REVISITING THE COMPETENCY TRAP

JERKER DENRELL
WARWICK BUSINESS SCHOOL, UNIVERSITY OF WARWICK, SCARMAN ROAD, COVENTRY CV4 7AL, WEST MIDLANDS, UNITED KINGDOM, DENRELL@WBS.AC.UK

GAEL LE MENS
UNIVERSITAT POMPEU FABRA

Abstract. We revisit the competency trap and reexamine when it occurs. We show that a bias against alternatives that improve with practice does not require that learning is myopic in the sense of lacking foresight or failing to explore. The same bias occurs even if learners engage in substantial exploration and have foresight. In fact, we demonstrate that even a rational and foresighted learner, who follow an optimal strategy for balancing exploration and exploitation, will learn to prefer alternatives with initially high payoffs that decrease with practice over alternatives, with identical expected values, that have initially low payoffs that increase with practice. Our results show that a bias against alternatives that improve with practice is due to an asymmetry in error correction rather than to myopic learning. The implication is that a wide range of selection systems, even optimally designed ones, will be biased against late-bloomers.

This is a pre-copyedited, author-produced version of an article accepted for publication in Industrial and Corporate Change following peer review. The version of record [Jerker Denrell, Gaël Le Mens, Revisiting the competency trap, Industrial and Corporate Change, Volume 29, Issue 1, February 2020, Pages 183–205, DOI: 10.1093/icc/dtz072] is available online at:
`https://academic.oup.com/icc/article-abstract/29/1/183/5675622`

1. Introduction

Jim was always interested in illustrations of how sensible learning processes could lead to suboptimal outcomes. Throughout his career, he collected, developed and honed a series of 'small ideas' (as he liked to call them) about flawed learning. Each idea is like a parable, providing theorists and practitioners with lessons about the limitations of adaptation processes in realistic settings. Like parables, small ideas consist of a simple storyline, containing only a few select elements chosen to emphasize the basic ideas. The purpose is not to be realistic, but to convey insight.

One of Jim's small ideas about flawed learning is the 'competency trap'; a parable about how competence in an activity may trap people into this activity (Levinthal and March, 1981; Leavitt and March, 1988). In its simplest version this is a story about an individual who has to select between two activities (Denrell and March, 2001). The individual is familiar with one of the alternatives, has tried it many times, and has become proficient in it. The other activity is new, unfamiliar, and the individual is less proficient in it. When the new activity becomes available, the individual decides to try it out, to check if it could be better. But the first time the individual tries the new activity it seems worse. The individual gives the new activity another try, but again performance is below that of the established activity. At this point, the individual has had enough and reverts back to the familiar activity. The problem is that if the individual had persisted with the new

activity, and had become more proficient in it, performance would have improved. The new activity does in fact have a higher potential performance than the old activity, but realizing this potential takes practice. By avoiding the novel activity based on poor initial performance, the individual will not practice enough and will never discover the higher potential of the new activity. The individual will continue believing that the new action is inferior and can claim that the data does indeed back this up. The irony is that this mistake occurs only if the individual is competent in the old activity. An individual unfamiliar with both activities, old and new, will not be biased and will discover the higher potential of the new activity. Like the little boy in Hans Andersen's tale of the 'Emperors New Clothes', the incompetent, inexperienced, and 'foolish' individual (March, 1971) is the only one who can see the truth.

The parable of the competency trap illustrates a simple reason for why seemingly objective comparisons can be misleading: what you see is not what you get. The problem is one of misplaced confidence in performance measures. Measures of short-run performance do not take into account the long-run potential of activities. If performance improves with practice, and the amount of practice differs between activities, short-run measures of the performance of different activities will be misleading as indicators of their long-run potential. The broader implication is that any learning or adaptation process which reduces the tendency to choose actions with poor short-run performance will be biased towards activities that do well initially, even if they have identical long-run potential. "The short run is privileged by organizational learning" (Levinthal and March, 1993, p. 101). Several other theorists have made similar claims that adaptive processes tend to be myopic and hence will not identify the best practices when those practices tend to do poorly initially (Nelson and Winter, 1982; Elster, 1984; Levinthal and March, 1993; Sterman, 2000).

The purpose of this paper is to revisit the small idea of the competency trap and reexamine when it occurs and why. Our focus is on when and why learning might be biased against alternatives with initially poor payoffs that increase with practice. Our intention is to sketch a new version of the parable of the competency trap; a version that illustrates why a simplistic interpretation of why the bias against the novel alternative occurs is not correct. The simplistic interpretation (an interpretation we ourselves used to hold) is this: the mistake the individual made was to behave as if short-run performance reflects long-run potential. That is, the individual was myopic in the sense of ignoring or failing to consider the long-run (Levinthal and March, 1993, p. 101). As a result, the individual did not trade-off the possibility of a high long-run potential against the short-run performance disadvantage. The bias emerges, according to this argument, because the learner prematurely avoids alternatives with poor short-run performance. If the individual had taken into account the fact that the performance of the new activity may increase, the individual would have realized the need to explore and persist with the novel activity, even if its performance is poor initially, to find out if it might be superior. If such exploration was conducted, the argument suggests, then the learner would not necessarily be biased against the novel alternative. The bias is thus a result of the assumption that the learning process is myopic in the sense of ignoring or failing to consider the long-run resulting in insufficient exploration.

Here we show that the assumption that learning is myopic is not the (only) reason why learning is biased against alternatives with poor short-run performance. The models developed in this paper show that a bias against alternatives that require practice holds even for learning algorithms that are not myopic but do consider the long-run and do engage in substantial exploration. Hence, lack of exploration

is not the reason why the bias occurs. We also show that even a rational, non-myopic, learner, who is aware that payoffs may increase with practice, may also be biased against alternatives with initially low payoffs that increase with practice. The learner is rational in the sense the decision maker knows the structure of the task he or she is facing, prior beliefs about unknown parameters are correct, beliefs are updated according to Bayes rule, and the learner chooses the policy that maximizes total expected payoff. We show that such a learner will end up being biased against an alternative with initially low payoffs that increase with practice even when the learner is rational and, as such, knows that the initial payoffs may not be representative of the long-term payoffs.

Specifically, a rational learner will be less likely to choose such an alternative and more likely to choose an alternative with identical total expected value with a payoff profile that starts out high but decreases with practice. Stated differently, even a rational learner will behave like the individual in the parable of the competency trap and will favor the alternative with high initial performance compared to the alternative with high long-run but poor short-run performance.

This modeling exercise shows that the origin of the bias is not necessarily the assumption that learning is myopic in the sense that it ignores the long-run. Rather, there are 'structural features of decision-making' that lead to the bias. By 'structural features of decision-making,' 'we mean consequences of the process that are implicit in the decision model itself, rather than a result of behavioral applications of the model'(Harrison and March, 1984, p. 27). In this case, there is an asymmetry in the cost of correcting underestimation errors between alternatives whose payoffs increase versus decrease with practice (Denrell, 2007). If an alternative is initially believed to be good, the learner will find out whether the payoff decreases or not without intentionally exploring. If an alternative is initially poor, the only way a learner will find out more about this alternative is by intentionally exploring. That is, to learn more about an alternative with initially poor payoffs, and correct any initial underestimation, the learner has to choose an alternative different from the alternative which is believed to have a highest expected payoff in the next period. This asymmetry in the cost of correcting errors of underestimation between alternatives that increase or decrease is the source of the bias. Another way to formulate our results is to say that they show that learning is myopic in a deeper sense: even rational learning will have to rely on the observed payoffs to make decisions about whether the stop choosing an alternative or not. An initially increasing alternative, with poor payoffs initially, is at a disadvantage compared to an initially decreasing alternative with high payoffs initially, because the initially increasing alternative is more likely to be initially confused with an alternative that always generates poor payoffs.

Our results extend the applicability of the logic behind the competency trap to settings in which managers are not myopic but realize that there may exist favorable long-term consequences of pursuing an alternative with poor initial payoffs. In many settings, it is realistic to assume that managers do consider long-term effects. Most managers do not expect that R & D investments will pay off immediately. Most managers know that it takes some time before anything useful is generated and they expect to fund projects that initially do not generate anything valuable for the organization. Here we demonstrate that even if they do so, they will nevertheless be biased against alternatives with poor short-run performance.

More generally, the models developed in this paper provide an alternative explanation for (seemingly) myopic behavior. Several scholars have argued that managers tend to behave myopically when faced with inter-temporal decisions (Hayes

and Abernathy, 1980; Laverty, 1996), favoring activities with high short-term performance and avoid activities with identical net present value but with negative performance in the short-term. Scholars have attributed such managerial myopia to incentives based on short-term performance (Narayanan, 1985; Stein, 1989), to capital markets that emphasize the short-run (Jacobs, 1991; Porter, 1992), and to inconsistent managerial time preferences and weakness of will (Strotz, 1956; Elster, 1984; Postrel and Rumelt, 1992; Loewenstein, 1996; Bazerman, Tenbrunsel, and Wade-Benzoni, 1998). The models developed in this paper illustrate that rational learning can offer an alternative explanation of (seemingly) myopic behavior. Myopic behavior does not necessarily have to be explained by incentives that ignore long-term effects or by irrational behavior. We only need to assume that decision makers cannot be certain, ex ante, whether payoffs will improve with practice or not. Such an alternative explanation of myopic behavior offers new insight into the origins of myopic behavior. For example, it has been argued that myopic behavior may result from incentives that emphasize short-run profitability and ignores long-run effects. A learning perspective suggests that seemingly myopic behavior can emerge even if managers have incentives that do focus on long-run performance. Myopic behavior emerges in our model even if managers try to maximize total performance.

