Nearly-Optimal Dynamic Programming and Behavioral Rules*

by

Dale O. Stahl

Malcolm Forsman Centennial Professor
Department of Economics
University of Texas at Austin
stahl@eco.utexas.edu

April 2, 2015

ABSTRACT

The standard dynamic programming approach to exact optimization of sequential decision problems is extended to allow near-optimal optimization. A nearly-optimal solution is a probabilistic decision rule that satisfies a modified Bellman equation. The existence of a solution to this equation is proven. Every such solution defines a behavioral rule. Conversely, given any behavioral rule, there exists a dynamic programming problem for which the behavioral rule is a nearly-optimal solution. In light of this equivalence result, we argue economics can and should dispense with the optimization interpretation of behavior.

Keywords: dynamic programming, near optimization, behavioral rules

JEL Classifications: B4, C61

---

*The author is indebted to George Malaith, Max Stinchcombe, Tom Wiseman and anonymous referees for criticisms and comments. All errors and interpretations are the sole responsibility of the author.

1. Introduction.

Everyday life presents us with a multitude of decision problems which we perceive and deal with sequentially.[1] Dynamic programming is a well-established solution method for optimization of sequential decision problems. The standard approach assumes exact optimization. However, there are many situations for which it is not reasonable to believe that humans employ exact solutions. In this paper, we extend the standard dynamic programming approach to allow near-optimization. A nearly-optimal solution is a probabilistic decision rule that satisfies a modified Bellman equation. The existence of a solution to this equation is proven. This extension increases the range of behaviors that could be rationalized as the outcome of attempts at optimization.

We further show equivalence between nearly-optimal solutions to dynamic programming problems and behavioral rules. Every nearly-optimal solution is a behavioral rule, and for every behavioral rule and dynamic programming problem, there exists a disturbance of the reward function such that the behavioral rule is a nearly-optimal solution to the modified problem.

This equivalence result raises the question of whether the optimization approach has greater merit than the behavioral approach. We argue that we can dispense with the former and progress towards models with greater out-of-sample predictive success.

2. The Framework.

For simplicity, we assume decisions occur at discrete times, indexed by $t \in \{1, 2, \ldots\}$. Let ω_t denote the state (or information available) at time t. Let Ω denote the state space: the set of all possible (information) states for all t. Part of this information is the set of feasible actions, $A_t \equiv \mathcal{A}(\omega_t)$, where $\mathcal{A}(\ )$ denotes the projection onto the action subspace. Let $A \equiv \mathcal{A}(\Omega)$ denote the subspace of all possible actions for all t.

A *decision rule* is a function, α, that maps each state ω_t into an element of A_t; i.e. $\alpha(\omega_t) \in A_t$ is selected. A decision rule is evaluated *ex ante* in terms of its consequences which can extend into the future. Let $r(a_t, \omega_t)$ denote the perceived immediate reward from action a_t in state ω_t. This reward function is assumed to be part of the available information ω_t.

[1] I am referring to conscious decision making, not autonomic (subconscious) decisions made continuously and in parallel to conscious decision making.

Anticipating future consequences requires a model of the world, i.e. a function that maps the current action and state into a probability measure on future states. For a discrete model, let $f(\omega_{t+1} \mid a_t, \omega_t)$ denote the probability of state ω_{t+1}, conditioned on (a_t, ω_t). This function is also assumed to be part of the available information ω_t.

In the standard approach, we assume the Decision Maker (DM) chooses a decision rule α that maximizes the normalized expected discounted present value of immediate rewards generated by α given initial state ω_1:

$$V(\alpha, \omega_1) \equiv (1-\beta)\{r[\alpha(\omega_1), \omega_1] + \sum_{t\geq 2} \beta^{t-1} \int r[\alpha(\omega_t), \omega_t] \boldsymbol{f}(\omega_t \mid \alpha, \omega_1) d\omega_t\}, \quad (1)$$

where $\beta \in [0, 1)$ is the discount factor[2], and for $t \geq 2$

$$\boldsymbol{f}(\omega_t \mid \alpha, \omega_1) \equiv \int \cdots \int \prod_{s=2}^{t} f[\omega_s \mid \alpha(\omega_{s-1}), \omega_{s-1}] \, d\omega_2 \cdots d\omega_{t-1} . \quad (2)$$

Thus, a dynamic programming problem is defined by a triplet (r, f, β) and eqs(1-2).

It is insightful to rewrite eq(1) as

$$V(\alpha, \omega_1) = (1-\beta) r[\alpha(\omega_1), \omega_1] + \beta \int V(\alpha, \omega_2) f[\omega_2 \mid \alpha(\omega_1), \omega_1] d\omega_2 . \quad (3)$$

In other words, the discounted present value of rewards can be expressed as the immediate reward plus the expected discounted value of future rewards given the decision rule and the probability distribution of the state in period 2 conditional on the current state and action.

