. The desired world is w_n, which is in $C_e(n)$, since $S_{w_n} = S_n$. $\dashv$

Hence, expected retractions are forcible from convergent, stochastic methods pretty much as they are from deterministic methods (lemma 5). Retractions in chance are a lower bound on expected retractions (lemma 4). On the other hand, it can be shown that a stochastic, stalwart, Ockham method incurs expected retractions only when its current theory is no longer uniquely simplest with respect to the data (lemma 8), so such a method incurs at most n expected retractions or retractions in chance after the end of e in $C_e(n)$. Violating Ockham's razor or stalwartness adds some extra retractions in chance (and expected retractions) that an Ockham method would not perform in every nonempty complexity cell $C_e(n)$, as in the deterministic case (lemmas 6 and 7).

The worst-case errors of stochastic methods are closely analogous those in the deterministic case. Ockham methods produce no expected errors or errors in chance in $C_e(0)$ (lemma 10) and all methods produce arbitrarily many expected errors or errors in chance, in the worst case, in each nonempty $C_e(n)$ such that $n > 0$ (lemma 11).

The retraction times of stochastic methods are a bit different from those of deterministic methods. Retraction times in chance are closely analogous to retraction times in the deterministic case, except that one must consider the times of fractional retractions in chance. The relevant lower bounds are provided by lemmas 15 and 16 and the upper bounds are provided by lemma 17. Expected retraction times are a bit different. For example, a stochastic method that produces fewer than n expected retractions may still have a nonzero time for retraction $m > n$, if the mth retraction is very improbable. That disanalogy is actually exploited in the proof of theorem 4. To force expected retraction times to be arbitrarily late in $C_e(n)$, for $n > 0$, one may choose the delay time

m in lemma 2 to be large enough to swamp the small chance $1 - n\varepsilon$ that n retractions fail to occur (lemmas 13, 16). But the anomaly does not arise for stalwart, Ockham methods, which satisfy upper bounds agreeing with the deterministic case, so the logic of the Ockham efficiency argument still goes through (lemma 17).

8 Conclusion and Future Directions

According to theorem 4, the situation with stochastic methods is essentially the same as in the deterministic case—obvious, stochastic analogues of Ockham's razor and stalwartness are necessary and sufficient for efficiency and for being unbeaten, when losses include retractions, retraction times, and errors. Every deterministic method counts as a stochastic method, so deterministic, convergent, stalwart, Ockham methods are efficient over all convergent, stochastic methods. Therefore, the game of inquiry is different from the game "rock-paper-scissors" and many other games in that respect. In fact, flipping a fair coin sequentially to decide between the uninformative answer K and the current Ockham answer T is a bad idea in terms of expected retractions—it is a violation of stalwartness that generates extra retractions in chance and expected retractions at each time one does it, from the second flip onward. That resolves the main question posed in the introduction: whether deterministic, stalwart, Ockham strategies are still efficient in comparison with convergent, stochastic strategies. In fact, the Ockham efficiency argument survives with aplomb, whether expected losses or losses in chance are considered and for a variety of Pareto combinations of epistemic losses including total retractions, total errors, and retraction times.

The second ambition mentioned in the introduction concerns statistical inference, in which outputs are stochastic due to randomness in the data rather than in the method. Let the question be whether the mean μ of a normal distribution of known variance is 0 or not. According to statistical testing theory, one calls theory $T_{\mu=0}$ that $\mu = 0$ the *null hypothesis* and one fixes a bound α on the probability that one's test rejects $T_{\mu=0}$ given that $T_{\mu=0}$ is true. A statistical test at a given sample size N partitions possible values of the sample mean $\overline{X}$ into those at which $T_{\mu=0}$ is accepted and into those at which $T_{\mu=0}$ is rejected. The test has *significance* α if the chance that the test rejects $T_{\mu=0}$ is no greater than α assuming that $T_{\mu=0}$ is true. It is a familiar fact that such a test does not converge to the true answer as sample size increases unless the significance is tuned downward according to an appropriate schedule. However, there are many significance-level schedules that yield statistically consistent procedures. We propose that retraction efficiency can plausibly bound the rate at which α may be dropped to the rate at which sample variance decreases.

Retractions in chance and, hence, expected retractions arise unavoidably, in the following way, in the problem of determining whether or not $\mu = 0$.[11] Suppose that the chance that a statistical test M accepts $T_{\mu=0}$ at sample size N when $\mu = 0$ exceeds $1 - \varepsilon/2$, where $\varepsilon > 0$ is as small as you please. Then there is a sufficiently small $r > 0$ such that the chance that M accepts $T_{\mu=0}$ at sample size N given that $\mu = r$ still exceeds $1 - \varepsilon/2$. But as sample size is increased, one reaches a sample size N' at which the test M "powers up" and the chance that M *rejects* $T_{\mu=0}$ given that $\mu = r$ is greater than $1 - \varepsilon/2$. We have forced the test into a retraction in chance of more than $1 - \varepsilon$.

The preceding argument is exactly analogous to the proofs of the stochastic Ockham efficiency theorems, in which it is shown that any consistent method accrues at least one expected retractions in complexity class one. If one assumes, as is natural, that $C(0)$ contains just $\mu = 0$ and $C(1)$ contains all values of μ other than 0, then the number of forcible retractions in chance equals the complexity of the statistical hypotheses in question, just as in our model of inquiry.[12]

Generalizing the Efficiency Theorems to statistical inference requires, therefore, only three further steps: (1) proving that methods that prefer simpler statistical hypotheses approximate the theoretical lower loss bounds, (2) proving that methods that violate Ockham's razor do not approximate those bounds, and (3) generalizing (1) and (2) to multiple retractions.

The first step, we conjecture, is straightforward for one-dimensional problems like determining whether the mean μ of a normally distributed random variable is zero or not—if losses are considered in chance. It appears that expected retractions may be unbounded even for simple statistical tests because there are values of μ at which the chance of accepting the null hypothesis hovers around 1/2 for arbitrarily many sample sizes.[13] Retractions in chance are more promising (and also agree with standard testing theory, in which power is an "in chance" concept). Suppose statistical method M ignores the traditional logic of statistical testing, and accepts the complex hypothesis that $\mu \neq 0$ with high chance $1 - \alpha$, contrary to the usual practice of favoring the null hypothesis. If μ is chosen to be small enough, then M is forced, with high probability, to accept that $\mu = 0$ with arbitrarily high chance, if M converges in probability to the true answer. Thereafter, M can be forced *back* to $\mu \neq 0$ when $\mu = r$, for r suitably near to 0. Thus, M incurs an *extra* retraction, in the worst case, of nearly $1 - \alpha$, both in $C(0)$ and in $C(1)$.

The second and third steps, in contrast, are significantly more difficult, because statistical methods that converge to the truth in probability cannot help but produce random "mixtures" of simple and complex answers. Therefore, efficiency and adherence to Ockham's razor and to stalwartness can only be approximate in statistical inference.

9 Acknowledgements

This work was supported generously by the National Science Foundation under grant 0750681. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We thank Teddy Seidenfeld for urging the extension of the result to stochastic methods. We thank Cosma Shalizi for commenting on an earlier draft of this paper at the Formal Epistemology Conference in 2009. We thank the anonymous referee for requesting clarification on the connections with game theory and for unusually detailed comments. We thank the editor, Branden Fitelson, for encouraging clarification and for providing the space for it.

10 Appendix - Comparison with Game Theory

The model of scientific inquiry described above might be represented any number of ways as a game in the economist's sense. Thus, the reader might be interested in the relationship between our results and those typically found in game theory. We remark upon at least five important differences.[14]

First, as stated in the introduction, the most general equilibrium existence theorems of game theory yield little information about what the equilibria are like. In contrast, our results uniquely pick out a particular important class of strategies, namely, the Ockham ones, as uniquely optimal. Some game-theoretic results specify properties of the equilibria. For instance, Von Neumann's minimax theorem shows that, in equilibria for finite, two-person, zero-sum games, both players employ *minimax* strategies, i.e. strategies that minimize the maximum possible loss. Although that theorem appears especially relevant to our results, the worst-case loss vectors that we consider are with respect to *cells of a complexity based partition of worlds*, and not with respect to all possible worlds. There are no minimax (simpliciter) actions in our model of inquiry (for either the scientist or Nature in our model of inquiry) and, as a result, Von Neumann's theorem is of little help.

