

SOCIAL LEARNING AND THE ADOPTION OF INNOVATIONS

Charles F. Manski
Department of Economics and Institute for Policy Research
Northwestern University

June 2002

This research was supported in part by National Science Foundation grant SES-0001436. I am grateful to Francesca Molinari for research assistance and comments.

1. Introduction

Social scientists have long wanted to understand the manner in which decision makers learn about and choose innovations. A common scenario envisions an initial condition in which decision makers choose among a set of actions with known attributes. At some point, a new alternative yielding unknown outcomes becomes available. From then on, successive cohorts of decision makers choose among the expanded choice set, with later cohorts observing the experiences of earlier ones and possibly learning from them.

It has often been conjectured, and sometimes observed, that the fraction of decision makers choosing the new alternative increases with time in the manner of an S-shaped curve – first rising slowly, then rapidly, and finally converging to some limit value (e.g., Griliches, 1957). However, this certainly is not the only possible dynamic for adoption of an innovation. The fraction of decision makers choosing the new alternative could begin high and then decrease with time, or the time path could be non-monotone.

The shape of the time path of adoption should depend on the beliefs that decision makers hold when the innovation is introduced and on how those beliefs evolve as experience accumulates. Manski (2002) showed that processes of social learning can generate potentially complex time paths for the adoption of innovations. The present paper carries this work further. Section 2 reviews the general analysis of social learning in Manski (2002) and the specific findings on adoption of innovations. Section 3 reports a set of computational experiments that enrich understanding of the dynamics of information accumulation and decision making after introduction of an innovation. Section 4 examines the implications for the welfare of the population of decision makers. Section 5 concludes.

2. Social Learning from Private Experiences

Manski (2002) analyzes social interaction processes that stem from the successive endeavors of new cohorts of heterogeneous decision makers to learn from the experiences of past cohorts. The members of each new cohort observe the actions chosen and outcomes realized by past cohorts, and then make decisions that produce new experiences observable by future cohorts. I emphasize that decision makers face a basic identification problem, the *selection problem*, as they seek to learn from the experiences of others. The problem is that only the outcomes of chosen actions are observable; one cannot observe the outcomes that earlier decision makers would have experienced if they had selected other actions. The logical impossibility of observing counterfactual outcomes has long been recognized to pose a fundamental difficulty for empirical research in the social sciences. It is no less an impediment to social learning.

I study the dynamics of information accumulation and decision making when new cohorts have no prior knowledge of the outcomes associated with alternative actions, nor of the decision processes of past cohorts. I assume that new decision makers must choose their actions at a specified time and cannot revise their choices once made. Thus, they cannot undertake *learning-by-doing* and cannot otherwise wait for empirical evidence to accumulate before making decisions. These simplifying assumptions imply that each decision maker faces a single choice problem with predetermined information. Thus, dynamics emerge purely out of the process of social learning across successive cohorts. Individuals do not themselves face dynamic choice problems.

The analysis assumes only one regularity condition and one form of prior information. The regularity condition is that, for each feasible action, successive cohorts of decision makers share the same distribution of outcomes. The informational assumption is that decision makers know about this stationarity. The stationarity assumption implies that empirical evidence accumulates over time, each successive cohort being able to draw inferences from a longer history of past experiences.

A medical illustration may help to envision the process of adoption of innovations under study here. Suppose that each year persons who are newly diagnosed with an illness must choose a treatment. Originally, only one treatment was available. This treatment was the universal choice and so its properties could readily be learned by observation of its success rate in curing the illness. At some point, a new treatment with a priori unknown properties is introduced. From then on, persons diagnosed with the illness choose between the old treatment and the new one. There initially is no empirical evidence about the success rate of the new treatment, so the persons who first choose the innovation tend to be those who are either most “optimistic” (in a sense to be made precise) about its success rate or who have the lowest cost of treatment, relative to the alternative. As empirical evidence accumulates, persons who are less “optimistic” or who have higher cost of treatment may adopt the innovation. These processes tend to make the rate of adoption grow over time, but a countervailing force may exist to the degree that empirical evidence shows optimism about the new treatment to be unwarranted. Hence the rate of adoption of the new treatment may increase, decrease, or be non-monotone over time.

Section 2.1 summarizes the basic analysis of information accumulation. Section 2.2 considers how decision makers may use the available empirical evidence to choose actions. Section 2.3 applies the findings to the adoption of innovations. The propositions and corollaries stated below are taken directly from Manski (2002) and are proved there.

2.1. Information Accumulation

To begin, I formalize the idea of a succession of cohorts who learn from past experiences. Suppose that at each integer date $T \geq 1$, each member of a cohort J_T of decision makers must choose an action from a finite time-invariant choice set C . Each person $j \in J_T$ has a response function $y_j(\cdot): C \rightarrow Y$ that maps actions into outcomes, which take values in space Y . Let $z_j \in C$ denote the action chosen by person j . Then person

j realizes outcome $y_j \equiv y_j(z_j)$. The counterfactual outcomes $y_j(c)$, $c \neq z_j$ are unobservable.

To formalize needed distributional concepts, let each cohort J_T be a probability space, say (J_T, Ω_T, P_T) , with Ω_T the σ -algebra and P_T the probability measure. For each $c \in C$, let $P_T[y(c)]$ be the outcome distribution for action c in this cohort. $P_T[y(c)]$ is the outcome distribution that would be realized if a randomly drawn member of J_T were to choose c . It is not the distribution among members of J_T who actually choose c . That is $P_T[y(c)|z = c]$.

