Some People Never Learn, Rationally: Multidimensional Learning Traps and Smooth Solutions of Dynamic Programs

Simon Loertscher
University of Melbourne*

Andrew McLennan
University of Queensland†

October 8, 2020

Abstract

We give conditions under which the smoothness properties of the value and policy functions of a dynamic program at discount factor zero extend to small positive discount factors. We apply this to a model of Bayesian learning by a decision maker who does not know which of several parameters is true. In each period she chooses an action from an open subset of a Euclidean space, observes one of finitely many possible outcomes, and updates her beliefs. There is an action that is uninformative in the sense that when it is chosen, all parameters give the same distribution over outcomes, and consequently beliefs do not change. We give conditions under which a policy specifying an action as a function of the current belief results in a positive probability that the sequence of beliefs converges to a belief at which the uninformative action is chosen, so that learning is asymptotically incomplete. Our dynamic programming results imply that when myopically optimal behavior leads with positive probability to asymptotically incomplete learning, such a "learning trap" also exists when the discount factor is positive (so that optimal behavior takes the value of experimentation into account) but sufficiently small.

*Email: simonl@unimelb.edu.au.

†Email: a.mclennan@economics.uq.edu.au. McLennan's work was funded in part by Australian Research Council grant DP0773324. A collaboration with Mark Feldman was an important precursor to this project; his insights and influence are gratefully acknowledged. We have benefitted from helpful comments by anonymous referees and remarks by seminar participants at the University of Queensland, University of Melbourne, the University of British Columbia, Keio University, Kyoto University, the 2012 Shanghai Microeconomics Workshop, the 2013 Australasian Econometric Society Meetings, and SAET 2013.

Running Title: Some Never Learn

Keywords: Bayesian learning, learning traps, information cascades, dynamic programming, stochastic optimal control.

Journal of Economic Literature Classification Numbers C61, D83.

1 Introduction

This paper studies stationary discounted discrete time stochastic dynamic programs. In particular, we study programs in which the set of states is the set of possible beliefs about an underlying parameter that the decision maker is learning about from experience. There are two main contributions.

Consider a dynamic program in which the sets of states and actions are compact subsets of smooth manifolds, and the per period reward function is smooth (in senses to be specified later). For such a problem the value function is the unique fixed point of the Bellman operator, and the optimal actions for any state are those that maximize the sum of the current reward and the discounted expectation of the value of the state in the next period. When the discount factor is zero, the optimal actions are the solutions of the optimization problem given by the per period reward function, so there is a smooth optimal policy function if the per period reward function satisfies the relevant second order conditions strictly, and (in part by virtue of the envelope theorem) the value function inherits smoothness properties from the per period reward function. Our first main contribution is to give conditions under which the smoothness properties of the value and policy functions at discount factor zero also hold for the value and policy functions of small positive discount factors, and these functions vary continuously (in an appropriate topology) with the discount factor.

Now consider a model in which an agent learns about an underlying parameter with finitely many possible values. In each period the agent begins with a belief about the parameter, chooses an action, and observes an outcome. The parameter governs the statistical relationship between the action and the outcome, so at the end of each period the agent revises her beliefs via Bayes' rule. The per period reward is a function of the parameter and the action, so its expectation is a function of the action and the current belief. The agent's goal is to maximize the expectation of the discounted sum of rewards, so her problem is a discrete time stochastic dynamic program in which the state space is the simplex of possible beliefs about the parameter, and there is a trade off between current reward and the value of learning via experimentation. Our conceptual concern is whether the agent necessarily learns the value of the parameter asymptotically.

An action is *uninformative* if, when that action is chosen, the distribution over outcomes does not depend on the parameter, so that Bayesian updating after choosing this action does not revise the belief. A belief is *critical* if the unique action maximizing expected current reward for that belief is uninformative. That is, a belief is critical if the unique optimal action for that belief when the discount factor is zero is uninformative. Our second main contribution is to give conditions under which the uninformative action is optimal for the critical belief for some range of positive discount factors, and to show that when the initial belief is close to the critical belief there can be a positive probability that the sequence of beliefs generated by optimal behavior converges to the critical belief, so that optimal behavior does not lead to complete learning in the limit. We describe this phenomenon by saying that there is a *learning trap* at the critical belief. In brief, the general results for dynamic programs imply that the optimal policy function for a sufficiently small positive discount factor is close (in an appropriate topology) to the optimal policy when the discount factor is zero, and the desired conclusions follow from this if certain conditions are satisfied.

We now describe each of these contributions in somewhat more detail.

1.1 The Dynamic Programming Results

In our dynamic program the spaces of states and actions are compact subsets of smooth (actually, "smooth enough") manifolds that are the closures of their interiors. There is a stochastic transition that assigns a probability distribution over states tomorrow to each state-action pair today. For a given *discount factor*[1] $\delta \in (-1, 1)$ and a given initial state the problem is to choose a sequence of actions, conditional on available information, that maximizes the expectation of the usual sum of discounted rewards.

For the given δ there is a *value function* that maps each state to the optimized value of the program for δ and that initial state. This value function is the unique fixed point of an updating operator mapping a "candidate value function" to the function computing the maximized expectation, for each state, of the sum of the current period reward and the candidate value function's discounted valuation of the next period's state. Usually we think of this operator as a contraction with respect to the sup norm, but, in order to show that derivatives of the value function vary continuously as δ varies, we need to show that this operator is a contraction with respect to a norm that metrizes a finer topology that is sensitive to these derivatives. In a nutshell, our method for showing that the value function is "nice" is to construct a set of "nice" candidate value functions and a

[1] Our presentation points out that the mathematics does not depend on the discount factor being positive, but we do not claim to have economically interesting applications of negative discount factors.

"nice" norm such that the set of candidate value functions is complete with respect to this norm, the set is mapped to itself by the updating operator, and the restriction of the updating operator to this set is a contraction with respect to the metric. The contraction mapping theorem implies that the operator has a fixed point in the set, which is of course the value function. A parameterized version of the contraction mapping theorem will show that the value function varies continuously (with respect to our metric) as δ and other parameters vary. Possibly this program can be carried out in various ways. Our method constructs a set of candidate value functions that are C^r, and whose r^{th} order partial derivatives satisfy certain Lipschitz bounds. By virtue of reasoning recalling the Arzela-Ascoli theorem, this set of candidate value functions is complete with respect to an appropriate C^r metric.

Two technical assumptions are required by this method. First, we assume that for each state there is a unique action maximizing the reward function at which the second order conditions hold strictly. As the discount factor varies near zero, the sum of the reward function and the expectation of the value function at the next period state varies continuously, in the appropriate C^r topology, and consequently for discount factors in a range around zero this sum has a unique maximizing action for each state, at which the second order conditions hold strictly. For discount factors in this range the implicit function theorem implies that the optimal policy function is C^{r-1}, and that the optimal policy function varies continuously in the appropriate C^{r-1} topology. Even though the optimal policy is only C^{r-1}, the envelope theorem implies that the value of the problem as a function of the state is a C^r function that varies continuously, in the appropriate C^r topology, as the discount factor varies.

Second we need the transition mapping a state-action pair today to the distribution of tomorrow's state to be suitably smooth. This will be expressed indirectly. There is an operator mapping a candidate value function to the function computing the expectation of tomorrow's value, as a function of today's state-action pair. We require that this operator maps sets of C^r functions satisfying Lipschitz bounds on the top derivative to sets of C^r functions satisfying similar bounds. Since there are many possible models, it would be difficult to present useful conditions on primitives under which this condition holds, and we do not attempt to do so. Instead this condition must be verified "by hand" in each application.

1.2 The Learning Model

In our learning model there is a finite set of possible parameters, the set of actions is an open subset of a Euclidean space, and there is a finite set of outcomes. There is a C^1

function that maps each parameter-action pair to an interior probability distribution on the set of outcomes. In each period the decision maker chooses an action, observes an outcome, and updates her beliefs concerning the parameter using Bayes' rule.

We study a given C^1 policy function mapping beliefs to actions that is stationary, in the sense of being applied in every period. (The initial phase of our analysis does not depend on whether this is in any sense optimal.) This structure gives rise to a Markov process in which the state space is the set of possible beliefs, i.e., the set of probability distributions on the set of possible parameters. Basic properties of Bayesian updating imply that this process is a martingale: conditional on any belief today, the expectation of tomorrow's belief is today's belief.

An action is *uninformative* if, when that action is chosen, the distribution over outcomes does not depend on the parameter. A belief is *critical* if the policy maps it to an uninformative action. A critical belief is a stationary point of the dynamic process. We assume that the probabilities of the various outcomes, conditional on the various parameters, are C^1 functions of the action. Together with the fact that the policy is C^1, this implies that, for beliefs close to the critical belief, the learning process is accurately approximated by a process given by the derivatives of the policy and outcome probability functions.

Suppose that the derivative of the policy function at the critical belief is nonzero, but quite small. Then the action the policy function prescribes at a given prior belief near the critical belief is quite close to being uninformative, so the distribution of the posterior has a finite support (because there are finitely many outcomes) that is close to the prior. Roughly, if the posterior is closer to the critical belief than the prior, then there is less experimentation in the next period, so that the process slows down, and if the posterior is farther away, then the process speeds up. For example, if the difference between the posterior and the critical belief is one half of the difference between the prior and the critical belief, then the process going forward from the posterior is roughly the process going forward from the prior rescaled by a factor of one half. Similarly, if the difference between the posterior and the critical belief is three halves of the difference between the prior and the critical belief, then the process going forward from the posterior is approximately the process going forward from the prior rescaled by a factor of three halves. As this example suggests, moving closer by a certain distance has a more powerful effect than moving away by that distance, and this can create a "stochastic ratchet effect" that can (with high probability) draw the sequence of posteriors closer and closer to the critical belief.

One must also consider the extent to which learning moves beliefs in directions that are orthogonal to the difference between the prior and the critical belief. In the extreme case where all learning is of this sort, convergence to the critical belief is impossible. We consider the limiting learning process given by the derivatives of the policy and outcome functions at the sum of the critical belief and a unit vector in the space of vectors tangent to the simplex of possible beliefs. We will show that if, for each unit vector, the expectation of the logarithm of the distance from the posterior to the critical belief is negative (which is to say, less than the logarithm of the distance between the prior and the critical belief) then the sequence of logarithms of distances between the current belief and the critical belief is a supermartingale. When this is the case a version of the law of large numbers will imply that when the initial belief is near to the critical belief, there is a high probability that the sequence of posterior beliefs converges to the critical belief.

Of course this will not happen if the derivative of the policy function is large. For example, an extreme possibility is that for a prior close to the critical belief, every point in the support of the distribution of the posterior belief is further from the critical belief than the prior. More generally, if, for each unit vector, the expectation of the logarithm of the distance from the posterior to the critical belief is positive, our version of the law of large numbers will imply that the probability of convergence to the critical belief is zero.

Can there be a positive probability of convergence to the critical belief even when the policy function is optimal for a positive discount factor, so that learning about the parameter will be rewarded in the future? Suppose that the dimension of the space of beliefs is at least as large as the dimension of the space of actions, there is a single uninformative action, and the derivative of the policy function at each critical belief has full rank. Then the regular value theorem (a "coordinate free" version of the implicit function theorem) implies that the set of critical beliefs is a manifold. If, in addition, there is uniform bound on the norms of the derivatives of the policy function at critical beliefs, then a generalization of the argument described above implies that there is a positive probability that the sequence of posterior beliefs converges to the set of critical beliefs, so that learning is incomplete.

It is easy to construct examples of learning problems for which the optimal policy function for discount factor zero has the qualitative properties described above. Our dynamic programming result gives conditions under which the optimal policy for discount factor δ varies continuously, as δ varies near 0, in a suitable C^1 topology, so that these properties are also satisfied by the optimal policy functions for positive discount factors sufficiently close to zero. In this way we establish that learning traps are possible for positive discount factors.

1.3 Related Literature

Prior literature on smoothness of the value and policy functions seems to be quite sparse. Benveniste and Scheinkman (1979) give conditions under which the value function is differentiable at a point, and stronger conditions under which it is C^1. Blume et al. (1982) give various conditions under which the value function is C^r, and there is a unique optimal policy which is stationary and C^r. However, they restrict attention to dynamic programs for which there is a unique optimal policy for any discount factor (by virtue of their Theorem 1.1(iv)). In learning problems it is generally suboptimal to choose an action that is close to uninformative if the discount factor is close to one. In the example from McLennan (1984) that we study in Section 3 this implies that the optimal policy for a large discount factor must have a jump from one side of the uninformative action to the other, so that for some belief two different actions are optimal. Thus the results of Blume et al. (1982) are not applicable to our framework.

There is a much more extensive literature concerning learning and learning traps. Of course learning and experimentation are omnipresent in economic life. Firms need to design new products and hire new employees, consumers have to choose hairstylists and restaurants, and politicians have to pick policies in an uncertain world. For the central economic models (systems of markets, noncooperative games) the relevance of equilibrium (market clearing, Nash equilibrium) is to a greater or lesser extent justified by a presumption that we are now observing a steady state that emerged from a process of learning and adjustment. Whether a decision maker will eventually learn, or fail to learn, the underlying parameters, is a natural and relevant question.

Some of the oldest literature on learning traps concerns multiarmed bandits (Gittens and Jones (1974), Berry and Fristedt (1985), and references therein). Rothschild (1974) introduced this topic to economics, with subsequent contributions by McLennan (1984), Kihlstrom et al. (1984), Easley and Kiefer (1988), Aghion et al. (1991), Moscarini and Smith (2001), and many others. Bergemann and Välimäki (2008) provide a survey and summary. There is also a now quite extensive literature studying Bayesian learning in multiagent environments[2].

[2]In a self-confirming equilibrium (Fudenberg and Levine (1993)) of an extensive form game the agents may have incorrect beliefs concerning strategies at unreached information sets. The study of incomplete learning resulting from "information cascades" was initiated by Banerjee (1992) and Bikhchandani et al. (1992) and continued by Aoyagi (1998), Smith and Sørensen (2000), Keller et al. (2005), among many others. In such models initial experimentation reduces uncertainty, and agents are not compensated for the external benefits of experimentation. In contrast Bolton and Harris (1999, 2000) provide models

From a mathematical point of view there are several reasons a single decision maker might fail to experiment. If the space of possible actions is discrete, as in the bandit literature (e.g., Gittens and Jones (1974), Rothschild (1974), and Banks and Sundaram (1992)) and the literature on information cascades (Banerjee (1992) and Bikhchandani et al. (1992)), there may be a positive lower bound on the costs of experimentation in a single period. Warren and Wilkening (2012) present a model with a somewhat different structure in which a regulator follows a static policy as a result of uncertainty resulting from failure to experiment. In addition, there may be switching costs (e.g., Banks and Sundaram (1994)) which loom large in many labor market applications. When the space of actions has an uninformative action on its boundary, the expected loss in the current period, and the amount of information acquired resulting from moving away from that point, will typically be proportional to the distance moved (see Radner and Stiglitz (1984) for one formalization of this notion) and in many setups it is easy to see that not experimenting is optimal when the future is sufficiently heavily discounted.

Finally, there is the possibility that the relevant space of beliefs is (homeomorphic to) an open subset of a Euclidean space, and that there is a positive probability that optimal behavior will induce a sequence of beliefs that converges to a critical belief at which the optimal action is uninformative. McLennan (1984) and Harrison et al. (2012) consider the case in which the space of beliefs is one dimensional, because the unknown parameter has two possible values. It is easy to construct examples in which there is an action that is uninformative, in the sense that the distribution of outcomes does not depend on the unknown parameter when it is chosen, and this action is chosen by the myopically optimal policy in response to a certain critical belief. It can also easily happen that the myopically optimal policy does not allow the sequence of beliefs to go from one side of the critical belief to the other, because the amount of experimentation is never sufficient. McLennan (1984) shows that the optimal policy can have this property even when the discount factor is positive. That is, even when the decision maker cares about

in which the rewards of experimentation are enhanced by the presence of other experimenting agents. Several papers (Gale and Kariv (2003), Çelen and Kariv (2004), Gale and Kariv (2008), Acemoglu et al. (2014), Acemoglu et al. (2011), Arieli and Mueller-Frank (2014)) have explored environments with incomplete information, often with information transmission facilitated by networks. Jung (2018) analyzes a model in which concern about being found to be untruthful biases an expert sender toward messages that promote preferred policies that are uninformative. Multiagent learning models have found many applications, e.g., Li (2001) (status quo bias) Malueg and Tsutsui (1997) (R&D) Décamps and Mariotti (2004) (investment) Bergemann and Välimäki (1996) and Harrison et al. (2012) (strategic pricing) Callander (2011a) (product design) Callander (2011b) (policy experimentation) Strulovici (2010) (learning in repeated elections) Laslier et al. (2003) and Berentsen et al. (2008) (regulation and taxation) Piketty (1995) (social mobility and redistribution) Breen and Garcia-Penalosa (2002) (labor economics) Ali (2011) (behavioral economics) and Baker and Mezzetti (2012) (law and economics). See Smith and Sørensen (2011) for a survey and summary.

the future, it can be optimal to behave in a way that sometimes results in the true value of the parameter remaining unknown in the limit.

We will see that the framework presented here has the results of McLennan (1984) as a special case, but the underlying mechanism presented in this paper is quite different. Instead of there being a barrier that cannot be crossed, with the current belief on one side and the truth on the other, here the critical belief is an attractor in a probabilistic sense. Our results are not limited to a 1-dimensional space of beliefs: we are able to produce concrete examples with any finite number of parameters. On the other hand, as we explain in Section 3, there is some reason to expect that the existence of an uninformative action will not be generic unless the space of observations in each period is binary ('yes' or 'no,' 'success' or 'failure,' etc.) but of course this case arises in many models. In addition, if an action is played repeatedly (perhaps because it is legally mandated) before it becomes possible to vary it, it may become approximately uninformative.

1.4 Organization of the Remainder

Section 2 states Theorem 1, which is a version of the dynamic programming result that is sufficient to support the analysis of the learning examples. In this result the spaces of states and actions are compact subsets of Euclidean spaces and are the closures of their interiors, and the value and optimal policy function are shown to be Lipschitz (relative to a norm that is sensitive to partial derivatives up to order r) functions of the discount factor in some interval around zero. This result is proved in Appendix A, and the reader may proceed directly to there from the end of Section 2.

Theorem B1 in Appendix B is more general than Theorem 1 in two ways. First, it allows the spaces of states and actions to be subsets of smooth manifolds. This is more "principled" insofar as our mode of analysis certainly depends on differential calculus, but the particular geometry of Euclidean space should be irrelevant. Second, it shows that optimal value and policy functions are jointly Lipschitz (relative to a suitable norm) functions of the reward function and the discount factor. Thus we establish Lipschitz continuity with respect to almost any parameterization of the problem. The analysis depends on results established in Appendix A, which is a prerequisite, but it is possible for the reader to proceed to Appendix B after completing Section 2 and Appendix A.

Section 3 presents the general model of learning as a Markov process and describes the example studied in McLennan (1984), illustrating how learning may be asymptotically incomplete. Section 4 provides conditions (in the statement of Theorem 2) on the policy function under which there is a positive probability that the sequence of beliefs converges to the critical belief, and conditions (in the statement of Theorem 3) under which the probability of convergence to the critical belief is zero. It also states and proves the version of the law of large numbers that is used to prove Theorems 2 and 3.

Sections 5 and 6 develop concrete examples illustrating Theorem 2. The conditions given by Theorems 2 and 3 involve the logarithm function, and are not immediately tractable, so Section 5 derives more easily verified sufficient conditions for the conditions identified in Theorem 2 to hold in parameterized problems. Section 6 lays out concrete examples in which the conditions identified in Theorem 2 are actually satisfied by the myopically optimal policy. These examples allow any finite number of possible values of the unknown parameter, which is to say any dimension of the space of beliefs.

Section 7 discusses ways in which various assumptions might be relaxed, and other possible directions of generalization and extension, thereby concluding the paper.

2 Euclidean Dynamic Programming

After providing technical prerequisites, this section provides a statement of a "basic" version of the dynamic programming result, that has spaces of states and actions that are compact subsets of Euclidean spaces that are the closures of their interiors. This result suffices for the applications to the learning problem studied in the next several sections. The statement of Theorem 1 requires the introduction of several concepts that may serve to give a good conceptual picture of how the argument works. At the same time many technicalities related to manifolds are avoided.

2.1 Notation and Conventions for Probability

For any measurable space S, let $\Delta(S)$ be the set of probability measures on S. For $s \in S$ let δ_s be the Dirac measure of s, i.e., the element of $\Delta(S)$ that assigns all probability to s. If $\sigma \in \Delta(S)$, $E \subset S$ is measurable, and $f : S \to \mathbb{R}$ is an integrable function, then $\mathbb{P}_\sigma(E) = \sigma(E)$ and $\mathbb{E}_\sigma(f) = \int_S f \, d\sigma$. We usually write $\mathbb{P}(E)$ and $\mathbb{E}(f)$ in place of $\mathbb{P}_\sigma(E)$ and $\mathbb{E}_\sigma(f)$ if σ should be clear from context. If S is finite, an element of $\Delta(S)$ is treated notationally as a $[0,1]$-valued function on S, so that $\sigma(s)$ is the probability of s.

Whenever S is a topological space, it has the Borel σ-algebra, and $\Delta(S)$ is endowed with the weak* topology; recall that this is the weakest topology such that $\sigma \mapsto \int_S f \, d\sigma$ is a continuous function from $\Delta(S)$ to $\mathbb{R}$ whenever $f : S \to \mathbb{R}$ is continuous and bounded. If S is finite, elements of $\Delta(S)$ are denoted as functions from S to [0,1], or as vectors indexed by the elements of S, and $\Delta^\circ(S)$ is the set of measures with full support, that is, the set of $\sigma \in \Delta(S)$ such that $\sigma(s) > 0$ for all s.

2.2 Lipschitz Concepts

Most metrics are denoted by d, with the space to be inferred from context. Suppose that X and Y are metric spaces. With certain obvious exceptions (e.g., a Euclidean space viewed as a cartesian product of copies of $\mathbb{R}$) $X \times Y$ is endowed with the metric $d((x,y),(x',y')) = \max\{d(x,x'), d(y,y')\}$. When X is compact the space $C(X,Y)$ of continuous functions from X to Y is endowed with the metric

$$d(f,f') = \max_x d(f(x), f'(x)).$$

Note that $C(X,Y)$ is complete if and only if Y is complete.

Recall that a function $f : X \to Y$ is *Lipschitz with Lipschitz constant* Λ, or simply Λ-*Lipschitz*, if $d(f(x), f(x')) \le \Lambda d(x,x')$ for all $x, x' \in X$. Of course such a function is continuous. We say that f is *strictly* Λ-*Lipschitz* if it is Λ'-Lipschitz for some $\Lambda' < \Lambda$. It is *Lipschitz* if it is Λ-Lipschitz for some Λ, and it is *locally Lipschitz* if each $x \in X$ has a neighborhood U such that $f|_U$ is Lipschitz. It is obvious that a composition of two (locally) Lipschitz functions is (locally) Lipschitz, and that the restriction of a (locally) Lipschitz function to a subset of the domain is (locally) Lipschitz.

Lemma 2.1. *If X is compact, then a function $f : X \to Y$ is Lipschitz if and only if it is locally Lipschitz.*

Proof. Suppose that f is locally Lipschitz. There is a finite cover of X by open sets $U_1, \ldots, U_k$ such that each $f|_{U_i}$ is Λ_i-Lipschitz for some Λ_i. The Lebesgue number lemma implies that there is some $\varepsilon > 0$ such that for all $x, x' \in X$, if $d(x,x') \le \varepsilon$, then there is some i such that $x, x' \in U_i$. Let $M = \max_{x,x' \in X} d(f(x), f(x'))$ be the diameter of $f(X)$. Then f is $\max\{\Lambda_1, \ldots, \Lambda_k, M/\varepsilon\}$-Lipschitz. □

The following variant of the contraction mapping theorem expresses one of the main ideas of the proof of Theorem 1.

Lemma 2.2. *If $\alpha \in (0,1)$, $\Lambda > 0$, X is complete, $c : X \times Y \to X$ is Λ-Lipschitz, and for each $y \in Y$, $c(\cdot, y)$ is α-Lipschitz, then for each y there is a unique fixed point x^*_y of $c(\cdot, y)$, and the function $y \mapsto x^*_y$ is $\frac{\Lambda}{1-\alpha}$-Lipschitz.*

Proof. The existence of a unique fixed point x^*_y for each y is the assertion of the contraction mapping theorem. For $y, y' \in Y$, let $x_0 = x^*_y$, and define $x_1, x_2, \ldots$ inductively by setting $x_i = c(x_{i-1}, y')$. Then

$$d(x_0, x_1) = d(c(x_0, y), c(x_0, y')) \le \Lambda d(y, y'),$$

$d(x_i, x_{i+1}) \le \alpha d(x_{i-1}, x_i)$ for all $i = 1, 2, \ldots$, and $x_i \to x^*_{y'}$, so $d(x^*_y, x^*_{y'}) \le \Lambda d(y,y')/(1 - \alpha)$. □

The following definitions introduce two key concepts. Let X, X', Y, and Y' be metric spaces with X and X' compact.