Second, in our model of inquiry, the scientist's preferences cannot be represented by utilities. The chief difficulty is that the scientist's preferences involve *lexicographic* components: among all losses of inquiry, the scientist values eventually finding the truth highest and considers all other losses (e.g. minimization of errors and minimization of retractions) secondary. It is well-known that, in games in which players' preferences contain lexicographic components, even the simplest theorems guaranteeing the existence of equilibria fail.[15] Moreover, our players' preferences are not representable as utilities because they are also pre-ordered, and not totally ordered. That feature immediately threatens the existence of Nash equilibria in even the simplest games: consider, for example, a one-person game in which the only player has two actions, whose outcomes have incomparable value. Then there is no Nash equilibrium in the standard sense, as there is no action that is even weakly better than all others. One can show that in competitive games in which players' preferences are represented by vectors of real numbers with the Pareto ordering (again, such preferences do not have lexicographic components), there are "weak" Nash equilibria, in the sense that there are strategy profiles from which no player has reason to deviate.[16] However, the equilibria guaranteed by such proofs are "weak" in the sense that players may *not* prefer the equilibrium strategy profile to all others in which only his or her action were changed; they may have no preference whatsoever. In contrast, the result we obtain here is more analogous to a "strong" Nash equilibrium; the scientist prefers playing Ockham strategies to non-Ockham ones *and* that preference is strict!

Third, both the scientist and "Nature" have access to infinitely many actions in our model of inquiry. There are well-known results guaranteeing the existence of countably-additive equilibria in infinite games, but generally, such theorems also contain strong restrictions on the player's preference relations, in addition to assuming that they are representable by utilities. For instance, it is often assumed that players' utility functions are continuous or bounded functions with respect to an appropriate topology on the outcome space.[17] No such assumptions hold in our model: the scientist's losses

are potentially unbounded (even within complexity classes), and the obvious topology to impose on our outcome space does not yield continuous preference relations. If one permits players to employ merely-finitely additive mixed strategies, one can drop these assumptions on preference relations (but not the assumption that they are representable by utilities) and obtain existence of equilibria in zero-sum games.[18] However, the randomized strategies considered here are countably-additive, which makes our result even more surprising.

Fourth, in game-theory, if one player is permitted to employ mixed strategies (or behavior strategies), it is typical to assume that all players are permitted to do so. The model of inquiry presented here does not permit the second player, "Nature", to employ mixed strategies. That raises the question: Can one make sense of Nature employing "mixed strategies" and if so, does it change the result stated here? We do think, in fact, that one can reasonably interpret Nature's mixed strategies as a scientist's prior probabilities over possible worlds, and one can prove the existence of (merely finitely-additive) equilibria in particular presentations of our model of inquiry when represented as game.[19] However, the main result of this paper employs no such prior probabilities.

Fifth, and finally, the last major hurdle in representing our theorems as game-theoretic equilibria is the development of a more general theory of simplicity. The definition of simplicity stated in this paper is very narrow, allowing only for prior knowledge about which finite sets of effects might occur—knowledge about timing and order of effects is not allowed for. But nothing prevents nature from choosing a mixed strategy that implies knowledge about timing or order of effects (recall that nature's mixture is to be understood as the scientist's prior probability). Such knowledge may essentially alter the structure of the problem. For example, if nature chooses a mixing distribution according to which effect a is always followed immediately by effect b, then the sequence a,b ought properly to be viewed as a single effect rather than as two separate effects.[20] But if simplicity is altered by nature's choice of a mixing distribution, then so is Ockham's razor and, hence, what counts as an Ockham strategy for the scientist. Therefore, in order to say what it means for Ockham's razor to be a "best response" to Nature, it is necessary to define simplicity with sufficient generality to apply to every possible restriction of the set of worlds compatible with K to a narrower range of worlds. More general theories of simplicity than the one presented in this paper have been proposed and have been shown to support Ockham efficiency theorems (Kelly 2007d, 2008), but those concepts are still not general enough to cover *all* possible restrictions of W_K. Of course, a general Ockham efficiency theorem based on a general concept of simplicity would be of considerable interest quite independently of this exploratory discussion of relations to game theory.

11 Proofs

The proof of theorem 4 breaks down naturally into two principal cases. Assume that e of length l is in F_K, that M is a method, that σ is an output sequence of length l such that $p(M_{[e^-]} = \sigma) > 0$. In the *defeat* case, the last entry in σ is some informative answer T to Q that is not Ockham with respect to e (i.e., any justification for T derived from Ockham's razor is defeated by e). Thus, Ockham methods pick up a retraction at

e in the defeat case and non-Ockham methods may fail to retract at e. The *non-defeat* case holds whenever the defeat case does not.

Proof of theorem 4: We begin by proving the case of theorem 4 that corresponds to the second clause of theorem 3. Assume that Q_e has no short skeptical paths. We begin by showing that convergent methods that are stalwart and Ockham from (e,σ) onward are efficient from (e,σ) onward. Let stochastic method O be stalwart and Ockham from (e,σ). Let e in F_K of length l be given and let σ be an answer sequence of length l such that $p(O_{[e_-]}=\sigma)>0$. Let M converge to the truth in Q_e. Then for each n such that $C_e(n)$ is non-empty, we have:

$$O \leq^{\rho}_{e,\sigma,n} M[\sigma/e_-] \quad \text{and} \quad O \leq^{\widehat{\rho}}_{e,\sigma,n} M[\sigma/e_-], \quad \text{by lemmas 5 and 9;}$$
$$O \leq^{\varepsilon}_{e,\sigma,n} M[\sigma/e_-] \quad \text{and} \quad O \leq^{\widehat{\varepsilon}}_{e,\sigma,n} M[\sigma/e_-], \quad \text{by lemmas 10and11.}$$

Furthermore, these statements are trivially true if $C_e(n)$ is empty, so they hold for all n.

Let w be in $C_e(n)$ and let k be the number of retractions in σ. Apply lemma 13 with m set to $\max_i \mathrm{Exp}(\tau_{O,w,k+i} \mid O_{[e_-]}=\sigma)$ in order to obtain world w_m in $C_e(n)$ and local loss function $\gamma_m \leq \rho$. Let $n' \leq n$. The lower bounds for $\mathrm{Exp}((\mathsf{D}i)\ (\gamma_{M,w_m,i} \geq n') \mid M_{[e_-]}=\sigma)$ obtained from lemma 13 meet the upper bounds for $\mathrm{Exp}(\tau_{O,w,n'} \mid M_{[e_-]}=\sigma)$ obtained from lemma 17. Furthermore, γ is a function of w and $w_m \neq w_{m'}$ if $m \neq m'$, so there is a single γ such that $\gamma_m(M,w_m,i,\omega)=\gamma(M,w_m,i,\omega)$, for each m. Hence, $O \leq^{\tau}_{e,\sigma,n} M[\sigma/e_-]$.

The argument that $O \leq^{\widehat{\tau}}_{e,\sigma,n} M[\sigma/e_-]$ is similar. Let $\varepsilon>0$. Apply lemma 13 with m set to $\max_i \widehat{\tau}(O,w,k+i \mid O_{[e_-]}=\sigma)$ in order to obtain world $w_{m,\varepsilon}$ in $C_e(n)$ and local loss function in chance $\widehat{\gamma}_{m,\varepsilon} \leq \widehat{\rho}$. Then by lemmas 15 and 17, there exists open interval I of length ε such that for all u not in I, we have $\widehat{\tau}(O,w,u \mid O_{[e_-]}=\sigma) \leq (\mathsf{D}i)\ (\widehat{\gamma}_{m,\varepsilon}(M,w_{m,\varepsilon},u \mid O_{[e_-]}=\sigma)$. Therefore, if $\mathscr{L}$ is a subset of either $\{\rho,\varepsilon,\tau\}$ or $\{\widehat{\rho},\widehat{\varepsilon},\widehat{\tau}\}$, we have that $O \leq^{\mathscr{L}}_{e,\sigma,n} M[\sigma/e_-]$, for each n, so O is efficient with respect to $\mathscr{L}$.

It is immediate that efficiency from (e,σ) onward implies being unbeaten from (e,σ) onward.

To show that being convergent and unbeaten from (e,σ) onward implies being stalwart and Ockham from (e,σ) onward, assume that M is convergent but violates either Ockham's razor or stalwartness at (e',σ'), where (i) e' is in F_{K_e}, (ii) σ' is an answer sequence extending σ, and (iii) both e' and σ'_- have length l'. Let O be a convergent method that is always stalwart and Ockham.