The structure of this paper is as follows. In the next section we describe the task the manager is facing and discuss when a learning procedure is biased against a particular alternative. In section 3, we turn to the issue of what type of learning policies that generate a bias against alternatives with payoffs that are initially low but improve with practice. We show that this evaluative bias emerges even for policies that initially explore substantially. In Section 4, we show this evaluative bias also emerges if the manager chooses options according to an optimal policy—a policy in which belief updating follows Bayes' rule and choices maximize the total expected payoff. In Section 5, we note that there are settings in which a rational manager may favor alternatives with payoffs that are initially low but improve with practice. However, in section 6 we show that if we compare an initially increasing and an initially decreasing alternative, the optimal policy is, under quite general conditions, biased against the initially increasing alternative. Section 7 shows that we get similar results if we consider a setting with binomial payoffs (success or failure). Section 9 explores the implications for beauty, justice, and truth and section 10 concludes.

2. Task

We focus on a simple 'one-armed bandit' scenario in which a learner can choose, in each of T periods, either a sure alternative with a known payoff distribution or an uncertain novel alternative. The learner can only observe, in period t, the payoff of the alternative chosen in that period. We assume that the sure alternative is known to generate a payoff equal to zero in each period. The uncertain alternative generates a payoff drawn from a normal distribution, with variance one. The variance is known to the learner, but the mean is unknown. The mean may depend on the number of times it has been tried in the past. Let u_i, $i = 1, ..., T$, be the mean payoff of the uncertain alternative if it is the ith time this alternative is tried. Note that this set-up focuses on a comparison between a known alternative and an unknown alternative. The intention is to formalize the comparison in the initial parable between a novel alternative and an established, and presumably known, alternative.

Given this set-up, we wish to evaluate the claim that learning tends to be biased against alternatives with initially low payoffs that improve with practice. This claim

refers to a certain subclass of unknown alternatives, with a particular functional relationship between the mean payoff u_i and the trial number i: u_i is lower than the payoff of the known alternative (zero) in the first trials. The claim is that there is a 'bias' against this type of 'increasing' unknown alternatives.

To assess this claim, we must specify the learning and choice policy: how does the learner rely on observed payoffs to estimate the value of the unknown alternative and when will the learner choose the unknown alternative? Below we consider several different learning and choice policies, heuristic as well as rational. Before that, we need to discuss how 'bias' should be defined. In particular, when should a learning algorithm be said to be biased against alternatives with initially low payoffs that improve with practice?

First, a sensible definition of bias needs to take into account the expected total value of an alternative. For example, it is not strange that an unknown alternative with mean payoffs $u_1 = -5$, $u_2 = -1$, $u_3 = -1$ and $u_4 = -1$ is avoided when compared to a known alternative with a known payoff equal to zero. The expected value of this unknown alternative is negative in all periods, so it is not strange that it is often avoided. To make the comparison fair, we should compare alternatives with identical expected total payoffs.

A possible definition of bias is then that a bias occurs when a learning algorithm ends up choosing an unknown alternative with total expected value equal to zero ($\sum_{i=1}^{T} u_i = 0$) less than 50% of the time. This definition has a drawback: according to this definition, a bias would occur even for an uncertain alternative with constant mean payoffs equal to zero in each period ($u_i = 0$). The reason is the hot-stove effect (Denrell and March, 2001; Denrell, 2007): such an alternative may generate a poor payoff and the poor payoff leads to avoidance, which implies that the learner will not find out about the possibility of high payoffs. This definition is hence too wide, capturing risk (or ambiguity) aversion as well as bias against alternatives that improve with practice. The problem is that in a wide range of learning processes, there is a bias against unknown and uncertain alternatives.

To avoid this problem, we narrow the definition of a bias, and compare the fates of two types of unknown alternatives. Specifically, we compare two proportions: a) the proportion of times a learning algorithm ends up choosing an unknown alternative of type A (which may have payoffs that increase with practice) over a known alternative and b) the proportion of times a learning algorithm ends up choosing an unknown alternative of type B (which may have payoffs that decrease with practice) over a known alternative. Both types A and B are unknown and are thus disadvantaged compared to the known alternative. Our focus here, however, is on whether alternatives of a particular class are especially disadvantaged (i.e., chosen less frequently).

We compare unknown alternatives with identical total expected values, to make the comparison fair. To complete the definition, we also need to define over what periods "chosen less often" should be measured. Are we interested in choices over all periods $1, ..., T$ or only the last period? Here we follow Denrell and March (2001) and focus on the last period T. We are interested in whether the learner might eventually come to realize, after having observed the high later payoffs, that an alternative with initially low payoffs is indeed valuable. This completes our definition of *Weak* bias:

Definition 1. *Consider two finite horizon one-armed bandit problems with T periods. In the first problem the decision maker decides between a known alternative (with sure payoff of 0) and an uncertain alternative with unknown payoff distribution, A. In the second problem, the choice is between the same known alternative and another uncertain alternative with unknown payoff distribution, B.*

A learning and choice policy is weakly biased *against alternative A if the total expected payoffs of alternatives A and B are identical ($\sum_{i=1}^{T} u_{A,i} = \sum_{i=1}^{T} u_{B,i}$), but the proportion of times alternative A is chosen in period T is lower than proportion of times alternative Bis chosen in period T.*

The weak bias can occur in at least two ways that we wish to distinguish. Suppose A is such that $u_{A,1} = -1$, $u_{A,2} = -1$, $u_{A,3} = 5$ and B is such that $u_{B,1} = 1$, $u_{B,2} = 1$, $u_{B,3} = 1$. The total expected values of A and B are the same (equal to 3) but A is worse during all periods except for the last. A large class of learning policies that are backward-looking and only react to past payoffs, and have no knowledge of, or make any guesses about, the future will be weakly biased against A (see Elster, 1978, Ch. 2). The reason is that if the expected total payoffs are the same over all three periods, and the last payoff is the higher for A (5 versus 1), then the total expected payoff over the first two periods has to be lower for A. Only the first two periods impact the choice in period three, however (unless the learning algorithm knows about or anticipates the outcome in the last period). The reason for the weak bias in this case is thus simply that A has a lower total expected value during the periods that matter for the choice in the last period (periods one and two). The fact that it has a higher expected payoff in period three does not impact the choice in period three in any way, unless the learning algorithm has foresight or knows that the payoff in the last period could be high.

To distinguish this reason for a bias from other potential reasons, we define another type of bias: the *strong* bias. Its definition excludes the scenario we just described. For the strong bias, we require that the total expected payoffs of A and B over periods $1, ..., T-1$ be identical. This implies that the expected total payoffs are identical during the periods that precede the last choice. We also require that the total expected payoffs be the same in the last period, T, such that the total expected payoffs over all periods are the same. For example, suppose A is such that $u_{A,1} = -1$, $u_{A,2} = 2$, $u_{A,3} = 2$ and B is such that $u_{B,1} = 0.5$, $u_{B,2} = 0.5$, $u_{B,3} = 2$. The total expected payoffs of these two alternatives over periods one and two are both 1 and their expected payoffs in the last period are both 2. They also have the same total expected payoff over all three periods (equal to 3). If one of these alternatives is favored, we will say that there is a *strong* bias in favor of that alternative.

More generally, our definition of *Strong* bias is the following:

Definition 2. *Consider the setting of Definition 1. A learning and choice policy is strongly biased against alternative A over alternative B if $\sum_{i=1}^{T-1} u_{A,i} = \sum_{i=1}^{T-1} u_{B,i}$, and $\sum_{i=1}^{T} u_{A,i} = \sum_{i=1}^{T} u_{B,i}$, but the proportion of times alternative A is chosen over the known alternative in period T is lower than proportion of times alternative B is chosen over the known alternative in period T.*

The two requirements that $\sum_{i=1}^{T-1} u_{A,i} = \sum_{i=1}^{T-1} u_{B,i}$ and $\sum_{i=1}^{T} u_{A,i} = \sum_{i=1}^{T} u_{B,i}$ imply that the average payoffs in the last period have to be identical, $u_{A,T} = u_{B,T}$. Consider again the two alternatives A and B introduced just before the definition (A: $u_{A,1} = -1$, $u_{A,2} = 2$, $u_{A,3} = 2$, B: $u_{B,1} = 0.5$, $u_{B,2} = 0.5$, $u_{B,3} = 2$). Alternative A is initially poor but increasing. Alternative B is initially 'flat', at a moderate value of 0.5. Both alternatives have the same expected payoff (2) the third time they are chosen.

Obviously, we make no claim that this set of payoff profiles are commonly occurring or realistic in anyway. Rather, they are quite artificial. What we have done is defined a set of payoff profiles that have properties such that a bias cannot occur

simply because a) there is a bias against unknown alternative over known alternatives b) the learning algorithm only relies on payoffs observed before the period in which the choice is made.

In almost all cases, strong bias implies weak bias. In the following, we therefore focus on strong bias.

3. Myopic Learning Policies Are Not Necessary For Bias Emergence

To explore when learning leads to strong bias we compare the proportion of choices of the unknown alternative in two different scenarios

(1) 'Increasing scenario:' The learner chooses, in each of three periods, between a known alternative that generates a payoff equal to zero and an uncertain alternative with an unknown mean. In this scenario, the mean of the unknown alternative is initially lower than that of the known alternative but it increases with the number of times it is tried. Specifically, the mean payoff of the unknown alternative if it is tried one, two, or three times is $u_{A,1} = -1$, $u_{A,2} = 2$, $u_{A,3} = 2$. Whenever the uncertain alternative is chosen it generates a payoff drawn from a normal distribution with mean $u_{A,j}$ and variance one.
(2) 'Comparison scenario:' The learner chooses, in each of three periods, between a known alternative that generates a payoff equal to zero and an uncertain alternative with an unknown mean. In this scenario, the mean payoff of the unknown alternative if it is tried one, two, or three times is $u_{B,1} = 0.5$, $u_{B,2} = 0.5$, $u_{B,3} = 2$. Whenever the uncertain alternative is chosen it generates a payoff drawn from a normal distribution with mean $u_{B,j}$ and variance one.

Both unknown alternatives have the same total expected value over all three periods and over the first two periods. The unknown alternative in the 'increasing' scenario is characterized by an increasing mean payoff: the mean payoff is initially lower than that of the known alternative, but higher from the second time it is tried. Our focus is on the proportion of times the unknown alternative will be chosen in the last period, i.e., in period three. If there is a strong bias against the increasing alternative, then the unknown alternative would be chosen less often in period three in the 'increasing scenario' than in the 'comparison scenario.'

