# *The B.E. Journal of Theoretical Economics*

## Advances

| *Volume* 9, *Issue* 1 | 2009 | *Article* 9 |
| --- | --- | --- |

# Satisficing: A 'Pretty Good' Heuristic

Jonathan Brodie Bendor* Sunil Kumar†

David A. Siegel‡

*Stanford University, bendor_jonathan@gsb.stanford.edu

†Stanford University, kumar_sunil@gsb.stanford.edu

‡Florida State University, dsiegel@fsu.edu

**Recommended Citation**

Jonathan Brodie Bendor, Sunil Kumar, and David A. Siegel (2009) "Satisficing: A 'Pretty Good' Heuristic," *The B.E. Journal of Theoretical Economics*: Vol. 9: Iss. 1 (Advances), Article 9.
Available at: http://www.bepress.com/bejte/vol9/iss1/art9

# Satisficing: A ‘Pretty Good’ Heuristic*

Jonathan Brodie Bendor, Sunil Kumar, and David A. Siegel

## Abstract

One of the best known ideas in the study of bounded rationality is Simon’s satisficing; yet we still lack a standard formalization of the heuristic and its implications. We propose a mathematical model of satisficing which explicitly represents agents’ aspirations and which explores both single-person and multi-player contexts. The model shows that satisficing has a signature performance-profile in both contexts: (1) it can induce optimal long-run behavior in one class of problems but not in the complementary class; and (2) in the latter, it generates behavior that is sensible but not optimal. The model also yields empirically testable predictions: in certain bandit-problems it pins down the limiting probabilities of each arm’s use, and it provides an ordering of the arms’ dynamical use-probabilities as well.

**KEYWORDS:** satisficing, heuristics, bounded rationality

*We would like to thank Thom Baguley, Jose Bermudez, Dan Carpenter, Mandeep Dhami, Itzhak Gilboa, Dan Goldstein, Dale Griffin, Dave Kreps, Duncan Luce, Howard Margolis, Joel Sobel, Richard Thaler and seminar participants at the Graduate School of Business, Stanford University, for their helpful comments.

# I. Introduction

Although Simon's seminal papers on satisficing (Simon 1955, 1956) were published half a century ago, many properties of this famous heuristic remain unknown. This is a strange scientific situation. Bounded rationality is regarded by many economists and other social scientists as an important way to think about decision making, and the best known idea in this research program is probably that of satisficing. (Typing 'satisficing' in Google-Scholar produces about 16,800 hits, nearly half as many as the more general term of 'bounded rationality' (35,100 hits).)

This paper formalizes Simon's theory and analyzes the dynamical as well as the asymptotic properties of satisficing, in both decision-theoretic and multi-agent contexts. The model shows that although the heuristic does not converge to optimal behavior for all choice problems, it does yield 'sensible' behavior: e.g., the *ex ante* probabilities of choosing actions are ordered just as are the actions' success probabilities (the chance that a payoff meets an aspiration level). This holds both in the limit and, under some conditions, dynamically as well. Further, satisficing exhibits adaptively rational comparative static properties: the more likely an action's payoff is to meet the agent's aspiration level, the more often it will be chosen.

These adaptive features are not produced by reformulating the heuristic to make it work better: we take Simon's verbal formulation as a descriptive theory and let the chips—satisficing's performance—fall where they may.

The present paper is most closely related to Karandikar et al (1998) and Cho and Matsui (2005). Karandikar et al assume satisficing agents with endogenous aspirations. By focusing on a class of 2x2 games (including the prisoner's dilemma), they can prove a sharp result: if aspirations adjust sufficiently slowly and are subject to trembles, the players will cooperate most of the time. In-Koo Cho and Matsui also study endogenous aspirations, which evolve as a player's average payoffs. This specific functional form and an ingenious use of stochastic approximation methods enable them to characterize the players' behavior and aspirations for all symmetric 2x2 games.

Gilboa and Schmeidler's case-based theory (1995) is somewhat more distant in both motivation and setup, but their agent's behavior does have satisficing-like properties—search only if dissatisfied with the option in hand. Further, they note (p.624-625) that this search behavior will not invariably converge to optimality in repeated choice contexts; this is one of our themes as well.

Satisficing theory has also been used to study politics: for a satisficing model of party competition see Bendor, Mookherjee and Ray (2006); for one on turnout see Collins, Kumar and Bendor (forthcoming). Finally, Shor (2004) has examined satisficing empirically in nonstationary environments. The heuristic outperforms several others in predicting the behavior of experimental subjects. (However, discriminating among adaptive heuristics is difficult [Salmon 2001].)

The rest of the paper is organized as follows. Section II introduces general assumptions used throughout the paper. Section III examines decision theoretic contexts with exogenous aspirations and stationary payoff environments. It presents two different kinds of results. Proposition 1 characterizes the class of choice problems where satisficing can induce optimal behavior and the class for which it cannot. Propositions 2-4 examine the heuristic's sensible though non-optimal properties in the second class of problems. Together, these results display a central feature of satisficing: a mixed performance profile. Section IV extends this theme into nonstationary environments. Section V shows that satisficing continues to exhibit its signature mixed performance profile when aspirations are endogenous. Section VI examines multi-person situations with exogenous aspirations. Proposition $1'$, which parallels Proposition 1, and Proposition 7 identify and characterize settings in which satisficing can induce optimal behavior in multiple agents, as well as contexts in which it cannot. Proposition 8 examines the heuristic's sensible behavior in 2x2 games. Section VII concludes.

## II. General Assumptions

We study agents who makes choices in periods $t = 1, 2, 3, \ldots$; agent $i$ gets payoff $\pi_{i,t}$ in $t$. Because the following assumptions hold throughout the paper, they will not be restated in the propositions.

- *Finitely Many Decision Makers:* The set of agents is denoted by $N = \{1, \ldots, n\}$, with $n \geq 1$.
- *Finite Action Sets:* All action sets are finite, and each agent has at least two feasible actions. We denote the number of possible actions, common to all agents, $m$.
- *Unique Best Actions/Responses:* Each player has a unique best response to every fixed vector of partners' actions. Hence in single person settings the agent has a unique optimal action. Adapting standard game-theoretic terminology, we call such situations *weakly generic*. (All generic games are weakly generic but not conversely: e.g., pure coordination games.)

We also assume throughout that agents satisfice; because this is our main topic we make it an explicit assumption in every result. We define satisficing formally in the next section. Informally, an agent *satisfices* if her behavior has two properties: (1) if her current action yields a payoff that meets or exceeds her aspiration level then she continues to use that action; (2) if her payoff is below her aspiration level then she might search for a new action. These and the other assumptions together constitute a stochastic process; its states are the actions available to each agent.[1] Accordingly, we often phrase results in terms of *stable* (i.e., absorbing) *states*.

A few comments about the above assumptions are in order.

Assuming finite action sets enhances tractability while giving up little theoretical insight or empirical leverage. Because these sets can be extremely large and our measurement technologies are imperfect, a discernible difference between "extremely large but finite" and "infinite" is rare.

Assuming unique best responses eliminates indifference, which for present purposes is a distraction. It can also be justified on robustness grounds. For several actions to generate *exactly* the same payoff is a knife-edge condition: indifference would be broken by even the slightest payoff-shock.

It is worth noting that many of our results on multi-person contexts hold for asymmetric as well as symmetric games.

Finally, in most of the paper we assume that agents operate in a stationary environment: the set of players, their action sets, and the mappings from actions to payoffs are all fixed. This is a conventional assumption in most game theoretic models of repeat play and also in many psychological theories of adaptation, e.g., reinforcement learning.

Later we relax this assumption and investigate how satisficing performs when the environment can change in important, payoff-relevant ways. We do this partly to expand the models' empirical domain: though some task environments are at least approximately stationary, some are not. Further, an important theoretical issue is at stake: the putative robustness of fast and frugal heuristics (Todd and Gigerenzer 2000, p.736-37). It has been hypothesized (Gigerenzer 2001) that although a fast and frugal heuristic may be suboptimal in a fixed environment, it may perform well in the more challenging context of changing environments. We examine this claim by embedding satisficing in a nonstationary environment in section IV.

---

[1]When aspirations are endogenous then the state space includes aspirations as well as actions. Because we focus mostly on the latter, however, we abuse terminology and refer to actions (or action-vectors, for $n > 1$) as the states.

## III. Decision Theoretic Contexts with Exogenous Aspirations

Choice under uncertainty often involves a tension between exploration and persistence (March 1991). On the one hand, if an agent does not know which action is best then she will usually need to explore the problem space. Then restlessness—a willingness to search—is valuable. On the other hand, restlessness may also imply that the agent cannot "settle down" on the optimal action (what we call persistence) once she has stumbled across it.

This section focuses on how this tension plays out in a variety of settings. We find that problems may be classified as either well- or ill-matched to the satisficing heuristic. In the former, satisficing can resolve the tension between exploration and persistence; the heuristic is *preadapted* to problems in this set. It cannot resolve this tension for problems in the ill-matched set.

Consider an agent with four alternatives: $w$, $x$, $y$ and $z$. Option $w$ delivers poor payoffs with certainty, $x$ gives poor or fair payoffs, $y$'s are fair or good, and $z$'s are always good. The agent knows none of this and proceeds by satisficing. Suppose that her aspiration level were exogenously set to equal a fair payoff. Then a poor outcome could trigger search while the other two outcomes would not, so the agent might be satisfied with the suboptimal $y$. In contrast, if her aspiration level were set to good then only $z$ would always satisfy her, so the optimal result would obtain.

This example suggests that what a satisficing agent codes as satisfactory—her aspiration level—can strongly affect the outcome. This is partly due to a deterministic feature of the heuristic: if her aspiration level is set to fair, she will regard fair outcomes as satisfactory *with certainty*. (If the agent could be dissatisfied with fair payoffs then she could not get stuck on the suboptimal $y$.) This point leads directly to our assumption about satisficing under exogenous aspirations.

**(A1):** The agent has an aspiration level, $a$, such that for all $t$ (i) if $\pi_t \geq a$ then she satisfices, i.e., uses the same action in $t+1$, and (ii) if $\pi_t < a$ then she searches for a new option in $t+1$ with a probability of $\theta > 0$.

(A1) formalizes this paper's key ideas: satisficing (part (i)) and search triggered by dissatisfaction (part (ii)).[2]

[2] Extending (A1) to multi-attribute choice problems is straightforward: replace $a$ and $\pi_t$ by the vectors $(a_1, \ldots, a_k)$ and $(\pi_1, \ldots, \pi_k)$; satisficing occurs when $\pi_j \geq a_j$ for all $j = 1, \ldots, k$. This formalizes Simon's verbal theory: "Aspiration levels provide a computational mechanism for satisficing. An alternative satisfices if it meets aspirations along all dimensions" (Simon 1996, p.30).

Returning to our $(w, x, y, z)$ example, note that a decision maker using a rule that satisfies (A1) with sufficiently demanding requirements can avoid becoming permanently trapped on a suboptimal action. Specifically, an aspiration level above 'fair' and below 'good' provides just the right mix of exploration and persistence to yield optimality, eventually.

But achieving this mix is not always easy. Consider a second example: a two-arm bandit problem. The left arm pays off with a fixed probability $p < 1$; the right, with probability $q < p$. If the machine pays off then it yields a set amount of money; otherwise it gives nothing. Unlike the first example, here the optimal action sometimes fails, i.e., yields nothing. We clarify this important distinction by the following definition: an action is **perfect** if it gives the agent his maximum feasible payoff with certainty. Otherwise it is **imperfect**.

The restlessness that was so useful in avoiding a suboptimal result in the first example is the agent's downfall in this one: the imperfection of the optimal arm and the dynamic generated by (A1) conspire against securing the optimal result. Decision problems where all actions are imperfect are generally much harder for satisficing rules.