Consider first the case for expected losses, in which τ is in $\mathscr{L}$, which is a subset of $\{\rho,\varepsilon,\tau\}$. It must be shown that $O[\sigma'/e'_-] \ll^{\mathscr{L}}_{e',\sigma'} M$. By the preceding efficiency argument, $O[\sigma'/e'_-] \leq^{\mathscr{L}}_{e',\sigma'} M$, so it suffices to show that $O[\sigma'/e'_-] \ll^{\tau}_{e',\sigma'} M$, for which it suffices, in turn, to show that $M \not\leq^{\tau}_{e',\sigma',n} O[\sigma'/e'_-]$, for each n for which $C_{e'}(n)$ is non-empty. Suppose that $C_{e'}(n)$ is nonempty. Then lemma 16 provides a world w in $C_{e'}(n)$ such that either $\mathrm{Exp}(\tau_{M,w,k+1} \mid M_{[e'_-]}=\sigma') > l'$ or $\mathrm{Exp}(\tau_{M,w,k+n+2} \mid M_{[e'_-]}=\sigma')>0$. But by lemma 17, whether or not the defeat case obtains, we have that $\mathrm{Exp}(\tau_{O[\sigma'/e'_-],w',k+1} \mid O[\sigma'/e'_-]_{[e'_-]}=\sigma') \leq l'$ and $\mathrm{Exp}(\tau_{O[\sigma'/e'_-],w',k+n+2} \mid O[\sigma'/e'_-]_{[e'_-]}=\sigma')=0$, for each w' in $C_{e'}(n)$. There is, therefore, no choice of $\gamma \leq \rho$ such that

$\mathrm{Exp}((\mathrm{D}i)\ (\gamma_{O[\sigma'/e'_-],w',i} \geq k+1) \mid O[\sigma'/e'_-]_{[e'_-]} = \sigma') > l'$ or $\mathrm{Exp}((\mathrm{D}i)\ (\gamma_{O,w',i} \geq k+n+2) \mid O[\sigma'/e'_-]_{[e'_-]} = \sigma') > 0$, so $M \not\leq^{\tau}_{e',\sigma',n} O[\sigma'/e'_-]$.

Next consider the case for losses in chance, in which $\widehat{\tau}$ is in $\mathscr{L}$, which is a subset of $\{\widehat{\rho}, \widehat{\varepsilon}, \widehat{\tau}\}$. Follow the preceding argument down to the invocation of lemma 16. The same lemma, in this case, provides a world w in $C_{e'}(n)$ and an $\alpha > 0$ such that either $\widehat{\tau}(M, w, k+1 \mid M_{[e'_-]} = \sigma') > l'$ or $\widehat{\tau}(M, w, k+n+1+\alpha \mid M_{[e'_-]} = \sigma') > 0$. By lemma 12, there exists $\varepsilon > 0$ such that the preceding inequalities hold for each v such that $k+1-\varepsilon < v \leq k+1$ or $k+n+1+\alpha-\varepsilon < v \leq k+n+1+\alpha$, respectively. So by lemma 17, there is no open interval I in the real numbers that witnesses $M \leq^{\tau}_{e',\sigma',n} O[\sigma/e'_-]$.

Next, we prove the case of theorem 4 that corresponds to the first clause of theorem 3. Focus first on the case of expected losses. Note that "always" is the special case of "from (e,σ) onward" in which e, σ are both the empty sequence. Therefore, the case in which τ is in $\mathscr{L}$ drops out as a special case of the preceding argument. For the case in which ρ is in $\mathscr{L}$, it suffices to show that if every theory is correct of a unique effect set and if M ever violates Ockham's razor or stalwartness, then M is beaten in terms of ρ at the *first* violation of either principle. Suppose that M violates either Ockham's razor or stalwartness at (e,σ), so that $p(M_{[e_-]} = \sigma) > 0$. Further, suppose that (e,σ) is the first time that M violates Ockham's razor, so that there are no proper subsequences e' and σ' of e and σ where some violation occurs. Let O be a convergent, stalwart, Ockham method, and suppose $C_e(n)$ is nonempty. Then $M \not\leq^{\rho}_{e,\sigma,n} O[\sigma/e_-]$ by the defeat and non-defeat cases of lemmas 6 and 9. Suppose that stalwartness is violated at (e,σ). Then $M \not\leq^{\rho}_{e,\sigma,n} O[\sigma/e_-]$ by lemmas 7 and 9. Note that only the non-defeat case of lemma 9 applies in this case due to lemma 7. The argument based on losses in chance is similar and appeals to the same lemmas. $\dashv$

Lemma 3 (forcing retractions in chance) *Suppose that M converges to the truth in Q_e and that $(S_0, \ldots, S_n)$ is a skeptical path in K_e such that $c_e(S_n) = n$. Then for each $\varepsilon > 0$, there exists world w in $C_e(n)$ such that:*

$$\sum_{i=l+1}^{\infty} \widehat{\rho}(M, w, i \mid D) > n - \varepsilon.$$

Proof: Let $\varepsilon > 0$. Using the skeptical path $(S_0, \ldots, S_n)$, apply lemma 2 to obtain a world w in $C_e(n)$ and stages $l = s_0 < \ldots < s_{n+1}$ such that $s_0 = l$ and $s_{i+1} - s_i \geq m$ and $p(M_{w|s_{i+1}} = T_{S_i} \mid D) > 1 - \varepsilon/2n$, for each i from 0 to n. It follows that M incurs more than $1 - \varepsilon/n$ retractions in chance from $s_i + 1$ to s_{i+1} in w, since T_i drops in probability from more than $1 - \varepsilon/2n$ to less than $\varepsilon/2n$. Since there are at least n such drops, there are more than $n - \varepsilon$ retractions in chance. $\dashv$

In all the lemmas that follow, assume that e of length l is in F_K, that M is a method, that σ is an output sequence of length l such that $p(M_{[e_-]} = \sigma) > 0$, and that $p(D) > 0$.

Lemma 4 (losses in chance that bound expected losses) *1. $\widehat{\rho}(M, w \mid D) \leq \mathrm{Exp}(\rho_{M,w} \mid D)$;*

2. $\widehat{\varepsilon}(M, w \mid D) = \mathrm{Exp}(\varepsilon_{M,w} \mid D)$.

Proof: Let S be an arbitrary set of natural numbers.

$$\begin{aligned}\sum_{i\in S}\widehat{\rho}(M,w,i \mid D) &= \sum_{i\in S}\sum_{T\in\mathscr{T}} p(M_{w|i-1}=T \mid D)\ominus p(M_{w|i}=T \mid D)\\ &\leq \sum_{i\in S}\sum_{T\in\mathscr{T}} p(M_{w|i-1}=T\wedge M_{w|i}\neq T \mid D)\\ &= \sum_{i\in S}\mathrm{Exp}(\rho_{M,w,i} \mid D)=\mathrm{Exp}(\sum_{i\in S}\rho_{M,w,i} \mid D).\end{aligned}$$

Furthermore:

$$\sum_{i\in S}\widehat{\varepsilon}(M,w,i \mid D)=\sum_{i\in S}p(M_{w|i-1}\neq T_w \mid D)=\sum_{i\in S}\mathrm{Exp}(\varepsilon_{M,w,i} \mid D)=\mathrm{Exp}(\sum_{i\in S}\varepsilon_{M,w,i} \mid D).$$

$\dashv$

Lemma 5 (retractions: lower bound) *Suppose that Q_e has no short paths, that M converges to the truth in Q_e, and that $C_e(n)$ is non-empty. Then for each $\varepsilon>0$, there exists w in $C_e(n)$ such that:*

1. $\widehat{\rho}(M,w \mid M_{[e_-]}=\sigma)\geq n+1-\varepsilon$ *in the defeat case;*
2. $\widehat{\rho}(M,w \mid M_{[e_-]}=\sigma)\geq n-\varepsilon$ *otherwise.*

The same is true if $\widehat{\rho}(M,w \mid M_{[e_-]}=\sigma)$ is replaced with $Exp(\rho_{M,w} \mid M_{[e_-]}=\sigma)$.

Proof: Let $\varepsilon'>0$. In the defeat case, the last entry T in σ is not Ockham at e. Hence, there exists S_0 in K_e such that $c_e(S_0)=0$ and $T\neq T_{S_0}$. Extend e with just effects from S_0 until e' is presented such that $p(M_{e'}=T_{S_0} \mid M_{[e_-]}=\sigma)>1-\varepsilon'/2$, which yields nearly one retraction in chance from l to the end of e'. Since there are no short paths, there exists a skeptical path $(S_0,\ldots,S_n)$ in K_e such that $c_e(S_n)=n$. Apply lemma 3 to $(S_0,\ldots,S_n)$ with e set to e', ε set to $\varepsilon'/2$, and arbitrary $m>0$ to obtain another $n-\varepsilon'/2$ retractions in chance after the end of e', for a total of more than $n+1-\varepsilon'$ retractions in chance from $l+1$ onward. The non-defeat case is easier—just apply lemma 3 directly to $(S_0,\ldots,S_n)$ to obtain $n-\varepsilon$ retractions in chance. To obtain the results for expected retractions, apply lemma 4. $\dashv$

Lemma 6 (retractions: lower bound for Ockham violators) *Suppose that Q_e has no short paths, that M converges to the truth in Q_e, and that $C_e(n)$ is non-empty. Assume, further, that each theory is correct of a unique effect set, that M is logically consistent, and that M violates Ockham's razor for the first time at (e,σ). Then there exists w in $C_e(n)$ such that:*

1. $\widehat{\rho}(M,w \mid M_{[e_-]}=\sigma)>n+1$ *in the defeat case;*
2. $\widehat{\rho}(M,w \mid M_{[e_-]}=\sigma)>n$ *otherwise.*

The same is true if $\widehat{\rho}(M,w \mid M_{[e_-]}=\sigma)$ is replaced with $Exp(\rho_{M,w} \mid M_{[e_-]}=\sigma)$.