Definition 2.1. *A set $T \subset C(X, Y)$ is* Lipschitz bounded *if there is a compact $K \subset Y$ and a constant $\Lambda > 0$ such that for every $f \in T$, $f(X) \subset K$ and f is Λ-Lipschitz.*

Definition 2.2. *If $S \subset C(X, Y)$ and $\gamma : S \to C(X', Y')$ is an operator, we say that a set $W \subset S$ is* γ-compliant *if, for every Lipschitz bounded $T \subset W$, $\gamma(T)$ is Lipschitz bounded and $\gamma|_T$ is Lipschitz. We say that γ is* tame *if it is continuous and every $f \in S$ has a γ-compliant neighborhood $U \subset S$.*

As is well known, and described below, dynamic programming considers fixed points of operators. In our analysis these operators are viewed as recombinations of more basic operators. Subsection A1 provides the numerous results we need concerning how various methods of recombining tame operators give rise to new tame operators.

2.3 Basics of Dynamic Programming

Later in this subsection the space of *states* Ω and the space of *actions* A will be subsets of Euclidean spaces, but for the first part of the discussion they can be general compact metric spaces. Let $Q : \Omega \times A \to \Delta(\Omega)$ be a measurable *transition function.* A (stationary, deterministic) *policy* is a measurable function $\pi : \Omega \to A$. For a given *initial state* $\omega_0 \in \Omega$, a policy π and the transition function Q induce a probability measure on the space of infinite histories $(\omega_0, a_0), (\omega_1, a_1), \ldots$ in which the distribution of ω_t conditional on $(\omega_0, a_0), \ldots, (\omega_{t-1}, a_{t-1})$ is $Q(\omega_{t-1}, a_{t-1})$ and $a_t = \pi(\omega_t)$ almost surely. Fix a continuous *reward function* $u_0 : \Omega \times A \to \mathbb{R}$. For a *discount factor* $\delta \in (-1, 1)$ the expectation of $\sum_{t=0}^{\infty} \delta^t u_0(\tilde{\omega}_t, \tilde{a}_t)$ relative to the induced distribution on histories is well defined and finite.

We now introduce the key operators:

(a) For $u \in C(\Omega \times A)$ let $J(u) : \Omega \to \mathbb{R}$ be the function

$$J(u)(\omega) = \max_{a \in A} u(\omega, a).$$

Since A is compact, $J(u) \in C(\Omega)$, and it is straightforward to show that $J : C(\Omega \times A) \to C(\Omega)$ is 1-Lipschitz.

(b) For $V \in C(\Omega)$ let $K_Q(V) : \Omega \times A \to \mathbb{R}$ be the function

$$K_Q(V)(\omega, a) = \int_{\Omega} V(\omega')\, Q(\omega, a; d\omega').$$

If Q is continuous, then $K_Q(V) \in C(\Omega \times A)$, and the operator $K_Q : C(\Omega) \to C(\Omega \times A)$ is evidently 1-Lipschitz.

(c) For $\delta \in (-1, 1)$ and $u \in C(\Omega \times A)$ let $I_\delta(u) = u_0 + \delta K_Q(J(u))$.

Ma and Stachurski (2018) call I_δ the *refactored Bellman operator*, and they systematically study this and similar reformulations of the toolkit of dynamic programming. Since J and K_Q are 1-Lipschitz, I_δ is $|\delta|$-Lipschitz, so the contraction mapping theorem implies that it has a unique fixed point u_δ, which Ma and Stachurski call the *refactored value function*. An *optimal policy* is a function $\pi_\delta : \Omega \to A$ such that $u_\delta(\omega, \pi_\delta(\omega)) = \max_a u_\delta(\omega, a)$ for all ω.

The *value function* is $V_\delta = J(u_\delta)$. The *Bellman operator* for δ is $L_\delta : C(\Omega) \to C(\Omega)$ given by

$$L_\delta(V) = J(u_0 + \delta \cdot K_Q(V)).$$

Evidently L_δ is $|\delta|$-Lipschitz, so it has a unique fixed point. If u_δ is a fixed point of I_δ and $V_\delta = J(u_\delta)$, then V_δ is a fixed point of L_δ:

$$V_\delta = J(u_\delta) = J(I_\delta(u_\delta)) = J(u_0 + \delta K_Q(J(u_\delta))) = J(u_0 + \delta K_Q(V_\delta)) = L_\delta(V_\delta).$$

Conversely, if V_δ is a fixed point of L_δ and $u_\delta = u_0 + \delta K_Q(V_\delta)$, then u_δ is a fixed point of I_δ:

$$\begin{aligned} u_\delta = u_0 + \delta K_Q(V_\delta) = u_0 + \delta K_Q(L_\delta(V_\delta)) &= u_0 + \delta K_Q(J(u_0 + \delta K_Q(V_\delta))) \\ &= u_0 + \delta K_Q(J(u_\delta)) = I_\delta(u_\delta). \end{aligned}$$

2.4 Multivariate Calculus

We are accustomed to using partial derivatives to describe differential phenomena for functions between Euclidean spaces, but we also understand that the derivative of a C^1 function at a point is a linear function. It seems that, in principle, the second derivative should be a linear function mapping into a space of linear functions, but it is also possible to regard it as a bilinear function that (thanks to the equality of cross partials) is symmetric. Similarly, the third derivative can be regarded as a symmetric trilinear function, and so forth. Developing this perspective formally will allow a principled and canonical treatment of higher order derivatives and Lipschitz conditions imposed on them.

Let V and W be finite dimensional vector spaces. For $i = 1, 2, \ldots$ a function $\lambda : V^i \to W$ is *multilinear* if

$$\lambda(v_1, \ldots, v_{j-1}, \cdot, v_{j+1}, \ldots, v_i) : V \to W$$

is linear for all $j = 1, \dots, i$ and all $v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_i \in V$. Let $\mathrm{Mult}^i(V, W)$ be the space of such functions. Let $\mathrm{Mult}^0(V, W) = W$.

Let Σ_i be the set of permutations of the indices $1, \dots, i$. We say that $\lambda \in \mathrm{Mult}^i(V, W)$ is *symmetric* if $\lambda(v_{\sigma(1)}, \dots, v_{\sigma(i)}) = \lambda(v_1, \dots, v_i)$ for all $\sigma \in \Sigma_i$ and $v_1, \dots, v_i$. Let $\mathrm{Sym}^i(V, W)$ be the space of symmetric elements of $\mathrm{Mult}^i(V, W)$. Fix a finite degree of continuous differentiability $r \geq 1$. There is the direct sum

$$J^r(V, W) = \bigoplus_{i=0}^{r} \mathrm{Sym}^i(V, W).$$

When $W = \mathbb{R}$ we omit this argument from our notation, writing $\mathrm{Mult}^i(V)$, $\mathrm{Sym}^i(V)$, and $J^r(V)$ in place of $\mathrm{Mult}^i(V, \mathbb{R})$, $\mathrm{Sym}^i(V, \mathbb{R})$, and $J^r(V, \mathbb{R})$.

Suppose that $U \subset V$ is open, and that $f : U \to W$ is a function. We say that f is C^r if there are continuous functions $D^i : U \to \mathrm{Sym}^i(V, W)$ for $i = 0, \dots, r$ such that

$$\lim_{h \to 0} \frac{|f(a + h) - \sum_{i=0}^{r} D^i f(a)(h, \dots, h)|}{|h|^r} = 0$$

for all $a \in U$. If f is C^r we e define $\Delta^r f \in C(U, J^r(V, W))$ by setting[3]

$$\Delta^r f(a) = (D^0 f(a), \dots, D^r f(a)).$$

No Revisions Below

Remove the below if it is not needed.

Let $\mathrm{Hom}(V, W)$ is the set of linear transformations from V to W. We define $\mathrm{Hom}^i(V, W)$ inductively by setting $\mathrm{Hom}^0(V, W) = W$ and $\mathrm{Hom}^i(V, W) = \mathrm{Hom}(V, \mathrm{Hom}^{i-1}(V, W))$. Define $\xi^i : \mathrm{Hom}^i(V, W) \to \mathrm{Mult}^i(V, W)$ inductively by letting ξ_0 be the identity function and setting

$$\xi^i(\kappa)(v_1, \dots, v_i) = \xi^{i-1}(\kappa(v_1, \cdot))(v_2, \dots, v_i)$$

for $i \geq 1$. It is easy to show that ξ^i is a linear isomorphism.

Suppose that $U \subset V$ is open, and that $f : U \to W$ is a function. The derivative $Df(a) \in \mathrm{Hom}(V, W)$ at a point $a \in U$ is defined as usual. We say that f is C^0 if it is continuous. We define C^i functions and order i derivatives $\tilde{D}^i f : U \to \mathrm{Hom}^i(V, W)$ and $D^i f : U \to \mathrm{Mult}^i(V, W)$ inductively as follows. Let $\tilde{D}^0 f = f$. For $i \geq 1$ we say that f is C^i if it is C^{i-1}, $D(\tilde{D}^{i-1} f)(x)$ is defined at every $x \in U$, and $D(\tilde{D}^{i-1} f)$ is continuous. If this is the case we set $\tilde{D}^i f = D(\tilde{D}^{i-1} f)$ and $D^i f = \xi^i \circ \tilde{D}^i f$. Let $C^i(U, W)$ be the space of C^i functions from U to W. A consequence of the equality of cross partials is

[3]It will always be clear from context whether Δ refers to differentiation or to probability.

that the image of $D^i f$ is contained in $\mathrm{Sym}^i(V,W)$. Fix a finite degree of continuous differentiability $r \geq 1$, and suppose that $f \in C^r$. We define $\Delta^r f \in C(U, J^r(V,W))$ by setting[4]

$$\Delta^r f(a) = (D^0 f(a), \ldots, D^r f(a)).$$

Suppose that $\mathbf{e}_1, \ldots, \mathbf{e}_m$ and $\mathbf{f}_1, \ldots, \mathbf{f}_n$ are bases of V and W. We can write $f(x) = \sum_j f_j(a)\mathbf{f}_j$. If $v_1, \ldots, v_m$ are the linear coordinates of $v \in V$ with respect to the basis $\mathbf{e}_1, \ldots, \mathbf{e}_m$ (that is, $v = v_1\mathbf{e}_1 + \cdots + v_m\mathbf{e}_m$) induction on i gives the formula

$$D^i f_j(x)(v^1, \ldots, v^i) = \sum_{1 \leq j_1, \ldots, j_i \leq m} \frac{\partial^i f_j}{\partial a_{j_1} \cdots \partial a_{j_i}}(x) v^1_{j_1} \cdots v^i_{j_i}.$$

We now have

$$D^i f(x)(v^1, \ldots, v^i) = \sum_j D^i f_j(x)(v^1, \ldots, v^i)\mathbf{f}_j,$$

so $D^i f(x) \in \mathrm{Mult}^i(V,W)$. In addition the symmetry of cross partials gives the proof that $D^i f(x) \in \mathrm{Sym}^i(V,W)$. Thus we have expressed $D^i f(x) \in \mathrm{Sym}^i(V,W)$ as a linear function of the order i partials of f. Conversely, the equation

$$\frac{\partial^i f_j}{\partial a_{j_1} \cdots \partial a_{j_i}}(x) = D^i f_j(x)(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_i})$$

expresses the order i partials of f as linear functions of $D^i f(a)$.

No Revisions Below

Now suppose that V and W are inner product spaces, and that $\mathbf{e}_1, \ldots, \mathbf{e}_m$ and $\mathbf{f}_1, \ldots, \mathbf{f}_n$ are orthonormal bases. We define derived inner products for some of the spaces defined above. For each $h_1, \ldots, h_i$ with $1 \leq h_j \leq m$ for all j and each $k = 1, \ldots, n$ there is a $\mathbf{e}_{h_1} \otimes \cdots \otimes \mathbf{e}_{h_i} \otimes \mathbf{f}_k \in \mathrm{Mult}^i(V,W)$ given by

$$\mathbf{e}_{h_1} \otimes \cdots \otimes \mathbf{e}_{h_i} \otimes \mathbf{f}_k(\mathbf{e}_{g_1}, \ldots, \mathbf{e}_{g_i}) = \begin{cases} \mathbf{f}_k, & g_1 = h_1, \ldots, g_i = h_i, \\ 0, & \text{otherwise.} \end{cases}$$

Clearly the $\mathbf{e}_{h_1} \otimes \cdots \otimes \mathbf{e}_{h_i} \otimes \mathbf{f}_k$ are a basis of $\mathrm{Mult}^i(V,W)$. We define an inner product $\langle \cdot, \cdot \rangle_i$ on this space by setting

$$\langle \mathbf{e}_{1h_1} \otimes \cdots \otimes \mathbf{e}_{ih_i} \otimes \mathbf{f}_k, \mathbf{e}_{1h'_1} \otimes \cdots \otimes \mathbf{e}_{ih'_i} \otimes \mathbf{f}_{k'} \rangle_i = \begin{cases} 1, & h_1 = h'_1, \ldots, h_i = h'_i \text{ and } k = k', \\ 0, & \text{otherwise.} \end{cases}$$

Note that if $i = 0$, then $\langle \cdot, \cdot \rangle_0$ is just the inner product of W.

[4]It will always be clear from context whether Δ refers to differentiation or to probability.

Let $O(V)$ and $O(W)$ be the groups of orthogonal transformations (inner product preserving linear automorphisms) of V and W respectively. There is an action[5] of the group $O(V)^i \times O(W)$ on $\text{Mult}^i(V, W)$ given by

$$((o_1, \ldots, o_i, r)\lambda)(v_1, \ldots, v_i) = r(\lambda(o_1(v_1), \ldots, o_i(v_i)))$$

for $o_1, \ldots, o_i \in O(V)$ and $r \in O(W)$. The inner product $\langle \cdot, \cdot \rangle_i$ is invariant with respect to the action $O(V)$ on any one of the copies of V because the inner product $\langle \sum_g \alpha_g \mathbf{e}_{jg}, \sum_g \beta_g \mathbf{e}_{jg} \rangle = \sum_g \alpha_g \beta_g$ is invariant with respect to the action of $O(V)$, and similarly for the action of $O(W)$, so it is invariant with respect to the action of $O(V)^i \times O(W)$. Thus it is invariant under the action of $O(V) \times O(W)$ on $\text{Mult}^i(V, W)$ given by

$$((o, r)\lambda)(v_1, \ldots, v_i) = r(\lambda(o(v_1), \ldots, o(v_i)))$$

because this is the restriction of the action above to a subgroup. For us the important point is that the inner product $\langle \cdot, \cdot \rangle_i$ does not depend on our choices of orthonormal bases. We endow $J^r(V, W)$ with the inner product

$$\langle (\mu_0, \ldots, \mu_r), (\mu'_0, \ldots, \mu'_r) \rangle = \sum_{i=0}^{r} \langle \mu_i, \mu'_i \rangle_i,$$

which again does not depend on the choices of orthonormal bases.

Remark 2.1. *As a matter of elementary calculus, Δ^r is a linear operator: if $f, f' \in C^r(U, W)$ and $\alpha \in \mathbb{R}$, then $\Delta^r(f + f') = \Delta^r f + \Delta^r f'$ and $\Delta^r(\alpha f) = \alpha \Delta^r f$. Thus $\Delta^r C^r(U, W)$ is a linear subspace of $C(U, J^r(V, W))$.*

A difficulty is that while how the operator Δ^r operates on sums and scalar multiples is simple, this is not the case for compositions. Let X be a third vector space, let $U' \subset W$ be open, and let $f : U \to U'$ and $g : U' \to X$ be C^r. For any linear coordinate systems for V, W, and X, the standard rules from elementary calculus imply that the partial derivatives of $g \circ f$ up to order r are polynomial functions of the various partial derivatives of f and g up to order r. Above we saw that there is a linear isomorphism between $J^r(V, W)$ and the Euclidean space that contains all the partials of f up to order r, and the same is true of $J^r(V, W)$ and $J^r(V, W)$. Without going into additional detail (formally describing the Euclidean space containing the partials of f would be a thankless chore) by taking the relevant compositions of the various maps described here we obtain the following result.

[5]An *action* of a group G on a set X is a function $(g, x) \mapsto gx$ from $G \times X$ to X such that $ex = x$ for all x (where e is the identity of G) and $(gh)x = g(hx)$ for all $g, h \in G$ and $x \in X$. A function $f : X \to Y$ is *invariant* under the action if $f(gx) = f(x)$ for all $g \in G$ and $x \in X$.

Proposition 2.1. *There is a C^∞ function*

$$\tau^r_{V,W,X} : J^r(V,W) \times J^r(W,X) \to J^r(V,X)$$

(which we denote by τ^r when ambiguity is impossible) such that if $U \subset V$ and $U' \subset W$ are open, $f : U \to U'$ and $g : U' \to X$ are C^r, and $x \in U$, then

$$\Delta^r(g \circ f)(x) = \tau^r(\Delta^r f(x), \Delta^r g(f(x))).$$

Recall that for an arbitrary $S \subset V$, the standard definition of what it means for a function $f : S \to W$ to be C^r is that there is an open U containing S and a C^r function $\tilde{f} : U \to W$ such that $\tilde{f}|_S = f$. Let $C \subset V$ be a compact set that is the closure of its interior, and let $f : C \to W$ be a function. We will say that f is weakly[6] C^r if it is continuous, the restriction f° of f to the interior of C is C^r in the usual sense, and $\Delta^r f^\circ$ has a continuous extension to all of C, in which case (abusing notation slightly) we let $\Delta^r f$ denote this extension. For $T \subset W$ let $C^r_{\mathrm{w}}(C,T)$ be the set of weakly C^r functions $f : C \to T$. As before, we write $C^r_{\mathrm{w}}(C)$ in place of $C^r_{\mathrm{w}}(C,\mathbb{R})$.

We endow $J^r(V,W)$ with the norm derived from its inner product, and we endow $C^r_{\mathrm{w}}(V,W)$ with the sup norm:

$$\|f\| = \max_{x \in C} \|\Delta^r f(x)\|.$$

The next result implies that the metric derived from this norm is complete, so that $\Delta^r C^r_{\mathrm{w}}(V,W)$ is a Banach space.

Proposition 2.2. *If $\{f_i\}$ is a sequence in $C^r_{\mathrm{w}}(C,W)$, $g \in C(C, J^r(V,W))$, $\Delta^r f_i$ converges uniformly to g, and f is the order 0 component of g, then $f \in C^r_{\mathrm{w}}(C,W)$ and $\Delta^r f = g$.*

Proof. Since f and g are limits of uniformly convergent sequences of continuous functions, they are continuous. The proof of Theorem 4.3 (p. 61) of Hirsch (1976) demonstrates that the restriction of f to the interior of C is C^r (so $f \in C^r_{\mathrm{w}}(C,W)$) and that $\Delta^r(f)$ agrees with g on this domain, so (by continuity) $\Delta^r f$ agrees with g on all of C. □

[6]For an example of a function that is weakly C^∞ but not C^1 let $C = [-1,0] \cup \bigcup_{n=1}^\infty [\frac{1}{2n}, \frac{1}{2n-1}]$, and define $f : C \to \mathbb{R}$ by setting $f(t) = 0$ if $t \le 0$ and $f(t) = \frac{1}{2n}$ if $t \in [\frac{1}{2n}, \frac{1}{2n-1}]$. The Whitney (1934) extension theorem gives sufficient conditions for a function on a closed subset of V to have a C^r extension to all of V, and the sharpened form of the Whitney extension theorem given by Fefferman (2005) (applied with his set E equal to the interior of C) implies that a weakly C^r function on C with Lipschitz r^{th} order partial derivatives has an extension $\tilde{f} : V \to W$ whose r^{th} order partials are also Lipschitz.

Let V and W be finite dimensional inner product spaces, and let $U \subset V$ be open. A C^r function $f : U \to W$ is $C^{r,1}$ if the r^{th} order partials of f are locally Lipschitz. (As we explained earlier, for any choice of coordinate systems $D^r f$ is a related to these partials by a linear isomorphism, so a C^r function f is $C^{r,1}$ if and only if $\Delta^r f$ is locally Lipschitz.) For an arbitrary $S \subset V$, a function $f : S \to W$ is $C^{r,1}$ if there is an open U containing S and a $C^{r,1}$ function $\tilde{f} : U \to W$ such that $\tilde{f}|_S = f$. For $T \subset W$ let $C^{r,1}(S, W)$ be the space of $C^{r,1}$ functions $f : S \to T$. If $C \subset V$ is compact, let $C^{r,1}_{\text{w}}(C, T)$ be the set of $f \in C^r_{\text{w}}(C, T)$ such that $\Delta^r f$ is Lipschitz. Insofar as the limit of a uniformly convergent sequence of Lipschitz functions need not be Lipschitz, $C^{r,1}_{\text{w}}(C, W)$ is not a closed subspace of $C^r_{\text{w}}(C, W)$. However, for any Λ, the uniform limit of a sequence of Λ-Lipschitz functions is Λ-Lipschitz, so Proposition 2.2 implies that the set of $f \in C^r_{\text{w}}(C, W)$ such that $\Delta^r f$ is Λ-Lipschitz is a closed subset of $C^r_{\text{w}}(C, W)$.

We now establish two required technical results.

Lemma 2.3. *If $T \subset V$ and $U \subset W$ are open and $f \in C^{r,1}(T, U)$ and $g \in C^{r,1}(U, X)$, then $g \circ f \in C^{r,1}(T, X)$.*

Proof. Of course basic facts concerning differentiation imply that $g \circ f \in C^r(T, X)$. Since (Proposition 2.1) $\Delta^r(g \circ f)(x) = \tau^r \circ (\Delta^r f(x), \Delta^r g(f(x)))$, the fact that compositions of locally Lipschitz functions are locally Lipschitz implies that $D^r(g \circ f)$ is locally Lipschitz. □

Lemma 2.4 ($C^{r,1}$ Inverse Function Theorem)**.** *If $U \subset \mathbb{R}^m$ is open, $f : U \to \mathbb{R}^m$ is $C^{r,1}$, $x \in U$ and $Df(x)$ is nonsingular, then there is an open $V \subset U$ containing x such that $f|_V$ is invertible and $(f|_V)^{-1}$ is $C^{r,1}$.*

Proof. The inverse function theorem for C^r functions gives an open $V \subset U$ containing x such that $f|_V$ is invertible and $(f|_V)^{-1}$ is C^r. Since we can replace V with a smaller neighborhood of x, we may assume that $Df(x')$ is nonsingular for all $x' \in V$. Let $g = (f|_V)^{-1}$. Differentiating the equation $g \circ f = \mathbf{1}_V$ gives $Dg(y) = Df(g(y))^{-1}$ for all $y \in f(V)$. Using Cramer's rule, each $\frac{\partial g_i}{\partial y_j}(y)$ can be expressed as a rational function[7] of the partials of f at $g(y)$. By repeatedly differentiating this formula (at each stage substituting the previously computed values of the partials of g) each r^{th} order partial of g at y can be expressed as a rational function of the partials of f up to order r at $g(y)$. Since a rational function is locally Lipschitz on its domain of definition (that is, where the denominator does not vanish) and compositions of locally Lipschitz functions are locally Lipschitz, it follows that the r^{th} order partials of g are locally Lipschitz. □

[7] A *rational function* is a quotient of polynomial functions.

2.5 Smooth Dynamic Programs

We now assume that $\Omega \subset V$ and $A \subset W$ are compact subsets of finite dimensional inner product spaces V and W that are the closures of their interiors. Our results require additional assumptions on Q and u_0.

For a sufficiently differentiable $u : \Omega \times A \to \mathbb{R}$ we write $\partial_a u(\omega, a)$ and $\partial_\omega u(\omega, a)$ in place of $D(u(\omega, \cdot))(a)$ and $D(u(\cdot, a))(\omega)$, $\partial_{aa} u(\omega, a)$ in place of $D^2(u(\omega, \cdot))(a)$, and so forth. We say that $u_0 \in C^2_{\mathrm{w}}(\Omega \times A)$ *satisfies the standard conditions* if, for each $\omega \in \Omega$, there is a unique maximizer $\pi_0(\omega)$ of $u_0(\omega, \cdot)$, $\pi_0(\omega)$ is in the interior of A, and $\partial_{aa} u_0(\omega, \pi_0(\omega))$ is negative definite.