Define

$$V^*(\omega_1) \equiv max_{\alpha} V(\alpha, \omega_1) \quad (4)$$

Then from eq(3-4):

[2] When the time periods are heterogeneous, we would replace β by $e^{-r\Delta t}$, where Δt is the period duration.

$$V^*(\omega_1) \;=\; max_a\,[(1-\beta)r(a, \omega_1) + \beta \int V^*(z)f(z \mid a, \omega_1)dz]\,. \qquad (5)$$

This is the well-known Bellman equation for dynamic programming [Bellman, 1954; Bertsekas, 2000]. It is also well-known that, provided r() is bounded, and A is compact, there is a unique function V*( ) satisfying eq(5). Associated with V*( ) is an *optimal decision rule* $\alpha^*(\,)$ such that $V[\alpha^*(\omega), \omega] = V^*(\omega)$ for all $\omega \in \Omega$. Although V*( ) is unique, there can be multiple optimal decision rules, so a complete solution to the dynamic programming problem requires the addition of a selection rule for the cases of multiple optimal decision rules. Also, when there are multiple optimal solutions, then any probabilistic rule whose support is confined to $arg\ max_\alpha$ $V(\alpha, \omega_1)$ is also an optimal solution. Given a dynamic programming problem (r, f, β), let **O**(r, f, β) denote the set of optimal decision rules (including all optimal probabilistic rules).

2.1. Nearly-optimal Decision Rules.

There are many reasons why this dynamic programming model is not a reliable approximation to human behavior. The first one we will consider is the requirement of *exact maximization* assumed in eqs(4-5). Since it is not unreasonable to assume that humans are imperfect maximizers, it is imperative that we consider imperfect maximization.

There are many ways to specify near optimization. For example, we could define ε-optimization in which ε provides a notion of nearness. In this vein, $\varepsilon = 0$ would indicate exact optimization. For our main result, it turns out that it suffices to consider logistic optimization in which the choice of action is given by a probabilistic function of the logistic form:

$$p(a_t|\omega_t, u; \nu) \;=\; \frac{\exp\{\nu\,[u(a_t, \omega_t)]\}}{\sum_{b \in A_t} \exp\{\nu\,[u(b, \omega_t)]\}}\,, \qquad (6)$$

where $\nu \geq 0$ is the precision[3], and

$$u(a_t, \omega_t) \;\equiv (1-\beta)\, r(a_t, \omega_t) + \beta \int_\Omega \left[\int_{\mathcal{A}(z)} u(b, z)p(b|z, u)db\right] f(z|a_t, \omega_t)dz \qquad (7)$$

[3] Note that as $\nu \to \infty$, $p(a|\, \omega_t, u; \nu) \to 0$ for all $a \notin arg\ max_a\, u(a, \omega_t)$.

is the expected normalized discounted value of current action a_t in state ω_t, assuming that all future action choices will be generated by eq(6-7). Of course, for this model to be logically consistent, the function u( ) must be a fixed point of eq(7).

Definition 2. $p(\bullet\,|\bullet, u^*; \nu)$ is a ***logistic solution*** to the dynamic programming problem (r, f, β) with precision ν, if it is specified by eq(6) and u^* is a fixed point of eq(7).

Note that by construction, a logistic solution is time-consistent: i.e. if $\omega_t = \omega_{t'}$, $p(\bullet|\omega_t, u^*; \nu) = p(\bullet|\omega_{t'}, u^*; \nu)$.

3. Existence of a Fixed Point.

Unfortunately, logistic decision rules do not guarantee that eq(7) is a contraction mapping. Therefore, we are not able to prove the existence of a unique fixed point. We will first consider the simple case when A and Ω are finite sets. Then we will address the more general case.

Suppose A and Ω are finite sets. Let N and M denote the number of actions and states respectively. Hence, r and u are elements of R^{N+M}. Assume r() is bounded[4]. Then, u can be confined to a compact, convex subset $K \subset R^{N+M}$. Given a regular finitely-precise decision rule, eq(7) defines a continuous map, call it T, from K into K. By Brower's fixed point theorem, there is $u^* \in K$ such that $T(u^*) = u^*$.

3.2 Existence for Compact Metric Spaces.[5]

The casual reader may want to skip to section 3.3. In this section it is convenient to drop the "t" subscript on actions and states.

Assumption 1. A and Ω are compact metric spaces.

[4] This assumption is invoked to prove the existence of a fixed point. The definition of logistic solutions does not invoke boundedness. On the other hand, if a logistic solution exists, then $r(\bullet,\omega_t)$ is at least integrable.

[5] The author is indebted to Max Stinchcombe for guidance in the proof of this result.