To see this more formally, suppose (A1) holds. Consider the class of problems where the set of payoffs, $\Pi$, is compact, with the minimal payoff denoted $\underline{\pi}$ and the maximal, $\overline{\pi}$. Focus first on solving the exploration problem, so let her aspiration equal $\overline{\pi}$. Then it is easily established that no suboptimal action is stable, and even the optimal action is stable if *and only if* it is perfect. Hence, conquering the exploration problem via demanding aspirations makes it impossible for satisficing-type rules to stabilize on the optimal action when that alternative is imperfect. Exploration is maximized at the expense of persistence. (This property continues to hold if the decision maker can get inaccurate feedback about payoffs. Indeed, assuming—as we do throughout—that feedback is accurate makes it *harder* to establish that the optimal action is unstable.)

Now focus on solving the persistence problem: let the agent's aspiration equal his minimal payoff, $\underline{\pi}$. Then it is to easy to show that agents with this aspiration level and who adapt via (A1) master the art of persistence too well: the lack of exploration ensures that *all* actions—suboptimal as well as optimal—are stable. Persistence is maximized at the expense of exploration.

Since one extreme produces insufficient exploration and the other, insufficient persistence, one naturally wonders whether satisficing based on intermediate aspirations would lead eventually to optimization. This guess is half-right: it holds for some but not all choice settings. Our first proposition draws a bright line between these two sets. We define this bright line now.

**Definition 1:** The decision problem of a weakly generic situation is **well-matched** to the satisficing heuristic if the optimal action has a minimum payoff and it exceeds the minimum payoff of every other action. All other problems are called **ill-matched**.

The following are examples of each type. In problem 1 action $x$ is equally likely to give \$ 0 or \$ 2; action $y$ gives \$1 or \$2 with equal odds. In problem 2 each action delivers \$ 0 or \$ 1 with some probability. Problem 1 is well-matched to satisficing: the optimal action's minimal payoff exceeds the inferior action's. Problem 2 is ill-matched: the actions' minimal payoffs are the same. (If the payoff distributions are unbounded then no action has a minimal payoff; such problems are ill-matched to satisficing by Definition 1.)

The content of Proposition 1 justifies Definition 1's partitioning of problems into those that are well- and ill-matched to satisficing and reveals why problems that satisfy the definition's payoff-property merit the honorific of "well-matched". (The proof is in the appendix.[3])

**Proposition 1:** If the agent uses a satisficing rule that satisfies (A1), then there exists an aspiration level such that the optimal action is the unique stable state if *and only if* the problem is well-matched to the heuristic.

A corollary about ill-matched problems follows immediately from Proposition 1: if the optimal action is absorbing then so is some other action. Hence the process might get trapped in a suboptimal state. Equivalently, if no suboptimal action is absorbing then neither is the best one.

The distinction between well- and ill-matched problems helps us to understand how satisficing rules handle risky choices. Suppose in a two-armed bandit problem one alternative is risky, giving either $x$ or $y > x$; the other is riskless, yielding $z$ for sure. The problem is more interesting if neither alternative dominates, so assume $x < z < y$. If the riskless option is optimal then the problem is well-matched to the heuristic: any exogenous aspiration in $(x, z)$ ensures that the agent will always be satisfied by the optimal arm, yet sometimes the inferior arm will dissatisfy and so will trigger search. Hence if the agent is lucky enough to have an aspiration in this interval then he cannot get trapped on the suboptimal alternative; he will settle down on the optimal one. But *if the risky arm is best, then the problem is ill-matched to satisficing*. In that case, no aspiration level can fine-tune exploration and persistence: either both options are always satisfying or neither is.

[3] Proposition 1 presumes satisficing with a stationary search probability, as described in (A1). The proof establishes a stronger result which does not require this kind of stationarity.

In some contexts persistence is impossible. Suppose that payoffs are normally distributed. Then for any exogenously fixed aspiration, dissatisfaction is always possible. In such environments an agent who adapts via (A1) will play suboptimal actions infinitely often with probability one.

**Satisficing's Sensible Features**

Having established via Proposition 1 that satisficing's long-run behavior is intimately tied to problem-difficulty, we still must explore what happens when a problem is ill-matched to satisficing and the heuristic does not converge to optimizing in the limit. We begin this subsection by examining satisficing's short-run properties, i.e., its dynamics given ill-matched problems, turning later to its asymptotic behavior. Doing both will establish that satisficing rules behaves sensibly: they have several intuitively desirable performance properties in a variety of settings.

We begin by examining a large and common class of problems that, as noted earlier, are ill-matched to satisficing: those with unbounded payoffs. We will show that in a subset of these problems, the heuristic nevertheless behaves sensibly.

Two factors complicate the analysis of dynamics: initial conditions and the degree of inertia. Initial conditions can obviously affect the short term; they can make a satisficer appear to behave foolishly, but the real culprit is the initial conditions. Let the probability of choosing action $k$ at $t$ be $p_{k,t}$, and suppose that to qualify as a reasonable heuristic a rule should ensure that success probabilities determine the ordering of choice-probabilities: $p_{k,t}$ should exceed $p_{l,t}$ for all $t$ if and only if $k$ produces a payoff higher than the exogenous aspiration level more often than $l$ does. But guaranteeing that early choice-probabilities line up as do success probabilities is impossible for arbitrary starting distributions. For example, suppose that options $k$ and $l$ have success probabilities of 0.8 and 0.9, respectively. If the agent starts off completely disposed toward $k$, then $p_{k,t}$ will exceed $p_{l,t}$ in periods 1-3 for any $\theta$. Because initial biases are peripheral to our main concerns, we suppress them when analyzing dynamics by assuming that the decision maker begins without predispositions: $p_{k,0} = \frac{1}{m}$ for all $k$. Call this a **neutral start**.

Inertia is more central to our analysis. It is well-known that dynamical systems can be ill-behaved if they are insufficiently inertial. Satisficing's dynamics have this feature. Hence, several of the next proposition's results depend on this parameter. To state this dependence crisply, in the next result $\delta$ denotes $1 - f_1\theta - f_2\theta$, where $f_i$ $(i = 1, 2)$ is the probability that arm $i$ fails, i.e., gives a payoff below the (fixed) aspiration level.

**Proposition 2:** Consider a two-armed bandit $(m = 2)$ played by an (A1)-

type satisficer. The payoff distributions $\Pi_1$ and $\Pi_2$ are unbounded and differentiable. If $\Pi_1$ strictly first-order stochastically dominates $\Pi_2$ and the start is neutral then the following conclusions obtain for any exogenously fixed aspiration level.

- **(i)** *Ex ante*, for all $t > 0$, $p_{1,t} > p_{2,t}$.
- **(ii)** For all $t > 0$, $\frac{\partial p_{i,t}}{\partial f_i} \leq 0$; except for a few knife-edge cases (identified by the proof) the inequality is strict.
- **(iii)**
  - **(a)** If $\delta > 0$ then the propensity to use the better arm monotonically increases over time: *ex ante*, $p_{1,t+1} > p_{1,t}$ for all $t > 0$.
  - **(b)** If $\delta < 0$ then the *ex ante* propensity to use the better arm oscillates but increases "on the whole": [i] $p_{1,2t+1} > p_{1,2t}$; [ii] $p_{1,2t+2} < p_{1,2t+1}$; [iii] $p_{1,2t+2} > p_{1,2t}$ for any $t > 0$.

Proposition 2 brings into sharp relief what we mean by "sensible" or "good" performance. A decision rule in this context is sensible if its probabilities of selecting different actions are always ordered exactly as the actions' success chances are ranked (part (i)), and if an action degrades then the rule chooses that action less often (part (ii)). These properties codify what we intuitively desire from any adaptive rule. Satisficing does rather well vis-a-vis these criteria in the short-run in the two-action setting. When first-order stochastic dominance holds, Simon's conjecture is correct.

Mostly, anyway. Part (iii) depends, via $\delta$, on the frequency of errors and the probability of search. These produce an important dynamical property of satisficing. When failures are common, a restless decision maker is too reactive. Hence, her probability of picking the better arm *oscillates* around the asymptotic value of $\frac{f_2}{f_1+f_2}$. This is analogous to the behavior of an underdamped spring, as in driving a car with bad shock absorbers over a bumpy road. Indeed, it exemplifies a general feature of dynamical systems. We suspect that many kinds of feedback-driven adaptation exhibit this oscillation when a condition on a $\delta$-type parameter does not hold.

This oscillation occurs if and only if $1 - \theta f_1 - \theta f_2 < 0$. Thus when $\theta < \frac{1}{2}$, i.e., the satisficer is sufficiently inertial, the probability of choosing the better arm rises monotonically to its steady state value (part (iii)(a)). This too is sensible performance, though it is a less compelling criterion than those in parts (i) and (ii).

As the inequality in (i) is strict at every date and the $p_t$'s are continuous in the initial probabilities, this result is not knife-edged: the conclusion holds as long as $p_{1,0}$ and $p_{2,0}$ are sufficiently close. Since there are only two actions, (ii) also implies that the probability of choosing one of the actions is in every period increasing in the other action's failure rate.

The role of the neutral start in part (i) is in some respects unsurprising: naturally, an initial bias toward the inferior action can persist for a while. It is more interesting that oscillation yields counterexamples where the agent is initially biased toward the *superior* action. For example, if $f_1 < f_2$, $f_1, f_2 \approx 1$, $\theta = 1$ and $p_{1,0} = 1$, the agent will nearly always use action 2 in the following period. Though she mostly returns to using action 1 next, it is not certain. Once again, this prolonged decrease in the optimal action's likelihood arises from the agent's insufficient persistence, thus highlighting inertia's short-term importance (in addition to the effect of initial conditions).

Whether the properties reported by Proposition 2 hold for more than two actions is an open question; thus far, computational work (using more restrictive parameters and functional forms) has generated no counterexamples. We can, however, analyze the multi-alternative ($m > 2$) case if we restrict attention to asymptotics and put some additional structure on search.

When the agent has only two alternatives, search must lead to a unique option. When there are more than two actions one must make additional assumptions about search. As this is a study of satisficing and not of the peculiarities of different types of search[4] we try to eliminate the possibility that the search rule itself prevents the agent from discovering the optimal alternative. (A2) does this by ensuring that all options are connected by search, given a long enough string of failures. Many search rules have this property. For example, if the agent searches by random walk—if option $k$ fails today then tomorrow she tries $k - 1$ or $k + 1$—then (A2) holds.

(A2) uses the following notation: $q_{i,j}$ denotes the (stationary) probability that the process transitions from action $i$ to action $j$, conditional on the agent searching.[5]

**(A2):** (i) If all actions are imperfect then all states (actions) communicate: for any pair of actions $i$ and $r$, there is a sequence $i, i + 1, \ldots, r - 1, r$ such

[4]Recall that in the bounded rationality program, satisficing is a heuristic, not a strategy. Because it isn't a complete action-plan, it needn't specify how search should be conducted. Other heuristics can do that.

[5]That is, suppose in period $t$ the agent gets an unsatisfactory payoff from action $i$. With probability $1 - \theta$ she uses the same action in $t + 1$ anyway. With probability $\theta$ she searches. Given search, with probability $q_{i,j}$ she transitions to action $j$. Hence, the probability of transitioning from $i$ to $j$, if currently dissatisfied with $i$, is $\theta q_{i,j}$.

that $q_{i,i+1} \cdot \cdots \cdot q_{r-1,r} > 0$. (ii) The transition probabilities $q_{i,j}$ are stationary.