Lemma 2.5. *For $r \geq 2$, if $u_0 \in C^r_{\mathrm{w}}(\Omega \times A)$ satisfies the standard conditions and, for each $\omega \in \Omega$, $\pi_0(\omega)$ is the unique maximizer of $u_0(\omega, \cdot)$, then $\pi_0 \in C^{r-1}_{\mathrm{w}}(\Omega, A)$.*

Proof. If ω is in the interior of Ω, the inverse function theorem implies that π_0 is C^{r-1} in a neighborhood of ω and

$$D\pi_0(\omega) = -\partial_{aa} u_0(\omega, \pi_0(\omega))^{-1} \circ \partial_{\omega a} u_0(\omega, \pi_0(\omega)).$$

The right hand side is C^{r-2}, and repeated differentiation of it gives formulas for $\Delta^{r-1}\pi_0 = (\pi_0, \Delta^{r-2} D\pi_0)$ that extend continuously to all of Ω. □

If $K_Q(C^r_{\mathrm{w}}(\Omega)) \subset C^r_{\mathrm{w}}(\Omega \times A)$ let $\mathcal{K}_Q : \Delta^r(C^r_{\mathrm{w}}(\Omega)) \to \Delta^r(C^r_{\mathrm{w}}(\Omega \times A))$ be the operator given by

$$\mathcal{K}_Q(\Delta^r V) = \Delta^r(K_Q(V)). \tag{2.1}$$

The transition function Q is said to be r^{th} *order smoothing* if $K_Q(C^r_{\mathrm{w}}(\Omega)) \subset C^r_{\mathrm{w}}(\Omega \times A)$ and $\mathcal{K}_Q$ is tame. This condition might hold for a variety of reasons, and must generally be verified in the context of each application, so we do not give conditions that imply it. As we will see later, it holds in our application of Theorem 1 because in that example $Q(\omega, a)$ is a distribution with finite support, and the location and probability of each support point is a C^∞ function.

Theorem 1. *If $r \geq 2$, $u_0 \in C^{r,1}(\Omega \times A)$ satisfies the standard conditions, and Q is r^{th} order smoothing, then there is an $\varepsilon > 0$ such that:*

(a) for each $\delta \in (-\varepsilon, \varepsilon)$ the discounted dynamic program with payoff u_0 and discount factor δ has value and refactored value functions $V_\delta \in C^{r,1}_{\mathrm{w}}(\Omega)$ and $u_\delta \in C^{r,1}_{\mathrm{w}}(\Omega \times A)$ and a unique stationary optimal policy $\pi_\delta \in C^{r,1}_{\mathrm{w}}(\Omega, A)$,

(b) u_δ satisfies the standard conditions, and

(c) $\Delta^r V_\delta$, $\Delta^r u_\delta$, and $\Delta^r \pi_\delta$ are Lipschitz functions of δ.

The proof of Theorem 1 is presented in Appendix A, and the interested reader may proceed there immediately. At this point we can describe the nature of the argument in very broad strokes. We would like to find a set $N \subset C^r_{\mathrm{w}}(\Omega \times A)$ containing u_0 and $\varepsilon > 0$ such that Lemma 2.2 can be applied to the map $(u, \delta) \mapsto I_\delta(u)$ from $N \times (-\varepsilon, \varepsilon)$ to $C^r_{\mathrm{w}}(\Omega \times A)$. This requires that the operator $K_Q \circ J|_N$ is Lipschitz, relative to some metric, and this condition must be derived from the properties of $J|_N$ and $K_Q|_{J(N)}$. Thus we want these operators to map weakly C^r functions to weakly C^r functions in a Lipschitzian manner. Due to the nature of the computations that arise when we compute the r^{th} derivative of a composition of C^r functions, this requires that we restrict attention to an N such that there is a Λ such that for all $u \in N$, $\Delta^r u$ is Λ-Lipschitz.

3 The Learning Model

We now describe the model of repeated experimentation and belief revision. Let Θ be a finite set of possible values of a *parameter* $\tilde{\theta}$ that is chosen by nature at the outset and does not change after that. The set of possible *beliefs* concerning $\tilde{\theta}$ is $\Omega = \Delta(\Theta)$.

There is a space A of *actions* that is a compact subset of a Euclidean space that is the closure of its interior. In each period $t = 0, 1, 2, \ldots$ the decision maker chooses an action $\tilde{a}_t$ from a set A, and observes an *outcome* $\tilde{y}_t$ that is an element of a finite set Y. For each $\theta \in \Theta$ there is a given function

$$q_\theta : A \to \Delta^\circ(Y)$$

specifying the probability distribution $q_\theta(a)$ over outcomes when θ is the parameter and action a is chosen. Let $q_\theta(y|a)$ denote the probability that y is observed when θ is the parameter and a is chosen.

When $\omega \in \Omega$ is the prior belief, action a is chosen, and outcome y is observed, the Bayesian posterior belief is $\beta(\omega, a, y) \in \Omega$ with components given by Bayes rule:

$$\beta_\theta(\omega, a, y) = \omega_\theta q_\theta(y|a)/q(y|\omega, a). \tag{3.1}$$

For $\omega \in \Omega$ let

$$q(y|\omega, a) = \sum_\theta \omega_\theta q_\theta(y|a)$$

be the probability of observing y when ω is the belief and a is chosen, so $q(\omega, a) \in \Delta(Y)$. The distribution of the posterior when the prior is ω and a is chosen is

$$Q(\omega, a) = \sum_y q(y|\omega, a)\delta_{\beta(\omega,a,y)} \in \Delta(\Omega).$$

A general property of Bayesian updating is that the sequence of beliefs constitutes a martingale, so the expectation of the posterior is the prior. Concretely, for all ω, a, and θ we have

$$\sum_y q(y|\omega,a)\beta_\theta(\omega,a,y) = \sum_y q(y|\omega,a)\big(\omega_\theta q_\theta(y|a)/q_\omega(y|\omega,a)\big) = \omega_\theta \sum_y q_\theta(y|a) = \omega_\theta.$$

The following will be proved in Appendix A, once the required technical results are available.

Proposition 3.1. *If $r \geq 1$ and, for all y and θ, $q_\theta(y|\cdot)$ is $C^{r,1}$, then Q is r^{th} order smoothing.*

We now assume that there is a *policy function* $\pi : \Omega \to A$ that is stationary in the sense that it governs the choice of action in every period. (Eventually we will be interested in policy functions that are optimal for a dynamic program, but for the time being this is irrelevant.) The policy function and an *initial state* or *prior belief* $\tilde{\omega}_0 \in \Omega$ determine stochastic processes $\{\tilde{\omega}_t\}$, $\{\tilde{a}_t\}$, and $\{\tilde{y}_t\}$ that are defined by requiring that, for all $t \geq 0$: (i) $\tilde{a}_t = \pi(\tilde{\omega}_t)$, (ii) conditional on $\tilde{\omega}_t$ and $\tilde{a}_t$ (and independent of $\tilde{\omega}_0, \ldots, \tilde{\omega}_{t-1}$, $\tilde{a}_0, \ldots, \tilde{a}_{t-1}$, and $\tilde{y}_0, \ldots, \tilde{y}_{t-1}$) the distribution of $\tilde{y}_t$ is $q(\tilde{\omega}_t, \tilde{a}_t)$, and (iii) $\tilde{\omega}_{t+1} = \beta(\tilde{\omega}_t, \tilde{a}_t, \tilde{y}_t)$.

An action a^* is *uninformative* if $q_\theta(y|a^*) = q_{\theta'}(y|a^*)$ for all $y \in Y$ and $\theta, \theta' \in \Theta$. Let A^* be the set of uninformative actions. If a^* is uninformative, $q(y|a^*)$ denotes the common value of $q_\theta(y|a^*)$, and $\beta(\omega, a^*, y) = \omega$ for all ω and y. A belief $\omega \in \Omega$ is *critical* (for π) if $\pi(\omega)$ is uninformative, so the set of critical beliefs is $\pi^{-1}(A^*)$. In particular, if $\tilde{\omega}_t$ is critical, then it is almost surely the case that, for all $s \geq t$, $\pi(\tilde{\omega}_s) = \pi(\tilde{\omega}_t)$ and $\tilde{\omega}_{s+1} = \tilde{\omega}_s = \tilde{\omega}_t$. Repeatedly choosing an uninformative action is a natural and "generic" possibility in many settings, especially if A is finite, but it is not our main focus.

Lemma 3.1. *If π and each $q_\theta(y|\cdot) : A \to (0,1)$ are continuous, then almost surely $\{\tilde{\omega}_t\}$ converges to a point in $\pi^{-1}(A^*) \cup \{\, \delta_\theta : \theta \in \Theta \,\}$.*

Proof. The continuity of $q_\theta(y|\cdot)$ implies that A^* is closed, after which the continuity of π implies that $\pi^{-1}(A^*)$ is closed. The martingale convergence theorem implies that $\{\tilde{\omega}_t\}$ converges with probability one. Consider a point $\omega \in \Omega \setminus (\pi^{-1}(A^*) \cup \{\, \delta_\theta : \theta \in \Theta \,\})$. Since $\pi(\omega)$ is not uninformative it gives rise to a positive probability of an outcome that results in a posterior that is different from ω, and continuity implies that there is an $\varepsilon > 0$ and a neighborhood U of ω such that $\mathbb{P}(\tilde{\omega}_{t+1} \in U | \tilde{\omega}_t = \omega') < 1 - \varepsilon$ for all $\omega' \in U$. Since A^* is closed, countably many such U cover $\Omega \setminus (\pi^{-1}(A^*) \cup \{\, \delta_\theta : \theta \in \Theta \,\})$, so if the probability that the limit of $\{\tilde{\omega}_t\}$ is not in $\pi^{-1}(A^*) \cup \{\, \delta_\theta : \theta \in \Theta \,\}$ was positive, for some such U the probability that $\{\tilde{\omega}_t\}$ entered U and never subsequently exited would be positive, which is not the case. ∎

We say that *learning is asymptotically complete* if $\sum_{\theta \in \Theta} \mathbb{P}(\tilde{\omega}_t \to \delta_\theta | \tilde{\omega}_0 = \omega_0) = 1$ for all ω_0, and otherwise *learning is possibly asymptotically incomplete.* Lemma 3.1 implies that if $\tilde{\omega}_t \neq \pi^{-1}(A^*)$ for all t, then the only possibility for asymptotically incomplete learning is a positive probability that $\{\tilde{\omega}_t\}$ converges to some critical belief outside of $\{\, \delta_\theta : \theta \in \Theta \,\}$. We say that $\omega^* \in \pi^{-1}(A^*) \setminus \{\, \delta_\theta : \theta \in \Theta \,\}$ is a *learning trap* if there is an initial belief $\omega_0 \neq \omega^*$ such that there is positive probability (conditional on $\tilde{\omega}_0 = \omega_0$) that $\{\tilde{\omega}_t\}$ converges to ω^*.

Remark: Uninformativeness may be thought of as a system of $(|\Theta| - 1)(|Y| - 1)$ equations. When the functions $q_\theta(y|\cdot)$ are C^1 and determined by independent random processes, we would typically expect A^* to be a submanifold of A of codimension $(|\Theta| - 1)(|Y| - 1)$. We would also typically expect the image of π to have an empty intersection with this submanifold unless

$$\dim A \leq \dim A^* + \dim \Omega = \dim A - (|\Theta| - 1)(|Y| - 1) + |\Theta| - 1.$$

In this sense the phenomenon studied here is "generic" if $|Y| = 2$ and not otherwise, and when $|Y| = 2$ the intersection of $\pi(\Omega)$ with A^* will typically consist of isolated points.

There are economic settings in which the equations governing uninformativeness may not be independent of each other. For example, suppose it becomes possible to vary a after many periods during which a^* was a status quo action, perhaps because a^* was legally mandated. At the time experimentation becomes possible the only parameters that have not been ruled out by experience are those that give the historically observed distribution of observations when a^* is chosen.

To illustrate these concepts we briefly review the instance of this phenomenon that was studied in McLennan (1984). In each period a customer enters a store, is quoted a price, and either purchases one unit or leaves without buying anything. There are two possible linear relationships between the price a and the probability of purchase, so $\Theta = \{\theta_1, \theta_2\}$ has two elements, and for each θ_i there is a price a_i that maximizes expected revenue, as shown in Figure 1. (Asymptotically incomplete learning for myopically optimal behavior when the demand curves are nonlinear is studied in great detail in Harrison et al. (2012).) There is a price $a^* \in (a_1, a_2)$ that is uninformative because the two inverse demand curves give the same probability of purchase.

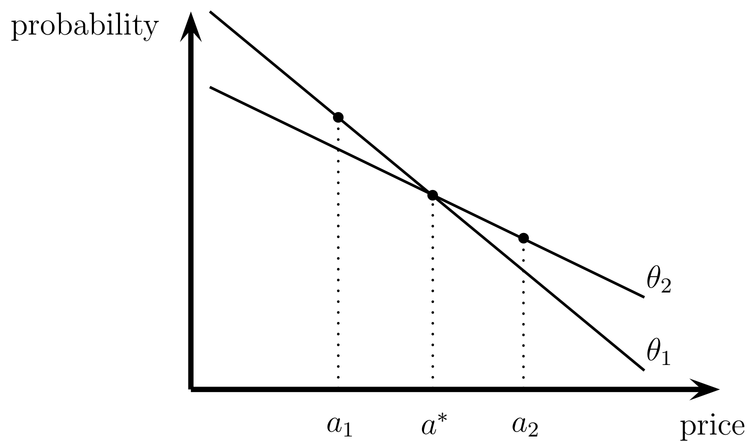

Figure 1

Suppose that π_0 is the myopically optimal policy. Elementary calculations show that π_0 is affine with $\pi_0(\delta_{\theta_1}) = a_1$ and $\pi_0(\delta_{\theta_2}) = a_2$, so there is a ω^* such that $\pi_0(\omega^*) = a^*$. It is not hard to find parameters such that for all ω between δ_{θ_1} and ω^*, for either observation after choosing $\pi_0(\omega)$, the posterior is also in this interval. (Harrison et al. (2012) stress that in somewhat more general settings this is also a frequent occurrence.) If the sequence of beliefs is trapped in the interval between δ_{θ_1} and ω^*, then the sequence of beliefs will almost surely converge to either δ_{θ_1} or ω^*. Using the fact that $\{\tilde{\omega}_t\}$ is a martingale, one can easily compute the probabilities of these limits. Note that there is a positive probability of converging to ω^* even when θ_1 is the true parameter. (Otherwise observing convergence to ω^* would lead one to believe that θ_2 is the true parameter.)

The challenge in McLennan (1984) was to show that it is possible that for all δ in some interval $[0, \overline{\delta})$, π_δ is close (in the C^1 sense) to π_0 and therefore has the qualitative features that imply that it is impossible to move between the two half intervals. This was accomplished using ad hoc methods. The results studied in subsequent sections are much more general and systematic.

The types of learning traps studied in this paper are possible for any finite number of parameters. Even for this example, the mechanism leading to asymptotically incomplete learning in this paper is more general and robust, since it does not depend on the sequence of beliefs being unable to move from one region in the space of beliefs to another.

4 Possibly Asymptotically Incomplete Learning

Fix $\omega^* \in \pi^{-1}(A^*) \cap \Delta^\circ(\Theta)$, and let $a^* = \pi(\omega^*)$. In this and the following two sections we study the possibility that ω^* is a learning trap. This is the case if and only if there is a positive probability that $\ln \|\tilde\omega_t - \omega^*\| \to -\infty$ when $\tilde\omega_0 \neq \omega^*$, and our main intuition is that the reasoning underlying the law of large numbers can be applied to the sum

$$\ln \|\tilde\omega_t - \omega^*\| - \ln \|\tilde\omega_0 - \omega^*\| = \sum_{s=0}^{t-1} \ln \frac{\|\tilde\omega_{s+1} - \omega^*\|}{\|\tilde\omega_s - \omega^*\|}.$$

In particular, suppose that the expectation of $\ln \frac{\|\tilde\omega_{t+1} - \omega^*\|}{\|\tilde\omega_t - \omega^*\|}$ conditional on $\tilde\omega_t$ is negative for all $\tilde\omega_t \neq \omega^*$ in some region around ω^*. (Note that this implies that ω^* has a neighborhood that contains no other point of $\pi^{-1}(A^*)$.) If $\tilde\omega_t$ is quite close to ω^*, then it is quite unlikely that the process will ever escape this region, and when it does not it will necessarily be drawn to ω^*. On the other hand, if the expectation of $\ln \frac{\|\tilde\omega_{t+1} - \omega^*\|}{\|\tilde\omega_t - \omega^*\|}$ conditional on $\tilde\omega_t$ is positive for all $\tilde\omega_t$ in some region around ω^*, then the probability that $\tilde\omega_t \to \omega^*$ is zero.

From this point forward we assume that π and each $q_\theta(y|\cdot) : A \to (0,1)$ are C^1. Since π is C^1, $\pi(\tilde\omega_t) - a^*$ is an approximately linear function of $\tilde\omega_t - \omega^*$, and since the functions $q_\theta(y|\cdot)$ are C^1, the distribution of $\tilde\omega_{t+1} - \omega^*$ is an approximately linear function of $\tilde\omega_t - \omega^*$ and $\pi(\tilde\omega_t) - a^*$. This leads us to expect that if $\{\tilde\omega_t\}$ and $\{\tilde\omega'_0\}$ are two versions of our stochastic process and $\tilde\omega'_0 - \omega^* = \alpha(\tilde\omega_0 - \omega^*)$, then the process $\{\tilde\omega'_t - \omega^*\}$ should be an approximate rescaling of $\{\tilde\omega_t - \omega^*\}$, and that in the limit as $\alpha \to 0$ the processes can be adequately characterized in terms of the derivatives of the functions π and q_θ.

We now provide precise substantiation of this intuition. It will be convenient to work with a system of polar coordinates. Let

$$H_0 = \{\, \tau \in \mathbb{R}^\Theta : \sum_\theta \tau_\theta = 0 \,\} \quad \text{and} \quad H_1 = \{\, \omega \in \mathbb{R}^\Theta : \sum_\theta \omega_\theta = 1 \,\}.$$

That is, H_1 is the hyperplane in $\mathbb{R}^\Theta$ that contains Ω, and H_0 is the parallel hyperplane through the origin. The primary role of H_0 is that it is the space of vectors tangent to H_1 at any of its points. The unit sphere in H_0 is

$$S = \{\, \sigma \in H_0 : \|\sigma\| = 1 \,\}.$$

For $(r, \sigma) \in [0, \infty) \times S$ let

$$\omega(r, \sigma) = \omega^* + r\sigma.$$

Of course any $\omega \in H_1 \setminus \{\omega^*\}$ is $\omega(r, \sigma)$ for a unique (r, σ). In this coordinate system the space of beliefs is

$$\hat\Omega = \{\, (r, \sigma) \in [0, \infty) \times S : \omega(r, \sigma) \in \Omega \,\}.$$

For each $y \in Y$ let $\tilde{q}(\cdot, y) : \hat{\Omega} \to (0,1)$ and $\tilde{\beta}(\cdot, y) : \hat{\Omega} \to \Omega$ be the functions

$$\tilde{q}(r, \sigma, y) = q(\omega(r, \sigma), \pi(\omega(r, \sigma)), y) \quad \text{and} \quad \tilde{\beta}(r, \sigma, y) = \beta(\omega(r, \sigma), \pi(\omega(r, \sigma)), y). \tag{4.1}$$

The rescaled (by $1/r$) amount by which beliefs are adjusted when y is observed is given by the function $\nu(\cdot, y) : \hat{\Omega} \to H_0$ defined by setting

$$\nu(r, \sigma, y) = \begin{cases} \sigma + \frac{1}{r}(\tilde{\beta}(r, \sigma, y) - \omega(r, \sigma)), & r > 0, \\ \frac{\partial \tilde{\beta}}{\partial r}(0, \sigma, y), & r = 0. \end{cases}$$

Lemma 4.1. *For each y, $\nu(\cdot, y)$ is continuous.*

Proof. Evidently $\nu(\cdot, y)$ is continuous at every (r, σ) with $r > 0$, and the restriction of $\nu(\cdot, y)$ to $\{0\} \times S$ is also continuous. If $\{(r_n, \sigma_n)\}$ is a sequence in $\{ (r, \sigma) \in \hat{\Omega} : r > 0 \}$ converging to $(0, \sigma)$, the mean value theorem implies that for each n there is a $r_n' \in (0, r_n)$ such that $\tilde{\beta}(r_n, \sigma_n, y) = \tilde{\beta}(0, \sigma_n, y) + r_n \frac{\partial \tilde{\beta}}{\partial r}(r_n', \sigma_n, y)$, so that

$$\nu(r_n, \sigma_n, y) = \sigma_n + \tfrac{1}{r_n}(\tilde{\beta}(r_n, \sigma_n, y) - \omega^* - r_n \sigma_n) = \frac{\partial \tilde{\beta}}{\partial r}(r_n', \sigma_n, y).$$

Since $\tilde{\beta}$ is C^1, $\nu(r_n, \sigma_n, y) \to \nu(0, \sigma, y)$. Thus $\nu(\cdot, y)$ is continuous at each $(0, \sigma)$. □

We now have

$$Q(\omega(r, \sigma), \pi(\omega(r, \sigma))) = \sum_y \tilde{q}(r, \sigma, y) \delta_{\tilde{\beta}(r, \omega, y)}.$$

The guiding intuition is that when r is small, this distribution should be well approximated by $\sum_y q(a^*, y) \cdot \delta_{\omega^* + r\nu(0, \sigma, y)}$. Let $G : \hat{\Omega} \to \Delta(H_0)$ be the function given by

$$G(r, \sigma) = \sum_y q(y | \omega(r, \sigma), \pi(\omega(r, \sigma))) \cdot \delta_{\nu(r, \sigma, y)}.$$

In particular, $G(0, \sigma)$ assigns probability $q(y|a^*)$ to each $\nu(0, \sigma, y)$ (recall that $q(y|a^*)$ is the common value of $q_\theta(y|a^*)$) and because G is continuous, $G(0, \cdot)$ is a good approximation of $G(r, \cdot)$ for small r. Below $\tilde{\nu}_{(r,\sigma)}$ denotes a random variable with distribution given by $G(r, \sigma)$.

We now arrive at two of our main results. Theorem 2 gives conditions under which learning is possibly asymptotically incomplete because $\mathbb{P}(\tilde{\omega}_t \to \omega^* | \tilde{\omega}_0 = \omega_0) > 0$ for all $\omega_0 \neq \omega^*$ in some neighborhood of ω^*. Theorem 3 gives conditions under which $\{\tilde{\omega}_t\}$ certainly does not converge to ω^*.

Theorem 2. *If, for all $\sigma \in S$,*

$$\mathbb{E}_{G(0,\sigma)}(\ln \|\tilde{\nu}_{(0,\sigma)}\|) < 0, \tag{4.2}$$

then there are $C, R > 0$ and $\gamma \in (0,1)$ such that

$$\mathbb{P}\big(\|\tilde{\omega}_t - \omega^*\| \geq R \text{ for some } t|\tilde{\omega}_0 = \omega_0\big) < C\gamma^{-\ln(\|\omega_0-\omega^*\|/R)}$$

for all $\omega_0 \neq \omega^*$ such that $\|\omega_0 - \omega^*\| < R$

Theorem 3. *If*

$$\mathbb{E}_{G(0,\sigma)}(\ln\|\tilde{\nu}_{(0,\sigma)}\|) > 0, \tag{4.3}$$

for all $\sigma \in S$, and $\omega_0 \neq \omega^*$, then $\mathbb{P}\big(\tilde{\omega}_t \to \omega^*\big|\tilde{\omega}_0 = \omega_0\big) = 0$.

These results give an important intuition concerning when asymptotically incomplete learning is possible. If $\nu(0,\sigma,y)$ is orthogonal to σ for all σ and y, then $\|\sigma+\nu(0,\sigma,y)\| \geq \|\sigma\| = 1$ and $\ln(\|\sigma+\nu(0,\sigma,y)\|) > 0$, so (4.3) holds and asymptotically incomplete learning is impossible. On the other hand, if, for all σ and y, $\nu(0,\sigma,y)$ is σ multiplied by a scalar in $(-1,\infty)$, then Jensen's inequality implies that (4.2) holds.