Let $\mathcal{F} \equiv \{ f(\cdot \mid a, \omega) \mid (a, \omega) \in A \times \Omega \}$ denote the family of probability density functions on Ω as a function of $A \times \Omega$. For any event $E \subset \Omega$, $\int_E f(z \mid a, \omega) dz$ gives the probability of the state transitioning to E, given current action and state (a, ω).

Let $\Phi \subset \mathcal{F}$ denote the subset of probability densities with the property that for each $\varphi \in \Phi$, for every $\varepsilon > 0$ there is a $\delta > 0$, such that $\|(a, \omega) - (a', \omega')\| < \delta$ implies that $\int_\Omega |\varphi(z \mid a, \omega) - \varphi(z \mid a', \omega')| \, dz < \varepsilon$. Φ is the set of uniformly continuous transition functions.

Assumption 2. The state transition function $f \in \Phi$.

Let $B_1(A \times \Omega)$ denote the family of real-valued functions on $A \times \Omega$ with sup norm not exceeding 1, and let $C_1(A \times \Omega)$ denote the subset of $B_1(A \times \Omega)$ consisting of continuous functions.

Assumption 3. $\{p(\bullet \mid \omega, u; \nu), \omega \in \Omega, u \in B_1(A \times \Omega)\}$ is the set of regular finitely-precise decision rules and is an equicontinuous family on $B_1(A \times \Omega)$.

This assumption clearly holds for the logistic specification, eq(6).

Given $\beta \in [0, 1)$, $r \in C_1(A \times \Omega)$ and $f \in \Phi$, define the mapping $T: B_1(A \times \Omega) \to B_1(A \times \Omega)$, by

$$T(u)(a, \omega) \equiv (1-\beta) r(a, \omega) + \beta \int_\Omega \left[\int_{\mathcal{A}(z)} u(b, z) \, p(b \mid z, u; \nu) db \right] f(z \mid a, \omega) dz . \qquad (8)$$

That is, $T(u) \in B_1(A \times \Omega)$, and at (a, ω), $T(u)$ takes the real value given by the r.h.s. of eq(8).

It is convenient to define

$$U(z, u; \nu) \equiv \int_{\mathcal{A}(z)} u(b, z) \, p(b \mid z, u; \nu) db. \qquad (9)$$

Then, we can rewrite eq(8) as

$$T(u)(a, \omega) \equiv (1-\beta)r(a, \omega) + \beta \int_{\Omega} U(z, u; \nu)\, f(z|a, \omega)dz\,. \qquad (10)$$

Lemma 1. $\{T(u), u \in B_1(A\times\Omega)\}$ is a bounded equicontinuous family.

PROOF: Clearly, $|U(z, u; \nu)| \leq 1$ for all $z \in \Omega$ and $u \in B_1(A\times\Omega)$. Then, since $f \in \Phi$, for every $\varepsilon > 0$ there is a $\delta' > 0$, such that $\|(a, \omega) - (a', \omega')\| < \delta'$ implies that

$$\left|\int \left[U(z, u; \nu)[f(z|a, \omega) - f(z|a', \omega')\right]dz\right| \leq \int |f(z|a, \omega) - f(z \mid a', \omega')|dz < \varepsilon.$$

Since $r \in C_1(A\times\Omega)$, for the same ε there is a δ'' such that $\|(a, \omega) - (a', \omega')\| < \delta''$ implies that $|r(a, \omega) - r(a', \omega')| < \varepsilon$. Then, letting $\delta = min\{\delta', \delta''\}$, $\|(a, \omega) - (a', \omega')\| < \delta$ implies that $|T(u)(a, \omega) - T(u)(a', \omega')| < \varepsilon$. Obviously the ε and δ values do not depend on u. Q.E.D.

Lemma 2. T is a continuous mapping.

PROOF:

$$T(u)(a, \omega) - T(u')(a, \omega) = \beta \int \left[\int [u(b, z)\, p(b|z, u; \nu) - u'(b, z)\, p(b|z, u'; \nu)]db\right] f(z \mid a, \omega)dz$$

$$= \beta \int \left[\int [u(b, z) - u'(b, z)]p(b \mid z, u; \nu) + u'(b, z)\, [p(b|z, u; \nu) - p(b|z, u')]db\right] f(z|a, \omega)dz.$$

Hence, $|T(u)(a, \omega) - T(u')(a, \omega)| \leq \beta\left[\int \|u - u'\| + \sup_z \left|\int [p(b|z, u; \nu) - p(b|z, u'; \nu)]db\right|\right]$.

By the equicontinuity property of $p(\bullet \mid z, \bullet; \nu)$, for every $\varepsilon > 0$ there is a $\delta \in (0, \varepsilon/2)$ such that $\|u - u'\| < \delta$ implies that $\sup_z \left|\int [p(b|z, u; \nu) - p(b|z, u'; \nu)]db\right| < \varepsilon/2$. Therefore, $|T(u)(a, \omega) - T(u')(a, \omega)| < \varepsilon$. Q.E.D.