Before examining satisficing's steady-state properties we must first establish that we are investigating a well-behaved (ergodic) process: one with a unique limiting distribution and which converges to that limit from any starting point—any vector of initial probabilities over the system's states. It is straightforward to show that (A1) and (A2) ensure that the corresponding Markov chain is ergodic when $m \geq 2$, the payoff distributions are unbounded, and the aspiration level is exogenously fixed. So we can proceed to examine satisficing's asymptotic features.

For this purpose (A2) by itself is too general: it allows for biased search rules which, since they are reflected in transition probabilities, transmit their bias into the limiting distribution. Suppose, e.g., options $x$, $y$ and $z$ succeed with probabilities 0.5, 0.4 and 0.3, respectively. Because $x$ is best, a sensible decision rule would use it most often in the limit. However, biased search can make $z$, the worst alternative, the most likely in the limit, through no fault of satisficing. For example, if the agent always tries $z$ after a disappointing payoff at either $x$ or $y$ but tries $x$ or $y$ with equal probability after failing with $z$, then $z$'s limiting probability is 0.437 but $x$'s is only 0.307.

To avoid obscuring our understanding of satisficing's consequences, we require search to be unbiased in the following sense. Imagine that the agent is a gambler who faces $m$ slot machines arranged in a circle. Then Markovian search is unbiased if the probability distribution over new options, conditional on the agent's being dissatisfied, is unaffected by his current location. For example, search starting from machine $y$ is the same as search starting from $x$: it is merely shifted over to the new starting point. Visually, the search distribution looks the same from every vantage point. (A3) gives the formal definition of unbiased search. Many search rules—e.g., random walks with states arranged in a circle, so that $q_{1,m} > 0$ and $q_{m,1} > 0$—satisfy both (A2)(i) and (A3).

**(A3):** For a given integer $s$, the probability of a transition between states $k$ and $k + s \pmod m$ upon search, $q_{k,k+s}$, is the same for all $k$.

By invoking unbiased search we can offer a fairly strong defense of satisficing's long-run reasonableness when $m > 2$. (Because Proposition 3 analyzes what happens in the limit, presuming a neutral start is unnecessary; the result does not use that assumption.)

**Proposition 3:** Consider an (A1)-type satisficer with $m \geq 2$ actions. The payoff distributions are unbounded and differentiable. If (A2) and (A3) hold and the payoff distributions can be ordered by first-order stochastic dominance—$\pi_k \succ \pi_{k+1}$ for all $k = 1, \ldots, m-1$—then the following properties hold in the limit for any fixed aspiration level:

- **(i)** An action's steady-state probability is inversely proportional to its failure rate: for all $k$ and $l$, $\tilde{p}_k > \tilde{p}_l$ if and only if $f_k < f_l$. Further, $\tilde{p}_k = \frac{1}{f_k}c^{-1}$, where $c = \sum_{j=1}^{m} \frac{1}{f_k}$.
- **(ii)** $\frac{\partial \tilde{p}_k}{\partial f_k} < 0$ for all $k$.

Note that this specified environment is, by our definition, ill-matched to the satisficing heuristic. Hence, no exogenously fixed aspiration can induce the agent to settle down on the optimal action. Nevertheless, the result shows that satisficing ensures success-ordering in this environment.

Note that the steady state probabilities are independent of $\theta$, the chance that a dissatisfied agent searches. Thus, inertia has no effect in the limit: being more likely to leave a state due to low inertia is canceled out by being more likely to enter it when other states fail.

The exact expression for the steady-state probabilities given in (i) implies that, in the limit, the *relative* likelihood of two options, $\frac{\tilde{p}_k}{\tilde{p}_l}$, is just $\frac{f_l}{f_k}$. Hence, satisficing endogenously generates an interesting and normatively desirable relation between the best action's success rate and its probability of use: *the likelihood of optimizing rises as doing so matters more.* Actions that fail at nearly the same rate will be used similarly in the limit, but an action that is far better than others will be used far more often. And this intuitively appealing property arises endogenously. (Compare to Myerson (1978), where it is exogenously assumed.)

Because the steady-state probabilities are continuous in the failure rates, if an optimal action is nearly perfect then in the limit it will be used most of the time. Thus, in this sense *satisficing fails gracefully*—a term in reliability engineering that captures an important kind of robustness.

Although part (ii) differs from optimizing—if the best action becomes more failure-prone but remains the best of the lot then an optimizing decision maker is unaffected—it is not inconsistent with expected utility theory, which predicts that the probability of using an action in bandit problems is *weakly* decreasing in its failure rate.

As noted earlier, Proposition 3 involves a concession to tractability: it gives up dynamics in order to study situations with more than two actions. Our next result involves a related though substantively distinct compromise: Proposition 4 obtains dynamical results for the $m > 2$ case by sacrificing some generality regarding search. This sacrifice is necessary to avoid potential idiosyncrasies in the dynamics due to *local* properties of particular search rules. For instance, even with a neutral start the success-ordering property cannot hold dynamically for the entire class of unbiased search rules, i.e., those satisfying (A3). For example, suppose that search is an unbiased random walk, with $\theta$ close to one. There are six alternatives, $\{a, b, c, d, e, f\}$, arranged in a circle. If $c$ and $e$ are terrible, $d$ is mediocre and $a$, $b$ and $f$ are almost perfect, then at the end of period 1 the agent is more likely to try $d$ than $a$ because $d$ will receive on average half of the refugees from the poor $c$ and $e$, while the chance of trying $a$ will stay close to the initial probability of $\frac{1}{6}$ since the agent will rarely be dissatisfied with $a$'s neighbors. Hence, although the success-ordering property will hold eventually (because search is unbiased), it does not obtain early on in this setting due to the random walk's dependence on the relative locations of individual alternatives.

This analysis suggests that if search lacks local structure then we might recover the success-ordering property. The next result shows that this is so, if (as before) the agent is sufficiently inertial. The only unbiased search rule without local structure is blind search: for any given alternative, dissatisfaction triggers search with a uniform probability distribution over all other alternatives, i.e., if the agent chose action $i$ in $t$ and is dissatisfied, then in $t + 1$ he chooses $j$ with probability $\frac{1}{m-1}$ for all $j \neq i$.

**Proposition 4:** Suppose that, in addition to the hypotheses of Proposition 3, the start is neutral, search is blind and $\theta f_k < \frac{m-1}{m}$ for all $k$. If $f_k < f_l$ then, *ex ante*, $p_{k,t} > p_{l,t}$ for all $t > 0$.

The proof of Proposition 4 actually yields a stronger result: it tells us that if the agent ever manages to get the choice-probabilities lined up in an order that is at least as good as a neutral start, then satisficing plus blind search ensure that the success-ordering property holds thereafter, provided that he is sufficiently inertial. I.e., if at *any* date $\bar{t}$, $p_{k,\bar{t}} \geq p_{l,\bar{t}}$ and $f_k < f_l$, then $p_{k,t} > p_{l,t}$ for all $t > \bar{t}$. (The neutral start is an important special case of this more general result.)

Note that the inertia condition gets easier to satisfy the more options there are ($\frac{m-1}{m} \rightarrow 1$ as $m \uparrow \infty$); if $\theta < 1$ then the crucial inequality must hold for sufficiently big $m$.

## IV. A Satisficing Agent in a Changing Environment

Both optimal decision theory and psychological learning theory usually posit a stationary setting when analyzing repeated interactions: although the mapping from actions to outcomes may be stochastic, it is constant over time. This convention is justified more by tractability than realism: presuming that decision makers always operate in stationary settings is wildly implausible. Hence, investigating how strategies and heuristics fare when environments can change has a natural empirical motivation.

Nonstationarity can have dire implications for optimization. Optimal strategies for bandit problems tend to be fragile: figuring out how to optimize depends strongly on the assumption that the payoff probabilities of actions that are not used in a given period remain fixed. If this assumption does not hold then in general *no one knows how to achieve goals such as maximizing the sum of discounted expected payoffs*. Lacking a complete theory of optimal decision-making in changing environments, it makes sense to study the behavior of heuristics such as satisficing.[6]

Consider a two-armed bandit problem with normally distributed payoffs. The payoff distributions of each arm differ by a constant: $Y = X + c$, with $c > 0$. The process switches environments in a Markovian manner with a fixed probability $s \in (0, 1)$. As in the standard stationary bandit problem, the agent observes her realized payoff. This problem is obviously harder than the standard bandit: inferences in nonstationary contexts can be confounded in ways that cannot occur in stationary ones. For example, suppose that the decision maker begins to believe, based on initial experience, that the left arm, L, is better than the right, R (in what appears to be the current environment, say E1), and so keeps picking L. If L then gives a string of unsatisfying payoffs, she is uncertain whether her prior estimate of L (in E1) was off or whether the environment has changed.

However, as Simon emphasized, not knowing what is optimal does not make us powerless. We can still deploy heuristics; some might work rather well. How, then, does satisficing perform in this nonstationary setting? The changing payoff probabilities make dynamical analysis very difficult, but we can address what happens in the limit. We consider the interesting case where each action is superior in exactly one environment. Hence, apart from the knife-edge case of $s = \frac{1}{2}$, optimizing entails switching from one action to the other, not settling down forever on one action (as it is in stationary contexts). In particular, the left arm produces the superior distribution $Y$ in E1, but in E2 distribution $Y$ is produced by the right arm. (Accordingly,

[6]Scholars have proposed interesting conjectures on this topic:: "Simple heuristics can be successful for two reasons: they can exploit *environmental structure*, as the example above illustrates, and they can be *robust*, that is, generalize well to new problems and environments. If there is uncertainty in an environment, in the sense of some degree of unpredictability and changing environments, robustness becomes an issue" (Gigerenzer 2001, p.47; original emphasis). As we will see shortly, Gigerenzer's claim about robustness is too bold regarding the best-known simple heuristic (satisficing).

distribution $X$ is generated by R in E1 and by L in E2.)

Satisficing gets this important qualitative feature right: it never gets stuck on either arm. Instead, it adapts indefinitely, which is appropriate in this nonstationary context. However, the next result shows that satisficing's quantitative performance depends strongly on *the rapidity of environmental change*, parameterized here by $s$.

Let $\tilde{p}_{L,E1}$ denote the steady state probability that the agent selects L when the environment is E1, and similarly for the three other possibilities of (R,E1), (L,E2) and (R,E2). We are especially interested in the probability that the agent selects the better action: L in E1 and R in E2.

**Proposition 5:** Suppose the above assumptions about the nonstationary environment hold. If the agent satisfices via (A1) and the aspiration level is exogenously fixed, then the process has a unique limiting distribution which has the following properties.

- **(i)** $\frac{\partial \tilde{p}_{L,E1}}{\partial s} < 0$ and $\frac{\partial \tilde{p}_{R,E2}}{\partial s} < 0$.
- **(ii)** If $s = \frac{1}{2}$ then $\tilde{p}_{L,E1} + \tilde{p}_{R,E2} = \frac{1}{2}$.

Part (i) says that the chance that the agent makes the right, environmentally-contingent moves is monotonically decreasing in $s$. Thus, this heuristic finds rapidly changing environments to be difficult. We suspect that this holds more generally for humans' total cognitive capacities: rapidly changing environments probably confuse most people.

Parts (i) and (ii) together imply that if $s > \frac{1}{2}$ then for any search probability $\theta$, satisficing flunks the performance benchmark of tossing a fair coin: *they get it wrong over half the time*—$\tilde{p}(R, E1) + \tilde{p}(L, E2) > \frac{1}{2}$. Thus, this heuristic works poorly on these nonstationary problems.