Suppose that the hypotheses of Theorem 2 are satisfied. The continuity of $\nu(\cdot,y)$ and $\tilde{q}(\cdot,y)$, and the compactness of S, imply that for some $\mu' > 0$, $\mathbb{E}_{G(0,\sigma)}(\ln\|\tilde{\nu}_{(0,\sigma)}\|) < -\mu'$ for all σ. The continuity of $\nu(\cdot,y)$ and $\tilde{q}(\cdot,y)$, the compactness of S, and the finiteness of Y, imply that there are $a', b' > 0$ such that $\ln\|\tilde{\nu}(0,\sigma,y)\| \in [-a', b']$ for all σ and y. Choose $a > a'$, $b > b'$, and $\mu \in (0,\mu')$. Continuity implies that there is an $R > 0$ such that

$$\mathbb{E}_{Q(\omega(r,\sigma),\pi(\omega(r,\sigma)))}(\ln(\|\tilde{\beta}(r,\sigma,y) - \omega^*\|/r)) = \mathbb{E}_{G(r,\sigma)}(\ln\|\tilde{\nu}_{(r,\sigma)}\|) < -\mu$$

for all $r \in [0,R]$ and σ, and

$$\ln(\|\tilde{\beta}(r,\sigma,y) - \omega^*\|/r) = \ln\|\nu(r,\sigma,y)\| \in [-a,b]$$

for all $r \in [0,R]$, σ, and y. Theorem 2 now follows from (a) of the following result (which is our version of the law of large numbers) if, for $\omega \neq \omega^*$, we set $\ell(\omega) = \ln(\|\omega-\omega^*\|/R)$. The same argument (with obvious modifications) shows that Theorem 3 follows from (b) of the result below.

Proposition 4.1. *Let Ω be a measurable space, let $Q : \Omega \to \Delta(\Omega)$ be a Markov kernel, and let $\{\tilde{\omega}_t\}_{t=0}^{\infty}$ be the Markov process generated by Q, conditional on $\tilde{\omega}_0$. Suppose that $a, b > 0$ and $\ell : \Omega \to \mathbb{R}$ is a measurable function such that*

$$\mathbb{P}_{Q(\cdot|\omega)}\big[\ell(\omega) - a \leq \ell(\omega') \leq \ell(\omega) + b\big] = 1$$

for all ω such that $\ell(\omega) \leq 0$. Let $p : \Omega \to [0,1]$ be the function

$$p(\omega_0) = \mathbb{P}\big[\ell(\tilde{\omega}_t) \geq 0 \text{ for some } t|\tilde{\omega}_0 = \omega_0\big].$$

(a) *If* $\mu \in (0,a)$ *and* $\mathbb{E}_{Q(\cdot|\omega)}\big[\ell(\omega')\big] \le \ell(\omega) - \mu$ *for all* ω *such that* $\ell(\omega) \le 0$*, then there are numbers* $C > 0$ *and* $\gamma \in (0,1)$ *such that* $p(\omega_0) < C\gamma^{-\ell(\omega_0)}$ *for all* ω_0 *such that* $\ell(\omega_0) < 0$*.*

(b) *If* $\mu \in (0,b)$ *and* $\mathbb{E}_{Q(\cdot|\omega)}\big[\ell(\omega')\big] \ge \ell(\omega) + \mu$ *for all* ω *such that* $\ell(\omega) \le 0$*, then* $p(\omega_0) = 1$ *for all* ω_0 *such that* $\ell(\omega_0) \le 0$*.*

Proof. For each $T = 0,1,2,\dots$ let $p_T : \Omega \to [0,1]$ be the function

$$p_T(\omega_0) = \mathbb{P}\big[\ell(\tilde{\omega}_t) \ge 0 \text{ for some } t = 0,\dots,T|\tilde{\omega}_0 = \omega_0\big].$$

(a) We will prove, by induction, that $p_T(\omega_0) < C\gamma^{-\ell(\omega_0)}$ for all T and all ω_0 such that $\ell(\omega_0) < 0$, which of course implies the claim. Clearly $p_0(\omega_0) = 0$ if $\ell(\omega_0) < 0$, so we may assume that the claim has already been proven with $T-1$ in place of T.

Choose $\gamma \in (0,1)$ such that

$$\frac{b+\mu}{a+b}\gamma^a + \frac{a-\mu}{a+b}\gamma^{-b} < 1.$$

To see that this is possible observe that, by Taylor's theorem applied to $\gamma^a = e^{a\ln\gamma}$ and $\gamma^{-b} = e^{-b\ln\gamma}$, the left hand side is

$$\begin{aligned} &= \frac{b+\mu}{a+b}(1 + a\ln\gamma + O((a\ln\gamma)^2) + \frac{a-\mu}{a+b}(1 - b\ln\gamma + O((b\ln\gamma)^2) \\ &= 1 + \mu\ln\gamma + O((a\ln\gamma)^2) + O((b\ln\gamma)^2). \end{aligned}$$

Clearly we can choose $C > 0$ large enough that

$$\frac{a-\mu}{a-\ell} + C\big(\frac{b+\mu}{a+b}\gamma^a + \frac{a-\mu}{a+b}\gamma^{-b}\big)\gamma^{-\ell} \le C\gamma^{-\ell}$$

whenever $0 \ge \ell \ge -b$.

If $-b \le \ell(\omega_0) \le 0$, then the probability distribution on $[\ell(\omega_0) - a, \ell(\omega_0) + b]$ that has mean at most $\ell(\omega_0) - \mu$ and which maximizes the probability of a nonnegative number is a sum of a mass point at 0 and a mass point at $\ell(\omega_0) - a$. Thus

$$p_1(\omega_0) = \mathbb{P}\big[\ell(\tilde{\omega}_1) \le 0|\tilde{\omega}_0 = \omega_0\big] \le \frac{a-\mu}{a-\ell(\omega_0)}$$

if $-b \le \ell(\omega_0) \le 0$, and of course $p_1(\omega_0) = 0$ if $\ell(\omega_0) \le -b$.

The induction assumption gives

$$\mathbb{E}\big[p_{T-1}(\tilde{\omega}_1)|\tilde{\omega}_0 = \omega_0\big] \le \mathbb{E}\big[C\gamma^{-\ell(\tilde{\omega}_1)}|\tilde{\omega}_0 = \omega_0\big].$$

Since $\gamma^{-\ell}$ is a convex function of ℓ, the right hand side is increased by replacing the distribution of $\ell(\tilde{\omega}_1)$ by a mean preserving spread, so it is maximized by the distribution on possible values of ℓ that places all mass on the endpoints and has expectation $\ell(\omega_0)-\mu$. Thus:

$$\mathbb{E}\big[p_{T-1}(\tilde{\omega}_1)|\tilde{\omega}_0=\omega_0)\big] \le C\big(\frac{b+\mu}{a+b}\gamma^{-\ell(\omega_0)+a}+\frac{a-\mu}{a+b}\gamma^{-\ell(\omega_0)-b}\big)$$
$$= C\big(\frac{b+\mu}{a+b}\gamma^{a}+\frac{a-\mu}{a+b}\gamma^{-b}\big)\gamma^{-\ell(\omega_0)}.$$

We have

$$p_T(\omega_0) \le \mathbb{P}\big[\ell(\omega_1) < \tfrac{\mu}{2}|\omega_0\big] + \mathbb{E}\big[p_{T-1}(\omega_1)|\omega_0\big].$$

If $\ell(\omega_0) \le -b$, then $p_1(\omega_0) = 0$ and the inequality above gives $p_T(\omega_0) < C\gamma^{-\ell(\omega_0)}$. If $0 \ge \ell(\omega_0) \ge -b$, then

$$p_T(\omega_0) \le \frac{a-\mu}{a-\ell(\omega_0)} + C\big(\frac{b+\mu}{a+b}\gamma^{a}+\frac{a-\mu}{a+b}\gamma^{-b}\big)\gamma^{-\ell(\omega_0)} < C\gamma^{-\ell(\omega_0)}.$$

(b) For $t = 0, 1, \ldots$ let $h_t : \mathbb{R} \to [0, 1]$ be the function

$$h_t(\ell) = \begin{cases} 0, & \ell \le -\mu t, \\ \frac{\ell+\mu t}{b+\mu t}, & -\mu t < \ell < 0, \\ 1, & \ell \ge 0. \end{cases}$$

Since, for all ℓ, $h_t(\ell) \to 1$ as $t \to \infty$, it suffices so show that $p_t(\omega_0) \ge h_t(\ell(\omega_0))$ for all t and ω_0. Of course this is the case when $t = 0$, so may assume it has already been established for t, and we will show that it holds with $t+1$ in place of t. Of course the claim is automatic if $\ell(\omega_0) \ge 0$, so assume that $\ell(\omega_0) < 0$. Note that every point of the graph of $h_t|_{(-\infty,b]}$ lies on or above the graph of the function $\ell \mapsto \frac{\ell+\mu t}{b+\mu t}$, so the expectation of h_t with respect to any distribution on $[\ell - a, \ell + b]$ whose mean is at least $\ell + \mu$, is at least $\frac{\ell+\mu(t+1)}{b+\mu t}$. We now have

$$p_{t+1}(\omega_0) = \mathbb{E}_{Q(\cdot|\omega_0)}\big[p_t(\omega_1)\big] \ge \mathbb{E}_{Q(\cdot|\omega_0)}\big[h_t(\ell(\omega_1))\big] \ge \frac{\ell(\omega_0)+\mu(t+1)}{b+\mu t}$$
$$\ge \frac{\ell(\omega_0)+\mu(t+1)}{b+\mu(t+1)} = h_{t+1}(\ell(\omega_0)). \qquad \square$$

5 An Algebraic Simplification

Due to the relative intractability of the logarithm function, conditions (4.2) and (4.3) are rather difficult to deal with. Our strategy in the next section is to study a class of examples parameterized by a parameter $\alpha \ge 0$. We will show that examples with the desired properties exist by showing that the derivatives of the relevant functions at $\alpha = 0$ have properties that imply that the example is satisfactory when $\alpha > 0$ is small enough.

In this section we compute the relevant derivatives of the expressions in (4.2) and (4.3) with respect to α, and in the next section we study the particular examples of interest.

In order to keep the notation compact we do not make α an explicit argument of the functions that have already been defined. Specifically, henceforth α will be an implicit argument of $q_\theta(y|a)$, $q_\omega(y|a)$, $\beta_\theta(\omega, a, y)$, $\hat{\beta}(r, \sigma, y)$, $\nu(r, \sigma, y)$, and $G(r, \sigma)$. For any function f of α and other variables, $\partial_\alpha f$ and $\partial_{\alpha\alpha} f$ denote the first and second partial derivative of f with respect to α, evaluated at $\alpha = 0$. We assume that for each θ and y, $q_\theta(y|a^*)$ is a constant $q(y|a^*)$ that does not depend on either θ or α, so that a^* is uninformative for all α. We also assume that each $q_\theta(y|\cdot)$ is jointly C^2 as a function of a and α, and that all $a \in A$ are uninformative when $\alpha = 0$. Intuitively, decreasing α corresponds to slowing down learning.

Our goal now is to develop algebraic versions of the tests provided by Theorems 2 and 3 by differentiating with respect to α at $\alpha = 0$. All the preceeding results hold for each α, and in addition $\sum_y q(y|a^*)\nu(0, \sigma, y) = 0$ holds by virtue of the martingale nature of Bayesian updating and the continuity given by Lemma 4.1. Differentiating this equation twice gives

$$\sum_y q(y|a^*)\partial_\alpha \nu(0, \sigma, y) = 0 \quad \text{and} \quad \sum_y q(y|a^*)\partial_{\alpha\alpha} \nu(0, \sigma, y) = 0. \tag{5.1}$$

(The second differentiation is valid because the first equation holds for all α, and not just when $\alpha = 0$.)

The mean in (4.2) and (4.3) is

$$M(\sigma) = \mathbb{E}_{G(\sigma,0)}(\ln \|\tilde{\nu}\|) = \sum_{y \in Y} q(y|a^*) \ln \|\nu(0, \sigma, y)\|.$$

We will need the following fact whose proof is a matter of elementary calculus, and is omitted.

Lemma 5.1. *If v is a vector in an inner product space, w is a C^2 function from $(-\varepsilon, \varepsilon)$ to this space with $w(0) = 0$, and $g : \mathbb{R} \to \mathbb{R}$ is the function $g(\alpha) = \ln \|v + w(\alpha)\|$, then*

$$g'(0) = \frac{\langle v, w'(0)\rangle}{\|v\|^2} \quad \text{and} \quad g''(0) = \frac{\|v\|^2(\|w'(0)\|^2 + \langle v, w''(0)\rangle) - 2\langle v, w'(0)\rangle^2}{\|v\|^4}.$$

Using these results (with $v = \sigma$ and $w(\alpha) = \nu(\sigma, 0, y) - \sigma$), the fact that $\|\sigma\| = 1$ if $\sigma \in S$, and (5.1), we compute that

$$\partial_\alpha M(\sigma) = \sum_y q(y|a^*)\langle \sigma, \partial_\alpha \nu(0, \sigma, y)\rangle = \Big\langle \sigma, \sum_y q(y|a^*)\partial_\alpha \nu(0, \sigma, y)\Big\rangle = 0$$

and

$$\partial_{\alpha\alpha} M(\sigma) = \sum_y q(y|a^*)\big(\|\partial_\alpha \nu(0,\sigma,y)\|^2 - 2\langle \sigma, \partial_\alpha \nu(0,\sigma,y)\rangle^2\big). \tag{5.2}$$

For small α we have $M(\sigma) \approx \partial_{\alpha\alpha} M(\sigma)\alpha^2$, by Taylor's theorem. Combining this with the continuity of $\nu(\cdot, y)$ and the compactness of S gives:

Proposition 5.1. *If $\partial_{\alpha\alpha} M(\sigma) < (>)\, 0$ for all $\sigma \in S$, then*

$$\sum_{y \in Y} q(y|a^*) \ln \|\nu(0,\sigma,y)\| < (>)\, 0$$

for all $\sigma \in S$ when α is sufficiently small.

The formula for $\partial_{\alpha\alpha} M(\sigma)$ conforms to our intuition concerning the circumstances under which incomplete learning is possible. Specifically, in order for $M(\sigma)$ to be negative, it must be the case that learning predominantly moves beliefs in directions that are parallel to σ. This is necessarily the case when Ω is 1-dimensional, and in this sense these results provide insights that cannot be obtained from the models studied in McLennan (1984) and Harrison et al. (2012).

6 A Class of Examples

Our goal in this section is to provide a class of concrete examples with possibly asymptotically incomplete learning. Let $\Theta = \{\theta_1, \ldots, \theta_n\}$. The space of actions A is $\Delta(\Theta)$ (Thus A is, in a sense, "the same" as Ω, but this has no conceptual significance.) Let $a^* = (\frac{1}{n}, \ldots, \frac{1}{n})$ be the barycenter. The possible outcomes are success and failure, so $Y = \{\mathbf{S}, \mathbf{F}\}$.

The examples are parameterized by a number $\alpha \geq 0$ and an $n \times n$ interior bistochastic matrix[8] $\ell = (\ell_{ij})$. For $i = 1, \ldots, n$ let

$$q_{\theta_i}(\mathbf{S}|a) = \tfrac{1}{2} + \alpha\big(\tfrac{1}{n^2} - \sum_j \ell_{ij} a_j^2\big). \tag{6.1}$$

We restrict attention to α that are small enough that $0 < q_{\theta_i}(\mathbf{S}|a) < 1$ for all i and $a \in A$. Note that $q_{\theta_i}(\mathbf{S}|a^*) = \frac{1}{2}$ for all i, so a^* is indeed uninformative for all α. Similarly, every action is uninformative when $\alpha = 0$.

We assume that the reward for success is 1 while the reward for failure is 0, regardless of the action. Thus the reward function $u_0 : \Omega \times A \to \mathbb{R}$ is given by

$$u_0(\omega, a) = \sum_\theta \omega_\theta q_\theta(\mathbf{S}|a) = \tfrac{1}{2} + \alpha\big(\tfrac{1}{n^2} - \sum_i \omega_{\theta_i} \sum_j \ell_{ij} a_j^2\big).$$

[8]Recall that this means that $\ell_{ij} > 0$ for all i and j, $\sum_j \ell_{ij} = 1$ for all i, and $\sum_i \ell_{ij} = 1$ for all j.

The analysis of myopically optimal behavior uses other functions which we introduce now. Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be the linear transformation whose matrix is ℓ, so $L(y) = (L_1(y), \ldots, L_n(y))$ where $L_i(y) = \sum_j \ell_{ij} y_j$. Let $\lambda : \mathbb{R}^n \to \mathbb{R}^n$ be the linear transformation whose matrix is the transpose of ℓ, so $\lambda(y) = (\lambda_1(y), \ldots, \lambda_n(y))$ where $\lambda_j(y) = \sum_i \ell_{ij} y_i$. Let $\Lambda = L \circ \lambda : \mathbb{R}^n \to \mathbb{R}^n$, so $\Lambda(y)$ is the vector with i^{th} component

$$\Lambda_i(y) = \sum_j \ell_{ij} \lambda_j(y) = \sum_j \ell_{ij} \sum_k \ell_{kj} y_k = \sum_k \langle \ell^i, \ell^k \rangle y_k \tag{6.2}$$

where ℓ^i is the i^{th} row of ℓ. Let $\lambda^0 = \lambda|_{H_0}$ and $L^0 = L|_{H_0}$. Note that

$$\sum_j \lambda_j(y) = \sum_j (\sum_i \ell_{ij} y_i) = \sum_i y_i = \sum_j y_j = \sum_i (\sum_j \ell_{ij} y_j) = \sum_i L_i(y).$$

Therefore the images of λ^0 and L^0 are contained in H_0, so $\Lambda^0 = L^0 \circ \lambda^0 : H_0 \to H_0$ is a well defined linear transformation.

For $\omega \in \Omega$ let

$$\iota(\omega) = \big(\sum_j \tfrac{1}{\lambda_j(\omega)}\big)^{-1}.$$

Let $\pi_0 : \Omega \to A$ be the function

$$\pi_0(\omega) = \iota(\omega)\big(\tfrac{1}{\lambda_1(\omega)}, \ldots, \tfrac{1}{\lambda_n(\omega)}\big). \tag{6.3}$$

Let $\omega^* = (\frac{1}{n}, \ldots, \frac{1}{n})$, and note that $\lambda_j(\omega^*) = \frac{1}{n}$ for all j and $\iota(\omega^*) = \frac{1}{n^2}$, so $\pi_0(\omega^*) = a^*$.

Lemma 6.1. *π_0 is the unique myopically optimal policy for u_0, and u_0 satisfies the standard conditions.*

Proof. The first order condition for optimality is that for all j and j', increasing a_j and decreasing $a_{j'}$ by the same amount does not have a first order effect, which means that $\sum_i \omega_{\theta_i} \ell_{ij} a_j = \sum_i \omega_{\theta_i} \ell_{ij'} a_{j'}$, i.e., $\lambda_j(\omega) a_j = \lambda_{j'}(\omega) a_{j'}$. Therefore the unique point in A satisfying the first order conditions is the positive scalar multiple of $(1/\lambda_1(\omega), \ldots, 1/\lambda_n(\omega))$ whose components sum to one, but the reward function is a strictly concave function of a because the ℓ_{ij} are positive, so this point is in fact the unique maximizer. It is obvious that $\partial_{aa} u_0(\omega, a)$ is negative definite for all ω and a, so u_0 satisfies the standard conditions. □

In view of this and Proposition 3.1, Theorem 1 now implies that for each r there is an $\varepsilon_r > 0$ such that for all $\delta \in (-\varepsilon_r, \varepsilon_r)$ the optimal policy π_δ for discount factor δ is $C^{r,1}$ and π_δ is a Lipschitz (relative to the norm of $\Delta^r C^r(\Omega, A)$) function of δ. We hasten to add that (because $\varepsilon_r \to 0$ is possible) it does *not* follow that there is an $\varepsilon_\infty > 0$ such that π_δ is C^∞ for all $\delta \in (-\varepsilon_\infty, \varepsilon_\infty)$, although we do suspect that this is the case.

The remainder of this section computes a closed form expression for $\partial_{\alpha\alpha} M(\sigma)$, and then shows that it has the desired properties. We begin with a technical result.

Lemma 6.2. *ω^* is a critical point of ι.*

Proof. Applying standard rules of differentiation and $\lambda_i(\omega^*) = 1/n$, for $\sigma \in H_0$ we have

$$D\iota(\omega^*)\sigma = -\Big(\sum_i \frac{1}{\lambda_i(\omega^*)}\Big)^{-2}\Big(\sum_i -\frac{D\lambda_i(\omega^*)\sigma}{\lambda_i(\omega^*)^2}\Big)$$

$$= n^{-2}\sum_i\sum_j \ell_{ji}\sigma_j = n^{-2}\sum_j \sigma_j = 0. \qquad \square$$

Proposition 6.1. *For all $\sigma \in S$,*

$$\partial_{\alpha\alpha} M(\sigma) = \frac{16}{n^4}\cdot\big(\|\Lambda^0(\sigma)\|^2 - 2\langle\sigma, \Lambda^0(\sigma)\rangle^2\big).$$

Proof. We will show that

$$\nu(0,\sigma,\mathbf{S}) = \sigma + \tfrac{4\alpha}{n^2}\Lambda^0(\sigma) \quad \text{and} \quad \nu(0,\sigma,\mathbf{F}) = \sigma - \tfrac{4\alpha}{n^2}\Lambda^0(\sigma). \tag{6.4}$$

The claim follows from this and equation (5.2). By definition $\nu(0,\sigma,\mathbf{S}) = \partial_r\tilde\beta(0,\sigma,\mathbf{S})$ and $\tilde\beta(r,\sigma,\mathbf{S}) = \beta(\omega^* + r\sigma, \pi_0(\omega^*+r\sigma), \mathbf{S})$. For $\omega \in \Omega$ let $f(\omega) = \beta(\omega, \pi_0(\omega), \mathbf{S})$. Then $\nu(0,\sigma,\mathbf{S}) = D(\omega^*)\sigma$.

Substituting the formula (6.3) for $\pi_0(\omega)$ into (6.1) gives

$$q_{\theta_i}(\mathbf{S}|\pi_0(\omega)) = \tfrac{1}{2} + \alpha\big(\tfrac{1}{n^2} - \iota(\omega)^2\sum_j \tfrac{\ell_{ij}}{\lambda_j(\omega)^2}\big).$$

Multiplying this by ω_{θ_i} and summing, then recognizing that $\sum_i \ell_{ij}\omega_{\theta_i} = \lambda_j(\omega)$, gives

$$q_\omega(\mathbf{S}|\pi_0(\omega)) = \tfrac{1}{2} + \alpha\big(\tfrac{1}{n^2} - \iota(\omega)^2\sum_j \tfrac{1}{\lambda_j(\omega)^2}\sum_i \ell_{ij}\omega_{\theta_i}\big) = \tfrac{1}{2} + \alpha\big(\tfrac{1}{n^2} - \iota(\omega)\big).$$

Substituting these equations into the formula (3.1) for the Bayesian posterior gives

$$\beta_{\theta_i}(\omega, \pi_0(\omega), \mathbf{S}) = f_{\theta_i}(\omega) = \frac{\omega_{\theta_i}\big(1 + 2\alpha\big(\tfrac{1}{n^2} - \iota(\omega)^2\sum_j \tfrac{\ell_{ij}}{\lambda_j(\omega)^2}\big)\big)}{1 + 2\alpha\big(\tfrac{1}{n^2} - \iota(\omega)\big)}.$$

Since ω^* is a critical point of ι (Lemma 6.2) the derivative of f_{θ_i} at ω^* is the derivative of the numerator divided by the denominator evaluated at ω^*, which is one, so

$$Df_{\theta_i}(\omega^*)\sigma = \Big[\sigma_{\theta_i}\Big(1 + 2\alpha\big(\tfrac{1}{n^2} - \iota(\omega)^2\sum_j \tfrac{\ell_{ij}}{\lambda_j(\omega)^2}\big)\Big) - \omega_{\theta_i}2\alpha D\big[\iota(\omega)^2\sum_j \tfrac{\ell_{ij}}{\lambda_j(\omega)^2}\big]\sigma\Big]\Big|_{\omega=\omega^*}.$$

Since $\iota(\omega^*) = \frac{1}{n^2}$, $\lambda_j(\omega^*) = \frac{1}{n}$, and $\sum_j \ell_{ij} = 1$, the first of the two terms reduces to σ_{θ_i}. Again taking advantage of the fact that ω^* is a critical point of ι, we have

$$D\big[\iota(\omega)^2 \sum_j \tfrac{\ell_{ij}}{\lambda_j(\omega)^2}\big]\sigma\Big|_{\omega=\omega^*} = -\iota(\omega^*)^2 \sum_j \tfrac{2\ell_{ij}\lambda_j(\sigma)}{\lambda_j(\omega^*)^3} = -\tfrac{2}{n}\sum_j \ell_{ij}\lambda_j(\sigma) = -\tfrac{2}{n}\Lambda_i^0(\sigma).$$

Multiplying this equation by $-\omega^*_{\theta_i} 2\alpha = -2\alpha/n$, we arrive at

$$\nu(0,\sigma,\mathbf{S}) = Df(\omega^*)\sigma = \sigma + \tfrac{4\alpha}{n^2}\Lambda^0(\sigma).$$

A similar calculation (or the martingale property) gives the second equation in (6.4). □

To produce an example for which (4.2) holds for all $\sigma \in S$ it now suffices to give an interior bistochastic matrix $\ell = (\ell_{ij})$ such that Λ^0 is a scalar multiple of Id_{H_0}. For example, for sufficiently small $\varepsilon > 0$ we can set

$$\ell_{ij} = \begin{cases} 1-(n-1)\varepsilon, & i = j \\ \varepsilon, & i \neq j. \end{cases}$$

We claim that $\Lambda^0 = (1 - n(2\varepsilon - n\varepsilon^2))\mathrm{Id}_{H_0}$. To see this first observe that

$$\langle \ell^i, \ell^j \rangle = \begin{cases} 1-(n-1)(2\varepsilon - n\varepsilon^2), & i = j \\ 2\varepsilon - n\varepsilon^2, & i \neq j. \end{cases}$$

Now, using equation (6.2) and the fact that $\sum_k \sigma_{\theta_k} = 0$, for $\sigma \in S$ and each i we compute that

$$\begin{aligned} \Lambda_i^0(\sigma) = \sum_k \langle \ell^i, \ell^k \rangle \sigma_{\theta_k} &= (1-(n-1)(2\varepsilon - n\varepsilon^2))\sigma_{\theta_i} + \sum_{j\neq i}(2\varepsilon - n\varepsilon^2)\sigma_{\theta_j} \\ &= (1 - n(2\varepsilon - n\varepsilon^2))\sigma_{\theta_i}. \end{aligned}$$

We have now produced a concrete example for which $\partial_{\alpha\alpha} M(\sigma) < 0$ for all $\sigma \in S$. By Proposition 5.1, for small α (4.2) holds, so the myopically optimal policy gives a positive probability of convergence to ω^*. As we explained at the end of Section 4, after we have shown that the optimal policy varies continuously in the C^1 topology, it follows that the optimal policy for discount factor δ also gives a positive probability of convergence to ω^* if $|\delta|$ is sufficiently small.