Let Y denote the closure of the convex hull of $\{T(u), u \in B_1(A\times\Omega)\}$.

Theorem. Under Assumptions 1 - 3, given $r \in C_1(A\times\Omega)$, $f \in \Phi$, $\beta \in [0,1)$ and $\nu \geq 0$, there is a $u^* \in Y$, such that $T(u^*) = u^*$.

PROOF: By the Arzela-Ascoli theorem, Y is a compact, convex set. Clearly, $T(Y) \subset Y$. Since T is a continuous function on Y, the claim is an immediate consequence of Schauder's fixed point theorem.[6] Q.E.D.

3.3 Further Characterizations of a Fixed Point u*.

Given a fixed point u*:

$p(\bullet | \bullet, u^*; \nu)$ is the corresponding logistic solution;

$U(\omega_t, u^*; \nu) \equiv \int_{A(\omega_t)} u^*(a, \omega_t)\, p(a \mid \omega_t, u^*; \nu) da$ is the normalized discounted present value in state ω_t of this logistic decision rule;

$$u^*(a_t, \omega_t; \nu) = (1-\beta) r(a_t, \omega_t) + \beta \int_\Omega U(z; u^*; \nu)\, f(z|a_t, \omega_t) dz \qquad (11)$$

is the normalized discounted present value in state ω_t from taking action $a_t \in \mathcal{A}(\omega_t)$ now, and thereafter continuing to use this decision rule.

Taking the expected value of both sides of eq(11) w.r.t. $p(\bullet | \omega_t, u^*; \nu)$:

$$U(\omega_t, u^*; \nu) = (1-\beta)\, \bar{r}\,(\omega_t, u^*) + \beta \int_\Omega U(z, u^*; \nu)\, \bar{f}\,(z|\omega_t, u^*)\, dz\,, \qquad (12)$$

where $\bar{r}\,(\omega_t, u^*; \nu) \equiv \int_{A(\omega_t)} r(a\,, \omega_t)\, p(a\,|\omega_t, u^*; \nu) da$, and

$\bar{f}\,(z \mid \omega_t, u^*; \nu) \equiv \int_{A(\omega_t)} f(z|a, \omega_t)\, p(a \mid \omega_t, u^*; \nu) da$.

Eq(12) reveals that $U(\omega, u^*; \nu)$ is the normalized expected discounted present value of the probabilistic decision rule at state ω_t, given the fixed point u*; i.e. the expected current

---

[6] E.g. see Khamsi, et.al. (2001).

reward plus the discounted expected future stream of rewards generated by the modified state transition function $\bar{f}()$. In other words, U() solves the Bellman equation for the dynamic process in which the state transition function, $\bar{f}$, is the composition of f() and $p(\bullet|\omega_t, u^*; \nu)$, and the reward function, $\bar{r}$, is the expected reward w.r.t. the logistic choice function $p(\bullet|\omega_t, u^*; \nu)$.

Each of the functions $p(\bullet|\omega_t, u^*; \nu)$, $u^*(a_t, \omega_t; \nu)$ and $U(\omega_t, u^*; \nu)$ entail the precision parameter ν, and thus one can do comparative statics w.r.t. ν. For example, $\nu = 0$ implies uniformly random behavior, and in fact T() is a contraction map, so there is a unique value function U(). In the limit as $\nu \to \infty$, T() is a contraction map, so there is a unique value function. By continuity of T(), there must be a unique value function for ν sufficiently small and sufficiently large. For "moderate" values of ν, it is an open question whether T() has more than one fixed point. Since there is a unique solution for ν sufficiently large, the limit of any converging sequence of $p(\bullet|\omega, u^*; \nu)$ as $\nu \to \infty$ is obviously unique and an exact solution.

Note that both the exact dynamic programming model, eq(5), and the logistic model, eq(12), embody rational expectations about the future and hence time consistency. The only difference is the precision of choice. Indeed, the difference in values, $[V^*(\omega) - U(\omega, u^*; \nu)]$, is a good measure of the loss due to finite precision. As $\nu \to \infty$, by continuity, the probability that $p(\bullet|\omega, u^*; \nu)$ puts on all actions except the *arg max*$_a$ $u^*(a, \omega; \nu)$ will converge to 0, and hence $U(\omega,u^*; \nu)$ will converge to $V^*(\omega)$. On the other hand, not all exact probabilistic solutions can be reached in this manner. Accordingly, given a dynamic programming problem (r, f, β), we denote the set of logistic solutions for all $\nu \geq 0$ by **LDP**(r, f, β). Then, we define the set of all *nearly-optimal* solutions by **NO**$(r, f, \beta) \equiv$ **O**(r, f, β) + **LDP**(r, f, β). We will see later that our choice of the logistic form places no restrictions of the range of behavior in **NO**(r, f, β).