To clarify how satisficing's performance depends on the rate of environmental change, let us compare two examples: in example 1 satisficing does well; in example 2, badly. In example 1 we assume that L almost always succeeds in E1 while R nearly always fails; in E2 the success probabilities are reversed. The chance that the environment changes is only 0.001. In this parametric setting a satisficer will almost always make the right choice in the limit. Here's why. L almost always works in E1, so as long as the process persists at E1—which is likely since $s = 0.001$—the agent has little reason to leave L: the satisficing slogan of "if it ain't broke don't fix it" is well-adapted to L's quality in E1. If that slogan is coupled to a high propensity to search given a failure ($\theta$ close to one) then the agent won't tarry long in the deficient R. Since the environment rarely changes, the satisficer can settle down in the above effective pattern without often being thrown for a loop by a rude switch to E2 (where the once-reliable L suddenly starts failing). And once the process *does* switch to E2, the agent can settle down into the opposite but equally comfortable habit.

Now consider a setting where the environment changes constantly ($s = 1$), with the same success probabilities as before. Then whenever the agent experiences a success, say with L in E1, *satisficing will probably lead her to a failure tomorrow*: she'll continue with L, thinking it unbroken, but the rapidly changing environment plays havoc with this implicit inductive belief. Nor is this ameliorated by greater restlessness (bigger $\theta$'s). On the contrary: if satisficing is bang-bang ($\theta = 1$), then the good states become extremely unlikely, even if the agent started off choosing correctly. The befuddled agent will eventually get into a bad cycle of failing with one arm, switching to the other, only to be foiled by the environmental change, switching again, and so on.

A heuristic based on the inductive premise that tomorrow will be like today is ill-suited to a world in which tomorrow differs sharply from today. "If it ain't broke, don't fix it" is a bad slogan when what works today fails tomorrow. (We are not claiming that *all* heuristics perform badly when environments change rapidly. For example, in the relatively simple context of Proposition 5, if the environment changes constantly ($s = 1$) then an anti-satisficing heuristic—"search if today's payoff is good, stay put if it is bad"—will do well. Of course, such rules have their own weaknesses: for them the $s = 0$ setting is difficult.)

## V. Endogenous Aspirations

Some scholars argue that a necessary condition for taking a theory of rules of thumb seriously is that it demonstrates that the rules satisfy optimality criteria (e.g., Feinberg 2004, p.11). Such scholars might object to the preceding results on the grounds that agents with exogenously fixed aspirations are insufficiently adaptive: they do not learn to set "reasonable" aspirations. Hence, so the critique goes, it is unsurprising that they often fail to converge to optimal behavior.

Reconsider, for example, the bandit example of p.4. This problem is well-matched to satisficing: there exist a set of exogenous aspirations that would discriminate between the optimal action and all others. However, this happy outcome requires luck: the exogenously fixed aspiration must be in a specified interval. If, say, it is too high—above the maximal possible payoff—then search will be excessive: the agent will never settle down on any option, not even the best one. In this setting, the assumption of exogenous aspirations is vulnerable to the criticism that most humans would eventually learn what is feasible: they would not indefinitely aspire for an impossibly high payoff.

Alternatively, if we view satisficing as a descriptive theory then positing exogenous aspirations is methodologically suspect for a different reason: since aspirations drive many of the results, to leave them exogenous makes the formulation ad hoc in a key respect (Elster 1986, p.26). In either case, it is important to examine satisficing with endogenous aspirations.[7]

In what follows we accommodate the idea that people adapt their aspirations based on their experience in a rather general way. All that we require (for reasons of tractability) is that aspiration-adjustment doesn't go on forever. Apart from that, any kind of rule is admissible.

**(A4):** There is a fixed date $T$ and a finite $a$ (which may depend on $a_0$ and the history of play in $t = 1, \ldots, T$) such that $a_t = a$ with probability one for all $t > T$.

For example, suppose $a_t$ is the agent's average payoff from all previous dates. Under this rule aspiration-adjustment will tend to become increasingly sluggish. Plausibly, then, the agent might stop adjusting after 10,000 periods, as the changes become so small as to be trivial.

Note that the process need not be ergodic. In the above example, the sample of payoffs the agent receives in the first 10,000 periods determine her long-run aspiration level. Hence different sample paths generate different long-run aspirations, which in turn produce different choice-probabilities.

Because (A4) is so general (it permits, e.g., stochastic adjustments)[8], it allows for many rules that reflect the intuition that people learn which aspirations are reasonable: they adapt their aspirations based on their experience—the payoffs they have gotten. Given this assumption on aspiration-adjustment, we can now re-investigate satisficing. The first step is to redefine the heuristic so that it satisfies (A4). This requires only a minor modification of (A1). We highlight the assumptions' similarity by labeling the new one (A1′).

[7]Clearly, models with exogenous aspirations and those with endogenous ones may score differently on these two criteria, i.e., the normative and the descriptive. Borgers and Sarin prove that in some contexts—low initial aspirations—"endogenous aspiration level adjustments will be harmful for the decision maker" (2000, p.924). But they also point out that, descriptively, models with endogenous aspirations have done better than those with exogenous aspirations on empirical tests (p.922).

[8]Note also that the date that aspirations converge to a fixed level could be a random variable. (A4) requires only that there exist *some* finite date $T$ after which the agent no longer adjusts.

**(A1′):** The agent has an aspiration level, $a_t$, such that for all $t$ (1) if $\pi_t \geq a_t$ then she satisfices, i.e., uses the same action in $t+1$, and (ii) if $\pi_t < a_t$ then she searches for a new option in $t+1$ with probability $\theta > 0$.

(A1′) thus mimics (A1): the agent changes her action with some probability if and only if she is dissatisfied given her current aspiration.

The next result shows that satisficing's mixed performance profile—sensible though often suboptimal behavior—continues to hold in decision-theoretic contexts in which aspirations adjust (for finitely many periods). Because long-run aspirations are affected by the agent's adjustment process until date $T$, in this subsection aspirations are part of the state space. Hence the phrase "the stochastic process" pertains to the agent's actions and aspiration levels.

**Proposition 6:** Suppose the agent satisfices via (A1′), with $m \geq 2$ actions, and (A4) governs aspiration-adjustment. The payoff distributions have unbounded support and continuous densities, with finite means and variances. If the payoff distributions can be ordered by first-order stochastic dominance then (1) the agent plays suboptimal actions infinitely often with probability one but also (2) in the limit $\tilde{p}_k \geq \tilde{p}_{k+1}$ for all $k = 1, \ldots, m-1$.

It is worth iterating that satisficing's mixed long-run performance profile does not require a well-behaved stochastic process. It can hold when aspirations adjust in ways that produce multiple limiting distributions (via, e.g., the dependence of the steady-state aspiration level on $a_0$).

Although the proposition does not identify conditions that are necessary for success-ordering to hold in the limit, one can show by an example that satisficing's nemesis—permanently excessive restlessness—can again cause problems. Since aspirations are endogenous here, the property of excessive restlessness includes the speed of aspiration-adjustment. To see why, consider a two-armed bandit with binary payoffs, $l$ and $h$. Suppose that aspirations adjust fully to payoffs: $a_{t+1} = \pi_t$. Hence, after period one there are only two feasible aspiration levels, $l$ and $h$ and only four states in the Markov chain, $\{L, l\}$, $\{L, h\}$, $\{R, l\}$ and $\{R, h\}$. The agent satisfices via (A1') but significantly we do not require that action-adjustment be rapid: $\theta$ can be anything in $(0, 1]$. Hence, because the agent could adjust actions sluggishly (a $\theta$ close to zero), the effect that we will see in a moment arises exclusively from overly fast aspiration-adjustment which lasts indefinitely.

Under the above assumptions the Markov process is ergodic, and simple algebra shows that success ordering holds in the limit if and only if $(1 - f_R)f_R > (1 - f_L)f_L$, where arm $L$ is optimal and $f_i$ denotes the probability that arm $i$ generates a low payoff. If the inequality goes the other way then the suboptimal arm is chosen more than half the time in the long run.

(Calculations of the probabilities in this example's limiting distribution are available from the authors upon request.)

The intuition for this result is as follows. The agent will be disappointed if and only if her aspiration is $h$ and her payoff is $l$. Exactly one path produces this. The path involves two periods. In $t$ her aspiration is $l$ and she gets a payoff of $h$. Satisfied with the action chosen in $t$, she chooses that action again in $t+1$; simultaneously, however, her aspiration rises to $h$. She then gets a payoff of $l$ in $t+1$, which dissatisfies her. This path occurs on arm $i$ with probability $(1-f_i)f_i$. Hence, the larger this product, the more likely she is to experience disappointment. This product—the variance of a Bernoulli random variable—is maximized at $f = \frac{1}{2}$. Hence, whichever arm generates less variability is less likely to disappoint the agent, and so is more likely to be chosen in the limit.

In contrast, if aspirations adjusted increasingly slowly (e.g., equaling the average payoff received by the agent), then disappointment in the limit would be created not by the above period-to-period comparison but by comparing today's payoff to a statistic that was becoming highly inertial. This comparison eventually depends mostly on the probability of a low payoff today, and if (as in the present example) the optimal action first order stochastically dominates the alternative, the probability of disappointment must be higher with the suboptimal action.[9]

## VI. Multiple Decision Makers, Exogenous Aspirations

So far we have focused on the simplest setting, with only a single decision maker, but this was more for ease of exposition than any substantive reason. The tension between exploration and persistence reappears in more complex contexts, and the type of problem—well- or ill-matched—remains of paramount importance.

There is, however, a conceptual issue to address. In decision theoretic situations, 'optimal' is well-defined; thus, questions about the convergence of satisficing to optimal outcomes can be well-posed. When $n > 1$, the notion of optimality is famously unclear. So we take a general route. Although there is no universally agreed upon definition of optimality for $n > 1$, one can stipulate desired action-vectors, for any a specific game. Accordingly, we treat the problem abstractly by partitioning the set of stage game action-vectors into two disjoint and nonempty subsets: "good" action-vectors (G)

[9] It is easy to prove this rigorously under (A4), but we suspect that the logic goes through if aspirations adjust forever but the amount of adjustment goes to zero as $t \to \infty$.

and "bad" (B) ones. Different criteria can be used to define G (and hence B): for the PD, one potential G-set could be the set of Nash equilibria actions while another one could be the set of actions that produce Pareto outcomes. Thus, optimality in this context means that agents in the long-run use action-vectors in a given G-set, and *only* these.

In the following definition, $\underline{\pi_i}(G)$ denotes player $i$'s minimal payoff from the set G and $\pi_i(b)$ denotes a realized payoff to $i$ from a particular element $b \in B$.

**Definition 2:** A game is called **well-matched** (for a specific definition of G) to satisficing if for every $b \in B$ there exists at least one $i$ such that $\pi_i(b) < \underline{\pi_i}(G)$. Otherwise it is called **ill-matched**.

The next result parallels Proposition 1. Hence, it justifies our stipulated definition of "well-matched" for multi-agent contexts.

**Proposition 1′:** Suppose $n > 1$ and all players adapt via an (A1)-type rule. Then there exists a set of aspiration levels such that every element of G is stable and every element of B is unstable if and only if the problem is well-matched to satisficing.