To explore when this happens, suppose first that the learner adopts a myopic learning rule, in the sense that the learning rule only chooses the unknown alternative in any given period if it seems superior to the known alternative. Specifically, suppose the unknown alternative is chosen in period one but only chosen in period $t > 1$ if the average observed payoff is equal to or higher than zero.[1] This rule implies the learner never trades-off the short-run versus the long-run. In particular, the learner never explores: he or she never chooses the unknown alternative when it is believed to be worse than the known alternative in order to learn more about the payoff distribution of the unknown alternative.

This myopic rule does generate a (strong) bias against the alternative with initially poor but increasing payoffs. Simulations show that in the 'increasing' scenario, the learner chooses the unknown alternative 16 % of the time in period three (based on 100 000 simulations). In contrast, in the 'comparison' scenario, the learner chooses the unknown alternative 62 % of the time in period three. In this latter case, the initially 'flat' alternative is much more likely to be chosen in the

[1]This policy implies that the unknown alternative is never chosen again if the known alternative is once chosen. The reason is that the average remains the same.

third period. There is thus a 'Strong Bias' against the unknown alternative of the 'increasing scenario' type.

Does this bias occur because learning is 'myopic'? To examine this, consider first what happens if the learner adopts a less myopic and more exploratory policy. Suppose that the learner keeps track of the average observed payoff of the unknown alternative and chooses the unknown alternative in period t if the average is larger than some cutoff c_t. Consider the policy ($c_2 = -1$, $c_3 = 0$), i.e., the learner chooses the unknown alternative in the second period if the first payoff, r_1, is equal or higher than $c_2 = -1$ and then chooses the unknown alternative again in the third period if the average of the first and second period payoffs, $.5r_1 + .5r_2$, is equal to or more than zero. This policy is exploratory, in the sense that the unknown alternative will be chosen in period two even if the learner believes that it is worse, on average, than the known alternative. Even though this policy is exploratory, it also generates a bias against alternatives with initially poor payoffs. Simulations show that in the 'increasing' scenario, the learner chooses the unknown alternative 47 % of the time in period three (based on 100 000 simulations). By contrast, in the 'comparison' scenario, the learner chooses the unknown alternative 75 % of the time in period three. The initially 'flat' alternative is again much more likely to be chosen in the third period.

Why does the bias occur? Figure 1 illustrates the underlying mechanism by plotting when the unknown alternative is chosen in the third period as a function of the first and second period payoffs. To be chosen in the third period, the first period payoffs, r_1, must be at or above -1. The average after two periods, $.5r_1+.5r_2$, must also be at or above zero. The unknown alternative is thus only chosen in the third period if the first and second period payoffs fall into region in the upper right corner. Figure 1 also plots the contours of the joint probability densities for the first and second period payoffs in the two scenarios. Because the first and second period payoffs are independently normally distributed with identical variance (one), the contours are circles centered on the averages ($u_{j,1}$ and $u_{j,2}$). The increasing alternative is centered at $r_1 = -1$ and $r_2 = 2$ while the initially flat alternative is centered at $r_1 = 0.5$ and $r_2 = 0.5$.

Figure 1 shows that, in the two scenarios, the unknown alternatives are equally likely to satisfy the constraint that the average after two observations is at or above zero. This is easily understood: i) the constraint only depends on the sum of the first and second period payoffs and ii) the distribution of the sum is identical since the $u_1 + u_2$ are identical and the random variables are normally distributed. Figure 1 also shows that the unknown alternative is less likely to satisfy the constraint that the first period payoff is at or above -1 in the 'increasing scenario.' Overall, then, the unknown alternative is less likely to simultaneously satisfy the two constraints in the 'increasing scenario.'

The reason for the bias against the initially increasing alternative is thus that it is less likely to satisfy the constraint that the first period payoff should be no less than -1. This argument clearly holds for policies other than ($c_2 = -1$, $c_3 = 0$). For example, the same argument holds for any policy of the form ($c_2 = -k$, $c_3 = 0$). For example, suppose ($c_2 = -2$, $c_3 = 0$). Simulations show that in the 'increasing' scenario, the learner chooses the unknown alternative 71 % of the time in period three (based on 100 000 simulations). By contrast, in the 'comparison' scenario, the learner chooses the unknown alternative 76 % of the time in period three.

The same basic argument holds for any policy of the form (c_2, c_3) whenever c_2 and c_3 are finite cutoff levels. For example, suppose the policy is ($c_2 = 1$, $c_3 = 1$). This policy only selects the unknown alternative if it is believed to be quite a lot better than the known alternative. Again, it generates a bias against the increasing

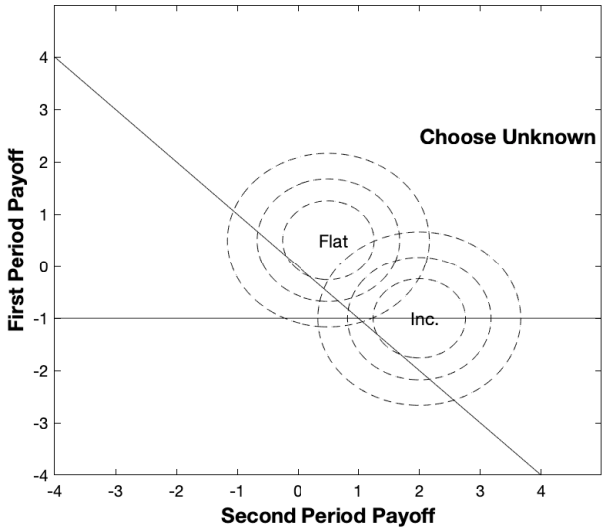

FIGURE 1. When the unknown alternative is chosen in the third period as a function of the first and second period payoffs when the policy is ($c_2 = -1$, $c_3 = 0$).

alternative (unknown alternative chosen 2 % of the time in the 'increasing scenario,' versus 26 % of the time in the 'comparison scenario'). The only case that does not lead to a bias against the initially increasing alternative is if there is no threshold after the first period ($c_2 = -\infty$). That is, suppose the learner always chose the unknown alternative at least two times. Then there is no bias: the proportion of times the unknown alternative is chosen in the third period is the same in both scenarios.

This conclusion is not limited to the particular payoff sequences we compared, nor to normally distributed alternatives, but holds generally for any 3 period payoff sequences with similar characteristics, as formulated in the following theorem:

Theorem 1. *Suppose that*

(1) *the unknown alternative belongs to one of two types, A and B,*
(2) *the payoff of the unknown alternative of type j, if chosen for the ith time, is drawn from a random variable with distribution $u_{j,i} + e_{j,i}$ where $e_{j,i}$ are for all i and j independently drawn from a density $f(e_{j,i})$ with infinite support, i.e., $e_{j,i} \in (-\infty, +\infty)$ and mean zero,*
(3) *$u_{A,1} < u_{B,1}$, $u_{A,1} + u_{A,2} = u_{B,1} + u_{B,2}$, and $u_{A,3} = u_{B,3}$,*
(4) *the learner always chooses the unknown alternative in period one, chooses it in period two if the first payoff is larger than c_2, and chooses it in the third period if the average observed payoff is larger than c_3.*

Then, unless $c_2 = -\infty$, this learning procedure is strongly biased against Alternative A.

A similar argument also holds if there are more than three periods. To state this argument, we first define when alternative A is 'cumulatively dominated' by alternative B in the sense that low payoffs tends to occur early on for alternative A:

Definition 3. *Suppose* $\sum_{i=1}^{T-1} u_{A,i} = \sum_{i=1}^{T-1} u_{B,i}$. *Moreover, for all periods* $j \in [1, T-2]$, $\sum_{i=1}^{j} u_{A,i} < \sum_{i=1}^{j} u_{B,i}$. *Then alternative A is strictly cumulatively dominated by alternative B.*

Theorem 2. *Suppose*

(1) *the unknown alternative belongs to one of two types, A and B,*
(2) *the payoff of the unknown alternative of type j, if chosen for the ith time, is drawn from a random variable with distribution* $u_{j,i} + e_{j,i}$ *where* $e_{j,i}$ *are for all i and j independently drawn from a density* $f(e_{j,i})$ *with infinite support, i.e.,* $e_{j,i} \in (-\infty, +\infty)$ *and mean zero,*
(3) *alternative A is strictly cumulatively dominated by alternative B.*
(4) *the learner always chooses the unknown alternative in period one and chooses it in period* $t > 1$ *if the average observed past payoff is larger than* $c_t = k_t$.

Then, unless all cutoffs $c_2, ..., c_{T-1}$ *are equal to* $-\infty$, *this learning procedure is strongly biased against Alternative A.*

These results imply that a bias against alternatives that are initially poor but improve with the number of selections will emerge for a large class of learning policies, not only 'greedy' or 'myopic' policies. As Theorem 2 shows, the bias emerges even if the learner explores during many initial periods. Suppose, for example, that there are 10 periods and the learner always choose the unknown alternative during periods one through eight, but then only chooses the unknown alternative in period nine if the average, after eight observed payoffs, is above some cutoff. Then the bias against alternatives with initially poor payoffs occurs. This is true even if the alternative with initially poor payoffs has higher average payoff, compared to the initially flat alternative, in period eight. The relevant payoff, however, is the sum of all payoffs during the first eight periods. The initially increasing alternative has a lower total expected payoff during the first eight periods (see the definition of cumulatively dominated). If there is some cutoff during these periods, the initially poor alternative will be at a disadvantage. The bias only disappears if the learner explores in all periods until the last.

4. Even a Farsighted Rational Learner may be Biased

Theorem 2 shows that a bias against alternatives with initially low payoffs hold for a wide range of policies, even those which initially engage in substantial exploration. But Theorem 2 only applies to a class of heuristic policies, that rely on the average payoff observed prior to each choice. Does the bias against alternatives with poor initial payoffs also hold if the learner is rational and anticipates that an alternative may be poor initially but improves with practice? A rational learner will not simply rely on the average payoff observed so far to decide whether the alternative should be discontinued or not. He or she will also make inferences about the type of the uncertain alternative based on the pattern of observed payoffs: if the payoff is initially low but increases with practice this suggests that the alternative of the kind that increases with practice. Such inferences can also impact whether the learner will explore or not: if the learner suspects that the alternative may improve with practice the learner may decide to explore during some initial periods, to check whether the payoffs improve. In this section, we demonstrate that even a rational learner may be biased against alternatives with low initial payoffs that improve with practice, even if the learner is aware that their payoff may increase. Whether or not a bias occurs, however, depends on the set of payoff sequences the learner believes are possible.