Since $\partial_{\alpha\alpha} M(\sigma) \leq 0$ when σ is an eigenvector, this class of examples does not admit the possibility that (4.3) might hold for all σ. If the eigenvalues are quite different, then (4.3) can hold for some σ. For example, let $n = 3$, and let

$$\ell^1 = (1,0,0), \quad \ell^2 = (0, \tfrac{1}{2}+\delta, 1-\delta), \quad \ell^3 = (0, 1-\delta, 1+\delta)$$

where $\delta > 0$ is small. (It will be obvious that perturbing slightly achieves $\ell_{ij} > 0$ for all i and j without disturbing the relevant properties.) The matrix of L is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} + 2\delta^2 & \frac{1}{2} - 2\delta^2 \\ 0 & \frac{1}{2} - 2\delta^2 & \frac{1}{2} + 2\delta^2 \end{pmatrix}.$$

The eigenvectors in H_0 are $(2, -1, -1)$ with corresponding eigenvalues 1, and $(0, 1, -1)$ with corresponding eigenvalue $4\delta^2$. We have

$$(z_1^2 + z_2^2)(z_1^2 + 16\delta^4 z_2^2) - 2(z_1^2 + 4\delta^2 z_2^2)^2 = z_2^4(-(z_1/z_2)^4 + (1 - 4\delta^2)(z_1/z_2)^2 - 16\delta^4).$$

If δ is small and $z_1/z_2 \cong 2\delta$, then this quantity is positive.

7 Concluding Remarks

We have provided an analysis of Bayesian learning traps that gives sufficient conditions for them to occur, and we have shown that if myopically optimal policies allow them, then so do the optimal policies of decision makers with small positive discount factors. We have given concrete methods to compute whether the relevant conditions hold, and have shown that for any finite space of possible parameters, there exist examples in which these conditions actually hold.

Some interesting questions remain unresolved. For example, can the conditions of Theorem 3 be satisfied by a myopically optimal policy? Also, can there be a positive probability of convergence to ω^* with discount factors arbitrarily close to one. In fact we expect that in the framework parameterized by α, reducing α is very similar to reducing the speed of learning, and thus may be thought of as analogous to dividing a period into many subperiods, so we expect that for any $\delta \in (0, 1)$, asymptotically incomplete learning will be possible when α is sufficiently small, but we have not managed to prove this. (In contrast, for a given problem with an uninformative action a^*, if δ is sufficiently close to one it cannot be optimal to choose a^* because if it was optimal to do so in a single period, it would be optimal to do so in every subsequent period, but the value in future periods of the information gleaned from a highly informative action exceeds the losses in the current period. A related result is Theorem 4.3 of Aghion et al. (1991), which asserts that as $\delta \to 1$, the discounted per period value of the problem approaches the discounted per period value of a decision maker who sees θ.)

Our analysis in this paper is restricted in several ways.

A natural direction of generalization is to situations in which the dimension of the space of beliefs may be greater than the codimension of the set A^* of uninformative actions. When these objects and the policy function are "well behaved," the set of beliefs mapped to uninformative actions will be a submanifold of Ω, and the question becomes whether there can be a positive probability that the sequence of beliefs converges to a point in this submanifold. One can anticipate certain additional technical complications, but at this point there seems to be little reason to expect the qualitative properties of the results to change.

A major direction for generalization is to consider the possibility that Y is infinite. In particular, the case of normally distributed shocks is of considerable interest. Again, significant additional complications can be foreseen, but at this point we are not aware of any insuperable obstacles.

Finally, an economically important possibility is that learning might be incomplete because there is a positive probability of convergence to a belief whose support is not all of Θ. As with the other extensions described above, this appears to present certain challenges, which most likely can be overcome.

A Proof of Theorem 1

This section continues Section 2, with the goal of proving Theorem 1. Appendix B generalizes this result to state and action spaces that are compact subsets of smooth manifolds, at the expense of additional technical complexity. That argument depends on the results in this section, so this section and Section 2 are prerequisites.

The next subsection studies the preservation of tameness by various operations that recombine operators, culminating in Proposition A3, which considers composition of C^r functions as an operator. We then use these tools to prove Proposition 3.1. After these preparations, the second subsection gives the bulk of the argument.

A1 Combinations of Tame Operators

Our argument is ultimately a matter of establishing the tameness of an operator mapping a neighborhood of the given single period reward function in $\Delta^r C^{r,1}_{\mathrm{w}}(\Omega \times A)$ to itself. Over the course of our work there will be numerous arguments showing that certain other operators are tame, or that tameness is preserved by certain operations. This subsection provides a basic toolkit for recombining such results. Throughout the following discussion, except as indicated, X, X', Y, Y', Z, Z', W, and W' are metric spaces, and X and X' are always compact.

The first of two results require no proof.

Lemma A1. *If $S \subset C(X, Y)$, $\gamma : S \to C(X', Y')$ is tame, and $S' \subset S$, then $\gamma|_{S'}$ is tame.*

Lemma A2. *If $g : X' \to Y'$ is Lipschitz, $S \subset C(X, Y)$, and $\gamma : S \to C(X', Y')$ is the constant operator $\gamma(f) = g$ for all f, then γ is tame.*

Let $C \subset X$ be compact, and let $\rho_C : C(X, Y) \to C(C, Y)$ be the operator $\rho_C(f) = f|_C$. Of course ρ_C is 1-Lipschitz, and it maps Lipschitz bounded sets to Lipschitz bounded sets, so:

Lemma A3. *ρ_C is a tame operator.*

Corollary A.1. *The identity function of $C(X, Y)$ is a tame operator.*

Lemma A4. *If $S \subset C(X, Y)$, $S' \subset C(X', Y')$, and $\gamma : S \to S'$ and $\gamma' : S' \to C(X'', Y'')$ are tame, then $\gamma' \circ \gamma$ is tame.*

Proof. Of course $\gamma' \circ \gamma$ is continuous. Fix $f \in S$, a γ-compliant neighborhood $U \subset S$ of f, and a γ'-compliant neighborhood $U' \subset S'$. Since γ is continuous, $U \cap \gamma^{-1}(U')$ is a neighborhood of f. For any Lipschitz bounded $T \subset U \cap \gamma^{-1}(U')$, $\gamma(T)$ and $\gamma'(\gamma(T))$ are Lipschitz bounded, and $\gamma' \circ \gamma|_T = \gamma'|_{\gamma(T)} \circ \gamma|_T$ is a composition of Lipschitz functions, so it is Lipschitz. Thus $U \cap \gamma^{-1}(U')$ is $\gamma' \circ \gamma$-compliant. □

We consider two sorts of cartesian products of functions. If $f \in C(X, Y)$ and $f' \in C(X, Y')$, $f \times f' : X \to Y \times Y'$ is the function $x \mapsto (f(x), f'(x))$. If $f \in C(X, Y)$ and $f' \in C(X', Y')$ let $f \otimes f' : X \times X' \to Y \times Y'$ be the function $(x, x') \mapsto (f(x), f'(x'))$. Our main use for theese constructions is as a proxy for the pair (f, f'), which allows us to avoid defining and analyzing tameness for operators acting on pairs of functions.

The proofs of the following two results are entirely parallel, so we only prove the first.

Lemma A5. *If $g : X \to Y'$ is Lipschitz, then the operator $\gamma : f \mapsto f \times g$ from $C(X, Y)$ to $C(X, Y \times Y')$ is tame.*

Proof. For any $f, f' \in C(X, Y)$, $d(f \times g, f' \times g) = d(f, f')$, so γ is 1-Lipschitz. Let $T \subset C(X, Y)$ be Lipschitz bounded. If $K \subset Y$ is a compact set such that $f(X) \subset K$ for all $f \in T$, then $K \times g(X)$ is compact and contains the image of $\gamma(f)$ for all $f \in T$. If every $f \in T$ is Λ-Lipschitz, then every element of $\{ f \times g : f \in T \}$ is Λ-Lipschitz, so $\{ f \times g : f \in T \}$ is Lipschitz bounded. Thus $C(X, Y)$ is γ-compliant. □

Lemma A6. *If $g : X' \to Y'$ is Lipschitz, then the operator $\gamma : f \mapsto f \otimes g$ from $C(X, Y)$ to $C(X \times X', Y \times Y')$ is tame.*

When $S \subset C(X,Y)$ and $\gamma : S \to C(X',Y')$ and $\gamma' : S \to C(X',Z')$ are operators, $\gamma \times \gamma' : S \to C(X', Y' \times Z')$ is the operator $(\gamma \times \gamma')(f) = \gamma(f) \times \gamma'(f)$.

Proposition A1. *For a nonempty $S \subset C(X,Y)$ two operators $\gamma : S \to C(X',Z)$ and $\gamma' : S \to C(X',Z')$ are tame if and only if $\gamma \times \gamma'$ is tame.*

Proof. First suppose that γ and γ' are tame. Of course $\gamma \times \gamma'$ is continuous. Fix $f \in S$ and neighborhoods $U, U' \subset S$ of f that are γ-compliant and γ'-compliant respectively. Suppose that $T \subset U \cap U'$ is Lipschitz bounded. Then $\gamma(T)$ and $\gamma'(T)$ are Lipschitz bounded, so there are compact sets $K \subset Z$ and $K' \subset Z'$ and Lipschitz constants Λ and Λ' such that for all $g \in T$, K and K' contain the images of $\gamma(g)$ and $\gamma'(g)$, $\gamma(g)$ is Λ-Lipschitz, and $\gamma'(g)$ is Λ'-Lipschitz. Therefore, for all $g \in T$, $K \times K'$ contains the image of $(\gamma \times \gamma')(g)$ and this function is $\max\{\Lambda, \Lambda'\}$-Lipschitz. In addition, there are constants L, L' such that $\gamma|_T$ is L-Lipschitz and $\gamma'|_{T'}$ is L'-Lipschitz, so $(\gamma \times \gamma')|_T$ is $\max\{L, L'\}$-Lipschitz. Thus $U \cap U'$ is $\gamma \times \gamma'$-compliant.

Now assume that $\gamma \times \gamma'$ is tame. Since $\gamma \times \gamma'$ is continuous, γ and γ' are continuous. Fixing $f \in S$, let $U \subset S$ be a $\gamma \times \gamma'$-compliant neighborhood of f, and let $T \subset U$ be Lipschitz bounded. There is a compact $K^* \subset Z \times Z'$ and a Λ^* such that for all $f \in T$, the image of $(\gamma \times \gamma')(f)$ is contained in K^* and $(\gamma \times \gamma')(f)$ is Λ^*-Lipschitz. Let $K = \{\, z : (z,z') \in K^* \,\}$ be the projection of K^* on Z. Since the projection $Z \times Z' \to Z$ is continuous, K is compact. For each $f \in T$, the image of $\gamma(f)$ is contained in K and $\gamma(f)$ is Λ^*-Lipschitz. Since $(\gamma \times \gamma')|_T$ is Lipschitz, $\gamma|_T$ is also Lipschitz. Thus U is γ-compliant and (by symmetry) γ'-compliant. □

If Z and Z' are compact, $S \subset C(X,Y)$ and $S' \subset C(X',Y')$, and $\gamma : S \to C(Z,W)$ and $\gamma' : S' \to C(Z',W')$ are operators, then $\gamma \otimes \gamma' : S \otimes S' \to C(Z,W) \otimes C(Z',W')$ is the operator given by $\gamma \otimes \gamma'(f \otimes f') = \gamma(f) \otimes \gamma'(f')$.

Proposition A2. *If Z, and Z' are compact, $S \subset C(X,Y)$ and $S' \subset C(X',Y')$ are nonempty, and $\gamma : S \to C(Z,W)$ and $\gamma' : S' \to C(Z',W')$ are tame operators, then $\gamma \otimes \gamma'$ is tame.*

Proof. Fix $f_0 \otimes f_0' \in S \otimes S'$. Let $U \subset S$ be a γ-compliant neighborhood of f_0, and let $U' \subset S'$ be a γ'-compliant of f_0'. Then $U \otimes U'$ is a neighborhood of $f \otimes f'$ in $S \otimes S'$. Let $T_0 \subset U \otimes U'$ be Lipschitz bounded, let $T = \{\, f : f \otimes f' \in T_0 \,\}$ and $T' = \{\, f' : f \otimes f' \in T_0 \,\}$, and let $T_1 = \{\, f \otimes f' : f \in T, f' \in T' \,\}$. There is a compact $K_0 \subset Y \times Y'$ that contains the image of every $f \otimes f' \in T_0$. Let K and K' be the projections of K_0 on Y and Y'. Let $\Lambda_0 > 0$ be such that every element of T_0 is Λ_0-Lipschitz. Then every element of T is Λ_0-Lipschitz with image contained in K, and similarly for T', so T and T' are Lipschitz bounded. Consequently $\gamma(T)$ and $\gamma'(T')$ are Lipschitz bounded and $\gamma|_T$ and $\gamma'|_{T'}$ are Lipschitz. It follows immediately that $\gamma \otimes \gamma'(T_1)$ is Lipschitz bounded and $\gamma \otimes \gamma'|_{T_1}$ is Lipschitz. Since $T_0 \subset T_1$, $\gamma \otimes \gamma'(T_0)$ is Lipschitz bounded and $\gamma \otimes \gamma'|_{T_0}$ is Lipschitz. Thus $U \otimes U'$ is $\gamma \otimes \gamma'$-compliant. □

The difficulty in proving the converse of the last result (which we do not need) arises as follows. Assume that $\gamma \otimes \gamma'$ is tame. Fixing $f_0 \otimes f_0' \in S \otimes S'$, let $\tilde{U} \subset S \otimes S'$ be a $\gamma \otimes \gamma'$-compliant neighborhood of $f_0 \otimes f_0'$. We may assume that $\tilde{U} = U \otimes U'$ where U and U' are neighborhoods of f_0 and f_0'. Let $T \subset U$ be Lipschitz bounded. If $T' \subset U'$ was Lipschitz bounded, then $T \otimes T'$ would be Lipschitz bounded and $\gamma \otimes \gamma'|_{T \otimes T'}$ would be Lipschitz, and we could infer from this that T was Lipschitz bounded and $\gamma|_T$ was Lipschitz. In particular, if U' necessarily contained a Lipschitz element f', then we could take $T' = \{f'\}$, but it seems that some such additional assumption is required.

Lemma A7. *If Y is compact, then the operator $\gamma : f \otimes g \mapsto g \circ f$ from $C(X,Y) \otimes C(Y,Z)$ to $C(X,Z)$ is tame.*

Proof. Let $T \subset C(X,Y) \otimes C(Y,Z)$ be Lipschitz bounded. There is a compact subset of $Y \times Z$ containing the image of every element of T; let K be its projection on Z. There is a $\Lambda > 0$ such that every $f \otimes g \in T$ is Λ-Lipschitz, which implies that f and g are Λ-Lipschitz. We can now see that for every element of $\gamma(T)$ is Λ^2-Lipschitz and its image is contained in K, so $\gamma(T)$ is Lipschitz bounded.

For any $f \otimes g, f' \otimes g' \in T$, the distance from $g \circ f$ to $g' \circ f'$ is not greater than the distance from $g \circ f$ to $g' \circ f$ plus the distance from $g' \circ f$ to $g' \circ f'$. The distance from $g \circ f$ to $g' \circ f$ is not greater than the distance from g to g', which is not greater than the distance from $f \otimes g$ to $f' \otimes g'$. The distance from $g' \circ f$ to $g' \circ f'$ is not greater than Λ times the distance from f to f', and the distance from f to f' is not greater than the distance from $f \otimes g$ to $f' \otimes g'$. Therefore $\gamma|_T$ is $(1+\Lambda)$-Lipschitz. We have shown that $C(X,Y) \otimes C(Y,Z)$ is γ-compliant. □

Lemma A8. *If $g : Y \to Z$ is locally Lipschitz, then the operator $\gamma_g : C(X,Y) \to C(X,Z)$ given by $\gamma_g(f) = g \circ f$ is tame.*

Proof. Let $T \subset C(X,Y)$ be Lipschitz bounded. If $K \subset Y$ is a compact set such that $f(X) \subset K$ for all $f \in T$, then $g(K)$ is compact and contains the image of $\gamma_g(f)$ for all $f \in T$. In addition, Lemma 2.1 implies that $g|_K$ is Λ'-Lipchitz for some Λ'. If every $f \in T$ is Λ-Lipschitz, then every element of $\gamma_g(T)$ is $\Lambda'\Lambda$-Lipschitz. Thus $\gamma_g(T)$ is Lipschitz bounded. Evidently $\gamma_g|_T$ is Λ'-Lipschitz. Thus $C(X,Y)$ is γ_g-compliant. □

Lemma A9. *$\gamma : f \otimes g \mapsto f$ is a tame operator from $C(X,Y) \otimes C(X',Y')$ to $C(X,Y)$.*

Proof. If $T \subset C(X,Y) \otimes C(X',Y')$ is Lipschitz bounded, then there is a compact set K containing the image of each of its elements and a $\Lambda > 0$ such that each of its elements is Λ-Lipschitz. For each $f \otimes g \in T$, f is Λ-Lipschitz and its image is contained in the projection of K on Y, so $\gamma(T)$ is Lipschitz bounded. Of course $\gamma|_T : f \otimes g \mapsto g$ is 1-Lipschitz. Thus $C(X,Y) \otimes C(X',Y')$ is γ-compliant. □

The following result is crucial.

Proposition A3. *If V, W, and X are finite dimensional inner product spaces, and $C \subset V$ and $D \subset W$ are compact sets that are the closures of their interiors, then the operator*

$$\gamma : \Delta^r C^r_{\mathrm{w}}(C,D) \otimes \Delta^r C^r_{\mathrm{w}}(D,X) \to \Delta^r C^r_{\mathrm{w}}(C,X)$$

given by $\gamma(\Delta^r f \otimes \Delta^r g) = \Delta^r(g \circ f)$ is tame.

Proof. From Proposition 2.1 we have $\Delta^r(g \circ f) = \tau^r \circ (\Delta^r f \otimes \Delta^r g)$, and Lemma A8 implies that the operator $\Delta^r f \otimes \Delta^r g \mapsto \tau^r \circ (\Delta^r f \otimes \Delta^r g)$ is tame. □

The proof of Proposition 3.1 applies many of the results above.

Proof of Proposition 3.1. Recall that $Q(\omega,a) = \sum_y q(y|\omega,a) \cdot \delta_{\beta(\omega,a,y)}$, so that

$$K_Q(V)(\omega,a) = \sum_y q(y|\omega,a) \cdot V(\beta(\omega,a,y)).$$

Since compositions of weakly $C^{r,1}$ functions are weakly $C^{r,1}$ (Lemma 2.3) it is clear that the functions $q(y|\cdot,\cdot)$ and $\beta(\cdot,\cdot,y)$ are $C^{r,1}$. In each case applying Lemma A7 (compositions of tame operators are tame) we have the following sequence of observations. Lemma A6 implies that for each $y \in Y$ the operator $\Delta^r V \mapsto \Delta^r V \otimes \Delta^r \beta(\cdot,\cdot,y)$ is tame, after which Proposition A3 implies that the operator $\Delta^r V \mapsto \Delta^r(V \circ \beta(\cdot,\cdot,y))$ is tame. Now Lemma A5 implies that the operator $\Delta^r V \mapsto \Delta^r q(y|\cdot,\cdot) \times \Delta^r(V \circ \beta(\cdot,\cdot,y))$ is tame. If μ is multiplication of two real numbers, then application of Lemma A6 and Proposition 2.1 implies that the operator

$$\Delta^r V \mapsto \Delta^r\big(\mu \circ (q(y|\cdot,\cdot) \times (V \circ \beta(\cdot,\cdot,y)))\big) = \Delta^r\big(q(y|\cdot,\cdot) \cdot (V \circ \beta(\cdot,\cdot,y))\big)$$

is tame. Finally, similar reasoning with addition in place of multiplication shows implies the tameness of the operator

$$\Delta^r V \mapsto \Delta^r\big(\sum_y q(y|\cdot,\cdot) \cdot (V \circ \beta(\cdot,\cdot,y))\big) = \Delta^r K_Q(V) = \mathcal{K}_Q(\Delta^r V).$$ □

A2 The Proof

We return to the setting provided by the hypotheses of Theorem 1. That is, V and W are finite dimensional inner product spaces, $\Omega \subset V$ and $A \subset W$ are compact subsets of V and W that are the closures of their interiors, $r \geq 2$, $u_0 \in C^{r,1}(\Omega \times A)$ satisfies the standard conditions, and $Q : \Omega \times A \to \Delta(\Omega)$ is r^{th} order smoothing. Let $\pi_0 : \Omega \to A$ be the myopic optimal policy function, so for each ω the unique maximizer of $u_0(\omega, \cdot)$ is $\pi_0(\omega)$.

We fix a compact neighborhood $Y \subset \Omega \times A$ of the graph of π_0. For each ω let

$$Y_\omega = \{\, a : (\omega, a) \in Y \,\}$$

be the "ω-slice" of Y. We require that Y is small enough that $\partial_a^2 u_0(\omega, a)$ is negative definite for all $(\omega, a) \in Y$, and that for each ω, $\partial_a u_0(\omega, a) \neq 0$ for all $a \in Y_\omega \setminus \{\pi_0(\omega)\}$.

Let U be the set of $u \in C^r_{\mathrm{w}}(\Omega \times A)$ such that:

(a) $\partial_a^2 u_0(\omega, a)$ is negative definite for all $(\omega, a) \in Y$.

(b) For each ω:

 (i) There is a unique maximizer of $u(\omega, \cdot)$ which is in the interior of Y_ω.

 (ii) $\partial_a u_0(\omega, a) \neq 0$ for all $a \in Y_\omega$ other than this maximizer.

Of course $u_0 \in U$.

Lemma A10. *U is open in $C^r_{\mathrm{w}}(\Omega \times A)$.*

In order to give a high level overview of the proof of Theorem 1 we defer the proofs of the supporting results until after that argument has been stated.

