GAMES AND ECONOMIC BEHAVIOR **15,** 1–26 (1996)
ARTICLE NO. 0056

# Case-Based Optimization*

Itzhak Gilboa

*KGSM-MEDS, Northwestern University, Leverone Hall, Evanston, Illinois* 60208

and

David Schmeidler

*Department of Statistics, Tel Aviv University, Tel Aviv* 69978, *Israel*; *and Department of Economics, Ohio State University, Columbus, Ohio* 43210-1172.

Received October 19, 1993

An aspiration level adjustment rule is "realistic" if the aspiration level is (almost always) set to be an average of its previous value and the best average performance so far encountered. It is "ambitious" if the aspiration level is set to exceed the maximal average performance by some constant infinitely often. We analyze a case-based decision maker with a realistic-but-ambitious aspiration level adjustment rule facing a multi-armed bandit repeatedly. Though unaware of the payoff distributions corresponding to the arms of the bandit, the decision maker will asymptotically choose only expected-utility maximizing acts. *Journal of Economic Literature* Classification Numbers: C6, C61, D7, D72, D8, D81, D83. 

## 1. INTRODUCTION

Case-based decision theory (CBDT) is an approach to decisions under uncertainty that emphasizes the role of one's past experiences in decision making. Rather than assuming, as does classical expected utility theory (EUT), that the decision maker (DM) behaves as if (s)he had a probability measure over some state space, with respect to which expected utility is maximized, CBDT uses past "cases" and a similarity function to determine their "relevance" to the decision problem at hand.

It turns out that the mathematical formulation of CBDT calls for a "default value," which is also interpreted as the "aspiration level" of the decision maker.

* We are grateful to Ehud Kalai, Jean-Marc Tallon, Nicolas Vieille, and two anonymous referees for stimulating conversations and comments. The first author gratefully acknowledges partial financial support from the Alfred P. Sloan Foundation.



0899-8256/96 $18.00

Behaviorally it is defined as the level of utility beyond which the DM does not appear to experiment with new alternatives; rather, if this level is attained, (s)he is “satisficed” in the sense of March and Simon (1958).

Thus case-based decision makers are not as “rational” as expected utility ones. They satisfice rather than optimize, they count on their experience rather than attempt to figure out what will be the outcomes of available choices, they follow what appears to be the best alternative in the short run rather than plan for the long run, and they use whatever information they happened to acquire rather than intentionally experiment and learn from a growing experience.

There are many applications for which we find this boundedly rational image of a decision maker quite plausible, at least in comparison with the way decision makers are portrayed by EUT. Especially in novel situations, people often find it hard to specify the space of states-of-the-world in a satisfactory way, let alone to form a prior over it. Many if not most of the decisions taken by governments, for instance, are made in a complex environment which cannot be said to have been encountered before in precisely the same way. History repeats itself, but typically with a twist.

It is therefore not suprising that the rational commandments of EUT, appealing as they are, are hard to follow. One does not have enough trials to figure out all the possible eventualities, not to mention their frequencies. There is little point to invest in learning and experimentation, since the knowledge acquired will be obsolete before it gets to be used; similarly, planning carefully for the long run may prove a futile endeavor.

However, when the environment is more or less “fixed,” and the situation may be modeled as a repeated choice problem, case-based decision makers do appear to be a little too naïve and myopic. While we argue (in Gilboa and Schmeidler, 1995) that CBDT is not designed for these situations to begin with, we show here that the “irrational” or “shortsighted” aspect of CBDT may disappear if one has an appropriate rule for updating one’s aspiration level.

In this paper we propose two properties of aspiration-level adjustment rules, which we find descriptively plausible in general. We show that in the special case of a repeated choice problem, these properties also guarantee optimality. Thus, these properties can also be supported on normative grounds. While there are many rules which may guarantee optimal choice in this special case, we hope to convince the reader that the rules discussed here are also fairly intuitive.

We assume that the aspiration-level adjustment rule is both “realistic” and “ambitious.” Here “realism” means that the aspiration level is set closer to the best average performance. “Ambitiousness” can take one of two forms: it may simply imply that the initial aspiration level is high; alternatively, it may be modeled as suggesting that the aspiration level is “pushed up” from time to time. We devote the next few paragraphs to explaining and motivating these properties.

We model “realism” by assuming an “adaptive” rule, which sets the new aspiration level at some weighted average of the old one and the maximal average performance so far encountered. Thus, if all the acts that were attempted in the past failed to perform up to expectations, the latter would have to be scaled down. Conversely, if the aspiration level is exceeded by some acts' performance, it is gradually increased. (As will be clear in what follows, the specific adjustment rule is immaterial; it is crucial, however, that it is gradually pushing the aspiration level toward the actual best average performance.[1])

While we do not provide an axiomatic derivation of this property, we would like to motivate it from several distinct viewpoints. First, suppose that we read the aspiration level as an answer to the question, What can you reasonably hope for in this problem? If a moderately rational decision maker is to provide the answer, some adaptation of the aspiration level to actual performance seems inevitable. Second, the updating rule appears to be psychologically plausible: people seem to be able to adapt to circumstances. One may wish to distinguish between scaling-up and scaling-down, but the main point is that over the long run the aspiration level is adjusted. A third, related point is that it is in some sense "optimal" to adjust the aspiration level: assume that you can choose an aspiration level for an individual (say, your child), where you care both about his/her "objective" performance and about his/her "subjective" happiness as measured by the aspiration level (cognitively interpreted). When the aspiration level is set too high, decreasing it avoids the subjective psychological cost of constant frustration, and the objective cost of discarding good choices; when it is too low, increasing it avoids objectively sub-optimal choices.

The "ambitiousness" property may have two separate (though compatible) meanings: *static* ambitiousness simply states that the initial aspiration level is relatively high. How high "high enough" is will inevitably depend on the environment. At any rate, a high initial aspiration level reflects the fact that our decision maker is "aggressive" and entertains great expectations. Whether the decision maker's initial aspiration level is high enough will depend on a variety of psychological, sociological, and perhaps also biological factors; while our "optimism" assumption may not be universally true, it is not blatantly implausible either. (See Shepperd, Ouellette, and Fernandez (1996) for related empirical evidence.)

The second meaning the "ambitiousness" assumption may take is *dynamic*, that is, that the decision maker never quite loses hope. Specifically, we will assume that at certain decision periods, the aspiration level is set to exceed the best average performance by a certain constant. In order to make this compatible with realism, we will allow these decision periods to become more and more infrequent. (As a matter of fact, for the optimality result we will require that the update periods have a limit frequency of zero.) That is, the longer one's memory is, the less one tends to increase the aspiration level in this somewhat arbitrary manner. However, dynamic ambitiousness requires that these update

[1] See subsection 4.4 below for technical details.

periods never end. Regardless of the memory's length, a dynamically ambitious decision maker still sometimes stops to ask, Why can't I do better than that?

As in the case of static ambitiousness, the claim of this assumption to descriptive validity can be qualified at best. Indeed, we are not trying to claim that all decision makers are realistic-but-ambitious, just as we do not believe that all people necessarily choose optimal acts in repeated problems. The main point is that the properties of realism and ambitiousness correspond to some general intuition, and they make sense beyond the special case in which a certain problem is encountered over and over again. In this special case, however, they also ensure optimal choice.

The "repeated problem" we discuss in this paper is akin to the "multi-armed bandit problem" (See Gittins, 1979): our decision maker is repeatedly faced with finitely many options, each of which is guaranteed to yield an independent realization of a certain random variable (with finite expectation and variance). Our results are as follows: first assume that the decision maker is realistic and statically ambitious. Then, given the distributions governing the various "arms," there exists a high enough initial aspiration level such that, with arbitrarily high probability, the limit frequency of the expected-utility maximizing acts will be 1. Thus the initial aspiration level depends both on the given distributions and on the desired probability with which this frequency will indeed be 1.