4.1. **Model setup.** As before, we consider a one-armed bandit problem. The learner can choose, in each of $T = 3$ periods, either a known alternative or an unknown alternative. The learner only observes, in period t, the payoff of the alternative chosen in that period. We assume that the known alternative is known to generate a payoff equal to zero in each period.

The unknown alternative generates a payoff drawn from a normal distribution with variance one. The mean of the unknown alternative depends on the number of times it has been selected in the past. Let u_i, $i = 1, ..., T = 3$, be the mean payoff of the unknown alternative if it is the ith time this alternative is selected. We assume that the unknown alternative is one of three possible types:

(1) Type A: Initially Increasing: $u_{A,1} = -1$, $u_{A,2} = 2$, and $u_{A,3} = 2$.
(2) Type B: Initially Flat: $u_{B,1} = 0.5$, $u_{B,2} = 0.5$, and $u_{B,3} = 2$.
(3) Type C: Always Poor: $u_{C,1} = -2$, $u_{C,2} = -2$, and $u_{C,3} = -2$.

The type of the unknown alternative is drawn at the start of period one and remains the same during all three periods. Each type is equally likely to be drawn (drawn with probability 1/3).

Finally, we assume the learner is rational: he or she updates beliefs using Bayes rule and selects alternatives according to an optimal policy that maximizes the (undiscounted) sum of the payoffs during the $T = 3$ periods: $\sum_{t=1}^{T=3} r_t$, where r_t is the payoff generated in period t.

4.2. **Comment.** We have assumed that the learner knows that the unknown alternative can be of the type that has an initially low average payoff but which increases with practice. This is not the only possibility, however. If it were - if the unknown alternative was known to be of the initially increasing type - then the learner would know the expected value of the unknown alternative and would know that it was always best to select it over the known alternative. The learner also knows that the unknown alternative may always be poor (of type C). It is this possibility that makes the decision problem non-trivial and makes learning essential: based on the observed payoffs the learner has to try to infer the type of the unknown alternative. Finally, there is the possibility that the unknown alternative is initially flat (type B). This type has the same total expected payoff as the initially increasing alternative (type A), both over all three periods and in periods one and two. We are interested in whether a rational learner will display a bias against the initially increasing alternative over the initially flat alternative

4.3. **Optimal policy.** The optimal policy maximizes the (undiscounted) sum of the payoffs during the $T = 3$ periods, $\sum_{t=1}^{T=3} r_t$. It can be computed by backward induction, starting from the last period. It works as follows (see Appendix B for the formal details):

Suppose the learner has chosen the unknown alternative in periods one and two. Given the observed payoffs, it is easy to compute the expected payoff from choosing the unknown alternative in period three. The expected payoff is simply the sum, over the three possible types, of the product of $u_{j,3}$ (the expected payoff the third selection given that the alternative is of type j) and the probability that the unknown alternative is of type j given the observed payoffs. The optimal policy, in period three, is then to select the unknown alternative if its expected payoff is above zero, the value of the known alternative. The expected payoff, at the beginning of period three, is thus the maximum of zero and the period three expected payoff of the unknown alternative.

We can then calculate the value of choosing the unknown alternative in the second period, given the observed payoff in the first period. The value of choosing the unknown alternative in the second period is the sum of the immediate payoff

plus the expected value in the third period. The immediate payoff in period two is the expected payoff given the observed payoff in the first period. To find the expected value in the third period, we have to integrate over the possible values of the second period payoff. We now know the value of choosing the unknown alternative in period two and three, given the observed payoff in period one. We can then find the value of the observed payoff in period one that makes the learner indifferent between choosing the unknown or the known alternative in period two.

Finally, note two features of the optimal policy. First, the unknown alternative should be chosen in the first period because, intuitively, i) the expected value from choosing it during all three periods is zero, and ii) there is a positive probability that the learner concludes, after two periods, that the unknown alternative is unlikely to be of the always poor type, in which case the expected value from choosing the unknown is higher than the known. Second, the unknown alternative is only chosen in period three if it has been chosen in the period two. Intuitively, this is because the immediate payoff of choosing the unknown alternative in period three after avoiding it in period two is the same as the immediate payoff of choosing it in period two. If it is avoided in period two, it should also be avoided in period three. It may be chosen in period two, however, and then avoided in period three.

4.4. **The optimal policy generates a bias against initially poor alternatives.** If we calculate the optimal policy, we find it is of the following form: 1) Always choose the unknown alternative in period 1; 2) choose the unknown alternative in period two if the observed payoff in period one was more than -2.185; 3) if the unknown alternative has been chosen twice, then choose the unknown alternative in period three if the payoffs in the first and second period are sufficiently large. Figure 2 illustrates the form of the optimal policy. The unknown alternative is only chosen in the third period if the first and second period payoffs fall into the region in the upper right corner. Figure 2 also plots the contours of the joint probability densities for the first and second period payoffs when the unknown alternative is initially increasing (type A) or when it is initially flat (type B). Because the first and second period payoffs are independently normally distributed with identical variance (one), the contours are circles centered on the averages ($u_{j,1}$ and $u_{j,2}$). The increasing alternative is centered at $r_1 = -1$ and $r_2 = 2$ while the initially flat alternative is centered at $r_1 = 0.5$ and $r_2 = 0.5$.

Figure 2 suggests that the initially increasing alternative is less likely to be chosen in the third period because it is less likely to satisfy the constraint on the first period payoff (being larger than -2.185). This is also the case: simulations show that the learner chooses the unknown alternative 87.5 % of the time in period three when it is of the 'initially increasing' type. In contrast, the learner chooses the unknown alternative 96.6 % of the time in period three when it is of the 'initially flat' type (based on 100 000 simulations). The unknown alternative is chosen more often in period three when it is of the 'initially increasing' type. There is thus a 'Strong Bias'.

This illustration shows that even an optimal policy can lead to a bias against an alternative with initially low average payoffs which increase with practice. Notice that this bias occurs even if the learner is aware that the unknown alternative may be of this type. Moreover, the bias also implies that the learner is more likely to underestimate the expected value of the unknown alternative if it is of the initially increasing type compared to when it is of the initially flat type. This happens even if the learner knows these two alternatives are equally likely and have equal expected total payoff. Why then does the bias occur? For very much the same reason that it occurs for the heuristic policies: the initially increasing alternative is less likely to satisfy the constraint that the initial payoff is above -2.185. Why

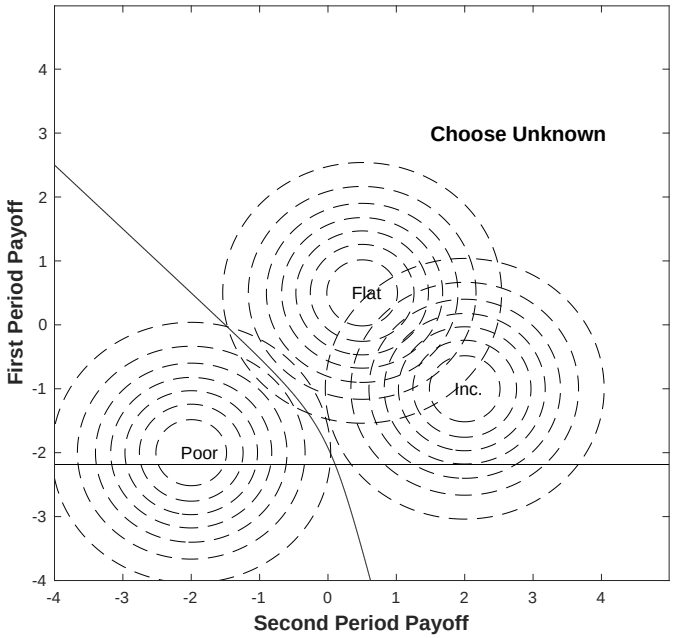

FIGURE 2. Region in which the unknown alternative is chosen in the third period as a function of the first and second period payoffs when the learner follows an optimal policy when the 'always poor' alternative has expected payoff -2.

then does the first period payoff have to satisfy this constraint? The reason is that alternatives which generate lower payoffs in the first period are likely to be of the always bad type, which the learner wants to avoid.

4.5. **A second period advantage for the initially increasing alternative.** A more detailed examination of the shape of the optimal policy, in Figure 2, shows a subtle difference between the reason why the bias occurs under the optimal policy and the reason it occurs under the heuristic policy considered in section 3.

In the heuristic policy considered in section 3, and illustrated in Figure 1, the initially increasing and the initially flat alternatives were equally likely to satisfy the constraint after two periods. The reason was that the choice only depended on the sum (average) of the past payoffs. In the third period, this sum was identically distributed for the initially increasing and the initially flat alternatives (because the sum of the expected payoffs over period one and two were equal).

Under the optimal policy, illustrated in Figure 2, the initially increasing alternative is *more* likely to satisfy the constraint after two periods than the initially flat alternative. The reason for this asymmetry is evident from Figure 2: Given two observations, the initially flat alternative is more likely to be confused with the always poor alternative than the initially flat alternative. This can be seen from the fact that the contour lines for the initially flat and the always poor alternatives overlap more than the contour lines do for the initially increasing and the always poor alternatives. The intuitive reason for this is that the initially increasing payoff has a high second period average payoff, which is very unlikely to be confused with a payoff generated by the always poor alternative. Thus, even if its initial

payoff is low, and thus makes it more likely to be confused with the always poor alternative after one observation, this alternative has a payoff profile that makes it more distinguishable from the always poor alternative after two observed payoffs than the initially flat alternative. Because the two alternatives differ in how similar they are to the always poor alternatives, the optimal policy is not symmetric in the first and second period payoffs. For example, the unknown alternative satisfies the constraint after the second period if second period payoff is -4 and the first period payoff is a bit larger than -2.5. If the first period payoff is -4, however, the second period payoff only has to be a bit larger than 0.8 to satisfy the constraint after the second period. The second period payoff thus matters more than the first.