Proof: Suppose that M violates Ockham's razor for the first time at (e,σ) so that for some T_S that is not Ockham at e, we have that $p(M_e = T_S \mid M_{[e_-]} = \sigma) = \alpha' > 0$. Consider the defeat case. Then the last entry T_S of σ is not Ockham at e. So there exists S_0 in K_e such that $c_e(S_0) = 0$ and $T_{S_0} \neq T_S$. Since each theory is true of at most one effect set and M was Ockham at e_- (since e is the first Ockham violation by M) and is no longer Ockham at e, it follows that S_e is not a subset of S. Since M is logically consistent, $p(M_e = T_S \mid M_{[e_-]} = \sigma) = 0$. But since T_S is the last entry in σ, we have that $p(M_{e_-} = T_S \mid M_{[e_-]} = \sigma) = 1$, so there is 1 retraction in chance already at e. Since there are no short paths, there exists skeptical path $(S_0, \ldots, S_n)$ such that $c_e(S_n) = n$. Choose $0 < \varepsilon' < \alpha'$ and let $\alpha = \alpha' - \varepsilon'$. Extend e with just the effects in S_0 until M produces T_{S_0} with chance $1 - \varepsilon'$. That entails a retraction in chance of at least α. Choose $0 < \varepsilon < \alpha$. The effects presented are still compatible with S_0, so one may apply lemma 3 to obtain w in which $n - \varepsilon$ more retractions in chance occur, for a total of $n + 1 + \alpha - \varepsilon > n + 1$ retractions in chance in w. The non-defeat case simply drops the argument for the first full retraction. For the expected case results, apply lemma 4. $\dashv$

Lemma 7 (retractions: lower bound for stalwartness violators) *Suppose that M converges to the truth in Q_e and that $C_e(n)$ is non-empty. Assume, further, that M violates the stalwartness property at (e,σ). Then the non-defeat case obtains and for each n such that $C_e(n)$ is non-empty, $\widehat{\rho}(M,w \mid M_{[e_-]} = \sigma) > n$ and $Exp(\rho_{M,w} \mid M_{[e_-]} = \sigma) > n$.*

Proof: Suppose that T is Ockham given e and that:

$$\begin{aligned} 0 &< p(M_{e_-} = T \wedge M_{[e_-]} = \sigma); \\ 1 &> p(M_e = T \mid M_{e_-} = T_S \wedge M_{[e_-]} = \sigma). \end{aligned}$$

The last entry in σ is T (by the first statement), so $p(M_{e_-} = T \mid M_{[e_-]} = \sigma) = 1$. By the second statement, $p(M_{e_-} = T \mid M_{[e]} = \sigma) < 1$. So $\widehat{\rho}(M,e,l \mid M_{[e_-]} = \sigma) = \alpha > 0$. Choose $\varepsilon > 0$ such that $\alpha > \varepsilon$, and apply lemma 3 to obtain w in $C_e(n)$ in which M has $n - \varepsilon$ more retractions in chance, for a total of $n + \alpha - \varepsilon > n$. For the expected case, apply lemma 4 $\dashv$

Lemma 8 *Suppose that method M is stalwart and Ockham from (e,σ) onward. Let w be in W_{K_e} and let $i > l$. Then the uniquely simplest theory in light of $w|(i-1)$ is no longer uniquely simplest at $w|i$, if:*

$$\textit{either } \widehat{\rho}(M,w,i \mid M_{[e_-]} = \sigma) > 0 \textit{ or } \mathrm{Exp}(\rho_{M,w,i} \mid M_{[e_-]} = \sigma) > 0.$$

Proof: By lemma 4, $\widehat{\rho}(M,w,i \mid M_{[e_-]} = \sigma) > 0$ implies that $Exp(\rho_{M,w,i} \mid M_{[e_-]} = \sigma) > 0$, so it suffices to consider the latter case. It follows that there exists random output sequence σ' of length $i+1$ with some theory T as penultimate entry and with final entry $T' \neq T$ such that $p(M_{[w|i]} = \sigma' \mid M_{[e_-]} = \sigma) > 0$. Hence, $p(M_{w|(i-1)} = T \mid M_{[e_-]} = \sigma) > 0$, so by the Ockham property, T is uniquely simplest for $w|(i-1)$. Also, since $p(M_{[e_-]} = \sigma) > 0$, we have that $p(M_{w|(i-1)} = T \wedge M_{[e_-]} = \sigma) > 0$. Furthermore, we have that:

$$p(M_{w|i} = T \mid M_{w|(i-1)} = T \wedge M_{[e_-]} = \sigma) < 1,$$

so by the stalwartness property, T is not uniquely simplest for $w|i$. $\dashv$

Lemma 9 (retractions: upper bound) *Suppose that M is stalwart and Ockham from (e,σ) onward, where $p(M_{[e_-]}=\sigma)>0$. Then:*

1. $\sup_{w\in C_e(n)} Exp\left(\rho_{M,w} \mid M_{[e_-]}=\sigma\right) \leq n+1$ *in the defeat case;*
2. $\sup_{w\in C_e(n)} Exp\left(\rho_{M,w} \mid M_{[e_-]}=\sigma\right) \leq n$ *otherwise.*

The same is true when $Exp\left(\rho_{M,w} \mid M_{[e_-]}=\sigma\right)$ is replaced by $\widehat{\rho}(M,w \mid M_{[e_-]}=\sigma)$.

Proof: The expected retraction case is an immediate consequence of lemma 8, allowing for an extra full retraction at e in the defeat case that stalwartness prevents in the non-defeat case. For the bound on retractions in chance, apply lemma 4. $\dashv$

Lemma 10 (errors: upper bound) *Suppose that M is Ockham from (e,σ) onward. Then:*

$$\sup_{w\in C_e(0)} Exp(\varepsilon_{M,w} \mid M_{[e_-]}=\sigma)=0.$$

The same is true when $Exp\left(\varepsilon_{M,w} \mid M_{[e_-]}=\sigma\right)$ is replaced by $\widehat{\varepsilon}(M,w \mid M_{[e_-]}=\sigma)$.

Proof: For all w in $C_e(0)$ and all $i \geq l$, the Ockham answer at $w|i$ is either K or T_w. Because M is Ockham from (e,σ) onward, it follows that M returns either T or K with probability one after l in w, thereby accruing no expected errors. For the error in chance case, apply lemma 4. $\dashv$

Lemma 11 (errors: lower bound) *If M converges to the truth in Q_e and $n>0$ and $C_e(n)$ is nonempty, then for each natural number m there exists w in $C_e(n)$ such that:*

$$\widehat{\varepsilon}(M,w \mid M_{[e_-]}=\sigma) > m.$$

The same is true when $\widehat{\varepsilon}(M,w \mid M_{[e_-]}=\sigma)$ is replaced by $Exp\left(\varepsilon_{M,w} \mid M_{[e_-]}=\sigma\right)$.

Proof. Suppose that $C_e(n)$ is nonempty and $n>0$. Let m be given. Then there exists a skeptical path $(S_0,\ldots,S_n)$ in K_e such that $c_e(S_n)=n$. Choose $\varepsilon>0$ and let $m'>m/(1-\varepsilon)$. Obtain w in $C_e(n)$ from lemma 2. Since the path is skeptical, $T_{S_{n+1}} \neq T_{S_n}$, so $T_{S_{n+1}}$ is incorrect of S_w. Since there are at least m' stages j along w at which $p(M_{w|j}=T_{S_{n+1}} \mid M_{[e_-]}=\sigma) > 1-\varepsilon$, it follows that $\widehat{\varepsilon}(M,w \mid M_{[e_-]}=\sigma) > m'(1-\varepsilon) > m$. For the bound on expected errors, apply lemma 4. $\dashv$

Lemma 12 *Suppose that $\widehat{\tau}(M,w,u \mid D)=j$. Then there exists $\varepsilon>0$ such that $\widehat{\tau}(M,w,v \mid D)=j$, for each v such that $u-\varepsilon<v\leq u$.*

Proof: Suppose that $\widehat{\tau}(M,w,u \mid D) = j$. Let $\varepsilon = u - \sum_{i=0}^{j-1} \widehat{\rho}(M,w,i)$. Then $\varepsilon > 0$, because $\widehat{\tau}(M,w,u \mid D) = j$ implies that $\sum_{i=0}^{j-1} \widehat{\rho}(M,w,i) < u$. Let $u - \varepsilon < v \leq u$. Then $\sum_{i=0}^{j-1} \widehat{\rho}(M,w,i) < v$. So $\widehat{\tau}(M,w,v \mid D) = j$. $\dashv$

In the following lemmas, assume that there are exactly k retractions in σ.