With this result in hand, satisficing's behavior in the 2-person symmetric PD becomes straightforward to analyze, especially when payoffs are deterministic. Using the conventional payoff matrix for the case of deterministic payoffs $\begin{bmatrix} & C & D \\ C & R,R & S,T \\ D & T,S & P,P \end{bmatrix}$, where $T > R > P > S$, we see that our definition codes the game as ill-matched to satisficing if G is defined in terms of Nash equilibria, since the "good" action-vector $\{C, C\}$ has each player getting $R > P$. Thus, mutual defection can be made stable only if an action-vector outside the stipulated G-set, mutual cooperation, is as well. A similar issue arises when G is defined as all action-vectors that produce Pareto optimal outcomes. However, if G is defined in terms of symmetric action-vectors that produce the maximal payoffs from this set, *then the game is well-matched to the satisficing heuristic*. If both players' aspirations are in $(P, R]$, then both will only be satisfied with the payoff from mutual cooperation. Thus, under plausible assumptions about search, the players will eventually stumble onto mutual cooperation, where they will stay (Macy and Flache 2002). Chicken is even simpler. Defining G in terms of pure Nash equilibria makes the game ill-matched to satisficing, but defining G as the set of Pareto outcomes makes it well-matched: there exist aspirations that stabilize any action-vector that generates a Pareto outcome and destabilize the Pareto-suboptimal outcome of mutual aggression.

That satisficing-induced stability in well-matched games is attuned more to Pareto optimal actions than to Nash equilibria is no accident: it illustrates a fundamental difference between satisficing and fully rational behavior. Even though satisficing's key assumptions are as individualistic as those underlying Nash equilibria, in games that are well-matched to satisficing the *collective* properties of Pareto optimality and Pareto superiority are more important than the benefits produced by one person's defecting. Proposition 7 formalizes this connection between the individualistic notion of satisficing and the collectivistic criteria of Pareto optimality and Pareto superiority. (To maintain a tight focus on this issue we restrict attention here to deterministic payoffs.)

**Proposition 7:** If $n > 1$ and payoffs are deterministic then conclusions (1)-(3) hold.

- **(1)** There always exist definitions of G that (a) make the game well-matched to satisficing and (b) include only action-vectors that yield Pareto optimal outcomes.
- **(2)** If players adjust by (A1)-type rules and G's definition makes the game well-matched then the following properties hold.
  - **(i)** At least one action-vector in G must yield a Pareto optimal outcome.
  - **(ii)** If any action-vector $x$ in G yields a Pareto suboptimal outcome, then all action-vectors that yield Pareto superior outcomes to that of $x$ are also in G.
- **(3)** Suppose the game is generic. If some action-vectors that yield Pareto outcomes aren't Nash equilibria, then there exist definitions of G that (a) make the game well-matched to satisficing, (b) contain only action-vectors that yield Pareto outcomes, and (c) contain no Nash equilibria.

The connection between Pareto superiority and satisficing in games well-matched to satisficing rests on a transitivity property shared by the Pareto criterion and satisficing *but not by Nash*. For a satisficer, if payoff $\pi^1$ is satisfactory and a second payoff $\pi^2$ exceeds $\pi^1$, then $\pi^2$ must also be satisfactory. The same transitive logic holds for a set of payoffs for the $n$ agents: if $(\pi_1^1, \ldots, \pi_n^1)$ is satisfactory for all the players, and each element of a second vector $(\pi_1^2, \ldots, \pi_n^2)$ exceeds the corresponding element of the first vector, then the second vector must also be satisfactory for everyone. This relation between the two payoff-vectors is precisely that of Pareto dominance. Well-matched problems are defined by the same logic: they are found whenever every good outcome satisfies everyone and every bad outcome dissatisfies someone.

In contrast, a game theoretically rational player does not compare the payoff he gets from a particular outcome to an internal standard. Instead, he uses a counterfactual, comparing it to what he *could* get were he to select another action while his partners stayed put. Hence, mutual defection in the PD is Nash, yet mutual cooperation, though Pareto-superior, is not: a game theoretically rational player in *that* outcome could do still better by unilaterally deviating. The absence of Nash equilibria in some G sets that make a game well-matched to satisficing is thus not a bizarre property, but rather a direct consequence of a fundamental difference between satisficing and optimizing (Gilboa and Schmeidler 1995, p.610). Casual attempts to merge the two–e.g., treating the former as a kind of optimal search—are misguided. It is also important to keep in mind that the nature of the problem has a major impact on the satisficing heuristic. Problems that are ill-matched to satisficing reveal the limited ability of aspiration-based mechanisms to discriminate among actions. As the folk theorems for adaptive processes (Bendor, Diermeier and Ting 2004) show, when aspirations are exogenous and payoffs are compact sets it is straightforward to stabilize *all* of a game's outcomes via satisficing. What is tricky is stabilizing *only* the 'good' ones. The simple heuristic of satisficing cannot do this trick in all task environments: its crudeness makes this impossible.[10]

Thus far, our results for stochastic payoffs (with $n > 1$) have been confined to payoff distributions with compact supports. We now consider contexts with unbounded distributions. We know that such environments are ill-matched to satisficing for $n = 1$; definition 2 indicates that they continue to be ill-matched in multi-person contexts. Hence, it is no surprise that such contexts cause problems for multiple agents who satisfice. For example, it is straightforward to prove that if payoffs are unbounded then all action-vectors are unstable, for any set of exogenously fixed aspirations. One can further

[10]However, even in difficult problems satisficing retains its desirable "inner environment" property of feasibility. And we are not proposing that, because satisficing falls short of optimizing in certain problem-environments, one should use an optimal strategy instead. As moral philosophers since Kant have said, ought implies can: if the problem is so hard that decision makers cannot figure out what *is* optimal then urging them to do that is pointless. Rather, they should think about using subtler heuristics that might perform better in the tricky environment at hand.

show (Bendor, Kumar and Siegel 2007) that the process will visit B-states infinitely often under in a fairly large array of circumstances, e.g., in weakly generic games where G is composed of Nash equilibria and $\theta_i < 1$ for all $i$.

Thus, there is bad news about satisficing's performance in games that are ill-matched to it. What about the other side of the heuristic's performance profile in these difficult contexts—its tendency to produce sensible (if not optimal) results in hard problems? To address this question we focus on 2x2 games and assume that the players adjust asynchronously: in each period exactly one player is 'active', i.e., makes a decision.[11] In every period each player is selected to be active with probability $\frac{1}{2}$; these realizations are independent over time.

Asynchronous adjustment makes the analysis more tractable: whereas the steady state distribution with synchronous adjustment is extremely difficult to understand (Bendor, Kumar and Siegel 2007), asynchronicity makes it possible to derive and interpret some closed form solutions.[12]

In the following result $\tilde{p}_{k,l}$ denotes the limiting probability of the outcome in which Row chooses action $k$ and Column, action $l$; $f_{k,l}$ and $g_{k,l}$ are Row's and Column's failure probabilities, respectively, given actions $k$ and $l$.

**Proposition 8:** Suppose $n = 2$ and $m = 2$. Both players satisfice via (A1), with a common $\theta > 0$; they adjust asynchronously with realized activity governed by an i.i.d. process. Payoff distributions are unbounded and differentiable. If they are ranked by first-order stochastic dominance then the following hold for any set of exogenously fixed aspiration levels.

- **(i)** $\frac{\partial \tilde{p}_{k,l}}{\partial f_{k,l}} < 0$ and $\frac{\partial \tilde{p}_{k,l}}{\partial g_{k,l}} < 0$ for all $k$ and $l$.
- **(ii)** If the game is also symmetric and has at least one pure Nash equilibrium, then in the limit any Nash outcome is more likely than is any non-Nash outcome.

Part (i) shows that satisficing responds in a sensible way to changes in payoffs. Consider, for example, a game of Chicken. If getting the short end of the stick gets stochastically worse for Row in the Nash equilibrium where Row is conciliatory and Column is aggressive, then Row will more often find

[11] We thank an anonymous referee for the extremely helpful suggestion of assuming asynchronous adjustment.

[12] A richer (but less tractable) model would allow for both synchronous and asynchronous adjustment. To show that our next result is robust, the appendix provides a model with both types of adjustment: it establishes that Proposition 8's conclusions continue to hold if the probability of synchronous adjustment is sufficiently low.

that experience dissatisfying. Hence, when that outcome occurs and Row is active, he will more often switch to being aggressive, and the probability of (conciliatory, aggressive) will fall. Note that this logic covers all outcomes, Nash and non-Nash alike. This makes sense. A satisficer may not know which outcomes are Nash equilibria; indeed, he may be ignorant of the concept. His behavior is governed by myopic adaptation to the relation between his realized payoffs and his aspiration level; it does not directly attend to the distinction between outcomes that are Nash equilibria and those that are not.[13]

This obliviousness makes part (ii) rather intriguing. In symmetric 2x2 games with pure Nash equilibria, satisficers tend to behave as if they were game theoretically rational.[14] Because unbounded payoffs make disappointment inevitable with any outcome[15], this pattern is only probabilistic. But it is a genuine tendency. It holds because aspiration levels and first-order stochastic dominance together create failure probabilities that are ranked appropriately, i.e., inversely to the payoff orderings that would drive the behavior of rational players. If payoff distribution $X$ stochastically dominates distribution $Y$ then a satisficer is more likely to be satisfied by $X$ than by $Y$. Suppose, then, that feasible actions are $\{a, b\}$ and the common payoff distribution in outcome $(a, a)$ stochastically dominates Row's in $(b, a)$ and Column's in $(a, b)$. This simultaneously implies that $(a, a)$ would be a Nash equilibrium for rational players and also that $f_{a,a} = g_{a,a} < f_{b,a} = g_{a,b}$. Since players adjust asynchronously, movement is governed by the dissatisfaction of a single player—the active one. For example, the movement between $(a, a)$ and $(b, a)$ is determined by Row's behavior. Since $f_{b,a} > f_{a,a}$, the process is more likely to go from $(b, a)$ to $(a, a)$ than vice versa: in short, movement is toward Nash. And in symmetric games, the strength of the movement is symmetric.

[13]Although responding to changes in one's own payoffs is intuitively sensible, this behavior yields a prediction that differs from that of classical game theory in games in which the unique Nash equilibrium is mixed: a fully rational player would alter choice-probabilities in response to changes in his partner's payoffs, not his own. Further, because "if it ain't broke don't fix it" produces serially correlated behavior, satisficers would differ from game theoretically rational mixers in a second respect. These differences in predictions are empirically testable. (See Bendor, Diermeier and Ting (2003b) for further discussion of these predictions.)

[14]As footnote 12 implies, part (ii) need not hold in the richer model (where both types of adjustment can occur) if synchronous adjustment is sufficiently likely.

[15]In weakly generic 2x2 games where satisficers have stationary $\theta$'s and adjust asynchronously it is trivial to show that non-Nash outcomes are visited infinitely often with probability one if $f_{k,l} g_{k,l} > 0$ for all $k$ and $l$.

This last feature matters: part (ii) of Proposition 8 does not hold in general for asymmetric 2x2 games. Consider, e.g., the game represented in the following matrix, in which the cell entries are the failure probabilities of Row and Column, respectively:[16] $\begin{bmatrix} & a & b \\ a & .02, .01 & .99, .99 \\ b & .01, .99 & .98, .01 \end{bmatrix}$ For Row, action $b$ dominates $a$; for Column, $b$ is the best response to itself. Hence, $(b, b)$ is the unique Nash equilibrium. Nevertheless, the stochastic process induced by satisficing spends about 97 percent of its time in $(a, a)$ in the limit.

The explanation is simple: once the process enters $(a, a)$ it tends to stay there a long time, but in all other states one of the players will probably be dissatisfied, and when the disgruntled player is active he will take the process out of that state. Symmetry precludes this. If, e.g., Row were very likely to be satisfied in $(b, a)$ then so must be Column in $(a, b)$; hence, the process would be less likely to enter $(a, a)$ than to leave it.

## VII. Conclusion

This model of satisficing yields testable predictions—in some contexts, quite specific ones. Our results show that the heuristic performs well in certain choice settings but badly in others. This performance-profile is probably not peculiar to satisficing. We hypothesize that heuristics that are often used by many people exhibit this variability. This performance-variation is directly related to Simon's general view of bounded rationality: information-processing constraints, latent when we face easy problems, show up in hard ones. Performance-variation, both for humans and for the heuristics we employ, is the result.