Our second result assumes a decision maker who is realistic but dynamically ambitious. We prove that for all given distributions and any initial aspiration level, the limit frequency of the optimal acts will be 1 with probability 1. This result is therefore stronger than the first, since it guarantees that *almost always* the "best" acts will almost always be chosen and that the same "algorithm" obtains optimality for *all* given distributions. Thus, dynamic ambitiousness is "safer" than static ambitiousness. Roughly, it is more important not to lose hope than to have great expectations.

The intuition behind both results can be explained easily in the deterministic case. Suppose that every time an act is chosen, it yields the same outcome. For the first theorem, assume that the decision maker starts out with a very high aspiration level. Thus all options seem unsatisfactory, and the decision maker switches from one to another, as prescribed by CBDT in case of negative utility values. Specifically, in this case the frequencies of choice are inversely proportional to the utility values. Hence, a high aspiration level prods the decision maker to experiment with all options with similar frequencies. On the other hand, as time goes by the aspiration level is updated toward the best average performance so far encountered. In the deterministic case, the "average performance" of an act is simply its utility value. Thus, in the long run the aspiration level tends to the maximal utility value; correspondingly, an act which achieves this value is almost satisficing and will be chosen with a limit frequency of 1.

Next consider a dynamically ambitious decision maker in a similar setup. Such a decision maker may have started out with "too low" an aspiration level;

thus (s)he may be choosing a suboptimal act, while the aspiration level is being adjusted *upward* toward the utility value of this act. However, if at some point the aspiration level is set above this utility value, this act is no longer satisficing, and the decision maker will try a new one. In the long run all acts will be tried, and the aspiration level will be "realistically" adjusted toward the maximal utility value. As opposed to the case of static ambitiousness, the aspiration level does not converge to this value, since it is pushed above it from time to time. Yet, these periods are assumed to have zero limit frequency, and thus the optimality result holds.

The general cases of both results, in which the available acts yield stochastic payoffs, are naturally more involved, but the proofs follow the same basic intuition.

We note here that both realism and ambitiousness are crucial for the optimality results. If our decision maker is realistic but not ambitious, (s)he may well choose a suboptimal act forever. In this case the choice is random in the following sense: an act is randomly selected at the first stage, and then it is chosen forever. On the other hand, if (s)he is, say, statically ambitious but not realistic, then all choices seem to him/her almost equally unsatisfactory; in this case one may show that the choice is close to random in the sense that all acts will have approximately the same frequency of being chosen. (See Gilboa and Schmeidler (1993).) By contrast, the combination of the two guarantees that all acts will be experimented with, but also that in the long run experimentation will give way to optimal choice.

In a sense, our results may be viewed as explaining the evolution of optimal (expected-utility maximizing) choice: a case-based decision maker who is both realistic and ambitious will "learn" to be an expected-utility maximizer. These results hold only in case the decision problem is repeated long enough in the "same" form. But this is precisely the case in which EUT seems the most plausible, i.e., when history is long enough to enable the decision maker to figure out what are the states of the world, and to form a (frequentist) prior over them. Furthermore, a case-based decision maker is more "open minded" than an expected utility maximizer. While the latter may have *a priori* beliefs whose support fails to contain the true distribution, the former simply does not entertain prior beliefs and thus cannot be wrong about them.

In the context of optimization problems, one may view our results as reinforcing a general principle by which global optimization may be obtained by local optimization coupled with the introduction of "noise." The annealing algorithms (Kirkpatrick *et al.* 1982) are probably the most explicit manifestation of this principle. Genetic algorithms (Holland, 1975) are another example, in which the adaptive process leads to a local optimum of sorts, and the "cross-over" one allows the algorithm to explore new horizons. Yet another example of the same principle may be found in evolutionary models in game theory such as are given by Foster and Young (1990), Kandori *et al.*, (1993), and Young (1993). In these models, a myopic best-response rule may lead to equilibria which are Pareto dominated ("local optima"), even in pure-coordination games. But the introduction of mutations provides the "noise" which guarantees (in such games) a high probability of a Pareto-dominating equilibrium (a "global optimum").

From this viewpoint, one may interpret our results as follows: the "realistic" nature of the aspiration-level adjustment rules induces convergence to a "local optimum," namely, to a high frequency of choice of the "best" acts among those that were tried often enough. The ambitiousness plays the role of the "noise," which prods the decision maker to choose seemingly suboptimal acts and, in the long run, to converge to a global optimum.

The annealing algorithms simulate physical phenomena; genetic algorithms and evolutionary game theory models are inspired by biological metaphors; by contrast, our process is motivated by psychological intuition. As mentioned above, we find this intuition valid beyond the specific model at hand.

The rest of this paper is organized as follows. Section 2 provides a brief overview of CBDT, as well as its alternative interpretation as a dynamic choice theory. It will hopefully serve to orient the reader and motivate the following sections. Section 3 provides the formal model and the main results. Section 4 concludes with some comments and variations. All proofs are relegated to the Appendix.

## 2. BACKGROUND

2.1. *Case-Based Decision-Theory—An Overview.* A full description of CBDT is certainly beyond the scope of this paper. The reader is referred to Gilboa and Schmeidler (1995) for detailed exposition, axiomatizations, variants, and theoretical discussions of CBDT, as well as for comparisons of it to expected utility theory for decision under uncertainty. In this section we will provide only a very sketchy outline of CBDT, which will hopefully suffice for the understanding of the following sections.

The primitives of CBDT are:

$P$—a set of decision *problems*

$A$—a set of available *acts*

$R$—a set of possible *results* (or outcomes)

The set of *cases* is defined to be

$$C = P \times A \times R.$$

That is, a "case" is a triple $(p, a, r)$, where $p$ is the problem encountered, $a$ is the act chosen by the decision maker, and $r$ is the result that was obtained in this case. We will assume that at any given point in time, a decision maker is equipped with some memory $M$, which is simply some subset of cases, and which will be interpreted as the set of problems the decision maker can remember.

CBDT postulates two main theoretical terms—"utility" and "similarity." As in the classical decision theory, the utility measures the desirability of the results and is thus a function

$$u\colon R \to \Re.$$

The notion of "similarity" is new and corresponds in many ways to that of "subjective probability" in expected utility theory. Similarity measures the extent to which one decision problem is similar to another; that is, it is a function

$$s\colon P \times P \to [0, 1].$$

Finally, we may describe the decision rule that is the heart of CBDT: Suppose that a decision maker, characterized by the utility $u$ and the similarity $s$, is faced with a decision problem $p$, while his/her memory is $M \subseteq C$. Then every possible act $a \in A$ is evaluated by the functional

$$U(a) = \sum_{(q,a,r)\in M} s(p, q)u(r),$$

and the decision maker will, according to CBDT, choose a maximizer of $U$.

A few comments are in order. First, notice that for two distinct acts $a, b \in A$, $U(a)$ and $U(b)$ are summations over *disjoint* sets of cases. Furthermore, for some acts this summation may be over an empty set, in which case its value is defined to be zero. This value is going to play a major role in the theory: one may think of it as the decision maker's "aspiration level." To be precise, this is the "default" (utility) value the decision maker seems to be attaching to an act that was never tried in the past (i.e., for which there are no cases in memory). If certain acts obtain higher $U$-value than zero, the decision maker is "satisficed" and will continue to choose among them without trying new acts and *without trying to maximize* $u$. Once all the acts that were tried in the past turned out to be unsatisfactory—that is, to have negative $U$ values—then the decision maker will choose a new act (assuming such exists), where the choice among these will be arbitrary.

In the formulation above, the aspiration level is implicitly assumed to be zero, where the utility function is correspondingly normalized. Since this paper focuses on the process by which the aspiration level is updated, it will be convenient to explicitly mention it. Let $H$ be the aspiration level, and redefine the functional $U$ as

$$U(a) = U_{p,M}(a) = \sum_{(q,a,r)\in M} s(p, q)[u(r) - H],$$

where $a \in A$.