Lemma 13 (expected retraction times: lower bound) *Suppose that Q_e has no short paths, that M converges to the truth in Q_e, and that $C_e(n)$ is nonempty. Let m be a positive natural number. Then there exists w in $C_e(n)$ and loss function $\gamma \leq \rho$ such that:*

1. $Exp\left((Di)\ (\gamma_{M,w,i} \geq k+1) \mid M_{[e_-]} = \sigma\right) \geq l$ *in the defeat case;*
2. $Exp\left((Di)\ (\gamma_{M,w,i} \geq j) \mid M_{[e_-]} = \sigma\right) > m$
   (a) *for all j such that $k+1 < j \leq n+k+1$ in the defeat case;*
   (b) *for all j such that $k < j \leq n+k$ in the non-defeat case.*

Proof: Let $m > 0$ be given. Consider the defeat case, in which the last entry T in σ is not Ockham at e. Hence, there exists S_0 in K_e such that $c_e(S_0) = 0$ and $T \neq T_{S_0}$. Let $p = p(M_e = T_{S_0} \mid M_{[e_-]} = \sigma)$. We now use p to construct a finite input sequence e', which we use in turn to construct w in $C_e(n)$ and $\gamma \leq \rho$. If $p = 1$, then set $e' = e$. If $p < 1$, then $p(M_{[e_-]} = \sigma \wedge M_e \neq T_{S_0}) > 0$, and one can choose $\varepsilon > 0$ sufficiently small so that:

$$pl + (1-p)(l+1)(1-\varepsilon) > l.$$

To see that ε exists, note that $pl + (1-p)(l+1) > l$ when $p < 1$. Let w' in $C_e(0)$ be such that $S_{w'} = S_0$. As M is convergent in Q_e, there exists $m' > m/(1-(n+1)\varepsilon)$ such that:

$$p(M_{w'|m'} = T_{S_0} \mid M_{[e_-]} = \sigma) \quad > \quad 1-\varepsilon.$$

Set $e' = w'|m'$. Since $C_e(n)$ is nonempty and Q_e has no short paths, there exists a skeptical path $(S_0,\ldots,S_n)$ in $K_{e'}$ such that $c_{e'}(S_n) = n$. Apply lemma 2 to $(S_0,\ldots,S_n)$, ε, and e' to obtain w in $C_{e'}(n)$ and stages $m' = s_0 < \ldots < s_{n+1}$ such that for all $0 \leq i \leq n$, one has $s_{i+1} - s_i > m'$ and $p(M_{w|j} = T_{S_i} \mid M_{[e_-]} = \sigma) \geq 1-\varepsilon$, for each j such that $s_{i+1} - m \leq j \leq s_{i+1}$. Let U be the set of all ω in Ω such that $\bigwedge_{i=0}^{n} M_{w|s_{i+1}}(\omega) = T_{S_i}$. Let ω be in U. Then since $T \neq T_{S_0}$ and $T_{S_i} \neq T_{S_{i+1}}$ for all $0 \leq i \leq n$, the random output sequence $M_{[w|s_n]}(\omega)$ has retractions at some positions $r_0,\ldots,r_n$, such that $l < r_0 = m' \leq s_0 < r_1 \leq s_1 < \ldots s_n < r_{n+1} \leq s_{n+1}$. Let γ be just like ρ except that for each ω in U, the function $\gamma(M,w,i,\omega)$ has value 0 at each stage i between $m'+1$ and s_{n+1} along $M_{[w|s_n]}(\omega)$ *except* at the $n+1$ stages $r_0,\ldots,r_n$. Note that the retraction at stage r_j is the $k+j+1$th retraction of M along w, as M retracts k times along e_-. Now by construction of w and m':

$$p(M_{w|m'} = T_{S_0} \mid M_{[e_-]} = \sigma \wedge M_e \neq T_{S_0}) \quad > \quad 1-\varepsilon.$$

So since $p(M_e \neq T_{S_0} \mid M_{[e_-]} = \sigma) = 1 - p$, it follows that:

$$p(M_{w|m'} = T_{S_0} \wedge M_e \neq T_{S_0} \mid M_{[e_-]} = \sigma) \quad > \quad (1-p)(1-\varepsilon').$$

Thus, if $p < 1$, we have:

$$\mathrm{Exp}\left((\mathsf{D}i)\ (\gamma_{M,w,i} \geq k+1) \mid M_{[e_-]} = \sigma\right) > pl + (1-p)(1-\varepsilon')(l+1) > l,$$

and if $p = 1$, the expectation is just $pl = l$. So w and γ satisfy condition 1. Moreover, by construction of γ and w:

$$\mathrm{Exp}\left((\mathsf{D}i)\ (\gamma_{M,w,i} \geq k+j+1) \mid M_{[e_-]} = \sigma\right) > m' \cdot (1-(n+1)\varepsilon) > m,$$

so world w and γ satisfy condition 2a. The argument for 2(b) is similar but easier, since in the non-defeat case one may skip directly to the existence of $(S_0, \ldots, S_n)$ in the preceding argument. $\dashv$

Lemma 14 (push) *If $\hat{\gamma}$ is is a local loss function in chance and $\widehat{\gamma}(M,w) \geq v$ and $u < v$, then:*

$$(Di)\ (\widehat{\gamma}(M,w,i \mid M_{[e_-]} = \sigma) \geq u) \leq (Di)\ (\widehat{\gamma}(M,w,i \mid M_{[e_-]} = \sigma) \geq v).$$

Furthermore, if $v > 0$, then for each natural number s, if $\sum_{i=1}^{s} \widehat{\gamma}(M,w,i) < v$, then

$$(Di)\ (\widehat{\gamma}(M,w,i \mid M_{[e_-]} = \sigma) \geq v) \geq s+1.$$

Proof: Immediate consequence of the definition of $(\mathsf{D}i)\ (\widehat{\gamma}(M,w,i \mid M_{[e_-]} = \sigma) \geq u)$. $\dashv$

Lemma 15 (retraction times in chance: lower bound) *Suppose that Q_e has no short paths, that M converges to the truth in Q_e, and that $C_e(n)$ is nonempty. Let m be a positive natural number. Then there exists $\widehat{\gamma} \leq \widehat{\rho}$ such that for all $\varepsilon > 0$ there exists world w in $C_e(n)$ such that:*

1. $(Di)\ (\widehat{\gamma}(M,w,i \mid M_{[e_-]} = \sigma) \geq u) \geq l$,

   for all u such that $k < u \leq k+n+1-\varepsilon$ in the defeat case;

2. $(Di)\ (\widehat{\gamma}(M,w,i \mid M_{[e_-]} = \sigma) \geq u) > m$,

   (a) *for all u such that $k+1 < u \leq n+k+1-\varepsilon$ in the defeat case;*

   (b) *for all u such that $k < u \leq n+k-\varepsilon$ in the non-defeat case.*

Proof: Let $\varepsilon, m > 0$. Consider the defeat case. The last entry T in σ is not Ockham at e. Hence, there exists S_0 in K_e such that $c_e(S_0) = 0$ and $T \neq T_{S_0}$. Since $C_e(n)$ is nonempty and Q_e has no short paths, there exists a skeptical path $(S_0, \ldots, S_n)$ in Q_e such that $c_e(S_n) = n$. Let $\varepsilon' < \varepsilon/2(n+1)$. Apply lemma 2 to obtain w in $C_e(n)$ such that there exist stages of inquiry $l = s_0 < \ldots < s_{n+1}$ such that for each i from 0 to n,

stage s_{i+1} occurs more than m stages after s_i and $p(M_{w|j} = T_{S_i} \mid D) > 1-\varepsilon'$, at each stage j such that $s_{i+1} - m \leq j \leq s_{i+1}$.

With respect to w, define $\widehat{\gamma}$ recursively as follows. Let $\widehat{\gamma}$ agree with $\widehat{\rho}$ except that (i) at stages s such that $l \leq s < s_1$, we let $\widehat{\gamma}(M,w,s \mid M_{[e_-]} = \sigma) = \min(a,b)$, where $a = \widehat{\rho}(M,w,s \mid M_{[e_-]} = \sigma)$ and $b = k+1 \ominus \sum_{i=0}^{s-1} \widehat{\gamma}(M,w,i \mid M_{[e_-]} = \sigma)$. The idea is that $\widehat{\gamma}$ accumulates fractional retractions greater than $k+1$ only after stage s_1, but s_1 occurs after a delay longer than m stages after stage $s_0 = l$.

By definition of $\widehat{\gamma}$, method M accumulates quantity k of $\widehat{\gamma}$ along e_-. Further, since $p(M_{w|(l-1)} = T \mid D) \geq 1$ and $T \neq T_{S_0} \neq \ldots \neq T_{S_n}$, method M accumulates at least $1-\varepsilon'$ quantity of $\widehat{\gamma}$ over stages s from $l-1$ to s_1 and at least $1-2\varepsilon'$ quantity of $\widehat{\gamma}$ over stages s such that $s_i < s \leq s_{i+1}$, for i from 1 to n. Thus:

$$(*) \quad \widehat{\gamma}(M,w) \geq k + (n+1) - 2(n+1)\varepsilon' > k + n + 1 - \varepsilon.$$

Let u be such that $k < u \leq k+n+1-\varepsilon$. By hypothesis, $\sum_{i=1}^{l-1} \widehat{\gamma}(M,w,i \mid M_{[e_-]} = \sigma) = k$. So by statement (*) and lemma 14, we have that $(\mathsf{D}i)$ $(\widehat{\gamma}(M,w,i \mid M_{[e_-]} = \sigma) \geq u) \geq l$, for all u such that $k < u \leq k+n+1-\varepsilon$. That establishes statement 1.