The analysis of information accumulation rests on two maintained assumptions:

Assumption 1 (Observability of Past Actions and Outcomes): Let $T \geq 1$. Before choosing actions, the members of cohort J_T observe the distributions $\{P_t(y, z), 1 \leq t \leq T-1\}$ of actions chosen and outcomes realized by earlier cohorts. ■

Assumption 2 (Stationarity of Outcome Distributions): For each $c \in C$, there exists a time-invariant probability distribution $P[y(c)]$ on the outcome space Y such that $P_T[y(c)] = P[y(c)]$, $\forall T \geq 1$. This stationarity of outcome distributions is common knowledge. ■

Assumption 1 asserts that members of each cohort can observe the experiences of past cohorts. Assumption 2 asserts the stationarity of outcome distributions that enables decision makers of each cohort to learn about their own outcome distributions by observing past experiences.

The main findings are

Proposition 1: Let Assumptions 1 and 2 hold. Let Γ denote the set of all probability distributions on Y . Let $T \geq 2$ and $c \in C$. The members of cohort J_T learn that

$$(1) P[y(c)] \in H(T, c) \equiv \bigcap_{1 \leq t \leq T-1} \{P_t(y|z=c)P_t(z=c) + \gamma_t \cdot P_t(z \neq c)\}, \gamma_t \in \Gamma.$$

The identification region for $\{P[y(c)], c \in C\}$ is $[H(T, c), c \in C]$. ■

Corollary 1: Let Assumptions 1 and 2 hold. Let $T \geq 2$ and $c \in C$. Let $\eta \in \Gamma$ be a specified probability distribution on Y . Given any measurable set $A \subset Y$, define

$$(2) \pi_{Tc}(A) \equiv \max_{1 \leq t \leq T-1} P_t(y \in A|z=c)P_t(z=c).$$

(a) Then $\eta \in H(T, c)$ if and only if $\eta(A) \geq \pi_{Tc}(A), \forall A \subset Y$.

(b) Let Y be countable. Then $\eta \in H(T, c)$ if and only if $\eta(y) \geq \pi_{Tc}(y), \forall y \in Y$.

(c) Let Y be countable. Let $S(T, c) \equiv \sum_{y \in Y} \pi_{Tc}(y)$. Then $H(T, c)$ contains a unique distribution if and only if $S(T, c) = 1$. When $S(T, c) = 1$, the unique feasible distribution is $\eta_{Tc}(y) \equiv \pi_{Tc}(y), y \in Y$.¹ ■

Proposition 1 shows that learning is a process of sequential reduction in *ambiguity*.² At date $T = 1$, decision makers have no knowledge at all of their outcome distributions. From $T = 2$ on, decision makers

¹ The event $S(T, c) > 1$ cannot occur under Assumptions 1 and 2; this event implies that $H(T, c)$ is empty. The distribution η_{Tc} is distinct from π_{Tc} , which is sub-additive and hence not a probability distribution; that is, $\eta_{Tc}(y) = \pi_{Tc}(y)$ for $y \in Y$ but $\eta_{Tc}(A) \geq \pi_{Tc}(A)$ for $A \subset Y$.

² A decision maker with a known choice set who wishes to maximize an unknown objective function is said to face a problem of choice under ambiguity. A common source of ambiguity is incomplete knowledge of a probability distribution describing a relevant population – the decision maker may know only that the distribution of interest is a member of some set of distributions. This is the generic situation of a decision maker who seeks to learn a population distribution empirically, but whose data and prior information do not suffice to identify the distribution. Thus, identification problems in empirical analysis induce ambiguity in decision making.

The study of choice under ambiguity dates back at least to Keynes (1921) and Knight (1921), who used the term *uncertainty*. Ellsberg (1961) introduced the term *ambiguity* and posed the problem in a particularly evocative way through a thought experiment requiring subjects to draw a ball from either of two urns, one with a known distribution of colors and the other with an unknown distribution of colors.

can use observations of past cohorts to learn about their outcome distributions. For each $c \in C$, the set $H(T+1, c)$ of distributions that are feasible at date $T+1$ is a subset of the corresponding set $H(T, c)$ at T . Because the process of information accumulation is monotone, it must converge to a *terminal information state*. That is, there necessarily exists a $[H(c), c \in C]$ such that

$$(3) \lim_{T \rightarrow \infty} [H(T, c), c \in C] = [H(c), c \in C].$$

The characterization of the set of feasible distributions given in Proposition 1 is simple but abstract. Corollary 1 gives a useful alternative characterization of $H(T, c)$. The basic finding is that a distribution is feasible if and only if the probability it places on each measurable subset of Y is no less than an easily computed lower bound. This characterization is particularly useful when the outcome space Y is countable. Then one need only consider the probability placed on each atom of Y . The Corollary shows that when Y is countable, the vector $[\pi_{Tc}(y), y \in Y]$ is a sufficient statistic for $H(T, c)$. Observe that $H(T, c)$ shrinks as $[\pi_{Tc}(y), y \in Y]$ increases.

2.2. Decision Making

Now consider how decision makers may behave in the setting described in Section 2.1. For $T \geq 2$ and $j \in J_T$, let $U_j(\cdot, \cdot): C \times Y \rightarrow \mathbb{R}^1$ denote the utility function that person j uses to evaluate actions. The utility $U_j[c, y(c)]$ that person j associates with action c may depend on the outcome $y(c)$, which the person does not know when facing the choice problem, as well as on attributes of c that the person does know.