The purpose of developing this concept of nearly-optimal solutions was not to suggest that it is a reasonable model for human decision making, but rather to allow for human error while maintaining the spirit of the neoclassical optimization model. This extension obviously increases the range of behavior than can be considered "optimization-based". How large this increase in behaviors is addressed in the next section.

4. Behavioral Rules.

A *behavioral rule* is a function, ρ, that maps each state ω_t into a probability measure on A_t; i.e. $\rho(\bullet \mid \omega_t) \in \Delta(A_t)$, and $\rho(a_t \mid \omega_t)$ denotes the probability that action $a_t \in A_t$ will be selected. Obviously, $\rho(a_t' \mid \omega_t) = 0$ for all $a_t' \notin A_t$. The set of all behavioral rules is denoted by $B(A \times \Omega)$.

Note that every nearly-optimal decision rule is a behavioral rule. Letting **NO**(*, β, f) denote the full range of nearly-optimal decision rules for all possible reward functions, we can state that **NO**(*, β, f) $\subseteq$ $B(A \times \Omega)$. The question we wish to address is whether $B(A \times \Omega) \subseteq$ **NO**(*, β, f). For any $\rho \in B(A \times \Omega)$, we need to find a reward function r_ρ such that **NO**(r_ρ, f, β) = ρ. The result is trivial if $\beta = 0$ since the problem is static. One of the contributions of this article is to show that the result holds when $\beta > 0$.

To begin the construction, take any $\rho \in B(A \times \Omega)$, and let $u(a, \omega_t)$ be any positive affine transformation of $ln[\rho(a_t \mid \omega_t)]$:

$$u(a_t, \omega_t) \equiv \sigma \, ln[\rho(a_t \mid \omega_t)] + c, \text{ for } a_t \in A_t \tag{13}$$

where $\sigma > 0$ and c is an arbitrary real number.

Then, inverting this transformation, we can express the behavioral rule in terms of u( ) as follows:

$$\rho(a_t \mid \omega_t) = \frac{\exp[u(a_t, \omega_t)/\sigma]}{\sum_{b \in A_t} \exp[u(b, \omega_t)/\sigma]}, \quad \text{for } a_t \in A_t.^{7} \tag{14}$$

The right-hand side of eq(14) is the familiar logistic function with precision $1/\sigma$. It arises directly from the specification of the behavioral rule and not from any structural model of the decision process. In other words, we started with an arbitrary behavior rule not necessarily

[7] We adopt the conventions that $ln(0) = -\infty$, and $exp(-\infty) = 0$. Note that since $\sum_{a \in A_t} \rho(a \mid \omega_t) = 1$, the denominator of eq(14) is always strictly positive. Hence, if $\rho(a \mid \omega_t) = 0$, by eq(13), $u(a, \omega_t) = -\infty$, but eq(14) still holds. When A_t is not countable, $\rho(a \mid \omega_t)$ is a probability density, and the summation in the denominator of eq(14) is replaced by an appropriate integral.

connected with any optimization-based procedure (let alone logistic), and we derived a logistic form for that rule.

On the other hand, we are free to interpret eq(14) as the probabilistic choice function of a DM with objective function $u(a_t, \omega_t)$ and extreme-valued additive noise with variance σ^2. We are not assuming that the DM attempts to optimize $u(a_t, \omega_t)$; rather we are pointing out only that the behavior rule can be interpreted **as if** such near-optimization were done. Further, every behavior rule can be interpreted this way (if we are so inclined), but that interpretation need not be objectively true.

Another implication of eqs(13-14) is that the specification of the logistic choice function in section 3 places no restrictions on the range of behavior so represented, since any behavioral rule can be expressed as a logistic choice function.

Next, given $u(a_t, \omega_t)$ from eq(13), and given any dynamic programming problem (r, β, f), define

$$\hat{r}(a_t, \omega_t) \equiv u(a_t, \omega_t) - \beta \int_{\Omega} \left[\int_{\mathcal{A}(z)} u(b, z)\, \rho(b \mid z) db \right] f(z \mid a, \omega_t)\, dz \,.^{8} \qquad (15)$$

Equivalently, we could have defined $\varepsilon(a_t, \omega_t) \equiv \hat{r}(a_t, \omega_t) - r(a_t, \omega_t)$ as a random noise term that separates the reward function $\hat{r}(a_t, \omega_t)$ perceived by the DM and the reward function $r(a_t, \omega_t)$ perceived by us as outside observers.