Statements 2(a) and 2(b) are trivially true when $n = 0$. Suppose that $n > 0$. Let u be such that $k+1 < u \leq k+n+1-\varepsilon$. By statement 1, $(\mathsf{D}i)$ $(\widehat{\gamma}(M,w,i \mid M_{[e_-]} = \sigma) \geq k+1) \geq l$. Furthermore, $\widehat{\gamma}$ accumulates no more than quantity $k+1$ before stage $s_1 > m$. So by statement (*) and lemma 14, statement 2(a) follows.

Now consider the non-defeat case when $n > 0$. Let $\varepsilon' > \varepsilon/2n$, and apply lemma 2 to obtain a world w in $C_e(n)$. Define $\widehat{\gamma}$ to accumulate nothing at each s along w such that $l \leq s < s_1$ and to agree with $\widehat{\rho}$ along w otherwise. Arguing as before, but without the first retraction due to the defeat case, obtain:

$$(\dagger) \quad \widehat{\gamma}(M,w) \geq k + n - 2n\varepsilon' > k + n - \varepsilon.$$

By hypothesis, $(\mathsf{D}i)$ $(\widehat{\gamma}(M,w,i \mid M_{[e_-]} = \sigma) \geq k) \geq l-1$. Furthermore, $\widehat{\gamma}$ accumulates no more than quantity k before stage $s_1 > m$. So by statement (†) and lemma 14, statement 2(b) follows. $\dashv$

Lemma 16 (retraction times: lower bound for violators) *Suppose that Q_e has no short paths, that M converges to the truth in Q_e, and that $C_e(n)$ is nonempty. Let m be a positive natural number. Then there exists w in $C_e(n)$ such that if $\widehat{\tau}(M,w,k+1 \mid M_{[e_-]} = \sigma) \leq l$, then there exists $\alpha > 0$ such that:*

1. *$\widehat{\tau}(M,w,k+n+1+\alpha \mid M_{[e_-]} = \sigma) > 0$ and $Exp\left(\tau_{M,w,k+n+2} \mid M_{[e_-]} = \sigma\right) > 0$, if Ockham's razor is violated at (e,σ);*
2. *$\widehat{\tau}(M,w,k+n+\alpha \mid M_{[e_-]} = \sigma) > 0$ and $Exp\left(\tau_{M,w,k+n+1} \mid M_{[e_-]} = \sigma\right) > 0$ and the non-defeat case obtains, if stalwartness is violated at (e,σ).*

Proof: Begin with the bounds for retraction times in chance. Suppose that M violates Ockham's Razor at e by producing theory T. Then $p(M_e = T \mid M_{[e_-]} = \sigma) > \alpha'$ for some $\alpha' > 0$, and moreover, there exists S_0 such that $T \neq T_{S_0}$ and $c_e(S_0) = 0$. Since there are no short paths and $C_e(n)$ is nonempty, there exists skeptical path $(S_0, \ldots, S_n)$

in Q_e such that $c_e(S_n) = n$. Choose ε such that $0 < \varepsilon < \alpha'/2n$. Apply lemma 2 to $(S_0, \ldots, S_n)$ to obtain w in $C_e(n)$ and stages $l = s_0 < \ldots < s_{n+1}$ such that $s_i - s_{i+1} > m$ and $p(M_{w|s_{i+1}} = T_{S_i} \mid M_{[e_-]} = \sigma) \geq 1 - \varepsilon$, for each i from 0 to n. Suppose that $\widehat{\tau}(M, w, k+1 \mid M_{[e_-]} = \sigma) \leq l$. Then, since there are only k retractions along e_-, there must be a full retraction in chance at $e = w|s_0$. Since $T \neq T_{S_0} \neq \ldots \neq T_{S_n}$, there is at least $\alpha' - \varepsilon$ retraction in chance by s_1 and another $1 - 2\varepsilon$ retraction in chance between s_i and s_{i+1}, for $1 \leq i \leq n$. So it follows that $\widehat{\rho}(M, w \mid M_{[e_-]} = \sigma) \geq k + 1 + \alpha' + n(1 - 2\varepsilon) > k + n + 1$. Therefore, there exists $\alpha > 0$ such that $\widehat{\tau}(M, w, k + n + 1 + \alpha \mid M_{[e_-]} = \sigma) > 0$.

Next, suppose that M violates stalwartness at e. Then since stalwartness is violated, it follows that the last entry of σ is some T_S that is Ockham at e, so S is uniquely simplest at e and we are in the non-defeat case. Since there are no short paths and $C_e(n)$ is nonempty, there exists skeptical path $(S = S_0, \ldots, S_n)$ in Q_e such that $c_e(S'_n) = n$. Choose ε such that $0 < \varepsilon < 1/2n$. Apply lemma 2 to $(S_0, \ldots, S_n)$ to obtain w in $C_e(n)$ and stages $l = s_0 < \ldots < s_{n+1}$ such that $s_i - s_{i+1} > m$ and $p(M_{w|s_{i+1}} = T_{S_i} \mid M_{[e_-]} = \sigma) \geq 1 - \varepsilon$, for each i from 0 to n. Suppose that $\mathrm{Exp}\left(\tau_{M,w,k+1} \mid M_{[e_-]} = \sigma\right) \leq l$. Then, again, $p(M_e = T_S \mid M_{[e_-]} = \sigma) = 0$, which is one full retraction in chance at $e = w|s_0$. By choice of w, there is another $1 - 2\varepsilon$ retraction in chance between s_i and s_{i+1}, for $1 \leq i \leq n$. Thus, $\widehat{\rho}(M, w \mid M_{[e_-]} = \sigma) \geq k + 1 + n(1 - 2\varepsilon) > k + n$. So there exists $\alpha > 0$ such that $\widehat{\tau}(M, w, k + n + \alpha) \mid M_{[e_-]} = \sigma) > 0$.

For expected retraction times, first consider the Ockham violation case. Let w be constructed exactly as in the Ockham violation case of the proof of lemma 16. Suppose that $\mathrm{Exp}\left(\tau_{M,w,k+1} \mid M_{[e_-]} = \sigma\right) \leq l$. Then there is a full retraction at e, so $\widehat{\tau}(M, w, k+1 \mid M_{[e_-]} = \sigma) \leq l$. So $\widehat{\tau}(M, w, k + n + 1 + \alpha \mid M_{[e_-]} = \sigma) > 0$, by lemma 16. Therefore, $\widehat{\rho}(M, w \mid M_{[e_-]} = \sigma) > k + n + 1$. Hence, $\mathrm{Exp}\left(\rho_{M,w} \mid M_{[e_-]} = \sigma\right) > k + n + 1$, by lemma 4. Therefore, there exists finite answer sequence σ' of length l' extending σ such that more than $k + n + 1$ retractions occur in σ' and $p(M_{[w|l']} = \sigma' \mid M_{[e_-]} = \sigma) > 0$. So at least $k + n + 2$ retractions occur in σ'. Hence, $\mathrm{Exp}\left(\tau_{M,w,k+n+2} \mid M_{[e_-]} = \sigma\right) > 0$.

The stalwartness violation case is similar. $\dashv$

Lemma 17 (retraction times: upper bound) *Suppose that M is stalwart and Ockham from (e, σ) onward, such that $p(M_{[e_-]} = \sigma) > 0$. Then for each w in $C_e(n)$:*

1. $\widehat{\tau}(M, w, u \mid M_{[e_-]} = \sigma) \leq l$ *if $u \leq k + 1$ in the defeat case;*
2. $\widehat{\tau}(M, w, u \mid M_{[e_-]} = \sigma) = 0$ *if $u > k + n + 1$ in the defeat case;*
3. $\widehat{\tau}(M, w, u \mid M_{[e_-]} = \sigma) = 0$ *if $u > k + n$ in the non-defeat case.*

Furthermore, for each $j \geq n$:

4. $\mathrm{Exp}\left(\tau_{M,w,k+1} \mid M_{[e_-]} = \sigma\right) \leq l$ *in the defeat case;*
5. $\mathrm{Exp}\left(\tau_{M,w,k+j+2} \mid M_{[e_-]} = \sigma\right) = 0$ *in the defeat case;*
6. $\mathrm{Exp}\left(\tau_{M,w,k+j+1} \mid M_{[e_-]} = \sigma\right) = 0$ *in the non-defeat case;*

Proof: Let w be in $C_e(n)$ and let M be stalwart and Ockham from (e,σ) onward. Let T be the last entry in σ. Consider the defeat case. Then T is not Ockham at e. So $p(M_{e^-} = T \mid M_{[e^-]} = \sigma) = 1$ and $p(M_e = T \mid M_{[e^-]} = \sigma) = 0$, by Ockham's razor. Thus, $\widehat{\tau}(M,w,u) \leq l$, for each $u \leq k+1$, which establishes statement 1. For statement 4, note that if σ' of length $l+1$ extends σ and is such that $p(M_{[e]} = \sigma') > 0$, then, because M is Ockham from (σ,e) onward and T is not Ockham at e, it follows that the last entry of σ' is not T. So σ' contains a retraction at stage l. Hence, $\mathrm{Exp}\left(\tau_{M,w,k+1} \mid M_{[e^-]} = \sigma\right) = l$.