One of the main features of CBDT is that it does not require the DM to "engage" in hypothetical reasoning: as opposed to expected utility thoery, where the very definition of an "act" involves hypothetical statements such as "If state $\omega$ occurs then I get $r$," in CBDT all the DM is required to "know" is the history of cases which *actually happened* and the utility he/she *actually experienced*. (The terms "engage" and "know" above are within quotation marks since one may choose a purely behavioral interpretation of the theory, according to which the DM does not have to "know" or to reason about anything.)

Without details we mention here that the decision rule of CBDT, together with the theoretical terms "utility" and "similarity," may be axiomatically derived from preferences, in a way which parallels the axiomatic derivations of "utility" and "probability," combined with the expected utility formula, in models such as Savage's (1954). (See Gilboa and Schmeidler, 1995, for one such axiom system, as well as additional discussions.)

The notion of a "case" will sometimes be interpreted in a broader fashion. For instance, a case in a decision maker's memory need not necessarily have been experienced by the same DM. It may well be a "story" told by someone else. Furthermore, it need not be a real case—it may be a hypothetical one, reflecting the DM's knowledge (or belief) about what would have occurred as a result of a possible choice.

Finally, let us briefly mention two variants of the basic CBDT model:

—*Averaged similarity*. Here one uses a functional similar to $U$ above, with the sole difference that for each act $a \in A$, the similarity coefficients $s(p, q)$ are normalized to sum up to 1. We denote this functional by $V$. Normalizing the aspiration level to be zero, it is defined by

$$V(a) = \sum_{(q,a,r)\in M} s'(p, q)u(r),$$

where

$$s'(p, q) = \frac{s(p, q)}{\sum_{(q',a,r)\in M} s(p, q')}$$

whenever the latter is well-defined (and zero otherwise).

—*Act similarity*. According to this model, acts may also be similar to each other, and the evaluation of an act $a$ depends not only on its own performance in the past, but also on that of similar acts. Thus, the similarity function is defined over problem-act pairs such that an act $a$ is evaluated by

$$U'(a) = U'_{p,M}(a) = \sum_{(q,b,r)\in M} s((p, a), (q, b))u(r)$$

(again, assuming $H = 0$).

In Gilboa and Schmeidler (1995) we axiomatize the first variant. The second is axiomatized in Gilboa and Schmeidler (1994).

2.2. *A Dynamic Theory of Choice*. As explained above, CBDT attempts to deal with decision problems under uncertainty, for which very little information is available. Thus, the decision maker's memory plays a major role in attempting to forecast the outcomes of various acts. Yet the same mathematical model may be interpreted in a different way, as proposed in Gilboa and Schmeidler (1993): rather than serving as a source of information, one's memory may enter one's utility directly. According to this interpretation, an act's desirability, whether under certainty or uncertainty, intrinsically depends on the previous cases in which it was chosen.

Thus the function $u$ has a slightly different interpretation: rather than the "utility" of an outcome, which is to be maximized by a supposedly rational decison maker, it is merely some derivative (with respect to time) of the utility $U$, which is by definition a memory-dependent aggregate. Thus, if $u$ is negative, the desirability of an act is lower the more it has been chosen. This may be taken to model boredom-averse, or change-seeking decision makers. Conversely, should $u$ be positive, an act is more desirable the more it has been experienced, thus exhibiting choice patterns which are consistent with habit formation. (See Gilboa and Schmeidler, 1993, for details.)

For the purpose of the present paper it will be useful to bear both interpretations in mind, since the optimality rule (CBDT combined with realism and ambitiousness) is motivated by some hybrid of the two. We discuss this point in Section 4.

## 3. MODEL AND RESULTS

Let $A = \{1, 2, \dots, n\}$ be a set of *acts* ($n \geq 1$). For $i \in A$ let there be given a distribution $F_i$ on $\Re$ (endowed with the Borel $\sigma$-algebra), to be interpreted as the (conditional) distribution of the utility yielded by act $i$'s whenever it is chosen. We assume that $F_i$ has finite expectation and variance, denoted $\mu_i$ and $\sigma_i$, respectively.

The underlying state space will be a subset of

$$S_0 = (\Re \times A \times \Re)^N,$$

where $N$ denotes the natural numbers. A state $\omega = ((H_1, a_1, x_1), (H_2, a_2, x_2), \dots) \in S_0$ will be interpreted as follows: for all $t \geq 1$, at period $t$, the aspiration level is $H_t$ at the beginning of the period, an act $a_t$ is chosen, and it yields a payoff of $x_t$. It will be convenient to define, for every $t \geq 1$, the projection functions

$$H_t, x_t\colon S_0 \to \Re \qquad \text{and} \qquad a_t\colon S_0 \to A$$

with the obvious meaning.

Next we define a function $C\colon S_0 \times A \times N \to 2^N$ to be the set of periods, up to a given time, at which a given act was chosen, according to a given state. That is,

$$C(\omega, i, t) = \left\{ j < t \mid a_j(\omega) = i \right\}.$$

We will similarly be interested in the number of times a certain act was chosen. Therefore, define a function $K\colon S_0 \times A \times N \to N \cup \{0\}$ by

$$K(\omega, i, t) = \#C(\omega, i, t).$$

We are mostly interested in the relative frequencies of the decision maker's choices. It will be convenient to define a function $f\colon S_0 \times A \times N \to [0, 1]$ to measure relative frequency up to a given time, i.e.,

$$f(\omega, i, t) = \frac{K(\omega, i, t)}{t}.$$

Dropping the time index will refer to the limit:

$$f(\omega, i) = \lim_{t \to \infty} f(\omega, i, t).$$

Finally, we will further abuse this notation by extending it to subsets of $A$: for $D \subseteq A$ we define

$$f(\omega, D, t) = \sum_{i \in D} f(\omega, i, t)$$

and

$$f(\omega, D) = \lim_{t \to \infty} f(\omega, D, t).$$

We now turn to define the CBDT functionals. Let $U\colon S_0 \times A \times N \to \Re$ be defined by

$$U(\omega, i, t) = \sum_{j \in C(\omega, i, t)} \left[ x_j(\omega) - H_t(\omega) \right].$$

We will also use the notation

$$V(\omega, i, t) = \frac{U(\omega, i, t)}{K(\omega, i, t)}.$$

(Thus, "$V(\omega, i, t)$ is well-defined" means "$K(\omega, i, t)$ is positive.") Since the values of both $U$ and $V$ depend on the aspiration level $H_t$, it will prove convenient to have a separate notation for the absolute average utility of each act. We denote

$$\overline{X}(\omega, i, t) = \frac{\sum_{j \in C(\omega,i,t)} x_j(\omega)}{K(\omega, i, t)}.$$

Note that $\overline{X}(\omega, i, t)$ is well-defined whenever $V(\omega, i, t)$ is and

$$\overline{X}(\omega, i, t) = V(\omega, i, t) + H_t(\omega).$$

We now wish to express the fact that the decision maker considered is a $U$-maximizer. We do this by restricting the state space as follows: define $S_1 \subseteq S_0$ by

$$S_1 = \left\{ \omega \in S_0 \mid a_t(\omega) \in \arg\max_{i \in A} U(\omega, i, t),\ \ \forall t \geq 1 \right\}.$$

Similarly, we further restrict the state space to reflect the fact that the aspiration level is updated in an adaptive manner. First define, for $t \geq 2$ and $\omega \in S_0$, the relative and absolute maximal average performance to be, respectively,

$$\overline{V}(\omega, t) = \max \{V(\omega, i, t) \mid i \in A,\ \ K(\omega, i, t) > 0\}$$

and

$$\overline{X}(\omega, t) = \max \left\{ \overline{X}(\omega, i, t) \mid i \in A,\ \ K(\omega, i, t) > 0 \right\}.$$

Next, for a given $\alpha \in (0, 1)$ and $H_1 \in \Re$ define the state space to be

$$\Omega = \Omega(\alpha, H_1) = \left\{ \omega \in S_1 \;\middle|\; \begin{array}{l} H_1(\omega) = H_1, \text{ and } \forall t \geq 2 \\ H_t(\omega) = \alpha H_{t-1}(\omega) + (1 - \alpha)\overline{X}(\omega, t) \end{array} \right\}.$$

Endow $S_0$ with the $\sigma$-algebra generated by the Borel $\sigma$-algebra on (each copy of) $\Re$ and $2^A$ on (each copy of) $A$. Let $\Sigma = \Sigma(\alpha, H_1)$ be the induced $\sigma$-algebra on $\Omega$. Finally, we turn to define the underlying probability measure. Given $\Omega$ and $\Sigma$, a probability measure $P$ on $\Sigma$ is *consistent with* $(F_i)_{i \in A}$ if for every $t \geq 1$ and $i \in A$, the conditional distribution of $x_t$ that it induces, given $a_t = i$, is $F_i$, and, furthermore, $x_t$ is independent (according to $P$) of the random variables $H_1, a_1, x_1, \ldots, H_{t-1}, a_{t-1}, x_{t-1}, H_t$. Notice that distinct measures on $\Sigma$, which are consistent with $(F_i)_{i \in A}$, can disagree only regarding the choice of an act $a_t$ where $\arg\max_{i \in A} U(\omega, i, t)$ is not a singleton.