By definition:

$$u(a_t, \omega_t) = \hat{r}(a_t, \omega_t) + \beta \int_{\Omega} \left[\int_{\mathcal{A}(z)} u(b, z)\, \rho(b \mid z) db \right] f(z \mid a_t, \omega_t)\, dz \,. \qquad (16)$$

Further, defining $V(\omega_t) \equiv \int u(a_t, \omega_t)\, \rho(a_t \mid \omega_t) da$, and substituting into eq(16):

$$V(\omega_t) = \int \hat{r}(a, \omega_t)\, \rho(a \mid \omega_t) da + \beta \int_{\Omega} V(z) \left[\int f(z \mid a, \omega_t) \rho(a \mid \omega_t) da \right] dz \,, \qquad (17)$$

[8] Note that despite the possible infinities in the integrand over $\mathcal{A}(z)$, by our convention the integrand is bounded and thus the integral exists.

which is equivalent to eq(12). Therefore, the arbitrary behavioral rule ρ is a solution to the dynamic programming problem ($\hat{r}$, f, β) defined by eqs (13), (14) and (15), and hence ρ is a nearly-optimal solution to the modified dynamic programming problem. Moreover, even the extreme case in which ρ is degenerate on one action in every state, it is trivial to specify a dynamic programming problem ($\hat{r}$, f, β) for which ρ is the exact solution. In other words, every behavioral rule can be rationalized as a nearly-optimal solution to a dynamic programming problem. In conclusion, **NO**(*, β, f) = B(A×Ω).

Note also that given any nearly-optimal solution $\hat{p}$ (• |•) to the dynamic programming problem (r, f, β), by using eq(13) we can construct another *as-if* utility function u() such that using eq(11): p(• |•, u; 1/σ) = $\hat{p}$ (• |•). In other words, even if the given near-optimal solution comes from a non-logistic decision rule, there is a corresponding dynamic programming problem with a logistic solution that generates the exact same behavior. Thus, it should be clear that the equivalence result does not depend on the logistic form of eq(13).

We have shown that near-optimization places no restrictions on behavior other than time-consistency.[9] This result contrasts with Harstad and Selten (2013) who argue that there is a fundamental difference between optimization-based behavior and boundedly-rational behavior.[10] Our equivalence result raises the question of whether the optimization interpretation provides any advantages over the behavioral interpretation.

4. Critique of the Optimization Interpretation.

We already criticized the exact maximization assumption of the exact dynamic programming approach as a model of human behavior, and we developed a near-optimal approach that does not assume exact maximization. However, there are additional reasons why even a nearly-optimal dynamic programming is not a good approximation of human behavior.

A serious criticism of all dynamic programming approaches is the computational complexity. First, there is the state space Ω which is potentially infinite (perhaps uncountably). Second, the action space A is also potentially infinite. Therefore, at best we can only compute an approximate solution to the Bellman equation. Further, all computational algorithms face the

[9] However, observe that if Ω includes t, then time-consistency is no restriction at all.

[10] Crawford (2013) provides further insightful comments.

curse of dimensionality. That is, the computational time is exponential in the number of grid points (or basis functions). It is far from clear that the human brain has the capability to carry out these computations in the relevant human time frame. Third, since eq(12) is not necessarily a contraction mapping, it is not clear that the iterative methods employed to solve eq(5) will work for logistic dynamic programming within the moderate range of precision.[11]

Another criticism of the optimization interpretation is that it does not address the question of how the DM acquires knowledge of the reward function and the transition function. There could be more to gain by better learning algorithms for these functions than by more precision in the optimization process. In principle, the actions taken to learn these functions can be fit into the decision-making framework. However, we face the paradox that the larger dynamic programming problem is more complex and therefore less likely to be optimized by a human DM.[12]

The last criticism to be raised here is the presumption that the DM makes a conscious choice. In contrast, there is substantial evidence from psychology that humans are not aware of all the factors that influence their choices. Whether or not humans have freewill is an open philosophical question that may never be resolved, so it is not clear why one should incorporate freewill as a core assumption in economic theory. Such metaphysical assumptions have no place in a science.

5. Implications.

The debate between behavioral descriptions and utility maximization dates back at least to 1938. The question then was whether the assumption of utility maximization places any restrictions on demand functions. Samuelson (1938) introduced the concept of *revealed preference* to capture the implications of observed choice behavior. Houthakker (1950) proved that there exists a utility function for which the observed behavior is optimal if and only if observed choice behavior satisfies the Strong Axiom of Revealed Preference. Underlying this approach is the focus on the existence of stable preferences; i.e. a function U(a) for all $a \in A$,

[11] However, it has been successful in some applications; e.g. Stahl (2015).

[12] An analogy is the decision problem of whether or not to carry out a calculation. This problem entails thinking about both the option of not carrying out the calculation and the option of carrying out the calculation, but the latter cannot be rationally evaluated without carrying out the calculation. Humans typically avoid this Buridan dilemma by simply making a guess.

independent of the state ω_t. The optimal choice depends upon the state only in so far as $A_t = \mathcal{A}(\omega_t)$ defines the budget set as a function of prices and income.