For statements 2 and 5, note that lemma 8 implies that M incurs expected retractions and retractions in chance at most at n positions $s_1 < \ldots < s_n$ along w. Thus, $\widehat{\tau}(M,w,u) = 0 \quad$ for each $u > k+n+1$, which establishes statement 2. For statement 5, each output sequence σ' of length greater than l has a retraction at position l followed by at most n more retractions. Thus, $\mathrm{Exp}\left(\tau_{M,w,k+j+2} \mid M_{[e^-]} = \sigma\right) = 0$ for each $j \geq n$.

For statements 3 and 6, drop retraction at e from the argument for statements 2 and 5. $\dashv$

Notes

[1]For discussion of the following, critical points, see (Kelly 2008, 2010) and (Kelly and Mayo-Wilson 2008).

[2]Nolan (1997), Baker (2003), and Baker (2007) claim that simpler theories are more explanatory. Popper (1959) and Mayo and Spanos (2006) both claim that simpler theories are more severely testable. Friedman (1983) claims unified theories are simpler, and finally, Li and Vitanyi (2001) and Simon (2001) claim that simpler theories are syntactically more concise.

[3]See (Forster and Sober 1994), (Vapnik 1998), (Hitchcock and Sober 2004), and (Harman and Kulkarni 2007).

[4]More precisely, in regression and density estimation, the predictive accuracy of the model-selection techniques endorsed by Forster, Sober, Harman, and Kulkarni are evaluated only with respect to the distribution *from which the data are sampled.* Thus, for example, one can approximate, to arbitrary precision, the joint density of a set of random variables and yet make arbitrarily bad predictions concerning the joint density when one or more variables are manipulated. The objection can be overcome by estimating from experimental data, but such data are often too expensive or unethical to obtain when policy predictions are most vital.

[5]See Jeffreys (1961) and Rosenkrantz (1977), respectively, for arguments that explicitly and implicitly assume that simpler theories are more likely to true.

[6]It is usually assumed that the data are received according to a Gaussian distribution centered on the true value of Y for a given value of X. Since our framework does not yet handle statistical inference, we idealize by assuming that the successive data fall within ever smaller open intervals around the true value Y.

[7]In this paper, empirical effects are stipulated. It is also possible to define what the empirical effects are in empirical problems in which they are not presupposed (Kelly 2007b, c). The same approach could have been adopted here.

[8]In Plato's dialogue *Meno*, knowledge is distinguished from true belief in terms of the former's stability—it is chained down by the evidence and does not run away. A similar moral is drawn by advocates of indefeasibility theories of knowledge (e.g., Lehrer 1990), according to which knowledge is true belief that true information would never defeat. We thank the anonymous referee for pointing out this the apparent conflict between delaying pain and accelerating retractions.

[9]For a simpler proof restricted to the deterministic case, cf. (Kelly and Mayo-Wilson 2010a), and similarly for theorems 2 and 3.

[10]In other words, $\{M_e : e \in F_K\}$ is a discrete, branching, stochastic process assuming values in $\mathscr{A}$.

[11] This argument was originally sketched, with some slight differences, by Kelly and Glymour (2004).

[12]For an outline of a more general theory of forceable retractions of statistical hypotheses, see (Kelly and Mayo-Wilson 2010b). There, we define a partial order $\preceq$ on sets of probability distributions that are

faithful to directed acyclic graphs (considered as causal networks), and show that any consistent procedure for inferring causal networks can be forced to accrue n expected retractions if there is a sequence of sets of distributions $A_1 \preceq A_2 \ldots \preceq A_n$ of length n. We expect the same partial-order can be employed in more general statistical settings.

[13]We are indebted to Hanti Lin for bringing this important point to our attention.

[14]All but the first issue are discussed in depth in Mayo-Wilson (2009).

[15]See Fishburn (1972) for a proof that Von-Neumann's theorem fails when players' preferences are non-Archimedean.

[16]See Mayo-Wilson (2009) for one proof; a second proof was suggested to us independently by both Teddy Seidenfeld and an anonymous referee, and involves extending pre-orders to total orders (which requires use of Zorn's Lemma for infinite games) and then applying standard game-theoretic theorems guaranteeing the existence of Nash equilibria in games in which players preferences are totally ordered.

[17]See, for example, Karlin (1959).

[18]The idea that purely-finitely additive strategies might be used to guarantee solutions in infinite games in which standard assumptions fail was first suggested by Karlin (1950), in which it was proved that equilibria exist in two person, zero sum games in which (a) pairs of players actions are points in the the unit square in $\mathbb{R}^2$, and (b) payoffs to both players were bounded. The theorem was extended by Yanoskaya (1970) and Heath and Sudderth (1972) for arbitrary two person-zero sum games in which one of the players payoffs is a bounded function when the other player's strategy is held fixed. Kadane, Schervish, and Seidenfeld (1999) drop the boundedness assumption. It is important to note that evaluation of losses in games in which players are permitted to employ finitely-additive strategies depends upon the order in which integration is specified, as Fubini's theorem fails for finitely-additive measures. Part of the importance of Yanoskaya, Kadane, Schervish, and Seindenfeld's result is that their formalism eliminates some arbitrariness in the specification of order of integration.

[19]Again, see Mayo-Wilson (2009). Interpreting Nature's mixed strategies for Nature as prior probabilities is not novel. It was suggested, to our knowledge, first by Wald (1950).

[20]The difficulties are exacerbated when scientist's prior probability (i.e. Nature's mixed strategy) is only finitely additive, as there is no obvious concept of "support" in that case, even over countable sets of worlds.

References

[1] Akaike, H. (1973) "Information theory and an extension of the maximum likelihood principle", *Second International Symposium on Information Theory*, pp. 267-281.

[2] Baker, A. (2003) Quantitative Parsimony and Explanatory Power, *British Journal for the Philosophy of Science* 54: 245-259.

[3] Baker, A. (2007) Occam's Razor in Science: A Case Study from Biogeography. *Biology and Philosophy.* 22: 193215.

[4] Cartwirght, N. (1999) *The Dappled World: A Study of the Boundaries of Science*, Cambridge: Cambridge University Press.

[5] Garey, M. and Johnson, D. (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness*, New York: W. H. Freman.

[6] Fishburn, P. (1972) "On the Foundations of Game Theory: The Case of Non-Archimedean Utilities." *International Journal of Game Theory.* Vol. 2, pp. 65-71.

[7] Forster, M. (2001) The New Science of Simplicity. (In A. Zellner, H. Keuzenkamp, and M. McAleer (Eds.) *Simplicity, Inference and Modelling.* (pp. 83-119). Cambridge: Cambridge University Press)

[8] Forster M. and Sober, E. (1994) How to Tell When Simpler, More Unified, or Less Ad Hoc Theories will Provide More Accurate Predictions. *The British Journal for the Philosophy of Science* 45, 1 - 35.

[9] Friedman, M. (1983) *Foundations of Spacetime Theories: Relativistic Physics and Philosophy of Science.* Princeton University Press.

[10] Gärdenfors, P., *Knowledge in Flux: Modeling the Dynamics of Epistemic States*, Cambridge: M.I.T. Press.

[11] Harman, G. and Kulkarni, S. (2007) *Reliable Reasoning: Induction and Statistical Learning Theory* (Cambridge: MIT Press).

[12] Heath, D. and Sudderth, W. (1972) "On a theorem of de Finetti, Oddsmaking and Game Theory." *Annals of Mathematical Statistics.* Vol. 43. No.6. 1972. pp. 2072-2077.

[13] Hempel, C. (1966) *Philosophy of Natural Science* (Englewood-Cliffs: Prentice Hall).

[14] Hitchcock, C. and Sober, E. (2004) "Prediction Versus Accommodation and the Risk of Overfitting." *British Journal for the Philosophy of Science.* 55: pp.

[15] Jeffreys H. (1961) *Theory of Probability.* Oxford: Clarendon Press.

[16] Kadane, J., Schervish, M., and Seidenfeld, T. *Rethinking the Foundations of Statistics.* Cambridge University Press, 1999.