We can finally formulate our first result:

THEOREM 1. *Let there be given* $A = \{1, \ldots, n\}$, $(F_i)_{i \in A}$ *as above,* $\alpha \in (0, 1)$ *and* $\varepsilon > 0$. *There exists* $H_0 \in \Re$ *such that for all* $H_1 \geq H_0$ *and every measure* $P$ *on* $(\Omega(\alpha, H_1), \Sigma(\alpha, H_1))$ *which is consistent with* $(F_i)_{i \in A}$,

$$P\left(\left\{ \omega \in \Omega \mid \exists f(\omega, \arg\max_{i \in A} \mu_i) = 1 \right\}\right) \geq 1 - \varepsilon.$$

Thus the theorem guarantees that, if we focus on those states $\omega$ at which there is a limit choice frequency for the set of expected utility maximizers, and it is 1, this set is measurable and has arbitrarily high probability provided the initial aspiration level is high enough.

Note that Theorem 1 cannot guarantee an aspiration level which is uniformly large enough for all given distributions $(F_i)_{i\in A}$. Indeed, it is obvious that any initial aspiration level may turn out to be too low. By contrast, our second result guarantees optimality for all given distributions, regardless of the initial aspiration level and with probability 1. The assumption which drives this much stronger conclusion is that the aspiration level is "pushed up" every so often. That is, that at a certain set of periods, which is infinite but sparse (i.e., has a zero limit frequency), the aspiration level is not adjusted by averaging its previous value and the best-average-performance value; rather, at these periods it is set to be at some level above the best-average-performance value, regardless of the previous aspiration level.

Formally, we define a new probability space as follows. Let there be given $H_1 \in \Re$ and $\alpha \in (0, 1)$ as above. Assume that $N_{\mathrm{A}} \subseteq N$ and $h > 0$ are given. $N_{\mathrm{A}}$ is interpreted as the set of periods at which the decision maker is ambitious. The number $h$ should be thought of as the aspiration-level increase. Define

$$\begin{aligned}\Omega &= \Omega(\alpha, H_1, N_{\mathrm{A}}, h)\\ &= \left\{ \omega \in S_1 \;\middle|\; \begin{array}{ll} H_1(\omega) = H_1, \text{ and } \forall t \geq 2 & \\ H_t(\omega) = \overline{X}(\omega, t) + h & \text{if } t \in N_{\mathrm{A}} \\ H_t(\omega) = \alpha H_{t-1}(\omega) & \\ \qquad\qquad + (1-\alpha)\overline{X}(\omega, t) & \text{if } t \notin N_{\mathrm{A}} \end{array} \right\}.\end{aligned}$$

Next, define $\Sigma = \Sigma(\alpha, H_1, N_{\mathrm{A}}, h)$ to be the corresponding $\sigma$-algebra. Similarly, a measure $P$ on $\Sigma$ is defined to be *consistent with* $(F_i)_{i\in A}$ as above.

We can now state:

THEOREM 2. *Let there be given* $A = \{1, \dots, n\}$, $(F_i)_{i\in A}$ *as above,* $\alpha \in (0, 1)$, $H_1 \in \Re$, $N_{\mathrm{A}} \subseteq N$, *and* $h > 0$. *If* $N_{\mathrm{A}}$ *is infinite but sparse, then for every measure* $P$ *on* $(\Omega(\alpha, H_1, N_{\mathrm{A}}, h), \Sigma(\alpha, H_1, N_{\mathrm{A}}, h))$ *which is consistent with* $(F_i)_{i\in A}$,

$$P\left(\left\{\omega \in \Omega \mid \exists f(\omega, \arg\max_{i\in A} \mu_i) = 1\right\}\right) = 1.$$

## 4. DISCUSSION

4.1. As briefly mentioned above, the adjustment rule is a hybrid of sorts: our decision makers choose acts by $U$-maximization; however, when it comes to adjusting their aspiration levels, they use the maximal $V$ value. This apparent inconsistency calls for an explanation.

Recall that, as described in Section 2 above, memory affects one's decisions in two ways: first, as a source of information, which is especially crucial for decisions under uncertainty; second, as a primary effect in a dynamic choice situation. Memory helps one to reason about the world, but also changes one's tastes.

Thus, there are two fundamental questions to which memory is key: first, "What do I want to do now?" and second, "What do I think of this act?" In answering the first question, memory plays a dual role: as a source of information and as a factor affecting preferences; in answering the second, memory only serves as a source of information. Correspondingly, we would like to suggest that $U$ offers an answer to the first, while $V$ answers the second.

Consider the following example: every day our decision maker has to choose a restaurant; this is a repeated choice, which may be thought of as decision under certainty or under uncertainty. The restaurant chosen will be a $U$-maximizer, allowing such behavior patterns as habit formation and boredom aversion. However, suppose our decision maker has a guest and is asked by him/her which is the best restaurant in town, namely, which restaurant should one go to if one has only one day to spend there (with no memory). Then, according to this interpretation, the decision maker will recommend a $V$-maximizing, rather than a $U$-maximizing act. Asked why (s)he is not choosing this restaurant him/herself, the decision maker may say, "Oh, I was there just yesterday." Having visited it recently, its $U$ value may have decreased (if our decision maker is change-seeking); however, the very fact it was recently chosen need not change its $V$ value.

The optimality rule discussed in his paper is therefore not as inconsistent as it may appear at first glance: our decision makers are $U$-maximizers in their choices. This means that memory enters their decision considerations not only as a source of information. With a high aspiration level, this also allows them to keep switching among the alternative acts and to continue "trying" acts whose past average performance happened to be poor. On the other hand, asking themselves, "What can I reasonably hope for?" or "What would I recommend to someone who hasn't tried any of the options?" they base their answer on $V$-maximizing acts. As we have shown, adjusting their aspiration level based on the maximal $V$ value also colors past experiences differently. In the long run, the dissatisfaction with $V$-maximizing acts decreases, and thus their relative frequency tends to 1.

4.2. Some readers will probably not be convinced by the above arguments. It certainly makes sense to consider two simpler alternatives, namely the "$U$-rule," which prescribes that decisions will be made so as to maximize $U$ and that the aspiration level will be adjusted according to $U$ as well, and the corresponding "$V$-rule." It is worth mentioning, however, that neither of these rules seems to guarantee optimal choice in the long run.[2]

4.3. The discussion in this paper focuses on finitely many alternatives. Indeed, with infinitely many alternatives CBDT, combined with a high aspiration level, does not make much sense in its original formulation. Furthermore, it certainly does not guarantee optimality: since every act has a default value (of the aspiration level), the decision maker will keep trying new (and arbitrarily chosen) acts indefinitely.

However, it is rarely the case that infinitely many acts are available without having some additional structure. For instance, prices and quantities may be modeled as continua, but then they are endowed with a natural metrizable topology. These cases are naturally modeled as CBDT with act similarity. (See subsection 2.1 above.) For instance, having set a price at \$20, a seller may have some idea about the outcomes that are likely to result from a price of \$20.01. Since these two acts are "similar," the past experience with one of them enters the evaluation of another.