Economists often assume that decision makers have rational expectations and choose actions that maximize expected utility. However, decision makers do not have rational expectations under Assumptions 1 and 2. Person j knows only that $\{P_T[y(c)], c \in C\}$, the vector of objective distributions of outcomes within

his cohort, is an element of the set $[H(T, c), c \in C]$ specified in Proposition 1.

How might a person behave in this setting? A pervasive idea in research on social learning has been that a person views himself as a member of some observable *reference group* and predicts that, if he were to choose a given action, he would experience an outcome drawn at random from the distribution of outcomes in this group. We can formalize this idea by assuming that person j views himself as a member of cohort J_T , predicts that his outcome under each action $c \in C$ is drawn from $P_T[y(c)]$, and aims to solve the problem

$$(4) \max_{c \in C} \int U_j[c, y(c)] dP_T[y(c)].$$

Problem (4) expresses a limited-information version of the usual rational expectations model, one in which person j conditions his expectations only on the information that he belongs to cohort J_T rather than on his full information set.

Elimination of Dominated Actions

Problem (4) may not be solvable if $[H(T, c), c \in C]$ contains multiple distributions. However, decision makers can eliminate actions that are dominated. Assumption 3 asserts that decision makers do not choose dominated actions:

Assumption 3 (Elimination of Dominated Actions): Let Assumptions 1 and 2 hold. Let $T \geq 2$ and $j \in J_T$. For $c \in C$ and $\gamma \in \Gamma$, let $\int U_j(c, y) d\gamma$ be the expected utility of action c if outcome y were distributed γ . Action $c' \in C$ is *dominated* if there exists another action $c'' \in C$ such that $\int U_j(c', y) d\eta' \leq \int U_j(c'', y) d\eta''$ for all $(\eta', \eta'') \in [H(T, c'), H(T, c'')]$ and $\int U_j(c', y) d\eta' < \int U_j(c'', y) d\eta''$ for some $(\eta', \eta'') \in [H(T, c'), H(T, c'')]$. Person j does not choose a dominated action. ■

The abstract characterization of dominated actions given in Assumption 3 becomes more transparent when the outcome space Y is countable and utility is bounded. In this case, Proposition 2 yields a simple description of the dominated actions.

Proposition 2: Let Y be countable. Let $T \geq 2$ and $j \in J_T$. Let $K_{0jc} \equiv \min_{y \in Y} U_j(c, y)$ and $K_{1jc} \equiv \max_{y \in Y} U_j(c, y)$ exist for all $c \in C$. Let

$$(5) \quad d \in \operatorname{argmax}_{c \in C} \sum_{y \in Y} \pi_{Tc}(y) \cdot U_j(c, y) + [1 - S(T, c)] \cdot K_{0jc}.$$

Action $c' \in C$ is dominated if

$$(6a) \quad \sum_{y \in Y} \pi_{Tc'}(y) \cdot U_j(c', y) + [1 - S(T, c')] \cdot K_{1jc'} < \sum_{y \in Y} \pi_{Td}(y) \cdot U_j(d, y) + [1 - S(T, d)] \cdot K_{0jd}$$

or if

$$(6b) \quad \sum_{y \in Y} \pi_{Tc'}(y) \cdot U_j(c', y) + [1 - S(T, c')] \cdot K_{1jc'} = \sum_{y \in Y} \pi_{Td}(y) \cdot U_j(d, y) + [1 - S(T, d)] \cdot K_{0jd}$$

and

$$(6c) \quad \min [S(T, c'), S(T, d)] < 1.$$

Action c' is undominated if neither (6a) nor (6b)-(6c) holds. ■

Proposition 2 shows that if utility is bounded, the accumulation of empirical evidence over time can enlarge the set of actions that decision makers may eliminate as dominated. Consider a sequence of decision

makers who share the same bounded utility function but who make decisions at successive dates. Let $c' \in C$. The upper bound on the expected utility of this action, given on the left side of (6a), decreases with T . The greatest lower bound of the expected utility of all actions, on the right side of (6a), increases with T . Hence action c' may be undominated at early dates but dominated later on. It is not possible for c' to be dominated early but undominated later.

Choice Among Undominated Actions

Assumption 3 leaves open how decision makers choose among undominated actions. There is no “optimal” way to make this choice, but many “reasonable” decision rules have been suggested over the years. In particular, Wald (1950) proposed the maximin rule, which solves the problem

$$(7) \max_{c \in C} \inf_{\eta \in H(T, c)} \int U_j[c, y(c)] d\eta.$$

Hurwicz (1951) suggested maximization of a weighted average of the minimum and maximum values of the objective function that are feasible for each action. Thus person j would solve the problem

$$(8) \max_{c \in C} \lambda_j \left\{ \inf_{\eta \in H(T, c)} \int U_j[c, y(c)] d\eta \right\} + (1 - \lambda_j) \left\{ \sup_{\eta \in H(T, c)} \int U_j[c, y(c)] d\eta \right\}$$

for some $\lambda_j \in [0, 1]$. Rule (8) provides a simple way of expressing degrees of pessimism and optimism; $\lambda_j = 1$ means that person j uses the maximin rule and $\lambda_j = 0$ that he uses the maximax rule.