Following the development of expected utility theory, it is common to allow preferences to be state-dependent [e.g. Anscombe and Aumann, 1963; and Karni, et al, 1983]. Indeed, it is hard to imagine why preferences (and behavior) would not depend on how hungry one is, what one ate last, whether one is sick or in good health, etc. Once we admit that factors other than the choice set matter, we are in the framework of this paper which has demonstrated there is no difference between behavioral rules and near-optimization.

Perhaps the feature lost by the behavioral interpretation that would be missed most is the ability to do welfare analysis. This ability is lost because the function $u(a, \omega_t)$ defined by eq(10) is not a utility function at all, and so using this function for welfare analysis cannot be justified. Hence, the concept of Pareto optimality (defined in terms of utility) cannot be applied. This problem has been discussed by Bernheim (2009), Bernheim and Rangel (2009), Rubinstein and Salant (2012), Fleurbaey and Schokkaert (2013), and Rabin (2013). The rest of this section offers a modest addition to this discussion.

One could attempt to recast Pareto optimality in terms of voting behavior. Consider a binary ballot between option a^0 and a^1, in which a vote *for* an option is cast iff the DM strictly prefers that option to the other, and the DM abstains if indifferent. Then, one could declare a^0 *Pareto optimal* if it receives some *for* votes for every feasible alternative a^1. This voting-based definition is equivalent to the standard utility-based definition <u>if and only if</u> each DM's voting behavior is degenerate: i.e. the probability measure on the three possible votes {*for* a^0, *for* a^1, abstention} puts probability 1 on one and only one option. In contrast, if any DM has non-degenerate probabilistic voting behavior, then no a^0 can be Pareto optimal, since for any alternative a^1 there is a positive probability that a^0 could receive no *for* votes. In other words, Pareto optimality becomes an empty concept.[13]

Nonetheless, policy analysis can still be conducted based on behavior. We can ask how likely is it that option a^1 would obtain more votes in a binary ballot with the status quo a^0, and use our best estimates of the behavioral rules to predict the outcome. A community of DMs

[13] Note that this conclusion also applies to neoclassical economic theory as soon as we allow non-degenerate probabilistic choice (e.g. standard discrete choice models). In these cases, the choices are probabilistic because of individual errors or because we (the outside observer) cannot observe the true utility function without error.

could adopt a constitution that declares a^1 the winner iff a^1 gets more than half (or 2/3, etc.) of the vote. Empirically the population distribution of behavior could allow for a median voter which would guarantee acyclicity.

Moreover one could still define a DM's willingness-to-pay (WTP) for an alternative a^1 to the status quo a^0, by hypothetically pitting augmented alternatives $a^1(\tau)$ against a^0, where τ is the amount of tax the DM would pay in the event alternative a^1 were adopted. In standard neoclassical economics, we would define DM i's WTP as that tax τ_i such that $u_i[a^1(\tau_i) \mid \omega_t] = u_i(a^0 \mid \omega_t)$. Under the behavioral rule interpretation, the condition on utilities could be replaced by the condition that the probability of voting for $a^1(\tau_i)$ over a^0 equals ½. These WTPs could then be aggregated and compared to costs, just as we do in standard neoclassical economics.[14] Hence, it is not clear that there would be any practical loss from dispensing with the optimization interpretation.

6. Evaluating Behavior.

If we dispense with the optimization interpretation of behavior, then how are we to evaluate alternative behavioral rules? First, we need to specify what happens given a behavioral rule ρ.

$$f^*(\omega_t, \ldots \omega_2 \mid \rho, \omega_1) \equiv \prod_{s=2}^{t} \left(\int f(\omega_s \mid a, \omega_{s-1}) \rho(a \mid \omega_{s-1}) da \right) . \tag{18}$$

This equation gives the joint probability of the sequence of states $\{\omega_t, \ldots \omega_2\}$ given the initial state ω_1 and the rule ρ.

The second step in the standard approach would be to specify an evaluation function that is a function of the sequence of states: $\mathcal{U}(\omega_t, \ldots \omega_1)$. Here to keep the notation simple we implicitly include a_{t-1} in the state description ω_t. Presumably there is some time T after which the DM is no longer alive, so we can truncate the sequence to $\{\omega_T, \ldots \omega_2\}$, and compute the expected value:

[14] For a critique of this aggregation approach in terms of neoclassical welfare economics, see Blackorby and Donaldson (1990).

$$E\mathcal{U}(\rho, \omega_1) \equiv \int \int \mathcal{U}(\omega_T, \ldots \omega_1) \boldsymbol{f}^*(\omega_T, \ldots \omega_2 | \rho, \omega_1) d\omega_T, \ldots d\omega_2 . \quad (19)$$

But how do we specify $\mathcal{U}(\omega_T, \ldots \omega_1)$? If we take this function to be the expected discounted present value of the reward function we derived in eq(12), then we will have the tautological result that ρ is a (finitely-precise) optimization of that evaluation function.