Let W^* be the dual of W, and let D be the set of C^{r-1} functions $d : Y \to W^*$ such that for each ω there is a unique $a \in Y_\omega$ such that $d(\omega, a) = 0$, a is in the interior of Y_ω, and $\partial_a d(\omega, a)$ is nonsingular. Let $P : U \to D$ be the operator given by

$$P(u)(\omega, a) = \partial_a u(\omega, a).$$

Let S be the set of $\pi \in C^{r-1}(\Omega, A)$ such that for each ω, $\pi(\omega)$ is in the interior of Y_ω. For $d \in D$ we define $\pi_d : \Omega \to A$ implicitly by requiring that $d(\omega, \pi_d(\omega)) = 0$. The implicit function theorem implies that $\pi_d \in C^{r-1}_{\mathrm{w}}(\Omega, A)$, so $\pi_d \in S$. Let $F : D \to S$ be the operator

$$F(d) = \pi_d.$$

In the following result tameness is with respect to relatively coarse norms of $C^r_{\mathrm{w}}(\Omega \times A, W^*)$ and $C^r_{\mathrm{w}}(\Omega, A)$, but even so the proof is relatively long and challenging.

Proposition A4. *F is locally Lipschitz and tame.*

Let B be the set of $u \otimes \pi \in U \otimes S$ such that $\partial_a u(\omega, \pi(\omega)) = 0$ for all ω. Let $H : B \to C^{r-1}_{\mathrm{w}}(\Omega)$ be the operator

$$H(u \otimes \pi) = u \circ (\mathrm{Id}_\Omega \times \pi).$$

Lemma A11. $H(B) \subset C^r_{\mathrm{w}}(\Omega)$.

When $u \in U$ is given, $\hat{u}$ denotes $\Delta^r u$, when $d \in D$ is given, $\hat{d}$ denotes $\Delta^{r-1} d$, and when $\pi \in S$ is given, $\hat{\pi}$ denotes $\Delta^{r-1}\pi$. If $\hat{u}$, $\hat{d}$, or $\hat{\pi}$ is given and there is no risk of confusion, u, d, or π will be the corresponding elements of U, D, or S. Let:

$$\mathcal{U} = \{\, \hat{u} : u \in U \,\};\ \ \mathcal{D} = \{\, \hat{d} : d \in D \,\};\ \ \mathcal{S} = \{\, \hat{\pi} : \pi \in S \,\};\ \ \mathcal{B} = \{\, \hat{u} \otimes \hat{\pi} : u \otimes \pi \in B \,\}.$$

Let $\mathcal{P} : \mathcal{U} \to \mathcal{D}$, $\mathcal{F} : \mathcal{D} \to \mathcal{S}$, $\mathcal{H} : \mathcal{B} \to \Delta^r(C^{r,1}(\Omega))$, and $\mathcal{G} : \mathcal{U} \to \Delta^r(C^r(\Omega \times A))$ be the operators:

$$\mathcal{P}(\hat{u}) = \Delta^{r-1}(P(u)); \quad \mathcal{F}(\hat{d}) = \Delta^{r-1}(F(d)); \quad \mathcal{H}(\hat{u} \otimes \hat{\pi}) = \Delta^r H(u \otimes \pi);$$

$$\mathcal{G}(\hat{u}) = \hat{u}_0 + \mathcal{K}_Q(\mathcal{H}(\hat{u} \otimes \mathcal{F}(\mathcal{P}(\hat{u})))).$$

Proposition A5. *$\mathcal{P}$ is tame.*

Proposition A6. *$\mathcal{F}$ is tame and locally Lipschitz.*

Proposition A7. *$\mathcal{H}$ is tame.*

Proposition A8. *$\mathcal{G}$ is tame.*

Proof of Theorem 1. Since U is open in $C^r_{\mathrm{w}}(\Omega \times A)$ and the norm of $\Delta^r C^r_{\mathrm{w}}(\Omega \times A)$ induces a finer topology, $\mathcal{U}$ is an open subset of $\Delta^r C^r_{\mathrm{w}}(\Omega \times A)$. Since $\mathcal{G}$ is tame, there is a $\mathcal{G}$-compliant neighborhood $\mathcal{W} \subset \mathcal{U}$ of $\hat{u}_0$. In a metric space every neighborhood of a point contains a closed neighborhood, so we may assume that $\mathcal{W}$ is closed.

Of course the image of $\hat{u}_0$ is bounded, so (in view of the definition of the norm of $\Delta^r C^r_{\mathrm{w}}(\Omega \times A)$) there is a neighborhood $\mathcal{T}'$ of $\hat{u}_0$ such that there is a compact neighborhood of the image of $\hat{u}_0$ that contains the image of every element of $\mathcal{T}'$. The closure of $\mathcal{T}'$ also has this property, so we may assume that $\mathcal{T}'$ is closed.

Choose $\Lambda > 0$ such that $\hat{u}_0$ is strictly Λ-Lipschitz. Let $\mathcal{U}^\Lambda$ be the set of $\hat{u} \in \mathcal{U}$ that are Λ-Lipschitz. The limit of a uniformly convergent sequence of Λ-Lipschitz functions is Λ-Lipschitz, so $\mathcal{U}^\Lambda$ is a closed subset of $\mathcal{U}$. Let

$$\mathcal{T} = \mathcal{W} \cap \mathcal{T}' \cap \mathcal{U}^\Lambda.$$

Since $\mathcal{T}$ is a closed subset of a Banach space, it is complete.

Of course $\mathcal{T}$ is a Lipschitz bounded subset of $\mathcal{W}$, so $\mathcal{G}(\mathcal{T})$ is Lipschitz bounded. Therefore there is a compact neighborhood of the image of $\hat{u}_0$ that contains the image of every element of $\mathcal{G}(\mathcal{T})$, and there are $\kappa, \mu > 0$ such that every element of $\mathcal{G}(\mathcal{T})$ is κ-Lipschitz and $\mathcal{G}|_{\mathcal{T}}$ is μ-Lipschitz. For sufficiently small $\varepsilon > 0$ it is the case, for all $\delta \in (-\varepsilon, \varepsilon)$, that $\hat{u}_0 + \delta\mathcal{G}(\mathcal{T}) \subset \mathcal{T}'$ and $\hat{u}_0 + \delta\mathcal{G}(\mathcal{T}) \subset \mathcal{U}^\Lambda$ (because $\hat{u}_0$ is $(\Lambda - \varepsilon\kappa)$-Lipschitz) so $\hat{u}_0 + \delta\mathcal{G}(\mathcal{T}) \subset \mathcal{T}$. In addition the map $(\hat{u}, \delta) \mapsto \hat{u}_0 + \delta\mathcal{G}(\hat{u})$ from $\mathcal{T} \times (-\varepsilon, \varepsilon)$ to $\mathcal{T}$ is $(1 + \mu)$-Lipschitz, and for each $\delta \in (-\varepsilon, \varepsilon)$ the map $\hat{u}_0 + \delta\mathcal{G}(\cdot) : \mathcal{T} \to \mathcal{T}$ is $\varepsilon\mu$-Lipschitz. If $\varepsilon\mu < 1$, Lemma 2.2 implies that each $\hat{u}_0 + \delta\mathcal{G}(\cdot)$ has a unique fixed point $\hat{u}_\delta$ that is a Lipschitz function of δ.

We now have $u_0 + \delta K_Q(H(u_\delta, F(P(u_\delta)))) = u_\delta$ and $J(u_\delta) = H(u_\delta, F(P(u_\delta)))$, so $I_\delta(u_\delta) = u_\delta$, which is to say that u_δ is the refactored value function. In addition, since $\mathcal{P}$, $\mathcal{H}$, and $\mathcal{F}$ are tame, by replacing ε with a smaller number if need be we can make $\hat{\pi}_\delta = \mathcal{F}(\mathcal{P}(\hat{u}_\delta))$ and $\hat{V}_\delta = \mathcal{H}(\hat{u}_\delta, \hat{\pi}_\delta)$ Lipschitz functions of δ. $\square$

We now give the proofs of the supporting results.

Proof of Lemma A10. We will show that a given $u \in U$ has a neighborhood $U' \subset C^r_{\mathrm{w}}(\Omega \times A)$ that is contained in U. For $\omega \in \Omega$ let $\pi(\omega)$ be the point in A where $u(\omega, \cdot)$ attains its maximum. Lemma 2.5 implies that π is C^{r-1}. There is a continuous $\beta : \Omega \to (0, \infty)$ such that for all ω the open $\beta(\omega)$-ball centered at $\pi(\omega)$ is contained in Y_ω. Let $\varepsilon > 0$ be small enough that $u(\omega, a) < u(\omega, \pi(\omega)) - \varepsilon$ and $\|\partial_a u(\omega, a)\| > \varepsilon$ for all ω and $a \in Y_\omega$ outside the open $\beta(\omega)$-ball centered at $\pi(\omega)$. Let U' be the set of $u' \in C^r_{\mathrm{w}}(\Omega \times A)$ such that $\partial_a^2 u'(\omega, a)$ is negative definite and $|u'(\omega, a) - u'(\omega, a)| < \varepsilon/2$ for all $(\omega, a) \in Y$. Evidently U' is an open subset of $C^r_{\mathrm{w}}(\Omega \times A)$. For $u' \in U'$ and $\omega \in \Omega$, any maximizer of $u(\omega, \cdot)$ must be in the ball of radius $\beta(\omega)$ centered at $\pi(\omega)$. If there were two such maximizers then the line segment between them would have a minimizer (relative to the line segment) of $u'(\omega, \cdot)$, say a, and $\partial_a^2 u'(\omega, a)$ could not be negative definite. Thus there is a unique maximizer. We have $\partial_a u'(\omega, a) \neq 0$ for all ω and $a \in Y_\omega$ outside the open $\beta(\omega)$-ball centered at $\pi(\omega)$. If $a \in Y_\omega$ is in the open $\beta(\omega)$-ball centered at $\pi(\omega)$, but is not the maximizer a^* of $u'(\omega, \cdot)$, then integration of $\partial_{aa} u'(\omega, \cdot)(a - a^*)$ along the line segment between a and a^* shows that $\partial_a u'(\omega, a)(a - a^*) \neq 0$. Thus $U' \subset U$. $\square$

Proof of Proposition A4. We first show that F is continuous. Aiming at a contradiction, suppose that $\{d_i\}$ is a sequence in D converging to $d \in D$ and there is an $\varepsilon > 0$ such that for each i there is some ω_i such that $\|\pi_{d_i}(\omega_i) - \pi_d(\omega_i)\| > \varepsilon$. Passing to a subsequence, assume that $\omega_i \to \omega$ and $\pi_{d_i}(\omega_i) \to a$. By continuity, $\pi_d(\omega_i) \to \pi_d(\omega)$, so

$\|a - \pi_d(\omega)\| \geq \varepsilon$. Since d_i converges to d uniformly, $d(\omega, a) = \lim d(\omega_i, \pi_{d_i}(\omega_i)) = 0$, contradicting the assumption that $d(\omega, \cdot)^{-1}(0)$ is a singleton.

We can now show that F is locally Lipschitz. Fix $d \in D$. We will work with an $\varepsilon > 0$ and a convex neighborhood $Z \subset D$ of d such that:

(a) a is in the interior of Y_ω whenever $\|a - \pi_d(\omega)\| \leq \varepsilon$.

(b) There is some $b > 0$ such that $\|\partial_a d'(\omega, a)w\| \geq b\|w\|$ for all $d' \in Z$, $w \in W$, $\omega \in \Omega$, and $a \in A$ such that $\|a - \pi_d(\omega)\| \leq \varepsilon$.

(c) There is some $c \in (0, b)$ such that $\|\partial_a d'(\omega, a) - \partial_a d'(\omega, a')\| < c$ for all $d' \in Z$, $\omega \in \Omega$, and $a, a' \in A$ such that $\|a - \pi_d(\omega)\|, \|a' - \pi_d(\omega)\| \leq \varepsilon$.

Such objects exist: if ε is sufficiently small, then (a) holds, and (b) and (c) hold if Z is a sufficiently small neighborhood of d.

For $d_0, d_1 \in Z$ and $\omega \in \Omega$, if $c : [0, 1] \to \mathbb{R}^n$ is the path $c(t) = (1 - t)\pi_{d_0}(\omega) + t\pi_{d_1}(\omega)$, then

$$\begin{aligned} d_0(\omega, \pi_{d_1}(\omega)) &= \int_0^1 \partial_a d_0(\omega, c(t)) \cdot (\pi_{d_1}(\omega) - \pi_{d_0}(\omega))\, dt \\ &= \partial_a d_0(\omega, \pi_{d_0}(\omega)) \cdot (\pi_{d_1}(\omega) - \pi_{d_0}(\omega)) \\ &\qquad + \int_0^1 \big(\partial_a d_0(\omega, c(t)) - \partial_a d_0(\omega, \pi_{d_0}(\omega))\big) \cdot (\pi_1(\omega) - \pi_{d_0}(\omega))\, dt. \end{aligned}$$

We can now apply $d_1(\omega, \pi_{d_1}(\omega)) = 0$, (b), and (c) to obtain

$$\begin{aligned} \|d_1 - d_0\| \geq \|d_1(\omega, \pi_{d_1}(\omega)) - d_0(\omega, \pi_{d_1}(\omega))\| &\geq \big\|\partial_a d_0(\omega, \pi_{d_0}(\omega)) \cdot (\pi_{d_1}(\omega) - \pi_{d_0}(\omega))\big\| \\ &\quad - \Big\| \int_0^1 \big(\partial_a d_0(\omega, c(t)) - \partial_a d_0(\omega, \pi_{d_0}(\omega))\big) \cdot (\pi_{d_1}(\omega) - \pi_{d_0}(\omega))\, dt \Big\| \\ &\geq (b - c)\|\pi_{d_1}(\omega)) - \pi_{d_0}(\omega)\|, \end{aligned}$$

so $\|\pi_{d_1} - \pi_{d_0}\| \leq \frac{1}{b-c}\|d_1 - d_0\|$.

We now show that F is tame. Fix $\delta \in (0, \varepsilon)$. Possibly after replacing Z with a smaller neighborhood of d, in addition to (a), (b), and (c) we may assume that $\|\pi_{d'} - \pi_d\| < \delta$ for all $d' \in Z$. We wish to show that Z is F-compliant. Suppose that $T \subset Z$ is Lipschitz bounded. We have already shown that $F|_T$ is Lipschitz, and our set up constrains the images of the functions $\pi_{d'}$ to lie in a compact set, so it remains to show that there is $\Lambda' > 0$ such that every element of $F(T)$ is Λ'-Lipschitz. We will show that there is a constant $C > 0$ such that if $d' \in Z$ is Λ-Lipschitz, then $\pi_{d'}$ is $C\Lambda$-Lipschitz.

Fix $\omega \in \Omega$. Since π_d is continuous there is a number $\zeta > 0$ small enough that $\delta + \|\pi_d(\omega_1) - \pi_d(\omega_0)\| < \varepsilon$ for all ω_0 and ω_1 in the open ζ-ball centered at ω. If $d' \in Z$ and $c : [0,1] \to A$ is the path $c(t) = (1-t)\pi_{d'}(\omega_0) + t\pi_{d'}(\omega_1)$, then

$$\begin{aligned} d'(\omega_0, \pi_{d'}(\omega_1)) &= \int_0^1 \partial_a d'(\omega_0, c(t)) \cdot (\pi_{d'}(\omega_1) - \pi_{d'}(\omega_0))\, dt \\ &= \partial_a d'(\omega_0, \pi_{d'}(\omega_0)) \cdot (\pi_{d'}(\omega_1) - \pi_{d'}(\omega_0)) \\ &\qquad + \int_0^1 (\partial_a d'(\omega_0, c(t)) - \partial_a d'(\omega_0, \pi_{d'}(\omega_0)))\, dt \cdot (\pi_{d'}(\omega_1) - \pi_{d'}(\omega_0)). \end{aligned}$$

Now (b) gives

$$\|\partial_a d'_\ell(\omega_0, \pi_{d'}(\omega_0)) \cdot (\pi_{d'}(\omega_1) - \pi_{d'}(\omega_0))\| \geq b\|\pi_{d'}(\omega_1) - \pi_{d'}(\omega_0)\|.$$

We have $\|\pi_{d'}(\omega_0) - \pi_d(\omega_0)\| \leq \delta$ and thus $\|\pi_{d'}(\omega_1) - \pi_d(\omega_0)\| \leq \varepsilon$, so $\|c(t) - \pi_d(\omega_0)\| \leq \varepsilon$ for all t, and we can apply (c) to the second term, obtaining

$$\|\partial_a d'(\omega_0, c(t)) - \partial_a d'(\omega_0, \pi_{d'}(\omega_0))\| < c.$$

Therefore

$$\|d'(\omega_0, \pi_{d'}(\omega_1)) - d'(\omega_1, \pi_{d'}(\omega_1))\| = \|d'(\omega_0, \pi_{d'}(\omega_1))\| \geq (b - c) \cdot \|\pi_{d'}(\omega_1) - \pi_{d'}(\omega_0)\|.$$

Now the argument follows the logic of the proof of Lemma 2.1. Since Ω is compact there are $\omega_1, \ldots, \omega_k$ and $\zeta_1, \ldots, \zeta_k$ such that the open balls around ω_i of radius ζ_i cover Ω, and if $d' \in Z$ is Λ-Lipschitz, then the restriction of $\pi_{d'}$ to each ball of radius ζ_i centered at ω_i is $\Lambda/(b-c)$-Lipschitz. The Lebesgue number lemma gives an $\alpha > 0$ such that for all $\omega_0, \omega_1 \in \Omega$, if $\|\omega_1 - \omega_0\| < \alpha$, then for some i the ball of radius ζ_i centered at ω_i contains both ω_0 and ω_1. Let $M = \max_{\omega_0, \omega_1 \in \Omega} \|\omega_1 - \omega_0\|$. We now see that if $d' \in Z$ is Λ-Lipschitz and $\|\omega_1 - \omega_0\| \geq \alpha$, then $\|\pi_{d'}(\omega_1) - \pi_{d'}(\omega_0)\| \leq M\Lambda \leq (M\Lambda/\alpha)\|\omega_1 - \omega_0\|$, so $\pi_{d'}$ is $\max\{\frac{1}{b-c}, \frac{M}{\alpha}\}\Lambda$-Lipschitz. ∎

Proof of Lemma A11. Suppose that $u \otimes \pi \in B$. On the interior of Ω the envelope theorem gives $D(H(u \otimes \pi))(\omega) = \partial_\omega u(\omega, \pi(\omega))$. Thus, on the interior of Ω, $DH(u \otimes \pi)$ is a composition of C^{r-1} functions, so it is C^{r-1}, and it extends continuously to all of Ω. Thus $H(u \otimes \pi)$ is C^r on the interior of Ω, and its derivatives up to order r extend continuously to all of Ω. ∎

Proof of Proposition A5. For $u \in U$ the passage from $\hat{u}$ to $\Delta^{r-1}P(u)$ is (in the language of partial derivatives) a matter of projecting away those partials of u that are only with respect to the components of ω (including u itself) and then restricting to Y. In our formalism the first process is a matter of discarding u and restricting the other multilinear functions representing the various derivatives of u to subspaces of their domains. These operations can be understood as composition with a projection, which is 1-Lipschitz, so this step is tame by Lemma A8. Lemma A3 implies that the second step is tame, so $\mathcal{P}$ can be understood as a composition of tame operators, which is tame by Lemma A4. $\square$

Proof of Proposition A6. Having shown that F is tame, in view of Proposition A1 it now suffices to show that each of the operators $d \mapsto D^s\pi_d$ $(s = 1, \dots, r-1)$ is tame and locally Lipschitz. Fix $d \in D$. A consequence of the implicit function theorem is the formula

$$D\pi_d(\omega) = -\partial_a d(\omega, \pi_d(\omega))^{-1} \cdot \partial_\omega d(\omega, \pi_d(\omega)),$$

which is derived by totally differentiating the formula $d(\omega, \pi_d(\omega)) = 0$ and solving for $D\pi_d(\omega)$. Let $f(L) = L^{-1}$ be the function mapping a nonsingular linear transformation to its inverse. Differentiating the formula $L \cdot f(L) = I$ gives the formula

$$Df(L)V = -L^{-1} \cdot V \cdot L^{-1}$$

for all linear $L, V : \mathbb{R}^n \to \mathbb{R}^n$ with L invertible. In particular, f is a locally Lipschitz function, and a formula for $D^s\pi_d(\omega)$ as a polynomial function of $\partial_a d(\omega, \pi_d(\omega))^{-1}$ and the various derivatives of d can be obtained by differentiating the formula above $s-1$ times, then substituting the previously obtained formulas for the lower order derivatives of π_d. For $\hat{d}'$ in a sufficiently small neighborhood of $\hat{d}$, there are consequently Lipschitz bounds on how rapidly $D^s\pi_{d'}$ varies as $\hat{d}'$ varies, so $\mathcal{F}$ is locally Lipschitz. Furthermore, Lipschitz bounded subsets of a sufficiently small neighborhood of $\hat{d}$ are mapped to Lipschitz bounded sets by the operator $\hat{d}' \mapsto D^s\pi_{d'}$. Thus $\mathcal{F}$ is tame. $\square$

Proof of Proposition A7. We have $\Delta^r H(u \otimes \pi) = H(u \otimes \pi) \times \Delta^{r-1} DH(u \otimes \pi)$, so Proposition A1 implies that $\mathcal{H}$ is tame if $\hat{u} \otimes \hat{\pi} \mapsto H(u \otimes \pi)$ and $\hat{u} \otimes \hat{\pi} \mapsto \Delta^{r-1} DH(u \otimes \pi)$ are tame. We show these in turn.

Since the projection $J^r(V, W) \to W$ onto the first component is linear, hence Lipschitz, $\hat{u} \mapsto u$ and $\hat{\pi} \mapsto \pi$ are tame (Lemma A8). Thus (Proposition A2) $\hat{u} \otimes \hat{\pi} \mapsto u \otimes \pi$ is tame. The projections $u \otimes \pi \mapsto u$ and $u \otimes \pi \mapsto \pi$ are tame (Lemma A9) and the constant operator $u \otimes \pi \mapsto \mathrm{Id}_\Omega$ is tame (Lemma A2). Now $u \otimes \pi \mapsto \mathrm{Id}_\Omega \times \pi$, $u \otimes \pi \mapsto u \otimes (\mathrm{Id}_\Omega \times \pi)$, and $u \otimes \pi \mapsto u \circ (\mathrm{Id}_\Omega \times \pi) = H(u \otimes \pi)$ are tame (Proposition A1 and Lemmas A9 and A7 respectively). Thus (Lemma A4) $\hat{u} \otimes \hat{\pi} \mapsto H(u \otimes \pi)$ is tame.

By the envelope theorem

$$DH(u \otimes \pi) = (\partial_\omega u + \partial_a u \circ D\pi) \circ (\mathrm{Id}_\Omega \times \pi) = \partial_\omega u \circ (\mathrm{Id}_\Omega \times \pi),$$

so (Lemma A4 and Proposition A3) $\hat{u} \otimes \hat{\pi} \mapsto \Delta^{r-1} DH(u \otimes \pi)$ is tame if $\hat{u} \otimes \hat{\pi} \mapsto \Delta^{r-1}\partial_\omega u \otimes \Delta^{r-1}(\mathrm{Id}_\Omega \times \pi)$ is tame. By virtue of Proposition A2 it suffices to show that $\hat{u} \otimes \hat{\pi} \mapsto \Delta^{r-1}\partial_\omega u$ and $\hat{u} \otimes \hat{\pi} \mapsto \Delta^{r-1}(\mathrm{Id}_\Omega \times \pi)$ are tame. Lemma A9 implies that $\hat{u} \otimes \hat{\pi} \mapsto \hat{u}$ and $\hat{u} \otimes \hat{\pi} \mapsto \hat{\pi}$ are tame, so (Lemma A4) it suffices to show that $\hat{u} \mapsto \Delta^{r-1}\partial_\omega u$ and $\hat{\pi} \mapsto \Delta^{r-1}(\mathrm{Id}_\Omega \times \pi)$ are tame. The first of these is tame because the passage from $\hat{u}$ to $\Delta^{r-1}\partial_\omega u$ is composition with a linear projection, which is Lipschitz (Lemma A8).