Bayesian decision theorists suggest that the decision maker assert a subjective distribution on the space of feasible outcome distributions and maximize subjective expected utility with respect to this distribution. Thus person j would solve the problem

$$(9) \max_{c \in C} \int \{ \int U_j[c, y(c)] d\eta \} dQ_{jc}(\eta),$$

where Q_{jc} is the subjective distribution that person j places on $H(T, c)$. Despite their widespread application in economic theory, Bayes decision rules have no particular normative force in the absence of credible prior information.

2.3. Learning About and Choosing Innovations

I now specialize the foregoing analysis to adoption of innovations. Assumption 4 specifies a model whose structure is very simple, yet which can generate a rich variety of dynamics.

Assumption 4: Assumption 3 holds. Moreover,

(a) The choice set is $C = (e, n)$. At date $T = 1$, all persons choose action e . The outcome space is $Y = \{0, 1\}$. The utilities that person j associates with actions e and n are $U_j[e, y(e)] = y(e)$ and $U_j[n, y(n)] = y(n) + u_j$, where $u_j \in \mathbb{R}^1$. Person j knows u_j before choosing an action.

(b) Person j uses the Hurwicz criterion (13) with parameter λ_j to choose among undominated actions.

(c) There exists a time-invariant probability distribution $P[y(e), y(n), u, \lambda]$ such that $P_T[y(e), y(n), u, \lambda] = P[y(e), y(n), u, \lambda], \forall T \geq 1$. The distribution of u is continuous. ■

Part (a) specializes Assumption 3 in various respects. Action e is the only pre-existing alternative, which all persons choose at $T = 1$. Action n is the innovation. The outcome y is binary, taking the value zero or one. Utility functions are separable in y . Parts (b) and (c) go beyond Assumption 3. Part (b) asserts that, to choose among undominated actions, each decision maker maximizes some weighted average of the lower and upper bounds on expected utility. Part (c) strengthens the stationarity condition asserted in Assumption

2. Successive cohorts of decision makers not only have the same outcome distributions, but have the same joint distributions of decision rules and outcomes.³ Requiring that $P(u)$ be continuous ensures that indifference between actions e and n occurs with probability zero, so the model yields well-defined choice probabilities.

The Dynamics of Choice

Proposition 3 describes the time path of adoption of the innovation.

Proposition 3: Let Assumption 4 hold. At each date $T \geq 2$,

$$(10) \quad P_T(z = n) = P\{\lambda \cdot \pi_{Tn}(1) + (1 - \lambda) \cdot [1 - \pi_{Tn}(0)] + u > P[y(e) = 1]\}. \quad \blacksquare$$

Proposition 3 shows how adoption of the innovation depends on $[\pi_{Tn}(1), \pi_{Tn}(0)]$, which weakly increase with time as empirical evidence accumulates. Consider date $T = 2$. No one chose action n at $T = 1$, so $\pi_{2n}(0) = \pi_{2n}(1) = 0$. Hence $P_2(z = n) = P\{1 - \lambda + u > P[y(e) = 1]\}$. From then on, $\pi_{(T+1)n}(1)$ and $\pi_{(T+1)n}(0)$ are given by the updating rule

$$(11a) \quad \pi_{(T+1)n}(1) = \max[\pi_{Tn}(1), P_T(y = 1 | z = n)P_T(z = n)]$$

$$(11b) \quad \pi_{(T+1)n}(0) = \max[\pi_{Tn}(0), P_T(y = 0 | z = n)P_T(z = n)].$$

Moreover,

$$(12a) \quad P_T(y = 1 | z = n) = P\{y(n) = 1 | \lambda \cdot \pi_{Tn}(1) + (1 - \lambda)[1 - \pi_{Tn}(0)] + u > P[y(e) = 1]\}$$

³ Part (c) does not strengthen the information condition asserted in Assumption 2. Here, as before, decision makers only know that outcome distributions are stationary.

$$(12b) \quad P_T(y = 1 | z = e) = P\{y(e) = 1 | \lambda \cdot \pi_{Tn}(1) + (1 - \lambda)[1 - \pi_{Tn}(0)] + u < P[y(e) = 1]\}.$$

Taken together, equations (10) - (12) show how $P[y(e), y(n), u, \lambda]$, the stationary distribution of outcomes and decision rules, determines the dynamics of learning and choice.

Under Assumption 3, the decision rule that persons would use if $P[y(n) = 1]$ were known is to choose n if $P[y(n) = 1] + u > P[y(e) = 1]$ and choose e otherwise. The rate of adoption of the innovation would be $p^* \equiv P\{P[y(n) = 1] + u > P[y(e) = 1]\}$. Observe that $\pi_{Tn}(1) \leq P[y(n) = 1] \leq 1 - \pi_{Tn}(0)$. Hence the actual fraction of a cohort who choose action n can be below or above p^* , depending on the population distribution of (u, λ) .

It is revealing to consider two extreme cases, one in which all decision makers use the maximin rule ($\lambda = 1$) and the other in which all use the maximax rule ($\lambda = 0$). If everyone uses the maximin rule, then $P_T(z = n) = P\{\pi_{Tn}(1) + u > P[y(e) = 1]\}$. Hence the fraction of a cohort who choose the innovation weakly increases with time, but always remains less than or equal to p^* . If all use the maximax rule, $P_T(z = n) = P\{1 - \pi_{Tn}(0) + u > P[y(e) = 1]\}$. In this case, the fraction who choose the innovation weakly decreases with time, but always remains greater than or equal to p^* .