Thus, given a metric topological space of acts and a similarity function (which is, say, monotonic in the metric), and assuming continuity of $u$, one would expect a similar optimality result to hold.

4.4. It almost goes without saying that our results do not hinge on the specific aspiration-level adjustment rule. First, the aspiration level need not be adjusted at every period, nor do the adjustment periods have to be deterministically set. All that is required is that there will be infinitely many of them with a high enough probability. Similarly, the "realistic" adjustment need not be done by a weighted average (with fixed weights). Generally, for Theorem 1 it is required only that (i) the adjustment process will guarantee convergence, i.e., that for all $a, b \in \Re$ and $\varepsilon > 0$, if $X(\omega, t) \in (a, b)$ for all $t \geq T_1$ for some $T_1$, then there will exist $T_2$ such that for almost all $t \geq T_2$, $H_t(\omega) \in (a - \varepsilon, b + \varepsilon)$; and (ii) the adjustment will not be too fast, i.e., that for all $R \in \Re$ and all $T_0 \geq 1$ there will be a number $H_0$ such that for all $H_1 > H_0$ and all $t \leq T_0$, $H_t > R$. For Theorem 2, one needs the convergence property and an increase in aspiration level over an infinite but sparse set of periods. Neither $h$ nor the set $N_{\mathrm{A}}$ need be

[2] Using the "$U$-rule," the aspiration level need not converge. (As a matter fact, it is not obvious what is the "right" way to define the aspiration level adjustment rule in this case.) Using the "$V$-rule," the decision maker may never retry certain alternatives which happened to have particularly low realizations in the first few periods. (We omit the simple examples.)

deterministic or exogneously given. Both may depend of the state $\omega$, on past acts, and on their results. It is essential, however, that for almost all $\omega$, $h$ is bounded away from zero (and not too large) and that $N_{\mathrm{A}}$ is infinite but sparse. Finally, one may assume that in the "ambitious" periods, the aspiration is set so as to exceed (by $h$) its own previous value, rather than the maximal average performance level.

4.5. Note that when the aspiration level is updated in our model, the $u$ value of past experiences is also updated. That is, outcomes which have been obtained in the past are re-evaluated according to the newly defined aspiration level. Thus we implicitly assume that the decision makers can "reflect" upon the outcomes themselves, sometimes realizing that they were not as unsatisfacotry as they seemed at the time.

Alternatively, one may assume that only the utility value of past experiences is retained in memory and that the original evaluation of an outcome will be forever used to judge the act which led to it. However, our first result does not hold in this case, since a very high initial aspiration level may make an expected-utility maximizing act have a very low $U$ value, to the extent that it may never be chosen again.

While one may argue for the psychological plausibility of the alterantive assumption, it seems that it is "more rational" to re-evaluate outcomes based on an adjusted aspiration level, rather than compare each outcome to a possibly different aspiration level. At any rate, the second result holds under the alternative assumption as well: having infinitely many periods in which the expected utility of any act is a negative number bounded away from zero guarantees that all acts will be chosen infinitely often with probability 1.

## APPENDIX: PROOFS

1. *Proof of Theorem* 1. A few words on the strategy of the proof are probably in order. The general idea is very similar to the deterministic case described in the Introduction: let the initial aspiration level be high enough so that each act is chosen a large enough number of times, and then notice that the aspiration level tends to the maximal expected utility. In the deterministic case, each act should be chosen at least once in order to get its average performance $\overline{X}$ equal to its expectation. In the stochastic case, more choices are needed, and a law of large numbers will be invoked for a similar conclusion. Thus the initial aspiration level should be high enough to guarantee that each of the acts is chosen enough times to get the average close to the expectation.

If the supports of the given distributions $F_i$ were bounded, one could find high enough aspiration levels such that all possible realizations of all possible choices seem similarly unsatisfactory. This would guarantee that, as long as the aspiration level is beyond a certain bound, all acts are chosen with similar frequencies, and therefore all of them will be chosen enough times for the law of large numbers to apply. However, these distributions need not have a bounded support. They are known only to have a finite variance. Thus the proof is slightly more involved, as we explain below.

Let us first assume w.l.o.g. (without loss of generality) that for some $r \leq n$,

$$\mu_1 = \mu_2 = \cdots = \mu_r > \mu_{r+1} \geq \mu_{r+2} \geq \cdots \geq \mu_n.$$

Furthermore, we assume that $r < n$ w.l.o.g. (the theorem is trivially true otherwise). Next denote

$$I = \arg\max_{i \in A} \mu_i = \{1, 2, \ldots, r\}$$

and

$$\delta = \frac{\mu_1 - \mu_{r+1}}{3}.$$

The number $\delta$ is so chosen that, if the average values are $\delta$-close to the corresponding expectations, then the maximal average value is obtained by a maximizer of the expectation.

We now turn to find the number of times which is needed to guarantee, with high enough probability, that the averages are, indeed, $\delta$-close to the expectations. Given $\varepsilon > 0$ as in the theorem and $i \in A$, let $K_i \geq 1$ be such that: for every $k \geq K_i$ and every sequence of i.i.d. random variables $X_i^1, X_i^2, \ldots, X_i^k$, each with distribution $F_i$,

$$\Pr\left(\left|\frac{1}{k}\sum_{j=1}^{k} X_i^j - \mu_i\right| \leq \delta\right) \geq (1-\varepsilon)^{1/2n},$$

where Pr is the measure induced by the distribution $F_i$. Notice that such $K_i$ exists by the strong law of large numbers (See, for instance, Halmos, 1950). Let $K = \max_{i \in A} K_i$.

We now turn to the construction of the initial aspiration level. As explained above, we would like to be able to assume that the $F_i$'s have bounded supports, in order to guarantee that each act is chosen at least $K$ times. We will therefore find an event with a high enough probability, on which the random variables $x_t$ are, indeed, bounded.

We start by finding, for each $i \in A$, bounds $\underline{b_i}, \overline{b_i} \in \Re$ such that, for any random variable $X_i$ distributed by $F_i$,

$$\Pr(\underline{b_i} \leq X_i \leq \overline{b_i}) \geq (1-\varepsilon)^{1/(4nK)},$$

where Pr is some probability measure which agrees with $F_i$. Notice that such bounds exist since $F_i$ has a finite variance. W.l.o.g. assume also that $\overline{b_i} > \mu_i + 2\delta$ for all $i \in A$. Next define

$$\underline{b} = \min_{i \in A} \underline{b_i} \qquad \text{and} \qquad \overline{b} = \max_{i \in A} \overline{b_i}.$$

The critical lower bound on the aspiration level (for the "experimentation" period, in which every act is chosen at least $K$ times) is chosen to be

$$R = 2\overline{b} - \underline{b}.$$

Let us define, for every $T \geq 1$, the event

$$B_T = \left\{ \omega \in \Omega \mid \forall t \leq T,\ \underline{b} \leq x_t(\omega) \leq \overline{b} \right\}.$$

Notice that, since the given measure $P$ is consistent with $(F_i)_{i \in A}$, $P(B_T) \geq (1 - \varepsilon)^{T/(4nK)}$. Hence, provided that $T$ is not too large, $B_T$ will have a high enough probability. In order to show that $T$ need not be too large to get enough ($\geq K$) observations of each act, we first show that, on $B_T$ and with sufficiently high aspiration level, the first $T$ choices are more or less "evenly" distributed among the acts:

CLAIM 1. *Let there be given $T \geq n$, and $\omega \in B_T$. Assume that for all $t \leq T$, $H_t(\omega) > R$. Then for all $i, j \in A$ and all $n \leq t \leq T$,*

$$K(\omega, i, t) \leq 2K(\omega, j, t).$$

*Proof.* Assume the contrary, and let $t_0$ be the minimal time $t$ such that $n \leq t \leq T$ and

$$K(\omega, i, t_0) > 2K(\omega, j, t_0)$$

for some $i, j \in A$. Notice that $K(\omega, a, n) = 1$ for all $a \in A$, and hence $t_0 > n$. It follows from minimality of $t_0$ that $a_{t_0-1}(\omega) = i$, i.e., that $i$ was the last act chosen.