5. A Rational Learner Can Favour an Initially Increasing Alternative

Unlike the heuristic policy, the optimal policy is not always biased against the initially increasing alternative. In fact, the optimal policy may be biased against the initially flat alternative and favour the initially increasing alternative. Such a reverse bias occurs if the optimal policy is a bit more exploratory than what was the case in the numerical example of the previous section. The optimal policy is more exploratory, in turn, if the always poor alternative is not as bad. To illustrate this, consider a set-up identical to that of the previous section except for the mean payoff of the Always Poor type. The three types are:

(1) Initially Increasing: $u_1 = -1$, $u_2 = 2$, and $u_3 = 2$.
(2) Initially Flat: $u_1 = 0.5$, $u_2 = 0.5$, and $u_3 = 2$.
(3) Always Poor: $u_1 = -1.5$, $u_2 = -1.5$, and $u_3 = -1.5$.

As before, the type of the unknown alternative is drawn at the start of period one and remains the same during all three periods. Each type is equally likely to be drawn (with probability 1/3).

The only difference to the setting of the previous section is that the Always Poor alternative is better than it was. This implies that the optimal policy will be more exploratory: the learner can afford to experiment more. This intuition can be confirmed by examining Figure 3 which shows the shape of the optimal policy for this problem. The optimal policy is to always choose the unknown alternative in period one, choose it after period one if the payoff in period one is more than -3.1, and choose it in period three if the first and second period payoffs are sufficiently high.

Simulations show that this optimal policy generates a bias against the initially flat alternative. In the 'increasing' scenario, the learner chooses the unknown alternative 96.6 % of the time in period three (based on 100 000 simulations). In contrast, in the 'comparison' scenario, the learner chooses the unknown alternative 94.5 % of the time in period three. The initially 'flat' alternative is thus less likely to be chosen in period three and there is a 'Strong Bias' against it. The reason for this bias in favor of the initially increasing alternative is that this alternative is less likely to be confused with the always poor alternative than the initially flat alternative, if the payoffs in both periods one and two have been observed. Hence, if the unknown alternative is chosen in periods one and two, then it is more likely to satisfy the constraint for selection in period three if it is of the initially increasing type. It is still less likely to satisfy period one constraint, but in this case this constraint only applies to payoffs below -3.1 and thus matters less.

This example illustrates a phenomenon that did not occur with the heuristic model: a bias in favor of the initially increasing alternative which occurs because it is less likely to look like an alternative which is always poor. It is important to note, however, that this advantage of the initially increasing alternative has nothing to do

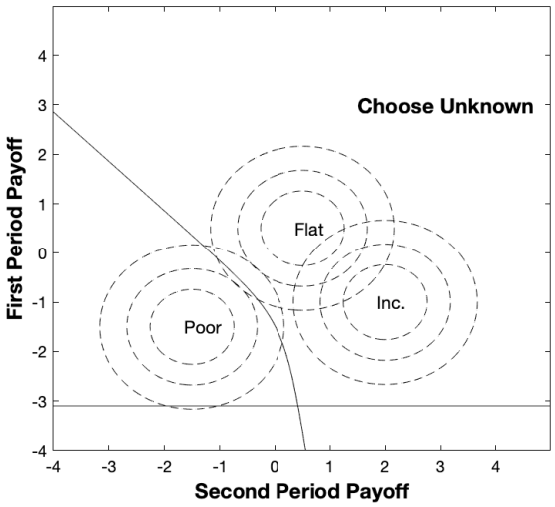

FIGURE 3. Region in which the unknown alternative is chosen in period three as a function of the first and second period payoffs when the learner follows an optimal policy and the 'always poor' alternative has expected payoff -1.5.

with the temporal order of the expected payoffs. The same advantage would hold for an alternative with expected payoffs of the first two periods in reverse order, i.e., high initial and low middle period payoff.

6. Rational Learners have a Bias Against Initially Increasing over Initially Decreasing Alternative

In sections 4 and 5, the unknown alternative was one of three types: initially low payoffs that increased with practice, initially flat payoffs, or consistently low payoffs. We showed that whether or not there is a bias against the initially increasing alternative over the initially flat alternative depended on how exploratory the optimal policy was. In this section we show that if we instead compare an initially increasing and an initially decreasing alternative, the optimal policy is, under quite general conditions, biased against the initially increasing alternative. That is, the proportion of times the unknown alternative will be chosen in period three is lower if the unknown alternative is of the initially increasing type compared to when it is of the initially decreasing type.

To illustrate this, suppose the unknown alternative can be of these three types, each equally likely:

(1) Initially Increasing: $u_1 = -1$, $u_2 = 2$, and $u_3 = 2$.
(2) Initially Decreasing: $u_1 = 2$, $u_2 = -1$, and $u_3 = 2$.
(3) Always Poor: $u_1 = -2$, $u_2 = -2$, and $u_3 = -2$.

Note that the mean payoffs for initially increasing and initially decreasing alternatives during the first two periods are identical numbers but in reverse order.

Figure 4 shows the shape of the optimal policy. As shown, the constraint after two observed payoffs is symmetrical in the first and second period payoffs: if

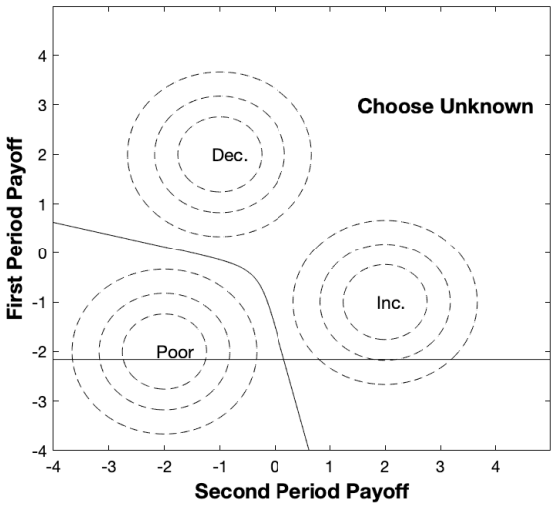

FIGURE 4. Region in which the unknown alternative is chosen in period three as a function of the first and second period payoffs when the learner follows an optimal policy.

$(r_1, r_2) = (a, b)$ satisfies the constraint then so does $(r_1, r_2) = (b, a)$. The implication is that the initially increasing and the initially decreasing alternatives are equally likely to satisfy the constraint after two observations. Because the initially increasing alternative is less likely to satisfy the constraint after the first payoff, it will be less likely to be chosen in period three. Simulations confirm this bias against the initially increasing alternative: if the unknown alternative is of the Initially Increasing type, the learner chooses it 86.5 % of the time in period three (based on 100 000 simulations). In contrast, if the unknown alternative is of the Initially Flat type, the learner chooses it 98.3 % of the time in period three.

The emergence of this bias can be demonstrated quite generally:

Theorem 3. *Suppose the unknown alternative can be of three possible types, each equally likely:*

(1) *Initially Increasing:* $u_{I,1} = b$, $u_{I,2} = a$, $u_{I,3} = a$,
(2) *Initially Decreasing:* $u_{D,1} = a$, $u_{D,2} = b$, $u_{D,3} = a$,
(3) *Always Poor:* $u_{P,1} = c$, $u_{P,2} = c$, $u_{P,3} = c$,

where $a > b > c$, $a > 0$, $a + b > 0$, *and* $c < 0$. *Then the optimal policy implies that the proportion of times the unknown alternative is chosen over the known alternative in period three is lower when it is of the Initially Increasing type than when it is of the Initially Decreasing type.*

The size of the bias depends on the value of the known alternative. If the value of the known alternative is lower, the optimal policy is more exploratory and the learner tends to persist longer with the unknown alternative and will be less likely to discard the initially increasing alternatives based on failure in early periods. For example, if the value of the known alternative is -1, the cutoff after the first period is -3.25. This is lower than when the known has a value of zero (in this case the cutoff is at -2.16). When the value of the known alternative is -1, the learner chooses the initially increasing type 97.9% of the time in period three and chooses the initially

decreasing type 99.2% of the time (based on 100 000 simulations). These results illustrate how even an optimal policy leads to a competency trap in the sense that more competent actors, who have a known alternative with a higher payoff, will be more biased against a new alternative with initially increasing payoff.

7. Binomial Payoffs

The same basic result - that an optimal policy has a bias against an alternative with increasing payoffs over an alternative with identical expected value but a decreasing sequence of payoffs - also hold if payoffs are binary, either equal to one or to zero. In this case it is also possible to compute the optimal policy if there are more than three periods. To illustrate this, consider the case with $T = 10$ periods. In each of ten periods the learner can choose between a known and an unknown alternative. The learner can only observe, in period t, the payoff of the alternative he or she chose in period t. Each alternative can succeed (generate a payoff equal to +1) or fail (generate a payoff equal to zero). The known alternative succeeds, in every period it is tried, with a known probability, $p_k = 0.5$. The probability that the unknown alternative will succeed may depend on the number times it has been chosen: the probability of a success when the unknown alternative is chosen for the ith time is equal to p_i, $i = 1, ..., 10$. The learner does not known the sequence p_i, $i = 1, ..., T$ but knows that it is equally likely to be of one of four types:

(1) Always Good: $p_i = 0.9$ for $i = 1, .., 10$.
(2) Initially Increasing: $p_i = 0.51 + (i-1)\frac{0.9-0.51}{9}$ for $i = 1, ..., 9$ and $p_{10} = 0.9$.
(3) Initially Decreasing: $p_i = 0.9 - (i-1)\frac{0.9-0.5}{9}$ for $i = 1, ..., 9$ and $p_{10} = 0.9$.
(4) Always Poor: $p_i = 0.1$ for $i = 1, .., 10$.