The Terminal Information State

The terminal information state is determined by the stationary distribution of outcomes and decision rules. Some sense of the possibilities is given by considering the special case in which all persons have the same value of λ , $y(n)$ is statistically independent of u , and u has support R^1 .

Let $P(\lambda = L) = 1$, where $L \in [0, 1]$. Let $b \equiv P[y(n) = 1]$. Let $p_t \equiv P_t(z = n)$, $t \geq 1$. By (16), statistical independence of $y(n)$ and u implies that for each $T \geq 2$,

$$(13a) \quad \pi_{Tn}(1) = P[y(n) = 1] \cdot \{\max_{1 \leq t \leq T-1} P_t(z = n)\} = b \cdot (\max_{1 \leq t \leq T-1} p_t)$$

$$(13b) \quad \pi_{Tn}(0) = P[y(n) = 0] \cdot \{\max_{1 \leq t \leq T-1} P_i(z = n)\} = (1 - b) \cdot (\max_{1 \leq t \leq T-1} p_t).$$

This and Proposition 3 yield

$$(14) \quad p_T = P\{L \cdot b \cdot (\max_{1 \leq t \leq T-1} p_t) + (1 - L) \cdot [1 - (1 - b) \cdot (\max_{1 \leq t \leq T-1} p_t)] + u > P[y(e) = 1]\} \\ = P\{(L + b - 1) \cdot (\max_{1 \leq t \leq T-1} p_t) + 1 - L + u > P[y(e) = 1]\}.$$

Inspection of (14) shows that the sign of $(L + b - 1)$ determines the qualitative dynamics of decision making and information accumulation. If $(L + b - 1)$ is positive, the probability of choosing the innovation increases with time. Thus $\max_{1 \leq t \leq T-1} p_t = p_{(T-1)}$ and equation (19) reduces to

$$(15) \quad p_T = P\{(L + b - 1) \cdot p_{(T-1)} + 1 - L + u > P[y(e) = 1]\}.$$

Recall that, by assumption, $p_1 = 0$. Hence equation (15) generates a monotone increasing rate of adoption whose limit as $T \rightarrow \infty$ is the value p' yielding the smallest solution to the equation

$$(16) \quad p' = P\{(L + b - 1) \cdot p' + 1 - L + u > P[y(e) = 1]\}.$$

The terminal information state is $[\pi_n(1) = b \cdot p', \pi_n(0) = (1 - b) \cdot p']$. The value p' lies in the open interval $(0, 1)$. Hence the terminal information state is informative about b but does not completely identify it.

If $(L + b - 1)$ is non-positive, the situation is entirely different. Consider dates $T = 2, 3$, and 4 . The following hold:

$$(17a) \quad p_2 = P\{1 - L + u > P[y(e) = 1]\}$$

$$(17b) \quad p_3 = P\{(L + b - 1) \cdot p_2 + 1 - L + u > P[y(e) = 1]\}$$

$$(17c) \quad p_4 = P\{(L + b - 1) \cdot p_2 + 1 - L + u > P[y(e) = 1]\}.$$

Equation (17a) holds because $p_1 = 0$, (17b) because $p_2 \geq 0$, and (17c) because $p_3 \leq p_2$. Hence information accumulation ceases at $T = 3$ and the terminal information state is $[\pi_n(1) = b \cdot p_2, \pi_n(0) = (1-b) \cdot p_2]$.

Thus the qualitative dynamics of decision making and information accumulation depend critically on how decision makers choose among undominated actions. If they act pessimistically (i.e., $L > 1 - b$), the adoption rate of the innovation increases with time. If they act optimistically (i.e., $L \leq 1 - b$), the adoption rate begins high and then immediately falls to a steady state value. However decision makers behave, social learning takes place but remains incomplete.

3. Computational Experiments

Computational experiments with particular specifications of the time-invariant probability distribution $P[y(e), y(n), u, \lambda]$ can enhance understanding of the model of adoption of innovations studied in Section 2.3. The experiments reported here aim to illuminate how behavior under ambiguity affects information accumulation and decision making.

3.1. Experimental Design

The experiments fix many features of $P[y(e), y(n), u, \lambda]$ but consider a broad set of nine specifications for the distribution of the parameter λ , which governs behavior under ambiguity. The maintained assumptions are

- $P[y(e), y(n), u, \lambda] = P[y(e)] \cdot P[y(n)] \cdot P(u) \cdot P[\lambda | y(n)]$
- $P[y(e) = 1] = P[y(n) = 1] = 0.5$
- $P(u) \sim N(0, 1)$
- λ can take the values $\{0, 0.5, 1\}$.

Thus, the random variables $[y(e), y(n), u]$ are statistically independent and have specified marginal distributions. The distribution of λ may vary with the outcome $y(n)$, but all decision makers use one of three decision rules: the maximin rule ($\lambda = 1$), the maximax rule ($\lambda = 0$), or the intermediate rule giving equal weight to the lower and upper bounds on expected utility ($\lambda = 0.5$).