Consider the following bounds on the $U$ values of the two acts:

$$U(\omega, i, t_0 - 1) \leq K(\omega, i, t_0 - 1)(\overline{b} - H_{t_0-1}(\omega))$$

and

$$U(\omega, j, t_0 - 1) \geq K(\omega, j, t_0 - 1)(\underline{b} - H_{t_0-1}(\omega)).$$

The optimality of $i$ at stage $t_0 - 1$ implies

$$U(\omega, i, t_0 - 1) \geq U(\omega, j, t_0 - 1);$$

hence

$$K(\omega, i, t_0 - 1)(\overline{b} - H_{t_0-1}(\omega)) \geq K(\omega, j, t_0 - 1)(\underline{b} - H_{t_0-1}(\omega)).$$

Recalling that $H_{t_0-1}(\omega) > R \geq \overline{b} \geq \underline{b}$, this is equivalent to

$$\frac{K(\omega, j, t_0 - 1)}{K(\omega, i, t_0 - 1)} \geq \frac{\overline{b} - H_{t_0-1}(\omega)}{\underline{b} - H_{t_0-1}(\omega)}.$$

By minimality of $t_0$ we know that

$$\frac{K(\omega, j, t_0 - 1)}{K(\omega, i, t_0 - 1)} = \frac{1}{2}.$$

We therefore obtain

$$\underline{b} - H_{t_0-1}(\omega) \leq 2\left(\overline{b} - H_{t_0-1}(\omega)\right),$$

which implies

$$H_{t_0-1}(\omega) \leq 2\overline{b} - \underline{b} = R,$$

a contradiction.

We now set $T_0 = 2nK$ and will prove that—as long as the aspiration level is kept above $R$—after $T_0$ stages, each act will be chosen at least $K$ times on the event $B_{T_0}$. Formally,

CLAIM 2. *Let there be given $\omega \in B_{T_0}$ and assume that $H_t(\omega) > R$ for all $t \leq T_0$. Then for $i \in A$,*

$$K(\omega, i, T_0) \geq K.$$

*Proof.* If $K(\omega, i, T_0) < K$ for some $i \in A$, then by Claim 1, $K(\omega, j, T_0) < 2K$ for all $j \in A$. Then we get

$$T_0 = \sum_{j \in A} K(\omega, j, T_0) < 2nK = T_0,$$

which is impossible.

We finally turn to choose the required level for the initial aspiration level. Choose a value

$$H_0 = H_0(\varepsilon) > \underline{b} + 2\left(\frac{1}{\alpha}\right)^{T_0} (\overline{b} - \underline{b})$$

and let us assume for the rest of the proof that $H_1 \geq H_0$. We verify that this bound is sufficiently high in the following:

CLAIM 3. *Let there be given* $\omega \in B_{T_0}$ *and assume that* $H_1 \geq H_0$. *Then for all* $t \leq T_0$, $H_t(\omega) > R$.

*Proof.* For all $1 < t \leq T_0$,

$$H_t(\omega) \geq \alpha H_{t-1}(\omega) + (1-\alpha)\underline{b}$$

or

$$H_t(\omega) - \underline{b} \geq \alpha(H_{t-1}(\omega) - \underline{b}).$$

Hence

$$H_t(\omega) - \underline{b} \geq \alpha^t(H_1 - \underline{b}) > 2\alpha^t\left(\frac{1}{\alpha}\right)^{T_0}(\overline{b} - \underline{b}) \geq 2(\bar{b} - b)$$

and

$$H_t(\omega) > 2\overline{b} - \underline{b} = R.$$

Combining the above, we conclude that, for $H_1 \geq H_0$, $K(\omega, i, T_0) \geq K$ for all $\omega \in B_{T_0}$ and all $i \in A$. Furthermore, for a measure $P$, consistent with $(F_i)_{i\in A}$,

$$P(B_{T_0}) \geq (1-\varepsilon)^{T_0/(4nK)} = (1-\varepsilon)^{1/2}.$$

We now define the event on which the limit frequency of the expected-utility maximizing acts is 1: let $B \subseteq B_{T_0}$ be defined by

$$B = \left\{\omega \in B_{T_0} \;\middle|\; \begin{array}{l} \forall t \geq T_0, \quad \forall i \in A, \\ \left|\overline{X}(\omega, i, t) - \mu_i\right| < \delta \end{array}\right\}.$$

By the choice of $K$ and the independence assumption, we conclude that $P(B \mid B_{T_0}) \geq (1-\varepsilon)^{1/2}$, whence $P(B) \geq (1-\varepsilon)$.

The proof of the theorem will therefore be complete if we prove the following:

CLAIM 4. *Assume that* $H_1 \geq H_0$ *and let* $P$ *be a measure on* $(\Omega(\alpha, H_1), \Sigma(\alpha, H_1))$ *which is consistent with* $(F_i)_{i\in A}$. *Then, for* $P$*-almost all* $\omega$ *in* $B$,

$$\exists f(\omega, I) = 1.$$

(*Recall that* $I = \arg\max_{i\in A}\mu_i$.)

*Proof.* Given $\omega \in B$ and $\xi > 0$, we wish to show that, unless $\omega$ is in a certain $P$-null event (to be specified later), there exists a $T = T(\omega, \xi)$ such that for all $t \geq T$,

$$f(\omega, I, t) \geq 1 - \xi.$$

It is sufficient to find a $T = T(\omega, \xi)$ such that for some $i \in I$, for all $t \geq T$, and for all $j \notin I$,

$$\frac{K(\omega, j, t)}{K(\omega, i, t)} = \frac{f(\omega, j, t)}{f(\omega, i, t)} \leq \frac{\xi}{n(1-\xi)}.$$

We remind the reader that for all $t \geq T_0$ and all $a \in A$ we have

$$\left|\overline{X}(\omega, a, t) - \mu_a\right| \leq \delta.$$

Also, since $H_{T_0}(\omega) > R > \mu_1$, for all $t \geq T_0$ we have

$$H_t(\omega) > \mu_1 - \delta = \mu_{r+1} + 2\delta.$$

That is, the aspiration level will be adjusted toward the average performance of one of the expected-utility maximizing acts and will be bounded away from the expected utility and from the average performance value of suboptimal acts.

We will need a uniform bound on $H_t(\omega)$. To this end, note that for all $a \in A$ and $t \leq T_0$, $\overline{X}(\omega, a, t) < R$, by definition of the set $B_{T_0}$. For $t \geq T_0$, the same inequality holds since $\overline{X}(\omega, a, t) < \mu_a + \delta < \overline{b_a} \leq R$. Since $H_{t+1}(\omega)$ is a convex combination of $H_t(\omega)$ and $\overline{X}(\omega, t) = \max_{a\in A} \overline{X}(\omega, a, t) < R$, we conclude that for all $t \geq 1$, $H_{t+1}(\omega) \leq \max\{H_t(\omega), R\}$. By induction, it follows that for all $t \geq 1$ $H_t(\omega) \leq H_1$.

Let $O(\omega) \subseteq A$ be the set of acts which are chosen infinitely often at $\omega$. That is,

$$O(\omega) = \left\{ a \in A \mid K(\omega, a, t) \xrightarrow[t\to\infty]{} \infty \right\}.$$

We would first like to establish the fact that some expected-utility maximizing acts are indeed chosen infinitely often. Formally,

CLAIM 4.1. $O(\omega) \cap I \neq \varnothing$.