Note that the probability of a success for the initially increasing alternative starts at $p_i = 0.51$ and increases to 0.9 in period nine. The probability of a success for the initially decreasing alternative starts at $p_i = 0.9$ and decreases to $0.51 > 0.5$ in period nine. Both of these alternatives have the same success probability equal to 0.9 in the last period. Figure 5A plots how the success probabilities vary with the number of times the unknown alternative has been chosen. Note that the value 0.51 was chosen so that the success probability of the initially decreasing alternative never falls below the success probability of the known alternative, which is 0.5. Note also that both the initially increasing and the initially decreasing alternatives have the same success probability in the last period, to ensure that their expected payoffs are identical over both the first nine periods and over all ten periods, consistent with the scope of our definition of 'Strong Bias' (Definition 2).

We assume that the learner wants to maximize the (undiscounted) sum of all payoffs: $\sum_{t=1}^{10} r_t$, where r_t is the payoff (success or failure) obtained in period t. The optimal policy can again be computed by dynamic programming, starting from the last period. Given the optimal policy, we can simulate the proportion of time a learner who follows the optimal policy will choose the unknown alternative if it is of the initially increasing type or the initially decreasing type. Figure 5B plots how the choice probabilities evolve over time. As shown, a learner who follows an optimal policy is less likely to choose the unknown alternative when it is of the initially increasing type than when it is of the initially decreasing type. In period ten, the learner chooses the initially decreasing type almost all the time (97.2%) but only chooses the initially increasing type 75.6% of the time (based on 100 000 simulations). Again, this bias in favor of the initially decreasing alternative emerges even though these alternatives have identical expected values (both over the first nine periods and over all ten periods) and even though the learner knows these types are equally likely.

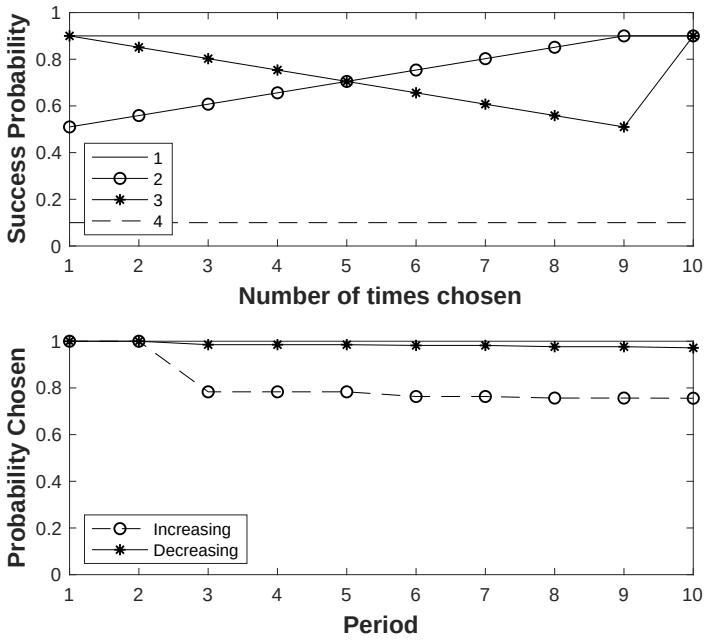

FIGURE 5. A) The success probabilities for the four different types of the unknown alternative. B) The probability of choosing the unknown alternative over the known alternative in period t if it is of the initially increasing or the initially decreasing type.

If the success probability of the known alternative is lower, the bias against the initially increasing alternative is less strong. For example, if $p_k = 0.4$, the learner chooses the initially decreasing type 98.4% of the time in period ten and chooses the initially increasing type 87.2% of the time (based on 100 000 simulations). If $p_k = 0.2$, the learner chooses the initially decreasing type 99.4% of the time in period then and chooses the initially increasing type 96.2% of the time (based on 100 000 simulations). Just as in the previous section, these results show that more competent actors, who have a known alternative with a high payoff, will be more biased against a new alternative with initially increasing payoff.

8. CAVEAT

The results in this paper rely on one important assumption: the learner does not get any information about the payoff sequences of the unknown alternative if the learner chooses the known alternative. This assumption may not hold: the learner may be able to observe the choices and payoffs of others. If individuals were able to learn about the unknown alternatives even if they did not choose it, the bias against the initially increasing alternative may disappear.

For example, consider the setting of section 3, with normally distributed payoffs. Suppose the learner chooses the unknown alternative in the first period, but only chooses it in the second period if the payoff in the first period is higher than zero. Suppose, however, that the learner will observe, in period two, what the payoff would have been from choosing the unknown alternative a second time, even if the learner chose the known alternative in period two. Finally, in the third period the

learner chooses the unknown alternative if the average observed payoff is above zero. In this case, there is no bias against an initially increasing alternative. To explain this, compare the two unknown alternatives: Initially Increasing ($u_{A,1} = -1$, $u_{A,2} = 2$, $u_{A,3} = 2$) and Initially Flat ($u_{B,1} = 0.5$, $u_{B,2} = 0.5$, $u_{B,3} = 2$). The distribution of the sum of the first two payoffs is identical, with expected value $u_{i,1} + u_{i,2} = 1$ and variance two. As a result, the probability that the average observed payoff is above zero after two observations is identical for the two alternatives. If the two initial payoffs are always observed, these two alternatives are equally likely to be chosen in period three.

Analyzing the case with a rational learner is more complex. Suppose the learner can observe, in period two, what the payoff would have been from choosing the unknown alternative a second time, even if the learner chose the known alternative in period two. The learner can then update his or her estimate of whether the unknown alternative is of the initially increasing type. Still, if the learner chooses the unknown alternative in the third period, it will only be the second time the learner selects this alternative, not the third, which may impact whether the learner chooses it or not.

9. Beauty, Justice, and Truth

We have shown that the bias against alternatives with increasing payoffs holds quite generally, for a large class of policies, not only greedy or myopic policies. Indeed, the bias occurs even if a learner is rational and farsighted, aware of the possibility that payoffs may be increasing. Our results show that the bias is more general than what prior results have suggested. Our results also show that the bias is not is not a consequence of the 'myopic' character of learning, if myopia is defined as lack of awareness of the possibility that payoffs may be increasing or if myopia is defined as a greedy policy which does not explore. Our results show that learning is myopic in a deeper sense: even rational learning will have to rely on the observed payoffs to make decisions about whether to stop choosing an alternative or not. An initially increasing alternative, with poor payoffs initially, is at a disadvantage compared to an initially decreasing alternative, with high payoffs initially. The disadvantage occurs because the initially increasing alternative is, initially, more likely to be confused with an alternative that always generates poor payoffs.

What are relevant practical implications of these results? Following Jim, we choose not to speculate on this. As Jim liked to say: "I am not now, nor have I ever been, relevant." Instead of commenting on relevance, we instead try to explain why the ideas have some beauty, do matter for justice, and contain a grain of truth.

9.1. **Beauty.**

The results are beautiful partly because the finding that even highly exploratory policies generate a bias against initially increasing alternatives is a bit surprising at first. Myopia, in the sense of a greedy policy or lack of foresight, is not needed. In hindsight, the reason for the bias seems clear: the initially increasing alternative is less likely to make the initial cutoff, if there is one. Figure 1 provides a simple visual proof. If two alternatives have the same probability distribution of the sum of the first two payoffs, they are equally likely to make the cut-off after two payoffs have been observed. To be selected in period three, however, an alternative also has to make the initial cut-off, after period one. If there is any initial cut-off, any alternative with a low expected value in period one will be at a disadvantage.

It is pleasing that the reason why the optimal policy generates a bias is essentially the same (see Figure 4). If two payoffs have been observed, the initially increasing and the initially decreasing alternatives are equally likely to make the cutoff and be chosen in period three. To be selected in period three, however, an alternative

also has to make the initial cut-off, after period one and any alternative with a low expected value in period one will be at a disadvantage.

The result that even the optimal policy generates a bias is initially surprising. Why would a rational decision-maker, who knows that an alternative may be poor initially, and knows that the expected values of the initially increasing and the initially decreasing alternatives are equal be biased against the initially increasing alternative? But the reason for the bias under the optimal policy turns out to be very similar to why a myopic policy generates a bias: as long as there is any initial cut-off, any alternative with a low expected value in period one will be at a disadvantage. This insight existed in conjectural form even before this paper. For example, Denrell and March (2001) state that

> ...one way of avoiding a bias against alternatives that require practice is rapid learning of competence and slow learning of choices -longer experimentation with alternatives that seem poor. The latter would clearly be a good idea save for the obvious complication that in a world in which seemingly poor alternatives most often are poor, extended periods of experimentation with apparently poor alternatives is usually a very costly strategy. (Denrell and March, 2001, p. 527)

That is, unless the learner persists for a long time with an alternative that requires practice it will on, average, appear to be worse than an alternative with opposite (decreasing) payoff profile.

The result that even the optimal policy generates a bias is also a bit mysterious, even if the mechanism is understandable. Why would it be rational to be biased against late bloomers? Why are decreasing alternatives advantaged? The reason is an asymmetry in the cost of correcting underestimation errors (Denrell, 2007). If an alternative is initially believed to be good, but may eventually decrease in payoff, the learner will find out whether the payoff decreases or not without exploring (Posen and Levinthal, 2012). If an alternative is initially poor, the only way a learner will find out more about this alternative is by exploring. That is, to learn more about an alternative with initially poor payoffs, and correct any initial underestimation, the learner has to choose an alternative different from the alternative which is believed to have a highest expected payoff in the next period. This asymmetry in the cost of correcting errors of underestimation between alternatives that increase or decrease is the source of the bias.