[17] Karlin, S. (1950)Operator Treatment of the Minimax Principle, in H. W. Kuhn and A. W. Tucker, eds, *Contributions to the Theory of Games.* Annals of Mathematics Studies, 24. Princeton, N.J: Princeton University Press, pp. 133-154.

[18] Karlin, S. *Mathematical Methods and Theory in Games, Programming and Economics,* Reading, Mass: Addison-Wesley Publishing Co., 1959.

[19] Kelly, K. (1996) *The Logic of Reliable Inquiry.* Oxford University Press.

[20] Kelly, K. (2002)Efficient Convergence Implies Ockham's Razor,*Proceedings of the 2002 International Workshop onvComputational Models of Scientific Reasoning and Applications*, Las Vegas, USA, June 24-27.

[21] Kelly, K. (2004) Justification as Truth-finding Efficiency: How Ockham's Razor Works, *Minds and Machines* 14: 485-505.

[22] Kelly, K. (2007a) A New Solution to the Puzzle of Simplicity, *Philosophy of Science* 74: 561-573.

[23] Kelly, K. (2007b)How Simplicity Helps You Find the Truth Without Pointing at it, V. Harazinov, M. Friend, and N. Goethe, eds. *Philosophy of Mathematics and Induction*, Dordrecht: Springer, pp. 321-360.

[24] Kelly, K. (2007c) "Ockham's Razor, Empirical Complexity, and Truth-finding Efficiency," *Theoretical Computer Science*, 383: 270-289.

[25] Kelly, K. (2007d) Simplicity, Truth, and the Unending Game of Science, *Infinite Games: Foundations of the Formal Sciences V*, S. Bold, B. Löwe, T. Rsch, J. van Benthem eds, Roskilde: College Press 2007 pp. 223-270.

[26] Kelly, K. (2008) Ockhams Razor, Truth, and Information, in *Philosophy of Information*, Van Benthem, J. Adriaans, P. eds. Dordrecht: *Elsevier*, 2008 pp. 321-360.

[27] Kelly, K. (2010) Simplicity, Truth, and Probability, in *Handbook on the Philosophy of Statistics*, Forster, M., and Bandyopadhyay, P., eds. Dordrecht: Kluwer.

[28] Kelly, K. and Glymour, C. (2004) Why Probability Does Not Capture the Logic of Scientific Justification, C. Hitchcock, ed., *Contemporary Debates in the Philosophy of Science*, Oxford: Blackwell, 2004 pp. 94-114.

[29] Review of Gilbert Harman and Sanjeev Kulkarni, *Reliable Reasoning: Induction and Statistical Learning Theory*, *Notre Dame Philosophical Reviews* http://ndpr.nd.edu/board.cfm.

[30] Kelly, K. and Mayo-Wilson, C. (2010a)"Ockham Efficiency Theorem for Random Empirical Methods." Technical Report ***, Department of Philosophy, Carnegie Mellon University. ⟨ http://www.andrew.cmu.edu/***⟩.

[31] Kelly, K. and Mayo-Wilson, C. (2010b)"Causal Conclusions that Flip Repeatedly and their Justification." *Under Review.*

[32] Lehrer, K. (1990) *Theory of Knowledge*, New York: Routledge.

[33] Kuhn, T. (1970)*The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.

[34] Lawlor, G. (2006) *Introduction to Stochastic Processes*, 2nd ed., New York: Chapman and Hall.

[35] Levi, I. (1976) *Gambling with Truth*, Cambridge: M.I.T. Press.

[36] Li, M. and Vitanyi, P. (1997) *An Introduction to Kolmogorov Complexity and Its Applications.* New York: Springer Verlag.

[37] Li, M. and Vitanyi, P. (2001) Simplicity, Information, and Kolmogorov Complexity, in A. Zellner, H. Keuzenkamp, and M. McAleer eds., *Simplicity, Inference and Modelling*, Cambridge: Cambridge University Press, pp. 83-119.

[38] Mayo, D. (1996) *Error and the Growth of Experiment Knowledge*, Chicago: University of Chicago Press.

[39] Mayo, D. and Spanos, A. (2006) Severe Testing as a Basic Concept in a Neyman-Pearson Philosophy of Induction, *British Journal for the Philsophy of Science*, 57: 323-357.

[40] Mayo-Wilson, C. (2009) A Game Theoretic Argument For Ockham's Razor. Master's Thesis, Department of Philosophy, Carnegie Mellon University..

[41] Nolan, D. (1997) Quantitative Parsimony, *British Journal for the Philosophy of Science,* 48: 329-343.

[42] Popper, K. (1959) The Logic of Scientific Discovery. London: Hutchinson.

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

[44] Reichenbach, H. (1938) *Experience and Prediction.* University of Chicago Press.

[45] Rissanen, J. (1983) A universal prior for integers and estimating by minimum description length. *The Annals of Statistics,* 11: 416-431.

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

[47] Rosenkrantz, R. (1977) *Inference, Method, And Decision: Towards A Bayesian Philosophy Of Science*, Boston: Reidel.

[48] Salmon, W. (1966) *The Foundations of Scientific Inference*, Pittsburgh: University of Pittsburgh Press.

[49] Savage, L. (1972) *The Foundations of Statistics*, New York: Dover.

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

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

[52] Schulte, O. (2001) Inferring Conservation Principles in Particle Physics: A Case Study in the Problem of Induction. *The British Journal for the Philosophy of Science,* 51, 771-806.

[53] Schulte, O. (2008) The Co-Discovery of Conservation Laws and Particle Families. *Studies in History and Philosophy of Modern Physics* 39: 288-314.

[54] Schulte, O., Luo, W., and Griner, R. (2007) Mind Change Optimal Learning of Bayes Net Structure. *20th Annual Conference on Learning Theory* (COLT), San Diego.

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

[56] Simon, H. (2001) Science Seeks Parsimony, not Simplicity: Searching for Pattern in Phenomena. (In A. Zellner, H. Keuzenkamp, and M. McAleer eds., *Simplicity, Inference and Modelling.* pp. 83-119, Cambridge: Cambridge University Press.

[57] Spirtes, P., Glymour, C. and Scheines, R. (2001) *Causation, Prediction, and Search*, 2nd. ed., Cambridge: M.I.T. Press.

[58] Vapnik, V. (1998) *Statistical Learning Theory*, New York: Wiley.

[59] van Fraassen, B. (1980) *The Scientific Image.* Oxford University Press.

[60] Wrinch, D. and Jeffreys, H. (1923) On Certain Fundamental Principles of Scientific Inquiry. *Philosophical Magazine,* 45: 368-374.

[61] Yanoskaya, E.B. (1970). The Solution of Infinite Zero-Sum, Two-Person Games with Finitely Additive Strategies. *Theory of Probability and Applications*. 15: 153-158.

[62] Wald, A. (1950). *Statistical Decision Functions*. New York: Wiley.

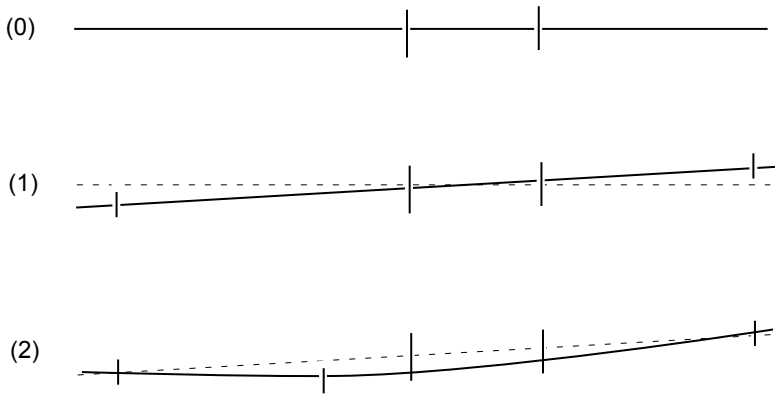

Figure 1: empirical effects and polynomial degree

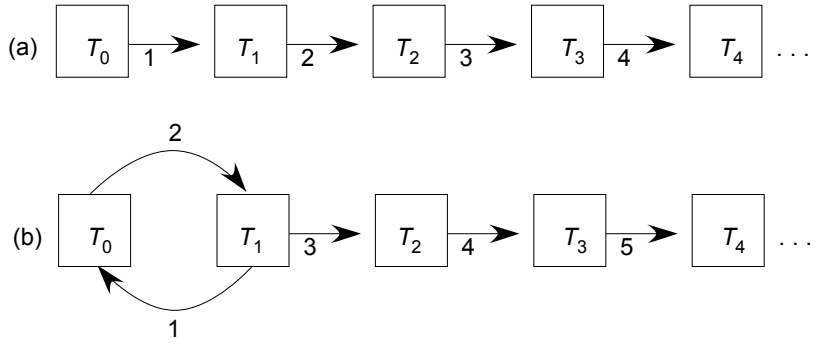

Figure 2: the Ockham efficiency argument