*Proof.* Let $\tilde{T} \geq T_0$ be such that for all $t \geq \tilde{T}$, $a_t(\omega) \in O(\omega)$. Assume the contrary, i.e., that $O(\omega) \cap I = \varnothing$. (In particular, $a_t(\omega) \notin I$ for all $t \geq \tilde{T}$.) For all $t \geq \tilde{T} \geq T_0$ we also know that

$$\overline{X}(\omega, j, t) < H_t(\omega) - \delta$$

for all $j \notin I$. Hence, for $j \notin I$,

$$\begin{aligned} U(\omega, j, t) &= K(\omega, j, t)V(\omega, j, t) \\ &= K(\omega, j, t)\left[\overline{X}(\omega, j, t) - H_t(\omega)\right] < -\delta K(\omega, j, t). \end{aligned}$$

This implies that $U(\omega, j, t) \xrightarrow[K(\omega,j,t)\to\infty]{} -\infty$. Thus, for all $j \in O(\omega)\backslash I$,

$$U(\omega, j, t) \xrightarrow[t\to\infty]{} -\infty.$$

On the other hand, consider some $i \in I \subseteq (O(\omega))^c$. Let $L$ satisfy $L > K(\omega, i, t)$ for all $t \geq 1$. Then

$$\begin{aligned} U(\omega, i, t) &= K(\omega, i, t)V(\omega, i, t) \\ &= K(\omega, i, t)\left[\overline{X}(\omega, i, t) - H_t(\omega)\right] > L(\underline{b} - H_1). \end{aligned}$$

It is therefore impossible that only members of $I^c$ would be $U$-maximizers from some $\tilde{T}$ on.

We now assume that for all $a \in O(\omega)$, $\overline{X}(\omega, a, t) \xrightarrow[t\to\infty]{} \mu_a$. By the strong law of large numbers, this is the case for all $\omega \in B$ apart from a $P$-null set.

Choose $\zeta > 0$ such that

$$\zeta < \frac{\xi\delta}{6n(1-\xi)}$$

and let $T_1 \geq T_0$ be such that for all $t \geq T_1$ and all $i \in O(\omega) \cap I$,

$$\left|\overline{X}(\omega, i, t) - \mu_1\right| < \zeta.$$

For all $t \geq T_1$ we also conclude that

$$\left|\overline{X}(\omega, t) - \mu_1\right| < \zeta$$

(where, as above, $\overline{X}(\omega, t) = \max_{a\in A} \overline{X}(\omega, a, t)$). It follows that the aspiration level, $H_{t+1}(\omega)$, which is adjusted to be some average of its previous value $H_t(\omega)$ and $\overline{X}(\omega, t)$, will also converge to $\mu_1$. To be precise, there is $T_2 \geq T_1$ such that for all $t \geq T_2$,

$$|H_t(\omega) - \mu_1| < 2\zeta.$$

We wish to show that there exists $T(\omega, \xi)$ such that for all $t \geq T(\omega, \xi)$, all $i \in O(\omega) \cap I$, and all $j \notin I$ the following holds:

$$\frac{K(\omega, j, t)}{K(\omega, i, t)} \leq \frac{\xi}{n(1-\xi)}.$$

It will be helpful to start with:

CLAIM 4.2. *For all $t \geq T_2$, all $i \in O(\omega) \cap I$, and all $j \notin I$, if $a_t(\omega) = j$, then*

$$K(\omega, j, t) < \frac{\xi}{2n(1-\xi)} K(\omega, i, t).$$

*Proof.* Let there be given $t$, $i$, and $j$ as above. Observe that

$$\begin{aligned} U(\omega, i, t) &= K(\omega, i, t)V(\omega, i, t) \\ &= K(\omega, i, t)\left[\overline{X}(\omega, i, t) - H_t(\omega)\right] \geq -3K(\omega, i, t)\zeta, \end{aligned}$$

while

$$\begin{aligned} U(\omega, j, t) &= K(\omega, j, t)V(\omega, j, t) \\ &= K(\omega, j, t)\left[\overline{X}(\omega, j, t) - H_t(\omega)\right] \leq -K(\omega, j, t)\delta. \end{aligned}$$

The fact that $a_t(\omega) = j$ implies that $U(\omega, j, t) \geq U(\omega, i, t)$. Hence

$$-K(\omega, j, t)\delta \geq -3K(\omega, i, t)\zeta$$

or

$$K(\omega, j, t) \leq \frac{3\zeta}{\delta} K(\omega, i, t).$$

However, the choice of $\zeta$ (as smaller than $\xi\delta/6n(1-\xi)$) implies that

$$\frac{3\zeta}{\delta} < \frac{\xi}{2n(1-\xi)}.$$

We have thus established that

$$K(\omega, j, t) < \frac{\xi}{2n(1-\xi)} K(\omega, i, t)$$

for any $t$ at which $j$ is chosen (i.e., $a_t(\omega) = j$).

We proceed as follows: let $T_3 \geq T_2$ be such that for all $t \geq T_3$, $a_t(\omega) \in O(\omega)$. Let $T_4 \geq T_3$ be large enough so that for all $t \geq T_4$, $a \in O(\omega)$ and $c \notin O(\omega)$,

$$K(\omega, c, t) \leq \frac{\xi}{n(1-\xi)} K(\omega, a, t).$$

Finally, let $T_5 > T_4$ be such that for all $a \in O(\omega)$, $K(\omega, a, T_5) > K(\omega, a, T_2)$. We now have

CLAIM 4.3. *For all $t \geq T_5$, all $i \in O(\omega) \cap I$, and all $j \notin I$,*

$$K(\omega, j, t) \leq \frac{\xi}{n(1-\xi)} K(\omega, i, t).$$

*Proof.* Let there be given $t$, $i$, and $j$ as above. If $j \notin O(\omega)$, the choice of $T_4$ concludes the proof. Assume, then, that $j \in O(\omega)$. Then, by choice of $T_5$, $j$ has been chosen since $T_2$. That is,

$$T_j^t \equiv \{s \mid T_2 \le s < t,\ a_s(\omega) = j\} \ne \emptyset.$$

Let $s$ be the last time at which $j$ was chosen before time $t$, i.e., $s = \max T_j^t$. Note that

$$K(\omega, j, t) = K(\omega, j, s+1)$$

and

$$K(\omega, i, t) \ge K(\omega, i, s+1).$$

Hence it suffices to show that

$$K(\omega, j, s+1) \le \frac{\xi}{n(1-\xi)} K(\omega, i, s+1).$$

By Claim 4.2 we know that

$$K(\omega, j, s) \le \frac{\xi}{2n(1-\xi)} K(\omega, i, s).$$

Since $s \ge T_2 \ge T_0$, $K(\omega, j, s) \ge K \ge 1$. This implies $(\xi/2n(1-\xi))K(\omega, i, s) \ge 1$. Next, observe that

$$\begin{aligned} K(\omega, j, s+1) &= K(\omega, j, s) + 1 \le \frac{\xi}{2n(1-\xi)} K(\omega, i, s) + 1 \\ &\le \frac{\xi}{2n(1-\xi)} K(\omega, i, s) + \frac{\xi}{2n(1-\xi)} K(\omega, i, s) \\ &= \frac{\xi}{n(1-\xi)} K(\omega, i, s) = \frac{\xi}{n(1-\xi)} K(\omega, i, s+1). \end{aligned}$$

This concludes the proof of Claim 4.3.

Thus $T_5$ may serve as the required $T(\omega, \xi)$. As a matter of fact, our claim regarding $T_5$ is slightly stronger than that we need to prove regarding $T(\omega, \xi)$. The latter should have the inequality of Claim 4.3 satisfied for *some* $i \in I$, while the former satisfies it for *all* $i \in O(\omega) \cap I$, and Claim 4.1 guarantees that this set indeed contains some $i \in I$.

At any rate, Claim 4.3 completes the proof of Claim 4, which, in turn, completes the proof of the theorem. ∎

2. *Proof of Theorem* 2. The general idea of the proof, as well as the proof itself, is quite simple: as long as the aspiration level is close to the average performance of an expected-utility maximizing act, the proof mimics that of Theorem 1. The problem is that the decision maker may "lock in" on suboptimal acts, which may be almost-satisficing or even satisficing, and not try the optimal acts frequently enough. However, the fact that the decision maker is "ambitious" infinitely often (in the sense of setting the aspiration level beyond the maximal average performance) guarantees that this will not be the case. Thus, the fact that $N_{\mathrm{A}}$ is infinite ensures that every act will be chosen infinitely often. On the other hand, the fact that it is sparse implies that these periods of "ambitiousness" will not change the limit frequencies obtained in the proof of Theorem 1.