In the present context, it turns out that certain problems are ill-matched to satisficing because they show up the *crudeness of the heuristic's discriminatory abilities*. In the canonical two-armed bandit, both arms sometimes fail to pay off: their minimal payoffs are the same. This presents problems for satisficing because this rule is partly failure-driven (search if payoffs don't meet aspirations), and in this class of bandit problems all actions fail sooner or later. Satisficing is too crude to be optimal even in the long run for such problems because it cannot cleanly distinguish what is suboptimal from what is *imperfect yet optimal*. Similarly, certain problems are difficult

[16] So, for example, in (a,a) .02 is Row's failure probability in that outcome; .01 is Column's.

for satisficing because they involve rapidly changing environments, which violates the heuristic's implicit inductive premise that what works today will work tomorrow.

However, although certain problems are indeed hard for satisficing, our results show that even in many such contexts it works reasonably well. For example, in decision theoretic bandit problems the heuristic can lead the agent to select actions in accord with their underlying and unknown success probabilities—both in the limit and, in certain plausible conditions, dynamically—even when the problem is ill-matched to satisficing. Further, even in nonstationary settings, the heuristic gets the choice problem right more than half the time provided that the environment does not change too rapidly. Thus, whereas full-rationality theories mostly use dichotomous evaluations—a strategy is either optimal or not—satisficing models, with their focus on 'good enough' performance, naturally incline toward more fine-grained evaluations: distinguishing only between the best and everything else is too coarse.

**Some Methodological Issues**

We have argued that when the optimal strategy is unknown, scholars should focus on adaptive rules. But which? There are many candidates, and Salmon (2001) has shown that empirically discriminating among alternative adaptive rules is often difficult, particularly with aggregate data. This suggests a methodological trade-off between satisficing and other adaptive models such as reinforcement learning, and between satisficing and more complex models that combine belief-learning and reinforcement (Camerer and Ho 1999). Satisficing models are quite tractable. In contrast, reinforcement learning theories typically involve mixed strategies; the state-spaces of the corresponding stochastic processes are usually much larger. (For example, the state space of a satisficing model of turnout (Collins et al forthcoming) increases linearly in the number of citizens; one that uses reinforcement learning (Bendor et al 2003a) increases exponentially.) Hence, satisficing results can often be derived analytically (Karandikar et al (1998), Collins et al forthcoming); reinforcement models often require computational methods, e.g., Roth and Erev (1995), Erev and Roth (1998), as do most of the predictions of Camerer and Ho's mixed models. To the extent that we value analytical results, satisficing models have an edge.

But this simplicity has a descriptive price. In comparing satisficing and reinforcement learning, this is clear on a priori grounds (Bendor, Mookherjee and Ray 2001, p. 162). Further, there is experimental evidence that Camerer and Ho's hybrid model does better empirically than purely backward-looking adaptive rules for micro-level data.

Hence, there is a trade-off between analytical tractability and descriptive accuracy. Scholars working with aggregate field data—the behavior of millions of consumers or voters—might decide to give up some of the latter to get more of the former. Experimentalists studying a handful of subjects might do the opposite. We suspect that fine-grained micro-level data will often reveal evidential problems in crisp-but-crude satisficing models.

However, even for micro-level data satisficing models might do well when subjects confront sufficiently hard problems. Consider two properties that probably make choices difficult for most people: nonstationary or multidimensional payoffs. Shor (2004, p.11-12) discovered that a satisficing model fit the behavior of *individual* subjects in a nonstationary choice environment better than did Erev-Roth reinforcement models. And because problems with multidimensional payoffs are probably harder than those that are clearly unidimensional, consumers in the former situation may be more likely to use satisficing or other noncompensatory rules, e.g., elimination-by-aspects (Tversky 1972) while those in the latter setting may be more likely to use compensatory rules, especially when taboo trade-offs (Fiske and Tetlock 1997; McGraw and Tetlock 2005) are involved.

# Appendix

**Notation.** The appendix uses some notation not introduced in the text. Let $\alpha$ denote an action, and $\mathcal{A}$ a vector of actions, one for each of the $n$ players. The term *wpp* means "with positive probability."

Because Proposition 1 assumes $n = 1$ whereas Proposition 1′ holds for $n \geq 1$, the former is a special case of the latter. Hence, the proof for Proposition 1′ (below) suffices to establish the validity of Proposition 1.

**Proof of Proposition 2**
**(i)** Since $m = 2$, $p_{i,t} = 1 - p_{j,t}$ at all times, the probabilities take a particularly simple form:

$$p_{i,t} = (1 - \theta f_1 - \theta f_2)p_{i,t-1} + \theta f_j = \delta p_{i,t-1} + \theta f_j. \tag{1}$$

Iteratively applying this yields

$$p_{i,t} = \frac{f_j}{f_1 + f_2} - \delta^t \left[\frac{f_j}{f_1 + f_2} - \frac{1}{2}\right], \tag{2}$$

implying that $p_{i,t} - p_{j,t} = \frac{f_j - f_i}{f_1 + f_2}(1 - \delta^t)$, from which (i) follows as $1 - \delta^t > 0$ for all $t$.

**(ii)** Differentiating (1) with respect to $f_i$ yields

$$\frac{\partial p_{i,t}}{\partial f_i} = \delta \frac{\partial p_{i,t-1}}{\partial f_i} - \theta p_{i,t-1},$$

which, when iterated, produces

$$\frac{\partial p_{i,t}}{\partial f_i} = \delta^t \frac{\partial p_{i,0}}{\partial f_i} - \theta \sum_{k=0}^{t-1} \delta^k p_{i,t-k-1}.$$

Since $p_{i,0} = \frac{1}{2}$, $\frac{\partial p_{i,0}}{\partial f_i} = 0$, and so $\frac{\partial p_{i,t}}{\partial f_i} = -\theta \sum_{k=0}^{t-1} \delta^k p_{i,t-k-1}$. As each $p_{i,t-k-1} \geq 0$ and $p_{i,0} > 0$, we have $\frac{\partial p_{i,t}}{\partial f_i} < 0$ as long as $\delta > 0$. If $\delta = 0$, then $\theta > 0$ and $\frac{\partial p_{i,t}}{\partial f_i} = -\theta^2 f_j$. This is never more than zero, and is strictly negative as long as the following set of parameters does not obtain: $t > 1, \theta = 1, f_i = 1$, and $f_j = 0$. Now let $\delta < 0$. Applying (2) to $\frac{\partial p_{i,t}}{\partial f_i}$, taking the sum, and performing some algebra yields

$$\frac{\partial p_{i,t}}{\partial f_i} = -\frac{f_j}{(f_i + f_j)^2}\left[1 - \delta^{t-1}\left(1 - \theta(f_i + f_j)\left(1 - \frac{t}{2f_j}(f_j - f_i)\right)\right)\right]. \quad (3)$$

The term outside the square brackets is always negative, so $\frac{\partial p_{i,t}}{\partial f_i} \leq 0$ whenever the expression inside the brackets is nonnegative, and the inequality is strict whenever the expression is strictly positive. This amounts to the condition

$$\delta^{t-1}\left[1 - \theta(f_i + f_j)\left(1 - \frac{t}{2f_j}(f_j - f_i)\right)\right] \leq 1, \quad (4)$$

for the weak inequality, with the strict inequality holding whenever (4) holds strictly. If (4) holds always, then the proof is done. In particular, if (4) holds for the values of $\theta, f_j, f_i \in [0, 1]$ and $t > 0$ that maximize the left-hand side, under the additional constraint that $\delta < 0$, then it holds for all values. There are two cases to consider: $t$ odd and $t$ even. For $t$ odd, we find that under these maxima (4) holds strictly. For $t$ even, we find that under these maxima (4) holds weakly always, and strictly as long as one of the following sets of parameter values does not obtain: $t = 2, f_i = \theta = 1$, and $f_j \in (0, 1]$, or $t$ even and $f_i = f_j = \theta = 1$. This proves the claim.

**(iii)** We show each result in turn.

(a): From (3), we have that $p_{i,t+1} - p_{i,t} = (1 - \delta)\delta^t \left[\frac{f_j}{f_1 + f_2} - \frac{1}{2}\right]$. If $\delta > 0$, then this is positive, as both $(1 - \delta)$ and $\left[\frac{f_j}{f_1+f_2} - \frac{1}{2}\right]$ are positive by assumption, and we are done.

(b): Note that the previous difference is also positive if $\delta < 0$ and $t$ is even, proving (i), and that the difference is negative if $\delta < 0$ and $t$ is odd, proving (ii). To show (iii), note that (3) also yields $p_{i,t+2} - p_{i,t} = (1-\delta^2)\delta^t \left[\frac{f_j}{f_1+f_2} - \frac{1}{2}\right]$, which is positive if $\delta < 0$ and $t$ is even. QED.

**Proof of Proposition 3**

**(i)** By (A1), $p_{i,t} = (1-\theta f_i)p_{i,t-1} + \sum_{j\neq i} \theta f_j q_{ji} p_{j,t-1}$. At the steady state, $p_{i,t} = p_{i,t-1} = \widetilde{p_i}$, which implies $\widetilde{p_i} = \frac{1}{f_i}\sum_{j\neq i} f_j q_{ji} \widetilde{p_j}$. The form of this equation is suggestive, so we guess that the steady state probabilities take the form $\widetilde{p_i} = \frac{1}{f_i}c^{-1}$, where $c = \sum_j \frac{1}{f_j}$ is a normalization factor. Trying this in the above equation yields

$$\frac{1}{f_i}c^{-1} = \frac{1}{f_i}\sum_{j\neq i} f_j q_{ji} \frac{1}{f_j}c^{-1} \Rightarrow 1 = \sum_{j\neq i} q_{ji}.$$

But this latter equality is true by the definition of a probability transition matrix, so our guess is correct. Using this, we can easily see that $\widetilde{p_i} > \widetilde{p_j}$ whenever $\frac{1}{f_i} > \frac{1}{f_j}$, or equivalently, $f_j > f_i$.

**(ii)** Using $\widetilde{p_i} = \frac{1}{f_i}c^{-1}$, we find that

$$\frac{\partial \widetilde{p_i}}{\partial f_i} = -\left(\frac{1}{f_i}\right)^2 c^{-1} + \left(\frac{1}{f_i}\right)^3 c^{-2} = \frac{-c^{-1}}{(f_i)^2}(1 - \frac{c^{-1}}{f_i}) = \frac{-c^{-1}}{(f_i)^2}(1-\widetilde{p_i}) < 0$$

for all $i$, as required. QED.

**Proof of Proposition 4**

Blind search implies $p_{i,t} = (1-\theta f_i)p_{i,t-1} + \sum_{j\neq i}\frac{\theta}{m-1} f_j p_{j,t-1}$, so that

$$p_{j,t+1} - p_{k,t+1} = (1-\theta f_j \frac{m}{m-1})p_{j,t} - (1-\theta f_k \frac{m}{m-1})p_{k,t}.$$

By assumption, $f_j < f_k$, and $\theta f_j < \frac{m-1}{m}$ for all $j$, so that this difference is positive whenever $p_{j,t} \geq p_{k,t}$. In particular, this latter condition is satisfied at time 0 by the assumption of a neutral start, implying that it holds for time 1 as well. Carrying through this logic inductively gives us our result for all $t$. QED.

**Proof of Proposition 5**

We first prove that under the hypotheses of the proposition the process has a unique limiting distribution. (We are not making the stronger claim that it must be ergodic. That requires aperiodicity, which is not ensured by the hypotheses of Proposition 5.) The proof will show that all nontransient states communicate, which in turn proves that a stationary finite Markov chain has exactly one limiting distribution.