The result that even the optimal policy generates a bias is also interesting because it implies that 'debiasing' people will not eliminate the bias (Le Mens & Denrell, 2011). Even an optimally designed system will show this bias. The only way to eliminate the bias is to engage in substantial exploration, so that there is no cut-off before the last period. Usually, this is not an optimal policy. The fact that rational learners will show the bias also has the interesting implication that irrational individuals may be most likely to avoid the bias. That is, the individual who do discovers the advantages of initially increasing alternatives may be those who overestimate them. Managers who have an irrational belief in the value of persistence will be more likely to discover when persistence does in fact pay off (Hirschman, 1967; Denrell and March, 2001). As the Polish philosopher Lezek Kolakowski noted, when crossing a desert occasional hallucinations may be necessary. Kolakowski writes that a strongly held conviction in the value of the eventual destination acts like

> ....a Fata Morgana which makes beautiful lands arise before the eyes of the members of a caravan and thus increases their efforts to the point where, in spite of all their sufferings, they reach the next tiny

> waterhole. Had such tempting mirages not appeared, the exhausted caravan would inevitably have perished in the sandstorms, bereft of hope. (Kolakowski, 1961, pp. 127-128, quoted in Hirschman, 1967, p. 32).

In a similar way, an entrepreneur with a strong conviction in the eventual success of his or her business, is more likely to persist despite initial failures. For example, Bowers interviewed one entrepreneur who has endured endless setbacks in his entrepreneurial career, but says that he is driven by his belief in his product. (Bowers, 2005, p.131). Even if this belief is initially irrational, the false belief may lead to persistence that allows the entrepreneur to confirm the belief. The model also implies that young people, without any good alternatives, are more likely to discover the potential value of alternatives with poor initial payoffs. Being young, they lack practice with any alternative. The value of the known alternative is thus low, which leads to exploration, which minimizes the chances of underestimating an alternative with initially increasing payoffs.

9.2. **Justice.** The results matter for justice because they show that late bloomers are at a disadvantage. The precocious individual, who is doing well initially, will be favored, even if the two individuals are equally productive over time. The results show that even an optimal selection system, designed to optimize the productivity of the chosen individuals, will have a bias against the late-bloomer. The bias is arguably unjust, because equally productive people, as measured over their lifetime or career, are not given equal chances. How can the injustice be eliminated? By avoiding early selection, i.e., by extending the period of initial exploration.

More generally, being aware of the possibility that payoffs increase with practice is important to avoid injustices due to premature stopping due to naive interpretation of poor early results. In an interview, Jim explained the implications of the competency trap as follows

> Some of my grandchildren say to me, "We're not very good at mathematics, so we're not going to take any more mathematics." I say, "Wait a minute. Mathematics is a practice sport. If you're not very good at it, you take more of it." That's counterintuitive, and it goes against the main logic of experiential learning, not to mention grandchildren's sentiments about control over their own lives. (James March, Interviewed by D. Coutu, Coutu, 2006, p. 86)

Jim notes that initial performance is not necessarily a good indicator of eventual performance, because math belongs to the set of activities for which performance increases with practice. Knowing this, and noting that their peers may differ in how much they have practiced, his grandchildren may interpret their experience differently and make different choices; choices that would not have considered if they had been sure that performance reflects a stable math talent.

9.3. **Truth.** When is the mechanism outlined here the correct, or most important, reason for seemingly myopic behavior, when is it of marginal importance, and when is it wrong? To examine these questions, one first needs to examine how this mechanism could be distinguished empirically from others. One possibility is to examine the impact of access to information about foregone payoffs, i.e., access to information about what the unknown alternative would have paid even in periods when one did not choose it. If the mechanism outlined here contributes to a bias against alternatives with initially increasing alternatives, access to such information should reduce the bias, following the arguments in section 8. One can also compare the strength of any bias in settings where decision-makers are informed about the

payoff profile of the unknown alternative to a setting where they do not know its payoff profile but have to learn it. In the former case, bias cannot be attributed to learning but may be the result of discounting. If the bias is equally strong in the two settings, learning does not seem to play an important role.

10. Conclusion

Perhaps this paper is an illustration of the competency trap. We started working on these models a long time ago. As a result of substantial practice with this type of modelling, and with the relevant literature, we have become relatively proficient in this activity. We know we can write this type of paper. Other possible theoretical or empirical projects may not seem as attractive in comparison. At least initially, we would not be as proficient and our effort would not seem to add much to the literature. Disappointed by such initial failures, we might continue to refine our models rather than branch out or try out some new techniques.

If this paper is an example of the competency trap, it is at least comforting to know that even rational learners can fall into the same trap, in the sense of being biased against an alternative with initially poor payoffs that may be increasing. Life is short and so are our careers; we cannot spend too much of it developing new tools we may not use very often or that may not suit us, especially if the payoff initially seems low. On the other hand, if the payoff from some alternative model or technique seems high early on, it does make sense to give it a try, even if it is possible that its promise will be illusory and its payoff will decrease over time.

References

Bazerman, M.H., Tenbrunsel, A.E. & Wade-Benzoni, K.A. (1998). Negotiating with Yourself and Losing: Understanding and Managing Conflicting Internal Preferences, Academy of Management Review, 23, 225-241.

Coutu, D. (2006). "Ideas As Art", Interview of James March by Diane Coutu, October, 84(10), 83-89.

Denrell, J. (2005). Why Most People Disapprove of Me: Experience Sampling in Impression Formation. Psychological Review, 112 (4): 951-978.

Denrell, J. (2007). Adaptive Learning and Risk Taking. Psychological Review, 114 (1): 951-978.

Denrell, J. & J. G. March. (2001). Adaptation as Information Restriction: The Hot Stove Effect. Organization Science, 12 (5): 523-538.

Einav, L. (2005). Informational asymmetries and observational learning in search. Journal of Risk and Uncertainty, 30, 241-259.

Elster, J. (1984). Ulysses and the Sirens: Studies in Rationality and Irrationality. 2nd Ed. Cambridge University Press, Cambridge.

Harrison, J. R., & March, J. G. (1984). Decision Making and Postdecision Surprises. Administrative Science Quarterly, 29(1), 26-42.

Hirschman, A. (1967). Development Projects Observed. Brookings Institution, Washington D.C.

Hayes, R.H. and Abernathy, W.J. (1980), Managing our way to economic decline, Harvard Business Review, Vol. 58 No. 4, pp. 67-77.

Jacobs, M. (1991). Short-Term America: The Causes and Cures of Our Business Myopia. Harvard Business Press: Boston, MA.

Laverty, K.J. (1996), Economic 'short-termism': the debate, the unresolved issues, and the implications for management practice and research, Academy of Management Review, Vol. 21 No. 3, pp. 825-60.

Le Mens, G., & Denrell, J. (2011). Rational learning and information sampling: On the "naivety" assumption in sampling explanations of judgment biases. Psychological review, 118(2), 379.

Levinthal, D. A., & March, J. G. (1993). The Myopia of Learning. Strategic Management Journal, 14(8), 95-112.

Levinthal, D. March, J.G. (1981). A model of adaptive organizational search. Journal of Economic Behaviorand Organization, 2(4), 307-333.

Levitt, B., & March, J. G. (1988). Organizational learning. Annual Review of Sociology, 14, 319-340.

Loewenstein, George. (1996). "Out of control: Visceral influences on behavior." Organizational Behavior and Human Decision Processes, 65, pp. 272-292.

March, J. G. (1996). Learning to be Risk Averse. Psychological Review, 103: 309-19.

March, J. G. (1971). The Technology of Foolishness. Civilokonomen (Copenhagen), 18(4), 4-12.

Narayanan, M. P. (1985). "Managerial Incentives for Short-term Results," Journal of Finance, 40, 1469-1484.

Nelson, R., S. G. Winter. (1982). An Evolutionary Theory of Economic Change. Harvard University Press, Cambridge, MA..

Porter, M. E. (1992). Capital Choices: The Causes and Cures of Business Myopia. Research Report to the U.S. Government's Council on Competitiveness, Washington D.C.

Postrel, Steven, and Richard P. Rumelt. (1992). "Incentives, Routines, and Self-Command." Industrial and Corporate Change 1,3: 397-425

Rosenbaum, J. E. (1984). Career mobility in a corporate hierarchy. New York: Academic Press, Inc.

Stein J. C. (1989). Efficient Capital Markets, Inefficient Firms: A Model of Myopic Corporate Behavior. The Quarterly Journal of Economics, 104(4), 655-669.

Sterman, J. (2000) Business Dynamics: Systems Thinking and Modeling for a Complex World. New York: Irwin/McGraw-Hill.

R.H. Strotz (1956). Myopia and inconsistency in dynamic utility maximization. Review of Economic Studies, 23(3), 165–180.

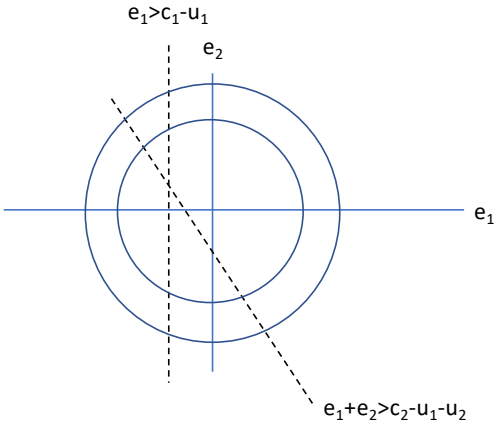

FIGURE 6. The area of the integral in the proof if Theorem 1.

APPENDIX A: PROOFS OF THEOREMS 1 AND 2

Proof of Theorem 1: Let j denote the type of the unknown alternative ($j = A$ or $j = B$). The probability of choosing the unknown alternative in period three is

$$P(r_1 + r_2 > c_3 \cap r_1 > c_2) = P(e_{j,1} + u_{j,1} + e_{j,2} + u_{j,2} > c_3 \cap u_{j,1} + e_{j,1} > c_2), \tag{1}$$

where $e_{i,j}$ are noise terms with mean zero independently drawn from density $f(e_{j,i})$. This can be rewritten as

$$P(e_{j,1} + e_{j,2} > c_3 - u_{j,1} - u_{j,2} \cap e_{j,1} > c_2 - u_{j,1}). \tag{2}$$

This probability, in turn, can be calculated as the following integral

$$\int_{(e_{j,1},e_{j,2})\in D} f(e_{j,1})f(e_{j,2})de_{j,1}de_{j,2} \tag{3}$$

where D is the region in which $e_{j,1}+e_{j,2} > c_3-u_{j,1}-u_{j,2}\cap e_{j,1} > c_2-u_{j,1}$. Figure 6 illustrates the region and also shows the contour lines for the joint normal density, $\frac{1}{2\pi}e^{-0.5e_1^2}e^{-0.5e_2^2}$. Nothing in this proof depends on the shape of these contour lines, however.