The nine specifications for the distributions $P[\lambda | y(n)]$ are

Case 1:	$P[\lambda = 0 y(n) = 0] = 1$	$P[\lambda = 0 y(n) = 1] = 1$
Case 2:	$P[\lambda = 0 y(n) = 0] = 1$	$P[\lambda = k y(n) = 1] = 1/3, k \in \{0, 0.5, 1\}$
Case 3:	$P[\lambda = 0 y(n) = 0] = 1$	$P[\lambda = 1 y(n) = 1] = 1$
Case 4:	$P[\lambda = k y(n) = 0] = 1/3, k \in \{0, 0.5, 1\}$	$P[\lambda = 0 y(n) = 1] = 1$
Case 5:	$P[\lambda = k y(n) = 0] = 1/3, k \in \{0, 0.5, 1\}$	$P[\lambda = k y(n) = 1] = 1/3, k \in \{0, 0.5, 1\}$
Case 6:	$P[\lambda = k y(n) = 0] = 1/3, k \in \{0, 0.5, 1\}$	$P[\lambda = 1 y(n) = 1] = 1$
Case 7:	$P[\lambda = 1 y(n) = 0] = 1$	$P[\lambda = 0 y(n) = 1] = 1$
Case 8:	$P[\lambda = 1 y(n) = 0] = 1$	$P[\lambda = k y(n) = 1] = 1/3, k \in \{0, 0.5, 1\}$
Case 9:	$P[\lambda = 1 y(n) = 0] = 1$	$P[\lambda = 1 y(n) = 1] = 1.$

Cases 1 and 9 express the extreme possibilities that all persons use the maximax rule (Case 1) or all use the maximin rule (Case 9). Cases 2 through 7 specify various forms of heterogeneity in behavior under ambiguity; some persons use the maximin rule, others use the maximax rule, and still others use the intermediate rule with $\lambda = 0.5$. The nine cases also express varied forms of dependence between λ and the outcome $y(n)$ that a person would experience if he were to choose the innovation; λ and $y(n)$ are statistically

independent in Cases (1, 5, 9), but are dependent in the other cases.

3.2. Findings

Figure 1 plots the time path for dates $T = 2, \dots, 10$ of adoption of the innovation in each of the nine cases specified above. Table 1 presents the same findings in numerical form. Figure 2 and Table 2 present the identification regions for $P[y(n) = 1]$; that is, $[\pi_{Tn}(1), 1 - \pi_{Tn}(0)]$, $T = 2, \dots, 10$.

The figures and tables show that the time path of adoption is monotone increasing in Cases (6, 9), monotone decreasing in Cases (1, 2, 4), slightly non-monotone in Cases (3, 7, 8), and flat in Case 5. The terminal information state is always reached by $T = 7$, but as soon as $T = 3$ in some cases.

The steady state rate of adoption of the innovation varies considerably across the nine cases, from a low of 0.3779 if all persons use the maximin rule to a high of 0.5613 if all persons use the maximax rule. If $P[y(n) = 1]$ were known, the rate of adoption would be

$$P\{P[y(n) = 1] + u > P[y(e) = 1]\} = P(u > 0) = 0.5.$$

Coincidentally, this also is the rate of adoption in Case 5. I say ‘‘coincidentally’’ because decision makers have only partial knowledge of $P[y(n) = 1]$ in Case 5; the adoption rate turns out to be 0.5 only because of the specific assumptions made about the distribution of λ .

The terminal information state also varies considerably across the nine cases. Cases 1 and 9 give the extreme outcomes of the learning process. If all decision makers use the maximax rule, they learn as soon as $T = 3$ that $P[y(n) = 1]$ lies in the interval $[0.3457, 0.6543]$. If all decision makers use the maximin rule, the terminal state is reached at $T = 7$ and they learn that $P[y(n) = 1]$ lies in the interval $[0.1889, 0.8111]$. The former and latter intervals for $P[y(n) = 1]$ have widths 0.3086 and 0.6222, respectively.

Thus, in these computational experiments, the way that decision makers behave under ambiguity very substantially affects both the steady state rate of adoption of the innovation and the amount of learning that takes place.

4. Welfare Analysis

How does incomplete knowledge of the outcomes associated with alternative actions affect the welfare of each cohort of decision makers? To address this question, I adopt the conventional perspective of public economics, in which the objective is to maximize a utilitarian social welfare function. Then we can compare the welfare realized when decision makers make choices under ambiguity with the welfare they would realize in other scenarios.

4.1. Welfare in Various Scenarios

Fix T and let J_T be the cohort of interest. In terms of the general discussion of decision making in Section 2.2, the utilitarian social welfare realized by this cohort is

$$(18) \quad W_T \equiv \int \left\{ \sum_{c \in C} U_j[c, y_j(c)] 1[z_j = c] \right\} dP_T(j),$$

where $1[\cdot]$ is the indicator function taking the value one if the condition in the brackets holds and zero otherwise. In the simple setting of Assumption 4, (18) reduces to

$$(19) \quad W_T \equiv \int \left\{ [y_j(n) + u_j] 1[z_j = n] + y_j(e) 1[z_j = e] \right\} dP_T(j)$$

$$= P_T[y(n) = 1, z = n] + E_T\{u \cdot 1[z = n]\} + P_T[y(e) = 1, z = e].$$

There are several welfare expressions with which W_T may be usefully compared. One is the ideal welfare that could be realized if decisions makers had perfect foresight (PF); that is, if each person j were to know his outcomes $[y_j(e), y_j(n)]$. Then person j would choose action n if $y_j(n) + u_j > y_j(e)$, and choose action e otherwise. Realized welfare would be

$$(20) W_{PF} \equiv \int \max [y(n) + u, y(e)] dP[y(e), y(n), u].$$

Another is the welfare that would be realized if decision makers were able to identify the reference-group (RG) outcome distributions $\{P[y(e)], P[y(n)]\}$ and solve problem (4). Then person j would choose action n if $P[y(n) = 1] + u_j > P[y(e) = 1]$. Realized welfare would be

$$(21) W_{RG} \equiv P\{y(n) = 1, u > P[y(e) = 1] - P[y(n) = 1]\} \\ + E\{u \cdot 1[u > P[y(e) = 1] - P[y(n) = 1]]\} + P\{y(e) = 1, u < P[y(e) = 1] - P[y(n) = 1]\}.$$