We establish this property under assumptions that are more general than those of the proposition.

We abbreviate $f_{L,E1}$ by $f_{L,1}$, etc.

Assume $\theta > 0$, $s > 0$, $0 < \max\{f_{L,1}, f_{R,1}\}$, $0 < \max\{f_{L,2}, f_{R,2}\}$, $1 > \min\{f_{L,1}, f_{R,1}\}$, $1 > \min\{f_{L,2}, f_{R,2}\}$. The agent satisfices via (A1). Everything is stationary-Markovian.

Since $s > 0$, no single state is absorbing. This immediately rules out the following types of state-classification configurations that would produce multiple limiting distributions: (a) all four states are absorbing; (b) one state is absorbing and the other three form a closed, communicating class. (c) One state is absorbing, two others form a closed, communicating class and the fourth state is transient. Patterns (a)-(c) *would* produce multiple limiting distributions, but each involves singleton absorbing states, which is impossible.

If all four states communicate, or if three do and the fourth is transient, then the basic condition—all nontransient states communicate—holds and we are done.

So it remains to consider the last type of pattern that could yield multiple invariant distributions: two closed pairs of states. (Note that since singleton states cannot be absorbing, the states within each pair communicate.)

Since $s > 0$, the process goes from each environment to the other wpp, so neither (L1, R1) nor (L2, R2) can be closed. Hence there are only two configurations of two pairs of closed subsets that must be examined. Configuration 1 involves (L1,L2) and (R1,R2); configuration 2 is (L1,R2) and (R1,L2).

Without loss of generality assume that $f_{R,1} = \max\{f_{L,1}, f_{R,1}\}$. Consider configuration 1 first. Since $f_{R,1}$ is $\max\{f_{L,1}, f_{R,1}\}$, it must exceed zero. So $\theta f_{R,1} s > 0$, whence the one-step transition from (R,1) to (L,2) occurs wpp. So (R1, R2) isn't closed, and configuration 1 is impossible.

Now consider configuration 2. We've assumed that $f_{R,1}$ is max of $\{f_{L,1}, f_{R,1}\}$; hence, since at least one of these two failure rates must be less than one, $f_{L,1} < 1$. So $(1 - f_{L,1})s > 0$, i.e., the process can go from (L,1) to (L,2) in one step wpp. Therefore (L1, R2) isn't closed, and configuration 2 is impossible. (Note that we haven't rule out that (R1,L2) is a closed (and communicating) subset, and indeed this occurs if $\theta = s = f_{R,1} = f_{L,2} = 1$. But since (R1,L2) is closed and can be reached in one step from (L1,R2), the latter pair of states is transient. So the key condition—that all *nontransient* states communicate—holds in this configuration.) QED.

With this guarantee of a unique limiting distribution in hand, we let *Mathematica* grind out the solutions (verified by hand) to the steady-state equations

$\tilde{p}_{L,E1} = s(f_{R,1} + f_{R_2}) + \theta(1 - 2s)f_{R,1}(f_{L,2} + f_{R,2}) \cdot c;$
$\tilde{p}_{R,E1} = s(f_{L,1} + f_{L_2}) + \theta(1 - 2s)f_{L,1}(f_{L,2} + f_{R,2}) \cdot c;$
$\tilde{p}_{L,E2} = s(f_{R,1} + f_{R_2}) + \theta(1 - 2s)f_{R,2}(f_{L,1} + f_{R,1}) \cdot c;$
$\tilde{p}_{R,E2} = s(f_{L,1} + f_{L_2}) + \theta(1 - 2s)f_{L,2}(f_{L,1} + f_{R,1}) \cdot c,$

where $c$, a normalizing constant that makes the limiting probabilities sum to one, equals $2(s(f_{L,1}+f_{L_2}+f_{R,1}+f_{R_2})+\theta(1-2s)(f_{L,2}+f_{R,2})(f_{L,1}+f_{R,1}))$.

Differentiating the relevant terms shows that $\frac{\partial \tilde{p}_{L,E1}}{\partial s} < 0$ and $\frac{\partial \tilde{p}_{R,E2}}{\partial s} < 0$ (part (i)), and by inspection it is easily seen that if $s = \frac{1}{2}$ then $\tilde{p}_{L,E1} + \tilde{p}_{R,E2} = \frac{1}{2}$ (part (ii)). QED.

**Proof of Proposition 6**

(1) By the assumption on $a_T$, for arbitrary $\delta > 0$ we have $P(a_T \geq -M_\delta) \geq 1-\delta$ for some $M_\delta$. Further, by the assumption of unbounded payoffs we have $P(\pi_k < -2M_\delta) \geq \delta'$ for some $\delta' > 0$. Therefore, $P(\pi_k < a_T) \geq \delta'(1-\delta) > 0$ for each $k$, which implies (1).

(2) By Proposition 3, for each realization $a$ of $a_T$ we have $P(\alpha_\infty = k|a) \geq P(\alpha_\infty = k+1|a)$. Suppose $a_T$ has distribution $F_T$. Then $\tilde{p}_k = \int P(\alpha_\infty = k|a)dF_T(a) \geq \int P(\alpha_\infty = k+1|a)dF_T(a) = \tilde{p}_{k+1}$, proving (2).

**Proof of Proposition 1′**

*Sufficiency.* Suppose the problem is well-matched to satisficing. Fix the vector of aspirations: $a_i = \underline{\pi_i}(G)$, for all $i$. Now consider an arbitrary element of $G$, say $g^\bullet$. Since $a_i \leq \pi_i(g^\bullet)$ for all $i$, every player is satisfied with the payoffs generated by $g^\bullet$. Hence that outcome is stable, and since $g^\bullet$ was arbitrary, so is every element of $G$. Now consider an arbitrary element of $B$, say $b^\bullet$. Because the problem is well-matched, there is at least one player, say $j$, such that $\pi_j(b^\bullet) < \underline{\pi_j}(G)$. Since $\underline{\pi_j}(G)$ equals $a_j$, wpp $j$ will be dissatisfied by $b^\bullet$; hence that outcome cannot be absorbing. Since $b^\bullet$ was arbitrarily selected, the same holds for all elements of $B$.

*Necessity.* We show by contradiction that if the problem is ill-matched to satisficing then there is no vector of aspirations such that every element of $G$ is stable and every element of $B$, unstable. Suppose such a vector did exist (for an ill-matched problem). Then, in order for every $g$ to be stable, it must be that $a_i \leq \underline{\pi_i}(G)$, for all $i$. Further, to ensure that every $b$ is unstable, for any particular $b$, say $b'$, there must be at least one player, say $j$, such that $a_j > \pi_j(b')$. Since $\underline{\pi_j}(G) \geq a_j$, this implies that $\underline{\pi_j}(G) > \overline{\pi}_j(b')$. But since this holds for every $b \in B$, the situation would satisfy the definition of a well-matched problem. Because it was posited that the problem was ill-matched, we have a contradiction. QED.

**Proof of Proposition 7**

(1) The proof is by construction. Pick a Pareto-optimal outcome in which player $i$ gets her maximal payoff in the game. Call this outcome $o^*$. Fix everyone's aspirations equal to their payoffs in $o^*$. Then clearly $o^*$ is stable, as is any other (necessarily Pareto optimal) outcome with the same vector of payoffs as $o^*$.

Now consider any Pareto-suboptimal outcome $o'$. By definition, at least one element of the vector of payoffs in $o'$ must be less than the corresponding element from $o^*$, implying that the person receiving that lesser payoff will be dissatisfied. Hence $o'$ is unstable. QED.

(2) If we prove (ii) then we eliminate the possibility that G contains some suboptimal outcomes but none that are Pareto-optimal. Thus, since G cannot be empty, proving (ii) would imply (i) as well. So we turn directly to proving (ii).

The proof is by contradiction. Suppose there exists a Pareto deficient outcome x in G, and there is an outcome y that is Pareto-superior to x yet which is not in G. Since B and G partition the set of stage game outcomes, y must be in B. But since this game is well-matched, there must be at least one player $j$ such that $j$'s payoff is less in outcome y than it is in *any* outcome in G, including in x. But this can't be, since y Pareto dominates x. Hence we have a contradiction. QED.

(3) From part (1) we know that one can always construct a G-set that is exclusively Pareto and which makes the game well-matched to satisficing. We apply this construction here to an outcome, $o^*$, which is Pareto but not Nash. Given that $a_i = \pi_i(o^*)$ for all $i$, all Pareto-deficient outcomes are unstable. So the only loose end to tie up concerns other Pareto-optimal outcomes that have the same vector of payoffs as $o^*$ yet which *are* Nash. But assuming genericity rules out such possibilities; i.e., G must be a singleton. Since $o^*$ is by choice not Nash, we are done. QED.

**Proof of Proposition 8**

Because either player is randomly selected to adjust, there are eight states: e.g., $(a, a)$ with Row active, $(a, b)$ with Column active, and so forth. The assumptions yield a stationary Markov process; the probability transition matrix is shown below. For notational convenience we write Row's failure probabilities as $f_1$, $f_2$, $f_3$ and $f_4$ and Column's as $g_1$, $g_3$, $g_2$ and $g_4$ for outcomes $(a, a)$, $(a, b)$, $(b, a)$, and $(b, b)$, respectively. (We reverse $g_2$ and $g_3$ in anticipation of the symmetric matrix of part (ii). With symmetric payoff lotteries, the above numbering gives $f_i = g_i$ for $i = 1, 2, 3, 4$.)

In the states listed in the transition matrix below, the overbar indicates which player is active. For example, in state $(\bar{a}, a)$, Row is active; in $(a, \bar{a})$, Column.

state in $t + 1$

| state in $t$ | $\bar{a}, a$ | $a, \bar{a}$ | $\bar{a}, b$ | $a, \bar{b}$ | $\bar{b}, a$ | $b, \bar{a}$ | $\bar{b}, b$ | $b, \bar{b}$ |
|---|---|---|---|---|---|---|---|---|
| $\bar{a}, a$ | $\frac{1-\theta f_1}{2}$ | $\frac{1-\theta f_1}{2}$ | 0 | 0 | $\frac{\theta f_1}{2}$ | $\frac{\theta f_1}{2}$ | 0 | 0 |
| $a, \bar{a}$ | $\frac{1-\theta g_1}{2}$ | $\frac{1-\theta g_1}{2}$ | $\frac{\theta g_1}{2}$ | $\frac{\theta g_1}{2}$ | 0 | 0 | 0 | 0 |
| $\bar{a}, b$ | 0 | 0 | $\frac{1-\theta f_2}{2}$ | $\frac{1-\theta f_2}{2}$ | 0 | 0 | $\frac{\theta f_2}{2}$ | $\frac{\theta f_2}{2}$ |
| $a, \bar{b}$ | $\frac{\theta g_3}{2}$ | $\frac{\theta g_3}{2}$ | $\frac{1-\theta g_3}{2}$ | $\frac{1-\theta g_3}{2}$ | 0 | 0 | 0 | 0 |
| $\bar{b}, a$ | $\frac{\theta f_3}{2}$ | $\frac{\theta f_3}{2}$ | 0 | 0 | $\frac{1-\theta f_3}{2}$ | $\frac{1-\theta f_3}{2}$ | 0 | 0 |
| $b, \bar{a}$ | 0 | 0 | 0 | 0 | $\frac{1-\theta g_2}{2}$ | $\frac{1-\theta g_2}{2}$ | $\frac{\theta g_2}{2}$ | $\frac{\theta g_2}{2}$ |
| $\bar{b}, b$ | 0 | 0 | $\frac{\theta f_4}{2}$ | $\frac{\theta f_4}{2}$ | 0 | 0 | $\frac{1-\theta f_4}{2}$ | $\frac{1-\theta f_4}{2}$ |
| $b, \bar{b}$ | 0 | 0 | 0 | 0 | $\frac{\theta g_4}{2}$ | $\frac{\theta g_4}{2}$ | $\frac{1-\theta g_4}{2}$ | $\frac{1-\theta g_4}{2}$ |