An alternative would be to take an evolutionary point of view, in which $\mathcal{U}(\omega_T, \ldots \omega_1)$ is the reproductive success of the DM. Then, given a population distribution of behavioral rules (perhaps indexed by DNA), we could construct a dynamic model of the evolution of behavioral rules. Rules that increase reproductive success would tend to increase as a proportion of the population of active rules.[15]

7. Conclusion.

We have extended the dynamic programming approach to allow for near-optimization, and we have established the existence of nearly-optimal solutions. Every such solution defines a behavioral rule which maps the state into a probability distribution on the available actions. Conversely, given any dynamic programming problem and any behavioral rule, we have shown that there is modified reward function for which the behavior is a nearly-optimal solution to the modified dynamic programming problem.

Moreover, this equivalence raises questions about the merits of the optimization interpretation and welfare analysis in particular. We have argued that we can and should dispense with the optimization interpretation. A further benefit of the behavioral approach is that it is a progressive research strategy that will lead to models that will survive based on their predictive success.

[15] There are many strands of literature that already pursue this course: e.g. agent-based modeling (Holland, et. al. 1991), computer science machine learning (Fürnkranz, et. al. 2012), evolution of finite automata (Miller, 1996, 2007), and empirical game theory (Stahl, 2000).

References

Anscombe, F. J. and Aumann, R. J. (1963), "A Definition of Subjective Probability", *The Annals of Mathematical Statistics*, **34** (1963), 199-205.

Bellman, Richard (1954), "The theory of dynamic programming", *Bulletin of the American Mathematical Society,* **60**, 503–516.

Bernheim, B. Douglas (2009). "Behavioral Welfare Economics." *Journal of the European Economic Association,* **7**, 267–319.

Bernheim, B. Douglas, and Antonio Rangel (2009). "Beyond Revealed Preference: Choice Theoretic Foundations for Behavioral Welfare Economics." *Quarterly Journal of Economics,* **124**, 51–104.

Bertsekas, D. P. (2000), *Dynamic Programming and Optimal Control,* (2nd ed.), Athena Scientific.

Blackorby, Charles, and David Donaldson (1990). "A review article: The case against the use of the sum of compensating variations in cost-benefit analysis," *Canadian Journal of Economics,* **23**, 471–94.

Crawford, Vincent P. (2013) "Boundedly Rational versus Optimization-Based Models of Strategic Thinking and Learning in Games." *Journal of Economic Literature,* **51**, 512-527.

Fleurbaey, Marc and Erik Schokkaert (2013), "Behavioral Welfare Economics and Redistribution." *American Economic Journal: Microeconomics,* **5***,* 180–205.

Fürnkranz, J., D. Gamber, and N. Lavrac (2012), *Foundations of Rule Learning*, Springer-Verlag.

Harstad, Ronald M., and Reinhard Selten (2013). "Bounded-Rationality Models: Tasks to Become Intellectually Competitive." *Journal of Economic Literature* **51**, 496-511.

Holland, J. H., and J. H. Miller (1991), "Artificial Adaptive Agents in Economic Theory," *American Economic Review,* **81**, 365-371.

Houthakker, H.S. (1950), "Revealed preference and the utility function", *Economica*, **17**, 159–174.

Karni, E., Schmeidler, D., and Vind, K. (1983). : "On State Dependent Preferences and Subjective Probabilities," *Econometrica*, **51**, 1021-1031.

Khamsi, Mohamed A., and William A. Kirk (2001), *An Introduction to Metric Spaces and Fixed Point Theory,*" Wiley & Sons.

Miller, J. H. and S. E. Sage (2007), *Complex Adaptive Social Systems: An Introduction to Computational Models of Social Life,* Princeton University Press.

Rabin, M. (2013), "Incorporating Limited Rationality into Economics," *Journal of Economic Literature,* **51***, 528–543.*

Rubinstein, Ariel, and Yuval Salant (2012). "Eliciting Welfare Preferences from Behavioral Data Sets." *Review of Economic Studies,* **79**, 375–87.

Samuelson, Paul A. (1938), "A note on the pure theory of consumer's behavior", *Economica*, **5**, 61–71.

_____ (1948), "Consumption theory in terms of revealed preference", *Economica*, 15, 243–253.

Stahl, D. (2000), "Rule Learning in Symmetric Normal-form Games," *Games and Economic Behavior,* **32**, 105-138.

_____ (2015), "Finitely-Precise Dynamic Programming and Portfolio Choice," *Computational Economics,* **45**, 397-405.