One more welfare expression with which W_T may usefully be compared is W_1 , the welfare realized at date $T = 1$, when all persons must choose the existing alternative. This is

$$(22) W_1 = P[y(e) = 1].$$

The welfare W_{PF} achievable with perfect foresight must be at least as large as each of the welfare expressions (W_T, W_{RG}, W_1) , whatever the distribution $P[y(e), y(n), u, \lambda]$ may be. The ranking of (W_T, W_{RG}, W_1) relative to one another depends on the form of this distribution. Inspection of (21) shows that $W_{RG} \geq W_1$

if u is statistically independent of $[y(e), y(n)]$, but this inequality does not necessarily hold under some forms of dependence between u and $[y(e), y(n)]$. It appears difficult to say much of anything a priori about the ranking of W_T relative to W_{RG} and W_1 . However, we can usefully calculate the various welfare expressions for the particular specifications of $P[y(e), y(n), u, \lambda]$ considered in the computational experiments of Section 3. Section 4.2 gives the findings.

4.2. Welfare in the Computational Experiments

The distributional assumptions of the computational experiments imply these values for the welfare expressions (W_{PF} , W_{RG} , W_1):

$$W_{PF} = 1.0114 \quad W_{RG} = 0.8989 \quad W_1 = 0.5.$$

The welfare W_T depends on the date T and on the specification for $P[\lambda|y(n)]$. I focus here on $T = 10$, which always yields the terminal information state. The results for the nine cases are:

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9
W_{10}	0.8942	0.8797	0.8552	0.9156	0.9012	0.8684	0.9434	0.9255	0.8801

Perhaps the most striking finding is that in each of the nine cases, W_{10} is reasonably close to the value of W_{RG} ; whereas $W_{RG} = 0.8989$, W_{10} takes values in the interval $[0.8552, 0.9434]$. In Section 3.2, we found that the way decision makers behave under ambiguity has strong quantitative effects on the rate of adoption of the innovation and on the terminal state of the learning process. Here, in contrast, we find only moderate quantitative effects on welfare. Moreover, realized welfare is relatively high in all cases, much closer to the

ideal W_{PF} than to the welfare that would be experienced in the absence of the innovation.

It is easy to see why Cases 3 and 7 yield the lowest and highest values for W_{10} . The ambiguity parameter λ and the outcome $y(n)$ are strongly dependent in these cases. In Case 3, the persons who behave optimistically, and so tend to choose the innovation, tend have bad draws of $y(n)$; those who behave pessimistically, and so tend to choose action e , tend to have good draws of $y(n)$. Hence the members of each new cohort J_T observe small values for the probabilities $[P_t(y = 1 | z = n), t < T]$ that earlier decision makers who chose the innovation had good outcomes. In Case 7, the nature of the dependency between λ and $y(n)$ is reversed.

One perhaps unintuitive finding is that W_{10} slightly exceeds W_{RG} in Case 5, where λ and $y(n)$ are statistically independent. This ranking is possible, albeit hard to explain, because decision makers who behave according to Assumption 4 make choices using knowledge of $[P_t(y | z = n), t < T]$. This information is not used in solving problem (4).

5. Conclusion

This paper has used computational experiments to shed further light on the theoretical analysis of the dynamics of social learning in Manski (2002). These experiments illustrate quantitatively the qualitative theme, advanced in my earlier paper, that social learning from private experiences is a process of *complexity within regularity*. The process is complex because the dynamics of learning and the properties of the terminal information state flow from the subtle interaction of information accumulation and decision making. Yet a basic regularity constrains how the process evolves, as accumulation of empirical evidence over time (weakly) reduces the ambiguity that successive cohorts face.

Theoretical analysis and computational experiments are valuable in understanding complex economic

processes, but I see a pressing need for new empirical research as well. In the present context, the critical empirical questions is how decision makers actually cope with ambiguity. It is clear that the way decision makers choose among undominated actions can critically affect the dynamics of learning and choice. An improved empirical understanding of decision making under ambiguity is necessary to guide theoretical and computational research in productive directions.

References

Griliches, Z. (1957), "Hybrid Corn: An Exploration in the Economics of Technological Change," Econometrica, 25, 501-522.

Hurwicz, L. (1951), "Some Specification Problems and Applications to Econometric Models," Econometrica, 19, 343-344.

Manski, C. (2002), "Social Learning from Private Experiences: The Dynamics of the Selection Problem," Review of Economic Studies, forthcoming.