The constant and identity operators $\hat{\pi} \mapsto \Delta^{r-1}\mathrm{Id}_\Omega$ and $\hat{\pi} \mapsto \hat{\pi}$ are tame (Lemma A2 and Corollary A.1) so (Proposition A1) $\hat{\pi} \mapsto \Delta^{r-1}\mathrm{Id}_\Omega \times \hat{\pi}$ is tame. The passage from $\Delta^{r-1}\mathrm{Id}_\Omega \times \hat{\pi}$ to $\Delta^{r-1}(\mathrm{Id}_\Omega \times \pi)$ is tame because it is accomplished by applying the linear function $J^{r-1}(V,V) \times J^{r-1}(V,W) \to J^{r-1}(V, V \times W)$ that takes the sum of the various components (Lemma A8). Thus (Lemma A4) $\hat{\pi} \mapsto \Delta^{r-1}(\mathrm{Id}_\Omega \times \pi)$ is tame. □

Proof of Proposition A8. We now know that $\mathcal{P}$, $\mathcal{F}$, and $\mathcal{H}$ are tame, and $\mathcal{K}_Q$ is tame by assumption. Therefore Lemma A4 implies that $\hat{u} \mapsto \mathcal{F}(\mathcal{P}(\hat{u}))$ is tame. Combining the tameness of the identity operator (Corollary A.1) and Proposition A2 shows that $\hat{u} \mapsto \hat{u} \otimes \mathcal{F}(\mathcal{P}(\hat{u}))$ is tame, and therefore (again by Lemma A4) $\hat{u} \mapsto \mathcal{K}_Q(\hat{u} \otimes \mathcal{F}(\mathcal{P}(\hat{u})))$ is tame. By assumption $\hat{u}_0$ is Lipschitz, so (Lemma A2) $\hat{u} \mapsto \hat{u}_0$ is tame, and applications of Proposition A2 and Lemma A8 (with addition as the Lipschitz function) imply that $\hat{u} \mapsto \hat{u}_0 + \mathcal{K}_Q(\hat{u} \otimes \mathcal{F}(\mathcal{P}(\hat{u}))) = \mathcal{G}(\hat{u})$ is tame. □

B Dynamic Programming on Manifolds

This appendix presents and proves Theorem B1, which generalizes Theorem 1 to dynamic programs in which the sets of states and actions are compact subsets of suitably smooth manifolds and are the closures of their interiors. The analysis is largely a matter of transporting the results of Appendix A to the more general setting, so that and Section 2 (but not Sections 3–7) are prerequisites.

B1 Smooth Manifolds

In this subsection we present basic definitions for C^r manifolds, and differential calculus for C^r functions between manifolds. These definitions should be understood as pertaining equally to $C^{r,1}$ manifolds and $C^{r,1}$ functions between then, insofar as the discussion is valid (with indicated modifications) if we replace C^r with $C^{r,1}$ everywhere.

A set $M \subset \mathbb{R}^k$ is an m-dimensional C^r *manifold*[9] if, for each $x \in M$, there are open

[9]A more principled definition of a smooth manifold specifies that it is a topological space that is covered by open sets with coordinate charts such that the "transition function" defined by any two such charts is smooth. Having manifolds be subsets of Euclidean spaces has various conveniences. The

sets $U \subset M$ and $V \subset \mathbb{R}^m$ such that $x \in U$ and there is a C^r *parameterization* $\psi : V \to U$, by which we mean that ψ is a C^r homeomorphism between V and U, and $D\psi(y)$ has rank m for all $y \in V$. (Here $\mathbb{R}^k$ is called the *ambient space* of M.) Standard constructions imply that ψ^{-1} is C^r, in the sense of having a C^r extension to a neighborhood of U in $\mathbb{R}^k$. We say that ψ^{-1} is a C^r *coordinate chart* for M.

The *tangent space* T_xM of a point $x \in U$ is the image $D\psi(\varphi(x))$, which is m-dimensional because $D\psi(\varphi(x))$ has full rank. (The chain rule easily implies that T_xM does not depend on the choice of φ.) Let $N \subset \mathbb{R}^\ell$ be an n-dimensional C^r manifold, let $f : M \to N$ be a C^r function, and let W be a neighborhood (in $\mathbb{R}^k$) of M for which there is a C^r function $\tilde{f} : W \to \mathbb{R}^\ell$ such that $\tilde{f}|_M = f$. The standard definition of the derivative of f at $x \in M$ is $Df(x) = D\tilde{f}(x)|_{T_xM}$, and we will have some use for this notion. However, for distinct $x, x' \in M$, $J^r(T_xM, \mathbb{R}^\ell)$ and $J^r(T_{x'}M, \mathbb{R}^\ell)$ are different spaces, and there is as yet no notion of distance between $\Delta^r f(x)$ and $\Delta^r f(x')$, so we take a different approach.

Let $\mathrm{Gr}_{k-m}(\mathbb{R}^k)$ be the Grassmann manifold of $(k-m)$-dimensional linear subspaces of $\mathbb{R}^k$. In standard and rather obvious ways this can be given the structure of a C^∞ manifolds. We endow the manifold M with a C^∞ function $\nu_M : M \to \mathrm{Gr}_{k-m}(\mathbb{R}^k)$ such that $T_xM + \mu_M(x) = \mathbb{R}^k$. We also endow M with a neighborhood $W_M \subset \mathbb{R}^k$ and continuous functions $\pi_M : W \to M$ and $\varpi_M : W \to \mathbb{R}^k$ such that $\pi_M(x) = x$ for all $x \in M$ and $\varpi_M(w) \in \nu_M(\pi_M(w))$ for all $w \in W$. Standard constructions based on the inverse function theorem imply the existence of such objects, and that π_M is C^r. For $x \in M$ we define

$$\tilde{\Delta}^r f = \Delta^r(f \circ \pi_M).$$

We say that f is $C^{r,1}$ if M and N are $C^{r,1}$ and $\tilde{\Delta}^r f$ is locally Lipschitz.

Lemma B1. *If (in addition to M and N) $P \subset \mathbb{R}^h$ is a p-dimensional $C^{r,1}$ manifold and $f : M \to N$ and $g : N \to P$ are C^r, then, for all $x \in M$,*

$$\tilde{\Delta}^r(g \circ f)(x) = \tau^r_{\mathbb{R}^k,\mathbb{R}^\ell,\mathbb{R}^h}(\tilde{\Delta}^r f(x), \tilde{\Delta}^r g(f(x))).$$

Proof. Using Proposition 2.1, we compute that

$$\tilde{\Delta}^r(g \circ f)(x) = \Delta^r(g \circ f \circ \pi_M)(x) = \Delta^r((g \circ \pi_N) \circ (f \circ \pi_M))(x)$$

$$= \tau^r_{\mathbb{R}^k,\mathbb{R}^\ell,\mathbb{R}^h}(\Delta^r(f \circ \pi_M)(x), \Delta^r(g \circ \pi_N)(f(x)))$$

$$= \tau^r_{\mathbb{R}^k,\mathbb{R}^\ell,\mathbb{R}^h}(\tilde{\Delta}^r f(x), \tilde{\Delta}^r g(f(x))).$$

□

easy Whitney embedding theorem (e.g., p. 24 of Hirsch (1976)) implies that restricting attention to submanifolds of Euclidean spaces is almost entirely without loss of generality.

Let C be a compact subset of M that is the closure of its interior. We will say that a function $f : C \to N$ is *weakly* C^r if it continuous, the restriction f° of f to the interior of C is C^r in the usual sense, and $\tilde{\Delta}^r f^\circ$ has a continuous extension to C, in which case, as before, we abuse notation slightly by letting $\tilde{\Delta}^r f$ denote this extension. For $T \subset N$ let $C^r_{\mathrm{w}}(C, T)$ be the set of such functions whose images are contained in T. As before, and we say that $f \in C^r_{\mathrm{w}}(C, T)$ is $C^{r,1}$ if $\tilde{\Delta}^r f$ is Lipschitz. Let $C^{r,1}_{\mathrm{w}}(C, T)$ be the set of such functions. We write $C^r_{\mathrm{w}}(C)$ and $C^{r,1}_{\mathrm{w}}(C)$ in place of $C^r_{\mathrm{w}}(C, \mathbb{R})$ and $C^{r,1}_{\mathrm{w}}(C, \mathbb{R})$.

We now establish the analogue of Proposition 2.2. That is, for a sequence of weakly C^r functions, if the associated sequences of derivatives converge, then the limiting function is weakly C^r and its derivatives are the limits of the associated sequences of derivatives.

Proposition B1. *If $\{f_i\}$ is a sequence in $C^r_{\mathrm{w}}(C, N)$ with $\tilde{\Delta}^r f_i \to g \in C(C, J^r(\mathbb{R}^k, \mathbb{R}^\ell))$, so that $f_i \to f \in C(C, N)$, then $f \in C^r_{\mathrm{w}}(C, N)$ and $\tilde{\Delta}^r(f) = g$.*

Proof. We first consider the special case in which C is contained in an open $U \subset M$ for which there is a C^r coordinate chart $\varphi : U \to \mathbb{R}^m$, and $f(C)$ is contained in an open $V \subset N$ for which there is a C^r coordinate chart $\psi : V \to \mathbb{R}^n$. For each i we have

$$\tilde{\Delta}^r(\psi \circ f_i \circ \varphi^{-1})(u) = \tau^r(\tilde{\Delta}^r \varphi^{-1}(u), \tilde{\Delta}^r f_i(\varphi^{-1}(u)), \tilde{\Delta}^r \psi(f_i(\varphi^{-1}(u)))).$$

for all u in the interior of $\varphi(C)$, and by continuity it holds for all u in $\varphi(C)$. (It is easy to see that the logic of Proposition 2.1 extends to the composition of any number of C^r functions.) Therefore

$$\tilde{\Delta}^r(\psi \circ f_i \circ \varphi^{-1})(u) \to \tau^r(\tilde{\Delta}^r \varphi^{-1}(u), g, \tilde{\Delta}^r \psi(f_i(\varphi^{-1}(u)))),$$

and Proposition 2.2 implies that $\psi \circ f \circ \varphi^{-1}$ is C^r on the interior of $\varphi(C)$, and that

$$\tilde{\Delta}^r(\psi \circ f \circ \varphi^{-1})(u) = \tau^r(\tilde{\Delta}^r \varphi^{-1}(u), g, \tilde{\Delta}^r \psi(f(\varphi^{-1}(u)))).$$

It follows that f is C^r on the interior of C, and that

$$\tilde{\Delta}^r f(x) = \tau^r\Big(\tilde{\Delta}^r \varphi(x), \tau^r\big(\tilde{\Delta}^r \varphi^{-1}(\varphi(x)), g(x), \tilde{\Delta}^r \psi(f(x))\big), \tilde{\Delta}^r \psi^{-1}(\psi(f(x)))\Big) = g(x)$$

for all $x \in C$.

To establish the general case we simply observe that C can be covered by finitely many compact subsets that are the closures of their interiors, that are contained in the domains of coordinate charts for M, and whose images under f are contained in the domains of coordinate charts for N. □

We note that $\tilde{\Delta}^r C^r(C)$ is a closed subspace of the Banach space $C(C, J^r(\mathbb{R}^k))$.

Lemma B2. *Let M and N be $C^{r,1}$ manifolds, and let $C \subset M$ be a compact set that is the closure of its interior. Let M' be a $C^{r,1}$ manifold, let $h : M' \to M$ be a $C^{r,1}$ diffeomorphism*[10]*, and let $C' = h^{-1}(C)$, Then the operator $\gamma_h : \tilde{\Delta}^r C^{r,1}(C, N) \to \tilde{\Delta}^r C^{r,1}(C', N)$ given by $\gamma_h(\tilde{\Delta}^r f) = \Delta^r(f \circ h|_{C'})$ is tame.*

Proof. Since h is $C^{r,1}$, $h|_{C'}$ and $\tilde{\Delta}^r h|_{C'}$ are Lipschitz, so Lemma A2 implies that the operators $\tilde{\Delta}^r f \mapsto h|_{C'}$ and $\tilde{\Delta}^r f \mapsto \tilde{\Delta}^r h_{C'}$ are tame. Corollary A.1 implies that $\tilde{\Delta}^r f \mapsto \tilde{\Delta}^r f$ is tame. Now Lemma A2 implies that $\tilde{\Delta}^r f \mapsto h|_{C'} \otimes \tilde{\Delta}^r f$ is tame, and Lemma A7 implies that $\tilde{\Delta}^r f \mapsto \tilde{\Delta}^r f \circ h|_{C'}$ is tame, so Proposition A1 implies that $\tilde{\Delta}^r f \mapsto \tilde{\Delta}^r h \times \tilde{\Delta}^r f \circ h|_{C'}$ is tame. We have $\tilde{\Delta}^r(f \circ h)(x') = \tau^r(\tilde{\Delta}^r h(x'), \tilde{\Delta}^r f(h(x')))$ for all x in the interior of C', and by continuity this formula also holds for boundary points, so Lemma A8 implies the claim. (Various appeals to the composition of tame operators being tame (Lemma A4) have been omitted.) □

The proof of the following result is quite similar, so we leave it to the reader.

Lemma B3. *Let M and N be $C^{r,1}$ manifolds, and let $C \subset M$ be a compact set that is the closure of its interior. Let N' be a $C^{r,1}$ manifold, and let $h : N \to N'$ be a $C^{r,1}$ diffeomorphism. Then the operator $\eta_h : \tilde{\Delta}^r C^{r,1}(C, N) \to \tilde{\Delta}^r C^{r,1}(C, N)'$ given by $\eta_h(\tilde{\Delta}^r f) = \Delta^r(h \circ f)$ is tame.*

B2 Smooth Dynamic Programs

We return to the dynamic programming setting. Fix an $r \geq 2$, and assume now that $M \subset \mathbb{R}^k$ and $N \subset \mathbb{R}^\ell$ are m- and n-dimensional $C^{r,1}$ manifolds, and that $\Omega \subset M$ and $A \subset N$ are compact sets that are the closures of their interiors, and there is a given transition function $Q : \Omega \times A \to \Delta(\Omega)$.

For $u \in C^r_{\mathrm{w}}(\Omega \times A)$ we usually write $\hat{u}$ in place of $\tilde{\Delta}^r u$, for $V \in C^r_{\mathrm{w}}(\Omega)$ we usually write $\hat{V}$ in place of $\tilde{\Delta}^r V$, and for $\pi \in C^{r-1}_{\mathrm{w}}(\Omega, A)$ we usually write $\hat{\pi}$ in place of $\tilde{\Delta}^{r-1}\pi$. If $\hat{u}$, $\hat{V}$, or $\hat{\pi}$ is given, then u, V, or π denotes the corresponding element of $C^r_{\mathrm{w}}(\Omega \times A)$, $C^r_{\mathrm{w}}(\Omega)$, or $C^{r-1}_{\mathrm{w}}(\Omega, A)$.

For $V \in C(\Omega)$ let $K_Q(V) : \Omega \times A \to \mathbb{R}$ be the function

$$K_Q(V)(\omega, a) = \int_\Omega V(\omega')\, Q(\omega, a; d\omega').$$

The transition function Q is said to be r^{th} *order smoothing* if $\tilde{K}_Q(C^r_{\mathrm{w}}(\Omega)) \subset C^r_{\mathrm{w}}(\Omega \times A)$ and $\tilde{\mathcal{K}}_Q$ is tame, where $\tilde{\mathcal{K}}_Q : \tilde{\Delta}^r(C^r_{\mathrm{w}}(\Omega)) \to \tilde{\Delta}^r(C^r_{\mathrm{w}}(\Omega \times A))$ is the operator given by $\tilde{\mathcal{K}}_Q(\hat{V}) = \tilde{\Delta}^r(\tilde{K}_Q(V))$.

[10]That is, h is a bijection and h and h^{-1} are $C^{r,1}$ functions.

We say that u *satisfies the standard conditions* if, for each $\omega \in \Omega$, there is a unique maximizer $\pi_{(u,0)}(\omega)$ of $u(\omega, \cdot)$, $\pi_{(u,0)}(\omega)$ is in the interior of A, and $\partial_{aa} u(\omega, \pi_{(u,0)})$ is negative definite. We call $\pi_{(u,0)}$ the *myopically optimal policy* of u. If $u \in C^r_{\mathrm{w}}(\Omega \times A)$ satisfies the standard conditions, then the logic of Lemma 2.5 easily generalizes to imply that $\pi_{(u,0)} \in C^{r-1}_{\mathrm{w}}(\Omega, A)$, and if, in addition, $u \in C^{r,1}_{\mathrm{w}}(\Omega \times A)$, then $\pi_{(u,0)} \in C^{r-1,1}_{\mathrm{w}}(\Omega, A)$.

The following result is the culmination of our work on dynamic programming. In addition to generalizing Theorem 1 to the manifold setting, it also shows how the per period reward can be taken as a parameter with respect to which the value function and optimal policy are Lipschitz functions.

Theorem B1. *Assume Q is r^{th} order smoothing, $u_0 \in C^{r,1}_{\mathrm{w}}(\Omega \times A)$ satisfies the standard conditions, $\hat{u}_0$ is strictly Λ-Lipschitz, and $\alpha > 0$. There is a neighborhood B of $\hat{u}_0$ in*

$$\{\, \hat{u} \in \Delta^r(\Omega \times A) : \hat{u} \text{ is } \Lambda\text{-Lipschitz}\}$$

and $\varepsilon > 0$ such that:

(a) For each $(\hat{u}, \delta) \in B \times (-\varepsilon, \varepsilon)$:

(i) u satisfies the standard conditions;

(ii) The discounted dynamic program with payoff u and discount factor δ has value and refactored value functions $V_{(u,\delta)} \in C^{r,1}_{\mathrm{w}}(\Omega)$ and $u_{(u,\delta)} \in C^{r,1}_{\mathrm{w}}(\Omega, A)$ and a unique stationary optimal policy $\pi_{(u,\delta)} \in C^{r-1,1}_{\mathrm{w}}(\Omega, A)$.

(b) The maps $(\hat{u}, \delta) \mapsto \hat{u}_{(u,\delta)}$, $(\hat{u}, \delta) \mapsto \hat{V}_{(u,\delta)}$, and $(\hat{u}, \delta) \mapsto \hat{\pi}_{(u,\delta)}$ are Lipschitz.

In its overall outline the proof follows the path of the proof of Theorem 1. We define three operators by generalizing our earlier definitions of P, F, and H, and the derived extensions of $\mathcal{P}$, $\mathcal{F}$, and $\mathcal{H}$. Composing an operator with a restriction operator can give an operator whose action is local, in the sense of being confined to the domain of a single coordinate chart. In the proofs of the analogues of Propositions A5, A6, and A7 the following result will be used to transfer the tameness of $\mathcal{P}$, $\mathcal{F}$, and $\mathcal{H}$ to the manifold setting.

Proposition B2. *Let X, X', Y, and Y' be metric spaces with X and X' compact, and let $\gamma : S \to C(X', Y')$ be an operator, where $S \subset C(X, Y)$. If $C_1, \ldots, C_k$ are compact subsets of X' whose interiors cover X' and $\rho_{C_j} \circ \gamma$ is tame for all j, then γ is tame.*

As in the body of the paper, we defer the proofs of the supporting results until after the main argument has been stated.

Let π_0 be the myopically optimal policy for u_0. Let $Y \subset \Omega \times A$ be a compact neighborhood of the graph of π_0. For each $\omega \in \Omega$ let

$$Y_\omega = \{ a \in A : (\omega, a) \in Y \}.$$

We require that Y is small enough that $\partial_{aa} u_0(\omega, a)$ is negative definite for all $(\omega, a) \in Y$ and, for each ω, $\partial_a u_0(\omega, a) \neq 0$ for all $a \in Y_\omega \setminus \{\pi_0(\omega)\}$.

Let $\tilde{U}$ be the set of $u \in C^r_{\mathrm{w}}(\Omega \times A)$ such that $\partial_{aa} u_0(\omega, a)$ is negative definite for all $(\omega, a) \in Y$ and, for each ω, there is a unique maximizer of $u(\omega, \cdot)$ which is in the interior of Y_ω, and $\partial_a u_0(\omega, a) \neq 0$ for all $a \in Y_\omega$ other than this maximizer. Of course $u_0 \in \tilde{U}$. Let $\tilde{\mathcal{U}} = \{ \hat{u} : u \in \tilde{U} \}$. For any $\Lambda' > 0$ let $\tilde{\mathcal{U}}^{\Lambda'}$ be the set of $\hat{u} \in \tilde{\mathcal{U}}$ that are Λ'-Lipschitz.

Lemma B4. *U is open in $C^r_{\mathrm{w}}(\Omega \times A)$.*

Let $\tilde{E} = \{ (x, y, \ell) : x \in M, y \in N, \ell \in T_y N^* \}$. (Here $T_y N^*$ is the dual of $T_y N$.) To see that $\tilde{E}$ is a $C^{r-1,1}$ manifold suppose that $\varphi : U \to \mathbb{R}^m$ and $\psi : V \to \mathbb{R}^n$ are C^r coordinate charts for open $U \subset M$ and $V \subset N$, and let

$$\zeta : \{ (x, y, \ell) \in \tilde{E} : (x, y) \in U \times V \} \to \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^{n*}$$

be the function $\zeta(x, y, \ell) = (\varphi(x), \psi(y), \ell \circ D\psi(y)^{-1})$. Evidently ζ is a C^{r-1} coordinate chart for $\tilde{E}$, and $\tilde{E}$ is covered by the domains of such charts. For $(x, y, \ell) \in \tilde{E}$ let $\varrho_{(x,y)}(x, y, \ell) = (x, y)$ and $\varrho_\ell(x, y, \ell) = \ell$.

Let $\tilde{D}$ be the set of C^{r-1} functions $d : Y \to \tilde{E}$ such that $\varrho_{(x,y)}(d(\omega, a)) = (\omega, a)$ and $\partial_a(\varrho_\ell \circ d)(\omega, a)$ is negative definite for all $(\omega, a) \in Y$, and for each ω there is a unique $a_\omega \in Y_\omega$ such that $d(\omega, a) = 0$, which is in the interior of Y_ω. For $d \in \tilde{D}$ we usually write $\hat{d}$ in place of $\tilde{\Delta}^{r-1} d$. Let $\tilde{P} : \tilde{U} \to \tilde{D}$ be the operator given by

$$\tilde{P}(u)(\omega, a) = (\omega, a, \partial_a u(\omega, a)).$$

Let $\tilde{S}$ be the set of $\pi \in C^{r-1}(\Omega, A)$ such that for each ω, $\pi(\omega)$ is in the interior of Y_ω. Let $\tilde{F} : \tilde{D} \to C(\Omega, A)$ be defined implicitly by requiring that

$$\varrho_\ell(d(\omega, \tilde{F}(d)(\omega))) = 0.$$

Let $\tilde{B}$ be the set of $u \otimes \pi \in \tilde{U} \otimes \tilde{S}$ such that $\partial_a u(\omega, \pi(\omega)) = 0$ for all ω. Let $\tilde{H} : \tilde{B} \to C^{r-1}_{\mathrm{w}}(\Omega)$ be the operator given by

$$\tilde{H}(u \otimes \pi) = u \circ (\mathrm{Id}_\Omega \times \pi).$$

Lemma B5. *$\tilde{H}(\tilde{B}) \subset C^r_{\mathrm{w}}(\Omega)$.*

Let:

$$\tilde{\mathcal{U}} = \{\, \hat{u} : u \in \tilde{U}\,\}; \qquad \tilde{\mathcal{D}} = \{\, \hat{d} : d \in \tilde{D}\,\}; \qquad \tilde{\mathcal{S}} = \{\, \hat{\pi} : \pi \in \tilde{S}\,\};$$

$$\tilde{\mathcal{B}} = \{\, \hat{u} \times \hat{\pi} : u \otimes \pi \in \tilde{B}\,\}.$$

Let $\tilde{\mathcal{P}} : \tilde{\mathcal{U}} \to \tilde{\mathcal{D}}$, $\tilde{\mathcal{F}} : \tilde{\mathcal{D}} \to \tilde{\mathcal{S}}$, $\tilde{\mathcal{H}} : \tilde{\mathcal{B}} \to \tilde{\Delta}^r(C^{r,1}(\Omega))$, and $\tilde{\mathcal{G}} : \tilde{\mathcal{U}} \to \tilde{\Delta}^r(C^{r,1}(\Omega \times A))$ be the operators given by:

$$\tilde{\mathcal{P}}(\hat{u}) = \tilde{\Delta}^{r-1}(\tilde{P}(u)); \qquad \tilde{\mathcal{F}}(\hat{d}) = \tilde{\Delta}^{r-1}(\tilde{F}(d)); \qquad \tilde{\mathcal{H}}(\hat{u} \otimes \hat{\pi}) = \tilde{\Delta}^r \tilde{H}(u \otimes \pi);$$

$$\tilde{\mathcal{G}}(\hat{u}) = \tilde{\mathcal{K}}_Q(\tilde{\mathcal{H}}(\hat{u} \otimes \tilde{\mathcal{F}}(\tilde{\mathcal{P}}(\hat{u})))).$$

Proposition B3. *$\tilde{\mathcal{P}}$ is a tame operator.*

Proposition B4. *$\tilde{F}(\tilde{D}) \subset \tilde{S}$ and $\tilde{\mathcal{F}}$ is a tame operator.*

Proposition B5. *$\tilde{\mathcal{H}}$ is a tame operator.*

Proposition B6. *$\tilde{\mathcal{G}}$ is a tame operator.*

Proof of Theorem B1. Since $\tilde{U}$ is open in $C^r_{\mathrm{w}}(\Omega \times A)$ and the norm of $\Delta^r C^r_{\mathrm{w}}(\Omega \times A)$ induces a finer topology, $\tilde{\mathcal{U}}$ is an open subset of $\Delta^r C^r_{\mathrm{w}}(\Omega \times A)$. Since $\mathcal{G}$ is tame, there is a $\mathcal{G}$-compliant neighborhood $\mathcal{W} \subset \mathcal{U}$ of $\hat{u}_0$. In a metric space every neighborhood of a point contains a closed neighborhood, so we may assume that $\mathcal{W}$ is closed.