Since the payoff distributions are unbounded, $f_i > 0$ and $g_i > 0$ for $i = 1, 2, 3, 4$. Hence, it is clear by inspection of the transition matrix that all states communicate. Therefore, the process has a unique limiting distribution. Computing the limiting distribution by hand is too tedious; Mathematica yields the following solution, where for simplicity we have collapsed the eight states to four, using the fact that $\tilde{p}_{a,a} = \tilde{p}_{\bar{a},a} + \tilde{p}_{a,\bar{a}}$ and so forth.

$$
\begin{aligned}
\tilde{p}_{a,a} &= \frac{2}{d}[f_3 f_4 g_3 + f_4 g_2 g_3 + f_2 f_3 g_4 + f_3 g_3 g_4] \\
\tilde{p}_{a,b} &= \frac{2}{d}[f_3 f_4 g_1 + f_1 f_4 g_2 + f_4 g_1 g_2 + f_3 g_1 g_4] \\
\tilde{p}_{b,a} &= \frac{2}{d}[f_1 f_4 g_3 + f_1 f_2 g_4 + f_2 g_1 g_4 + f_1 g_3 g_4] \\
\tilde{p}_{b,b} &= \frac{2}{d}[f_2 f_3 g_1 + f_1 f_2 g_2 + f_2 g_1 g_2 + f_1 g_2 g_3]
\end{aligned}
$$

where $d$, the normalizing factor, equals $2\Big([f_3 f_4 g_3 + f_4 g_2 g_3 + f_2 f_3 g_4 + f_3 g_3 g_4] +$

$[f_3f_4g_1+f_1f_4g_2+f_4g_1g_2+f_3g_1g_4]+[f_1f_4g_3+f_1f_2g_4+f_2g_1g_4+f_1g_3g_4]+[f_2f_3g_1+f_1f_2g_2+f_2g_1g_2+f_1g_2g_3]\Big)$.

**(i)** Consider first $\tilde{p}_{a,a}$. By inspection we see that its numerator does not depend on either of the failure probabilities associated with this outcome, $f_1$ and $g_1$. In contrast, $d$ is strictly increasing in $f_1$ and in $g_1$. Hence it follows immediately that $\frac{\partial \tilde{p}_{a,a}}{\partial f_1} < 0$ and $\frac{\partial \tilde{p}_{a,a}}{\partial g_1} < 0$. The same logic holds for the other three limiting probabilities.

**(ii)** Symmetry implies that $f_i = g_i$ for $i = 1, 2, 3, 4$. Replacing the $g$ terms by their corresponding $f$ terms gives the following solutions:

$$\begin{aligned}
\tilde{p}_{a,a} &= \frac{2}{d}[f_3f_4f_3 + f_4f_2f_3 + f_2f_3f_4 + f_3f_3f_4] \\
\tilde{p}_{a,b} &= \frac{2}{d}[f_3f_4f_1 + f_1f_4f_2 + f_4f_1f_2 + f_3f_1f_4] \\
\tilde{p}_{b,a} &= \frac{2}{d}[f_1f_4f_3 + f_1f_2f_4 + f_2f_1f_4 + f_1f_3f_4] \\
\tilde{p}_{b,b} &= \frac{2}{d}[f_2f_3f_1 + f_1f_2f_2 + f_2f_1f_2 + f_1f_2f_3]
\end{aligned}$$

which simplifies to

$$\begin{aligned}
\tilde{p}_{a,a} &= \frac{2}{d}[2(f_3)^2f_4 + 2f_2f_3f_4] \\
\tilde{p}_{a,b} &= \frac{2}{d}[2f_1f_3f_4 + 2f_1f_2f_4] \\
\tilde{p}_{b,a} &= \frac{2}{d}[2f_1f_3f_4 + 2f_1f_2f_4] \\
\tilde{p}_{b,b} &= \frac{2}{d}[2f_1f_2f_3 + 2f_1(f_2)^2]
\end{aligned}$$

As expected, given symmetric failure probabilities, $\tilde{p}_{a,b} = \tilde{p}_{b,a}$. Hence, we need only compare $\tilde{p}_{a,a}$ and $\tilde{p}_{b,b}$ to each other and to one of the off-diagonal outcomes. Start with $\tilde{p}_{a,a} > (<)\tilde{p}_{b,a}$. After cancellations, this reduces to $f_3(f_2 + f_3) > (<)f_1(f_2 + f_3)$ or simply $f_3 > (<)f_1$. Given weak genericity, $f_3 \neq f_1$. If $f_3 > f_1$ then $(a, a)$ is a Nash equilibrium and $\tilde{p}_{a,a} > \tilde{p}_{b,a}$; if the inequality is reversed then $(b, a)$ is Nash and $\tilde{p}_{b,a} > \tilde{p}_{a,a}$.

Similarly, the comparison between $\tilde{p}_{b,b}$ and $\tilde{p}_{a,b}$ reduces to $\tilde{p}_{b,b} = f_2f_3 + (f_2)^2 > (<)f_3f_4 + f_2f_4 = \tilde{p}_{a,b}$, which simplifies to $f_2(f_3 + f_2) > (<)f_4(f_3 + f_2)$.

If $f_2 > f_4$ then $(b, b)$ is a Nash equilibrium and $\tilde{p}_{b,b} > \tilde{p}_{a,b}$; if the inequality is reversed then $(a, b)$ is Nash and is more likely in the limit.

Finally, comparing $\tilde{p}_{a,a}$ to $\tilde{p}_{b,b}$ reduces to evaluating $2(f_3)^2 f_4 + 2f_2 f_3 f_4$ versus $2f_1 f_2 f_3 + 2f_1(f_2)^2$, respectively. Cancellations simplify this to $f_3 f_4(f_3 + f_2) > (<) f_1 f_2(f_3 + f_2)$, or just $f_3 f_4 > (<) f_1 f_2$. Suppose first that $(a, a)$ is a Nash equilibrium while $(b, b)$ isn't. The former implies that $f_1 < f_3$; the latter, that $f_2 < f_4$. Hence $f_1 f_2 < f_3 f_4$ and $\tilde{p}_{a,a} > \tilde{p}_{b,b}$. If $(b, b)$ is Nash while $(a, a)$ isn't, then $f_1 > f_3$ and $f_2 > f_4$, whence $\tilde{p}_{a,a} < \tilde{p}_{b,b}$. QED.

(If both or neither outcomes are Nash equilibria then Proposition 8 is silent. But it is interesting to note that their relative likelihood depends on the probability of moving from $(a, a)$ to $(b, b)$, which is $f_1 f_2$, versus the probability of going from $(b, b)$ to $(a, a)$, which is $f_4 f_3$. This is intuitive.)

## References


Bendor, Jonathan, Daniel Diermeier, and Michael Ting. 2003a. "A Behavioral Model of Turnout." *American Political Science Review* 97, 261-80.

Bendor, Jonathan, Daniel Diermeier, and Michael Ting. 2003b. "Recovering Behavioralism: Adaptively Rational Strategic Behavior with Endogenous Aspirations." In Ken Kollman, John Miller, and Scott Page (eds.), *Computational Models in Political Economy.* Cambridge: The MIT Press.

Bendor, Jonathan, Daniel Diermeier, and Michael Ting. 2004. "The Empirical Content of Behavioral Models of Adaptation." Presented at the Annual Meeting of the Midwest Political Science Association, Chicago.

Bendor, Jonathan, Dilip Mookherjee, and Debraj Ray. 2001. "Aspiration-Based Reinforcement Learning in Repeated Interaction Games: An Overview." *International Game Theory Review*, 3, 159-74.

Bendor, Jonathan, Dilip Mookherjee, and Debraj Ray. 2006. "Satisficing and Selection in Electoral Competition." *Quarterly Journal of Political Science*, 1, 171-200.

Bendor, Jonathan, Sunil Kumar, and David Siegel. 2007. "Satisficing: A *Pretty* Good Heuristic." Unpublished ms., Graduate School of Business, Stanford University.

Borgers, Tilman, and Rajiv Sarin. 2000. "Naive Reinforcement Learning with Endogenous Aspirations." *International Economic Review*, 41, 921-

50.

Camerer, Colin, and Teck Ho. 1999. "Experience-Weighted Attraction Learning in Normal Form Games." *Econometrica*, 67, 287-74.

Cho, In-Koo, and Akihiko Matsui. 2005. "Learning Aspiration in Repeated Games." *Journal of Economic Theory*, 124, 171-201.

Collins, Nathan, Sunil Kumar, and Jonathan Bendor. "The Adaptive Dynamics of Turnout." *Journal of Politics*, forthcoming.

Elster, Jon. 1986. "Introduction." In Elster (ed.), *Rational Choice.* New York: New York University Press.

Erev, Ido, and Alvin Roth. 1998. "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique Mixed Strategy Equilibria." *American Economic Review*, 88, 848-81.

Feinberg, Yossi. 2004. "Learning to Optimize while Optimizing Learning." Unpublished ms., Stanford University.

Fiske, Alan, and Philip Tetlock. 1997. "Taboo trade-offs: Reactions to transactions that transgress the domain of relationships." *Political Psychology*, 18, 255-97.

Gigerenzer, Gerd. 2001. "The Adaptive Toolbox." In G.Gigerenzer and R. Selten (eds.) *Bounded Rationality: The Adaptive Toolbox.* Cambridge MA: The MIT Press.

Gilboa, Itzhak, and David Schmeidler. 1995. "Case-Based Decision Theory." *Quarterly Journal of Economics*, 110, 605-39.

Karandikar, R., D. Mookherjee, D. Ray and F. Vega-Redondo. 1998. "Evolving Aspirations and Cooperation." *Journal of Economic Theory*, 80, 292-331.

McGraw, A. Peter, and Philip Tetlock. 2005. "Taboo Trade-Offs, Relational Framing, and the Acceptability of Exchanges." *Journal of Consumer Psychology*, 15, 2-15.

Macy, M., and A. Flache. 2002. "Learning dynamics in Social Dilemmas." *Proceedings of the National Academy of Sciences*, 99, 7225-36.

March, James. 1991. "Exploration and Exploitation in Organizational Learning." *Organization Science*, 2, 71-87.

Myerson, Roger. 1978. "Refinements of the Nash equilibrium concept." *International Journal of Game Theory*, 7, 73-80.

Roth, Alvin, and Ido Erev. 1995. "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term." *Games and Economic Behavior*, 8, 164-212.

Salmon, Timothy. 2001. "An Evaluation of Econometric Models of Adaptive Learning." *Econometrica*, 1597-1628.

Shor, Mikhael. 2004. "Learning to Respond: The Use of Heuristics in Dynamic Games." Unpublished Ms., Vanderbilt University.

Simon, Herbert. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics*, 69, 99-118.

Simon, Herbert. 1956. "Rational Choice and the Structure of the Environment." *Psychological Review*, 63, 129-38.

Simon, Herbert. 1996. *The Sciences of the Artificial*. $3^{rd}$ ed. Cambridge: MIT Press.

Todd, Peter, and Gerd Gigerenzer. 2000. "Authors Response." *Behavioral and Brain Sciences*, 23, 767-77.

Tversky, Amos. 1972. "Elimination by aspects: a theory of choice." *Psychological Review*, 79, 281-99.