In the formal proof it will prove convenient to take the following steps: we will restrict our attention to the event at which all acts, which are chosen infinitely often, have a limit average performance equal to their expectation. On this event we will show that the expected-utility maximizers among those acts have a limit choice frequency of 1. Finally, we will show that all acts are chosen infinitely often, whence the result follows.

We adopt some notation from the proof of Theorem 1. In particular, assume that for some $r < n$,

$$\mu_1 = \mu_2 = \cdots = \mu_r > \mu_{r+1} \geq \mu_{r+2} \geq \cdots \geq \mu_n,$$

and denote

$$I = \arg\max_{i \in A} \mu_i = \{1, 2, \ldots, r\}.$$

We will also use

$$O(\omega) = \left\{ a \in A \mid K(\omega, a, t) \xrightarrow[t \to \infty]{} \infty \right\}$$

and the new notation

$$I(\omega) = \arg\max \{\mu_i \mid i \in O(\omega)\}.$$

We would like to focus on the event

$$B = \left\{ \omega \in \Omega \mid \forall i \in O(\omega), \overline{X}(\omega, i, t) \xrightarrow[t \to \infty]{} \mu_i \right\}.$$

Since $A$ is finite, the strong law of large numbers guarantees that $P(B) = 1$ for any consistent $P$. Thus it suffices to show that for every $\omega \in B$, $f(\omega, I) = 1$. We do this in two steps: we show first that $f(\omega, I(\omega)) = 1$, and then that $I(\omega) = I$.

CLAIM 1. *For all* $\omega \in B$, $\exists f(\omega, I(\omega)) = 1$.

*Proof.* Let there be given $\omega \in B$, and denote $\mu = \mu_i$ for some $i \in I(\omega)$. Given the proof of Claim 4 in Theorem 1, it suffices to show that for every $\zeta > 0$, $|H_t(\omega) - \mu| < \zeta$ holds for all $t \notin N_0$ where $N_0 \subset N$ is sparse.

Let $\zeta > 0$ be given, and assume w.l.o.g. that $\zeta < \delta = (\mu - \mu_i)/3$ for all $i \notin I(\omega)$ and that $\zeta < h$. Let $T_1$ be such that for all $t \geq T_1$ and all $i \in O(\omega)$,

$$\left|\overline{X}(\omega, i, t) - \mu_i\right| < \zeta/2.$$

Let $T_2 \geq T_1$ be such that for all $t \geq T_2$, $i \in O(\omega)$, and $j \notin O(\omega)$, $\overline{X}(\omega, i, t) > \overline{X}(\omega, j, t)$. Thus, for $t \geq T_2$, if $t \notin N_{\mathrm{A}}$, $H_t(\omega)$ is adjusted "toward" $\overline{X}(\omega, t)$ which equals $\overline{X}(\omega, i, t)$ for some $i \in O(\omega)$, where the latter is close to $\mu$. Since for $t \in N_{\mathrm{A}}$, $H_t(\omega)$ is set to $\overline{X}(\omega, t) + h$, there exists $T_3 \geq T_2$ such that for all $t \geq T_3$,

$$|H_t(\omega) - \mu| < 2h.$$

We now wish to choose a number $k$, such that any sequence of $k$ periods following $T_3$, at which $H_t(\omega)$ is adjusted "realistically," i.e., as an average of $H_{t-1}(\omega)$ and $\overline{X}(\omega, t)$, will guarantee that it ends up $\zeta$-close to $\mu$.

Let $k > \log_\alpha(\zeta/4h)$. Define

$$N_A \oplus k = \left\{ t \in N \;\middle|\; \begin{array}{l} t = t_1 + t_2 \quad \text{where} \\ t_1 \in N_A \text{ and } 0 \leq t_2 \leq k \end{array} \right\}.$$

Note that for $t \geq T_3$, if $t \notin N_A \oplus k$, i.e., if $t$ is at least $k$ periods after the most recent "ambitous" update, we have

$$|H_t(\omega) - \mu| < \zeta.$$

Setting $N_0 = (N_{\mathrm{A}} \oplus k) \cup \{1, \ldots, T_2\}$ (and noting that it is sparse) completes the proof.

CLAIM 2. *For all $\omega \in B$, $I(\omega) = I$.*

*Proof.* It suffices to show that $O(\omega) = A$ for all $\omega \in B$. Assume, to the contrary, that for some $\omega \in B$, $j \in A$ and $L \geq 1$, $K(\omega, j, t) \leq L$ for all $t \geq 1$. Let $i \in O(\omega)$. For any $t \in N_{\mathrm{A}}$,

$$\begin{aligned} U(\omega, i, t) &= K(\omega, i, t)V(\omega, i, t) \\ &= K(\omega, i, t)\left[\overline{X}(\omega, i, t) - H_t(\omega)\right] < -hK(\omega, i, t). \end{aligned}$$

Let $T_4 \geq T_3$ be such that for all $t \geq T_4$, $a_t(\omega) \neq j$. Recall that for all $t \geq T_4$, $H_t(\omega) < \mu + 2h$. Consider $t \in N_{\mathrm{A}}$ such that $t \geq T_4$. Then

$$\begin{aligned} U(\omega, j, t) &= K(\omega, j, t)V(\omega, j, t) \\ &= K(\omega, j, t)\left[\overline{X}(\omega, j, t) - H_t(\omega)\right] > -LC, \end{aligned}$$

where $C = \mu + 2h - \overline{X}(\omega, j, T_3)$. That is, $U(\omega, j, t)$ is bounded from below. Since for a large enough $t \in N_A$, $U(\omega, i, t)$ is arbitrarily small for all $i \in O(\omega)$, we obtain a contradiction to $U$-maximization. Thus we conclude that $O(\omega) = A$. This concludes the proof of the claim and the theorem. ∎

## REFERENCES


FOSTER, D., AND YOUNG, H. P. (1990). "Stochastic Evolutionary Game Dynamics," *Theoret. Popul. Biol.* **38**, 219–232.

GILBOA, I., AND SCHMEIDLER, D. (1995). "Case-Based Decision Theory," *Q. J. Econ.* **110**, 605–639.

GILBOA, I., AND SCHMEIDLER, D. (1993). "Case-Based Consumer Theory," Discussion Paper no. 1025. Northwestern University.

GILBOA, I., AND SCHMEIDLER, D. (1994). "Act Similarity in Case-Based Decision Theory," forthcoming in *Econ. Theory*.

GITTINS, J. C. (1979). "Bandit Processes and Dynamic Allocation Indices," *J. Roy. Statis. Soc. B* **41**, 148–164.

HALMOS, P. R. (1950). *Measure Theory*. Princeton, NJ: Van Nostrand.

HOLLAND, J. H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor: Univ. of Michigan Press.

KANDORI, M., MAILATH, G. J., AND ROB, R. (1993). "Learning, Mutation and Long-Run Equilibria in Games," *Econometrica* **61**, 29–56.

KIRKPATRICK, S., GELLATT, C. D., AND VECCHI, M. P. (1982). "Optimization by Simulated Annealing," in *IBM Thomas J. Watson Research Center*. Yorktown Heights, NY.

MARCH, J. G., AND SIMON, H. A. (1958). *Organizations*. New York: Wiley.

SAVAGE, L. J. (1954). *The Foundations of Statistics*. New York: Wiley.

SHEPPERD, J. A., OUELLETTE, J. A., and FERNANDEZ, J. K. (1996). "Abandoning Unrealistic Optimism: Performance, Estimates and the Temporal Proximity of Self-Relevant Feedback," *J. Personality Social Psychol.* **70**, 844–855.

YOUNG, H. P. (1993). "The Evolution of Conventions," *Econometrica* **61**, 57–84.