The image of $\hat{u}_0$ is bounded, so there is a neighborhood $\mathcal{T}' \subset \mathcal{W}$ of $\hat{u}_0$ such that there is a compact set that contains the image of every element of $\mathcal{T}'$. The closure of $\mathcal{T}'$ also has this property, so we may assume that $\mathcal{T}'$ is closed.

Let $\tilde{\mathcal{U}}^\Lambda$ be the set of $\hat{u} \in \tilde{\mathcal{U}}$ that are Λ-Lipschitz. Since the limit of a uniformly convergent sequence of Λ-Lipschitz functions is Λ-Lipschitz, Proposition B1 implies that $\tilde{\mathcal{U}}^\Lambda$ is a complete metric space. Let $\tilde{\mathcal{T}} = \tilde{\mathcal{T}}' \cap \tilde{\mathcal{U}}^\Lambda$. Of course $\tilde{\mathcal{T}}$ is Lipschitz bounded, and since it is a closed subset of a complete metric space, it is complete.

Choose $\gamma > 0$ such that the closed ball $\mathbf{B}_{2\gamma}(\hat{u}_0)$ of radius 2γ centered at $\hat{u}$ is contained in $\tilde{\mathcal{T}}'$. Choose $\alpha > 0$ such that $\hat{u}_0$ is $(\Lambda - 2\alpha)$-Lipschitz. Let B be the set of elements of $\mathbf{B}_\gamma(\hat{u}_0)$ that are $(\Lambda - \alpha)$-Lipschitz. Since $\tilde{\mathcal{T}}$ is Lipschitz bounded, $\tilde{\mathcal{G}}(\tilde{\mathcal{T}})$ is Lipschitz bounded, so there is a compact set that contains the image of every element of $\tilde{\mathcal{G}}(\tilde{\mathcal{T}})$. For sufficiently small $\varepsilon > 0$ it is the case that $\delta\tilde{\mathcal{G}}(\tilde{\mathcal{T}}) \subset \mathbf{B}_\gamma(0)$ for all $\delta \in (-\varepsilon, \varepsilon)$. In addition, for some $\kappa, \mu > 0$ every element of $\tilde{\mathcal{G}}(\tilde{\mathcal{T}})$ is κ-Lipschitz and $\tilde{\mathcal{G}}|_{\tilde{\mathcal{T}}}$ is μ-Lipschitz. For sufficiently small $\varepsilon > 0$ every element of $\delta\tilde{\mathcal{G}}(\tilde{\mathcal{T}})$ is α-Lipschitz for all $\delta \in (-\varepsilon, \varepsilon)$. Choose $\varepsilon \in (0, 1/\mu)$ such that both these conditions hold.

Let $G : \tilde{\mathcal{T}} \times B \times (-\varepsilon, \varepsilon) \to \tilde{\mathcal{T}}$ be the function

$$G(\hat{u}', \hat{u}, \delta) = \hat{u} + \delta\tilde{\mathcal{G}}(\hat{u}').$$

The conditions developed above imply that G is Lipschitz and for each $(\hat{u}, \delta) \in B \times (-\varepsilon, \varepsilon)$, $G(\cdot, \hat{u}, \delta)$ is $\mu\varepsilon$-Lipschitz, hence a contraction. Applying Lemma 2.2, for each $(\hat{u}, \delta) \in B \times (-\varepsilon', \varepsilon')$ there is a unique fixed point $\hat{u}_{(u,\delta)} \in \tilde{\mathcal{T}}$ of $G(\cdot, \hat{u}, \delta)$, and $\hat{u}_{(u,\delta)}$ is a Lipschitz function of $(\hat{u}, \delta)$.

We now have

$$u + \delta K_Q(\tilde{H}(u_{(u,\delta)}, \tilde{F}(\tilde{P}(u_{(u,\delta)})))) = u_{(u,\delta)}$$

and

$$J(u_{(u,\delta)}) = \tilde{H}(u_{(u,\delta)}, \tilde{F}(\tilde{P}(u_{(u,\delta)}))),$$

so $I_\delta(u_{(u,\delta)}) = u_{(u,\delta)}$, which is to say that $u_{(u,\delta)}$ is the refactored value function for the problem with per period reward u and discount factor δ.

Let $\hat{\pi}_{(u,\delta)} = \tilde{\mathcal{F}}(\tilde{\mathcal{P}}(\hat{u}_{(u,\delta)}))$ and $\hat{V}_{(u,\delta)} = \tilde{\mathcal{H}}(\hat{u}_{(u,\delta)} \otimes \hat{\pi}_{(u,\delta)})$. Since $\tilde{\mathcal{P}}$, $\tilde{\mathcal{F}}$, and $\tilde{\mathcal{H}}$ are tame, after restricting to a smaller neighborhood of $(\hat{u}_0, 0, \hat{u}_0)$, $\hat{\pi}_{(u,\delta)}$ and $\hat{V}_{(u,\delta)}$ are Lipschitz functions of (u, δ). □

We now prove the supporting results.

Proof of Proposition B2. Fix $f \in S$. Let $W \subset S$ be a neighborhood of f such that for each j, W is $(\rho_{C_j} \circ \gamma)$-compliant. We claim that W is γ-compliant. Let $T \subset W$ be Lipschitz bounded. For each j there is a compact $K_j \subset Y'$ that contains the image of $\rho_{C_j}(\gamma(f))$ for all $f \in T$. If we set $K = \bigcup_j K_j$, then K is compact and contains the image of $\gamma(f)$ for all $f \in T$. For each j there is a constant $\Lambda_j > 0$ such that for every $f \in T$, $\rho_{C_j}(\gamma(f))$ is Λ_j-Lipschitz. Let $M = \max_{y,y' \in K} d(y, y')$. The Lebesgue covering lemma gives an $\varepsilon > 0$ such that for all $x, x' \in X'$ such that $d(x, x') < \varepsilon$ there is a j such that $x, x' \in C_j$. Evidently every $\gamma(f) \in \gamma(T)$ is $\max\{\Lambda_1, \ldots, \Lambda_k, M/\varepsilon\}$-Lipschitz. Thus $\gamma(T)$ is Lipschitz bounded. For $f', f'' \in T$ we have

$$d(\gamma(f'), \gamma(f'')) = \max_{x' \in X'} d(\gamma(f')(x'), \gamma(f'')(x')) = \max_j \max_{x' \in C_j} d(\gamma(f')(x'), \gamma(f'')(x'))$$
$$= \max_j d(\rho_{C_j}(\gamma(f')), \rho_{C_j}(\gamma(f''))).$$

Since each $\rho_{C_j} \circ \gamma|_T$ is Lipschitz, so is $\gamma|_T$. □

The reader should have no difficulty passing from the proof of Lemma A10 to a proof of Lemma B4, so we do not provide a second version of this argument.

Proof of Proposition B3. It suffices to show that a given $\hat{u} \in \tilde{\mathcal{U}}$ has a neighborhood such that the restriction of $\tilde{\mathcal{P}}$ to that neighborhood of is tame. Let π be the myopically optimal policy for u. We cover the graph of π with compact sets $Y_1, \ldots, Y_k \subset Y$ such that, for each i:

(a) $Y_i = \Omega_i \times A_i$ is a cartesian product of compact subsets of Ω and A that are the closures of their interiors.

(b) For each $\omega \in \Omega_i$, $\pi(\omega)$ is in the interior of A_i.

(c) There are $C^{r,1}$ coordinate charts $\varphi_i : U_i \to \mathbb{R}^m$ and $\psi_i : V_i \to \mathbb{R}^n$ with $\Omega_i \subset U_i \subset M$ and $A_i \subset V_i \subset N$.

For each i let $\tilde{D}_i$ be the set of C^{r-1} functions $d : Y_i \to \tilde{E}$ such that $\varrho_{(x,y)}(d(\omega, a)) = (\omega, a)$ for all (ω, a), and let $\tilde{P}_i : U \to \tilde{D}_i$ be the operator $\tilde{P}_i(u) = \tilde{P}(u)|_{Y_i}$. Let $\tilde{\mathcal{P}}_i$ be the operator given by $\tilde{\mathcal{P}}_i(\hat{u}) = \tilde{\Delta}^{r-1}(\tilde{P}_i(u))$. Evidently $\tilde{\mathcal{P}}_i = \rho_{Y_i} \circ \tilde{\mathcal{P}}$, so if we can show that each $\tilde{\mathcal{P}}_i$ is tame, Proposition B2 will imply that $\tilde{\mathcal{P}}$ is tame. Now $\tilde{\mathcal{P}}_i = \tilde{\mathcal{P}}'_i \circ \rho_{Y_i}$ where $\tilde{\mathcal{P}}'_i$ has the same definition as $\tilde{\mathcal{P}}$, restricted to the subdomain, so it suffices to show that $\tilde{\mathcal{P}}'_i$ is tame. Let P_i and $\mathcal{P}_i$ be the analogous operators defined in Appendix A for the sets $\varphi_i(\Omega_i)$ and $\psi_i(A)$, and let $\zeta_i : (x, y, \ell) \mapsto (\varphi_i(x), \psi_i(y), \ell \circ D\psi_i(y)^{-1})$ be the C^{r-1} coordinate chart for $\tilde{E}$ we saw earlier. We now have $\tilde{P}_i(u') = P_i(u' \circ (\varphi_i \times \psi_i)^{-1}) \circ \zeta_i$ and thus

$$\tilde{\mathcal{P}}_i(\hat{u}') = \gamma_{\zeta_i}(\mathcal{P}_i(\gamma_{(\varphi_i \times \psi_i)^{-1}}(\hat{u}')))$$

and thus the claim follows from Proposition A5 and Lemma B2. □

Proof of Proposition B4. Fix a $d \in \tilde{D}$. It suffices to show that there is a neighborhood $\tilde{D}' \subset \tilde{D}$ such that $\tilde{F}(\tilde{D}') \subset \tilde{S}$ and the restriction of $\tilde{\mathcal{F}}$ to $\{\, \hat{d} : d \in \tilde{D}' \,\}$ is tame. Let $\pi_d = \tilde{F}(d)$. Let $Y_1 = \Omega_1 \times A_1, \ldots, Y_k = \Omega_k \times A_k$ be a compact cover of the graph of π_d with the properties enumerated in the last proof, so for each i we have $C^{r,1}$ coordinate charts $\varphi_i : U_i \to \mathbb{R}^m$ and $\psi_i : V_i \to \mathbb{R}^n$ as above. Let $\zeta_i : (x, y, \ell) \mapsto (\varphi_i(x), \psi_i(y), \ell \circ D\psi_i(y)^{-1})$ be the C^{r-1} coordinate chart for $\tilde{E}$ we saw earlier. For each i let $\tilde{F}_i : \tilde{D}' \to C(\Omega_i, A_{j_i})$ be the operator $\tilde{F}_i(d') = \tilde{F}(d')|_{\Omega_i}$, and let $\tilde{\mathcal{F}}_i$ be the operator given by $\tilde{\mathcal{F}}_i(\tilde{\Delta}^r d') = \tilde{\Delta}^{r-1}(\tilde{F}_i(d'))$. Evidently $\tilde{\mathcal{F}}_i = \rho_{\Omega_i} \circ \tilde{\mathcal{F}}$, so if we can show that each $\tilde{\mathcal{F}}_i$ is tame, Proposition B2 will imply that $\tilde{\mathcal{F}}$ is tame. Now $\tilde{\mathcal{F}}_i = \tilde{\mathcal{F}}'_i \circ \rho_{\Omega_i \times A_{j_i}}$ where $\tilde{\mathcal{F}}'_i$ has the same definition as $\tilde{\mathcal{F}}$, restricted to the subdomain, so it suffices to show that $\tilde{\mathcal{F}}'_i$ is tame. Let F_i and $\mathcal{F}_i$ be the analogous operators defined in Appendix A for the sets $\varphi_i(\Omega_i)$ and $\psi_i(A)$. Tracing through the definitions, we find that

$$\tilde{F}_i(d') = \psi_i^{-1} \circ F_i(\zeta_i \circ d' \circ (\varphi_i \otimes \psi_i)^{-1}) \circ \varphi_i.$$

Therefore

$$\tilde{\mathcal{F}}_i(\hat{d}') = \eta_{\psi_i^{-1}}(\gamma_{\varphi_i}(\mathcal{F}_i(\eta_{\zeta_i}(\gamma_{(\varphi_i \otimes \psi_i)^{-1}}(\hat{d}'))))),$$

and thus the claim follows from Proposition A5 and Lemma B2. □

Proof of Lemma B5. For $u \otimes \pi \in \tilde{B}$ the envelope theorem gives $D(\tilde{H}(u \otimes \pi))(\omega) = D(u(\cdot, \pi(\omega)))(\omega)$. Thus $D\tilde{H}(u \otimes \pi)$ is a composition of C^{r-1} functions, and is C^{r-1}, so $\tilde{H}(u \otimes \pi)$ is C^r. □

Proof of Proposition B5. Fix $u \otimes \pi \in \tilde{B}$. It suffices to show that there is a neighborhood $\tilde{B}' \subset \tilde{B}$ such that the restriction of $\tilde{\mathcal{H}}$ to $\{\, \hat{u}' \times \hat{\pi}' : u' \otimes \pi' \in \tilde{B}' \}$ is tame. Let $Y_1 = \Omega_1 \times A_1, \ldots, Y_k = \Omega_k \times A_k$ be a compact cover of the graph of π with the properties enumerated in the proof of Proposition B3, so for each i we have $C^{r,1}$ coordinate charts $\varphi_i : U_i \to \mathbb{R}^m$ and $\psi_i : V_i \to \mathbb{R}^n$ as above. For each i let $\tilde{B}_i$ be the set of $(u', \pi') \in \tilde{B}$ such that $\pi'(\Omega_i)$ is contained in the interior of A_{j_i}, and let $\tilde{H}_i : \tilde{B}' \to C^r(\Omega_i, A_i)$ be the operator $\tilde{H}_i(u', pi') = \tilde{H}(u', \pi')|_{\Omega_i}$. Let $\tilde{\mathcal{H}}_i$ be the operator given by $\tilde{\mathcal{H}}_i(\hat{u}' \otimes \hat{\pi}') = \tilde{\Delta}^r(\tilde{H}_i(u', \pi'))$. Evidently $\tilde{\mathcal{H}}_i = \rho_{\Omega_i} \circ \tilde{\mathcal{H}}$, so if we can show that each $\tilde{\mathcal{H}}_i$ is tame, Proposition B2 will imply that $\tilde{\mathcal{H}}$ is tame. Now $\tilde{\mathcal{H}}_i = \tilde{\mathcal{H}}'_i \circ (\rho_{\Omega_i \times A_{j_i}} \otimes \rho_{\Omega_i})$ where $\tilde{\mathcal{H}}'_i$ has the same definition as $\tilde{\mathcal{H}}$, restricted to the subdomain, so it suffices to show that $\tilde{\mathcal{H}}'_i$ is tame. Let H_i and $\mathcal{H}_i$ be the analogous operators defined in Appendix A for the sets $\varphi_i(\Omega_i)$ and $\psi_i(A)$. Tracing through the definitions, we find that

$$\tilde{H}_i(\hat{u}' \otimes \hat{\pi}') = H_i((u' \circ (\varphi_i \otimes \psi_i)^{-1}) \otimes (\psi_i \circ \pi' \circ \varphi_i^{-1})) \circ \varphi_i.$$

If $\theta_{\hat{u}'}$ and $\theta_{\hat{\pi}'}$ are the projections $\hat{u}' \otimes \hat{\pi}' \mapsto \hat{u}'$ and $\hat{u}' \otimes \hat{\pi}' \mapsto \hat{\pi}'$, then

$$\tilde{\mathcal{H}}_i = \gamma_{\varphi_i} \circ \mathcal{H}_i \circ ((\gamma_{(\varphi_i \otimes \psi_i)^{-1}} \circ \theta_{\hat{u}'}) \otimes (\eta_{\psi_i} \circ \gamma_{\varphi_i^{-1}} \circ \theta_{\hat{\pi}'}))).$$

Lemma A8 implies that $\theta_{\hat{u}'}$ and $\theta_{\hat{\pi}'}$ are tame, so Lemmas B2, B3 and Proposition A2, together with the tameness of compositions of tame operators (Lemma A4), imply that $\tilde{\mathcal{H}}_i$ is tame. □

Proof of Proposition A8. We now know that $\tilde{\mathcal{P}}$, $\tilde{\mathcal{F}}$, and $\tilde{\mathcal{H}}$ are tame, and $\tilde{\mathcal{K}}_Q$ is tame by assumption. Therefore Lemma A4 implies that $\hat{u} \mapsto \tilde{\mathcal{F}}(\tilde{\mathcal{P}}(\hat{u}))$ is tame. Combining the tameness of the identity operator (Corollary A.1) and Proposition A2 shows that $\hat{u} \mapsto \hat{u} \otimes \tilde{\mathcal{F}}(\tilde{\mathcal{P}}(\hat{u}))$ is tame, and therefore (again by Lemma A4) $\hat{u} \mapsto \tilde{\mathcal{K}}_Q(\hat{u} \otimes \tilde{\mathcal{F}}(\tilde{\mathcal{P}}(\hat{u}))) = \tilde{\mathcal{G}}(\hat{u})$ is tame. □

References

Acemoglu, D., Bimpikis, K., and Ozdaglar, A. (2014). Dynamics of information exchange in endogenous social networks. *Theoretical Economics*, 9:41–97.

Acemoglu, D., Dahleh, M., Lobel, I., and Ozdaglar, A. (2011). Bayesian learning in social networks. *Review of Economic Studies*, 78:1201–1236.

Aghion, P., Bolton, P., Harris, C., and Jullien, B. (1991). Optimal learning by experimentation. *Review of Economic Studies*, 58:621–654.

Ali, S. N. (2011). Learning self-control. *Quarterly Journal of Economics*, 126:857–893.

Aoyagi, M. (1998). Mutual observabiblity and the convergence of actions in a multi-person two-armed bandit model. *J. Econ. Theory*, 82:405–424.

Arieli, I. and Mueller-Frank, M. (2014). Multi-dimensional social learning. Working Paper.

Baker, S. and Mezzetti, C. (2012). A theory of rational jurisprudence. *J. Polit. Economy*, 120:513–551.

Banerjee, A. V. (1992). A simple model of herd behavior. *Quarterly Journal of Economics*, 107:73–88.

Banks, J. S. and Sundaram, R. K. (1992). Denumerable-armed bandits. *Econometrica*, 60:1071–1096.

Banks, J. S. and Sundaram, R. K. (1994). Switching costs and the Gittens index. *Econometrica*, 62:687–694.

Benveniste, L. M. and Scheinkman, J. A. (1979). On the differentiability of the value function in dynamic models of economics. *Econometrica*, 47:727–732.

Berentsen, A., Bruegger, E., and Loertscher, S. (2008). Learning, public goods provision and the information trap. *Journal of Public Economics*, 92:998–1010.

Bergemann, D. and Välimäki, J. (1996). Learning and strategic pricing. *Econometrica*, 64:1125–1149.

Bergemann, D. and Välimäki, J. (2008). Bandit problems. In Durlauf, S. N. and Blume, L. E., editors, *The New Palgrave Dictionary of Economics (Second Edition)*. Palgrave Macmillan, New York.

Berry, D. and Fristedt, B. (1985). *Bandit Problems: Sequential Allocation of Experiments*. Monographs on Statistics and Applied Probability. Chapman and Hall, London.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992). A theory of fads, fashion, custom, and cultural change in informational cascades. *J. Polit. Economy*, 100:992–1026.

Blume, L., Easley, D., and O'Hara, M. (1982). Characterization of optimal plans for stochastic dynamic programs. *J. Econ. Theory*, 28:221–234.

Bolton, P. and Harris, C. (1999). Strategic experimentation. *Econometrica*, 67:349–374.

Bolton, P. and Harris, C. (2000). Strategic experimentation: The undiscounted case. In Hammond, P. and Myles, G., editors, *Incentives, Organization and Public Economics—Essays in Honour of Sir James Mirlees*, pages 53–68. Oxford University Press, Oxford.

Breen, R. and Garcia-Penalosa, C. (2002). Bayesian learning and gender segregation. *Journal of Labor Economics*, 20:899–922.

Callander, S. (2011a). Searching and learning by trial and error. *Amer. Econ. Rev.*, 101:2277–2308.

Callander, S. (2011b). Searching for good policies. *American Political Science Review*, 105:643–662.

Çelen, B. and Kariv, S. (2004). Observational learning under imperfect information. *Games and Economic Behavior*, 47:72–86.

Décamps, J.-P. and Mariotti, T. (2004). Investment timing and learning externalities. *J. Econ. Theory*, 118:80–102.

Easley, D. and Kiefer, N. (1988). Controlling a stochastic process with unknown parameters. *Econometrica*, 56:1045–1064.

Fefferman, C. L. (2005). A sharp form of Whitney's extension theorem. *Ann. of Math.*, 161:509–577.

Fudenberg, D. and Levine, D. K. (1993). Self-confirming equilibrium. *Econometrica*, 61:523–545.

Gale, D. and Kariv, S. (2003). Bayesian learning in social networks. *Games and Economic Behavior*, 45:329–346.

Gale, D. and Kariv, S. (2008). Learning and information aggregation in networks. In Durlauf, S. N. and Blume, L. E., editors, *The New Palgrave Dictionary of Economics (Second Edition)*. Palgrave Macmillan, New York.

Gittens, J. and Jones, D. (1974). A dynamic allocation index for the sequential allocation of experiments. In Gani, J., editor, *Progress in Statistics*, pages 241–266. North Holland, Amsterdam.

Harrison, J.-M., Keskina, N.-B., and Zeevi, A. (2012). Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution. *Management Science*, 58:570–586.

Hirsch, M. (1976). *Differential Topology*. Springer-Verlag, New York.

Jung, H. M. (2018). Receiver's dilemma. *J. Math. Econom.*, 75:116–124.

Keller, G., Rady, S., and Cripps, M. (2005). Strategic experimentation with exponential bandits. *Econometrica*, 73:39–68.

Kihlstrom, R. E., Mirman, L. J., and Postlewaite, A. (1984). Experimental consumption and the 'Rothschild effect'. In Boyer, M. and Kihlstrom, R. E., editors, *Bayesian Models in Economic Theory*. Elsevier, Amsterdam.

Laslier, J.-F., Trannoy, A., and van der Straeten, K. (2003). Voting under ignorance of job skills of unemployed: the overtaxation bias. *Journal of Public Economics*, 87:595–626.

Li, H. (2001). A theory of conservatism. *J. Polit. Economy*, 109:617–636.

Ma, Q. and Stachurski, J. (2018). Dynamic programming deconstructed. working paper, Australian National University.

Malueg, D. A. and Tsutsui, S. (1997). Dynamic R&D competition with learning. *Rand Journal of Economics*, 28:751–772.

McLennan, A. (1984). Price dispersion and incomplete learning in the long run. *Journal of Economic Dynamics and Control*, 7:331–347.

Moscarini, G. and Smith, L. (2001). The optimal level of experimentation. *Econometrica*, 69:1629–1644.

Piketty, T. (1995). Social mobility and redistributive politics. *Quarterly Journal of Economics*, 110:551–584.

Radner, R. and Stiglitz, J. (1984). A nonconcavity in the value of information. In Boyer, M. and Kihlstrom, R., editors, *Bayesian Models of Economic Theory*, pages 33–52. Elsevier, Amsterdam.

Rothschild, M. (1974). A two-armed bandit theory of market pricing. *J. Econ. Theory*, 9:185–202.

Smith, L. and Sørensen, P. (2000). Pathological outcomes of observational learning. *Econometrica*, 68:371–398.

Smith, L. and Sørensen, P. (2011). Observational learning. In Durlauf, S. and Blume, L., editors, *The New Palgrave Dictionary of Economics Online Edition*, pages 29–52. Palgrave Macmillan, New York.

Strulovici, B. (2010). Learning while voting: Determinants of collective experimentation. *Econometrica*, 78:933–971.

Warren, P. L. and Wilkening, T. S. (2012). Regulatory fog. *Journal of Economic Behavior and Organization*, 84:840–856.

Whitney, H. (1934). Analytic extensions of differentiable functions defined in closed sets. *Trans. Am. Math. Soc.*, 36:63–89.