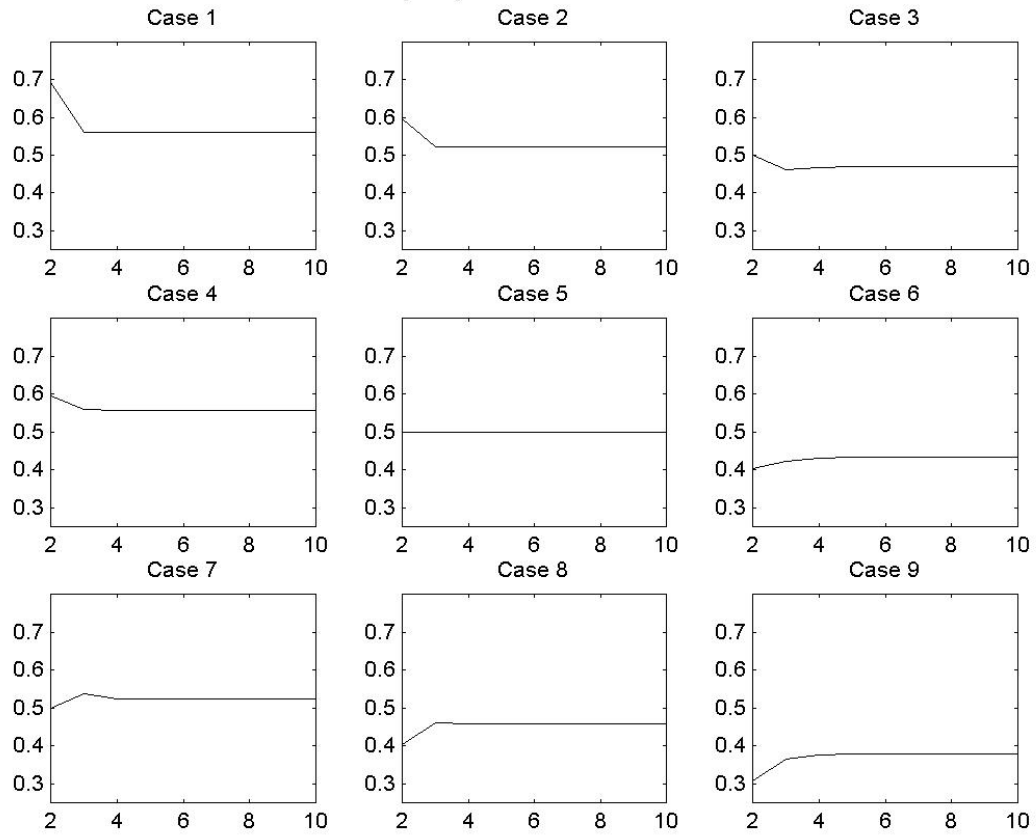
Wald, A. (1950), Statistical Decision Functions, New York: Wiley.

Table 1: Rate $P_T(z = n)$ of Adoption of the Innovation

T	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9
1	0	0	0	0	0	0	0	0	0
2	0.6915	0.5957	0.5000	0.5957	0.5000	0.4043	0.5000	0.4043	0.3085
3	0.5613	0.5212	0.4630	0.5588	0.5000	0.4231	0.5370	0.4599	0.3648
4	0.5613	0.5212	0.4683	0.5560	0.5000	0.4311	0.5246	0.4581	0.3754
5	0.5613	0.5212	0.4694	0.5560	0.5000	0.4326	0.5246	0.4580	0.3774
6	0.5613	0.5212	0.4695	0.5560	0.5000	0.4329	0.5246	0.4580	0.3778
7	0.5613	0.5212	0.4696	0.5560	0.5000	0.4330	0.5246	0.4580	0.3779
8	0.5613	0.5212	0.4696	0.5560	0.5000	0.4330	0.5246	0.4580	0.3779
9	0.5613	0.5212	0.4696	0.5560	0.5000	0.4330	0.5246	0.4580	0.3779
10	0.5613	0.5212	0.4696	0.5560	0.5000	0.4330	0.5246	0.4580	0.3779

Table 2: Identification region for $P[y(n) = 1]$

T	Case 1		Case 2		Case 3		Case 4		Case 5		Case 6		Case 7		Case 8		Case 9	
	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
2	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
3	0.3457	0.6543	0.25	0.6543	0.1543	0.6543	0.3457	0.75	0.25	0.75	0.1543	0.75	0.3457	0.8457	0.25	0.8457	0.1543	0.8457
4	0.3457	0.6543	0.25	0.6543	0.1824	0.6543	0.3457	0.7406	0.25	0.75	0.1824	0.75	0.3457	0.7807	0.2593	0.7994	0.1824	0.8176
5	0.3457	0.6543	0.25	0.6543	0.1877	0.6543	0.3457	0.7406	0.25	0.75	0.1877	0.75	0.3457	0.7807	0.2593	0.7976	0.1877	0.8123
6	0.3457	0.6543	0.25	0.6543	0.1887	0.6543	0.3457	0.7406	0.25	0.75	0.1887	0.75	0.3457	0.7807	0.2593	0.7976	0.1887	0.8113
7	0.3457	0.6543	0.25	0.6543	0.1889	0.6543	0.3457	0.7406	0.25	0.75	0.1889	0.75	0.3457	0.7807	0.2593	0.7976	0.1889	0.8111
8	0.3457	0.6543	0.25	0.6543	0.1889	0.6543	0.3457	0.7406	0.25	0.75	0.1889	0.75	0.3457	0.7807	0.2593	0.7976	0.1889	0.8111
9	0.3457	0.6543	0.25	0.6543	0.1889	0.6543	0.3457	0.7406	0.25	0.75	0.1889	0.75	0.3457	0.7807	0.2593	0.7976	0.1889	0.8111
10	0.3457	0.6543	0.25	0.6543	0.1889	0.6543	0.3457	0.7406	0.25	0.75	0.1889	0.75	0.3457	0.7807	0.2593	0.7976	0.1889	0.8111

$P[z = n]$ as a function of T 

Identification region for $P[y(n)=1]$ as a function of T 