Now, the line defined by $e_{j,1}+e_{j,2} = c_3-u_{j,1}-u_{j,2}$ is always the same whenever the sum of the expected payoffs, $u_{j,1}+u_{j,2}$, remains the same. But, the line defined by $e_{j,1} = c_2 - u_{j,1}$ will be pushed towards the right when $u_{j,1}$ is lower. Note that we assumed $u_{A,1} < u_{B,1}$. It follows that the line defined by $e_{A,1} = c_2 - u_{A,1}$ will be pushed towards the right compared to the line defined by $e_{B,1} = c_2 - u_{B,1}$. Hence, the region that satisfies $e_{A,1} + e_{A,2} > c_3 - u_{A,1} - u_{A,2} \cap e_{A,1} > c_2 - u_{A,1}$ is smaller than the region that satisfies $e_{B,1} + e_{B,2} > c_3 - u_{B,1} - u_{B,2} \cap e_{B,1} > c_2 - u_{B,1}$. The only exception is when $c_2 = -\infty$, in which case the value of $u_{j,1}$ does not matter.

Proof of Theorem 2: The proof is similar to the proof of Theorem 1. The probability of choosing the unknown alternative in period T is

$$P(r_1 + ... + r_{T-1} > c_{T-1} \cap ... \cap r_1 > c_2). \tag{4}$$

This can be written as

$$P(\sum_{i=1}^{T-1} e_i > c_{T-1} - \sum_{i=1}^{T-1} u_{j,i} \cap ... \cap e_1 > c_2 - u_{j,1}), \tag{5}$$

where $e_1, ..., e_{T-1}$ are independently distributed noise terms drawn from density $f(e_i)$. This probability can be calculated as the following integral

$$\int_{(e_1,...,e_T)\in D} \prod_{i}^{T-1} f(e_i) de_1 ... de_T, \tag{6}$$

where D is the region in which $\sum_{i=1}^{T-1} e_i > c_{T-1} - \sum_{i=1}^{T-1} u_{j,i} \cap ... \cap e_1 > c_2 - u_{j,1}$.

The plane defined by $\sum_{i=1}^{T-1} e_i > c_{T-1} - \sum_{i=1}^{T-1} u_{j,i}$ is identical for alternatives A and B whenever $\sum_{i=1}^{T-1} u_{A,i} = \sum_{i=1}^{T-1} u_{B,i}$, but the region over which the integral is computed is smaller for alternative A because $\sum_{i=1}^{j} u_{A,i} < \sum_{i=1}^{j} u_{B,i}$ for all periods $j \in [1, T-2]$. In summary, the probability of choosing the unknown alternative in period T, when it is A, is lower than when it is B. In other words, the learning procedure is strongly biased against Alternative A.

The only exception is when all cutoffs $c_2, ..., c_{T-1}$ are equal to $-\infty$, in which case the regions over which the integral is computed are identical. In this case, the probability of choosing the unknown alternative in period T, is the same whether it is A or B.

Appendix B: Calculation of the Optimal Policy

We first calculate the optimal policy in period three, given that the unknown alternative has been chosen in periods one and two. Let τ denote the type of the unknown alternative ($\tau \in 1, 2, 3$). Given the observed payoffs in period one, r_1, and two, r_2, the expected value of choosing the unknown alternative in period three is

$$EC_3(r_1, r_2) = p(\tau = 1 \mid r_1, r_2)u_{1,3} + p(\tau = 2 \mid r_1, r_2)u_{2,3} + p(\tau = 3 \mid r_1, r_2)u_{3,3}. \tag{7}$$

where $p(\tau = j \mid r_1, r_2)$ is the posterior probability that the unknown alternative is of type j given the observed payoffs and the prior probabilities, p_j. Using Bayes theorem and the properties of the normal distribution we have

$$p(\tau = j \mid r_1, r_2) = \frac{p_j e^{-0.5(r_1 - u_{j,1})^2} e^{-0.5(r_2 - u_{j,2})^2}}{\sum_{i=1}^{3} p_i e^{-0.5(r_1 - u_{i,1})^2} e^{-0.5(r_2 - u_{i,2})^2}} \tag{8}$$

The optimal policy, in period three given that the unknown alternative has been chosen in periods one and two, is to select the unknown alternative whenever $EC_3(r_1, r_2) > 0$.

Once we know the value of choosing the unknown alternative in period three, we can calculate the value of choosing the unknown alternative in period two, given an observed payoff in period one. Consider then period two and suppose the unknown alternative was selected in period one. The value of choosing the unknown alternative in period two is the sum of the immediate expected payoff plus the expected value in period three.

The immediate expected payoff in period two is the expected payoff given the observed payoff in period one, r_1. To calculate this, we first calculate the posterior probability that the unknown alternative is of type j given that the payoff observed in period one was r_1:

$$p(\tau = j \mid r_1) = \frac{p_j e^{-0.5(r_1 - u_{j,1})^2}}{\sum_{i=1}^{3} p_i e^{-0.5(r_1 - u_{i,1})^2}} \tag{9}$$

The immediate expected payoff of choosing the unknown alternative in period two is

$$\pi_2(r_1) = \sum_{j=1}^{3} p(\tau = j \mid r_1)u_{j,2}. \tag{10}$$

The expected value in period three of choosing the unknown alternative in period two is: $\max[EC_3(r_1, r_2), 0]$, i.e., either the unknown is chosen in period three or the known alternative is chosen, depending on which has the highest expected value. We need to calculate the expected value of $\max[EC_3(r_1, r_2), 0]$ given the observe first period payoff, r_1. To calculate this, we integrate over the possible values of r_2:

$$val(r_1) = \int_{r_2} \max[EC_3(r_1, r_2), 0] f(r_2|r_1) dr_2, \tag{11}$$

Here $f(r_2|r_1)$ is the posterior density of period two payoff given the observed payoff in period one. This conditional density equals $f(r_2 \mid r_1) = \sum_{j=1}^{3} p(\tau = j \mid r_1) f(r_2 \mid \tau = j)$, i.e., it is a mixture density, where the density of each type is weighted by its posterior density. Overall then,

$$val(r_1) = \sum_{j=1}^{3} \left(p(\tau = j \mid r_1) \int_{r_2} \max[EC_3(r_1, r_2), 0] f(r_2 \mid \tau = j) dr_2 \right). \tag{12}$$

The total value of choosing the unknown alternative in period two and following an optimal policy thereafter is

$$EC_2(r_1) = \pi_2(r_1) + val(r_1). \tag{13}$$

We can compute $EC_2(r_1)$ for each value r_1 by numerical integration. If the function $EC_2(r_1)$ only crosses the zero line once at $r_1 = r_1^*$, is negative below $r_1 = r_1^*$ and positive above, then the optimal policy is to select the unknown alternative whenever $r_1 > r_1^*$. That is, r_1^* is a threshold such that if period one payoff is larger than r_1^* then the optimal policy is to select the unknown alternative in period two but if the period one payoff is below r_1^* the optimal policy is to select the known alternative in period two.

Finally, we can compute the expected value of selecting the unknown alternative in period one. This value is

$$EC_1 = \sum_{j=1}^{3} u_{j,1} p_j + \sum_{j=1}^{3} \left(p_j \int_{r_2} \max[val(r_1), 0] f(r_1 \mid \tau = j) dr_1 \right), \tag{14}$$

where p_j is the prior that the unknown alternative is of type j. The first term is the expected payoff in period one and the second term is the expected value of following an optimal policy after period one. If $EC_1 < 0$ the optimal policy is to always select the known alternative, if $EC_1 \geq$ the optimal policy is to select the unknown alternative in period one.

Abstract C: Proof of Theorem 3

Proof of Theorem 3: We first show that there is just one value $r_1 < r_1^*$ such that if $r_1 < r_1^*$ then the known alternative is chosen in period two and if $r_1 > r_1^*$ then the unknown alternative is chosen in period 2. This is equivalent to showing that there is a unique r_1^* such that if $r_1 = r_1^*$, then $EC_2(r_1) = 0$ and if $r_1 < r_1^*$, then $EC_2(r_1) < 0$ and if $r_1 > r_1^*$, then $EC_2(r_1) > 0$. Using equations 9, 11, 13, we can write

$$EC_2(r_1) = \sum_{j=1}^{3} p(\tau = j \mid r_1) \left(u_{j,2} + g_j(r_1) \right), \tag{15}$$

where

$$g_j(r_1) = \int_{r_2} \max[EC_3(r_1, r_2), 0] f(r_2 \mid \tau = j) dr_2$$

From this, $EC_2(r_1) = 0$ if only if

$$\sum_{j=1}^{3} \frac{p(\tau = j \mid r_1)}{p(\tau = 3 \mid r_1)} \left(u_{j,2} + g_j(r_1)\right) = 0, \tag{16}$$

We can rewrite this equation as

$$\frac{e^{-0.5(r_1-b)^2}}{e^{-0.5(r_1-c)^2}} \left(b + g_1(r_1)\right) + \frac{e^{-0.5(r_1-a)^2}}{e^{-0.5(r_1-c)^2}} \left(a + g_2(r_1)\right) + \left(c + g_3(r_1)\right) = 0 \tag{17}$$

We wish to show that the left hand side of 17 is an increasing function of r_1. Consider first $g_j(r_1)$, which is a function of $EC_3(r_1, r_2)$. To demonstrate that $EC_3(r_1, r_2)$ is increasing in r_1, note that

$$EC_3(r_1, r_2) = (1 - p(\tau = 3 \mid r_1, r_2))a + p(\tau = 3 \mid r_1, r_2)c, \tag{18}$$

where

$$p(\tau = 3 \mid r_1, r_2) = \frac{1}{1 + \frac{e^{-0.5(r_1-a)^2} e^{-0.5(r_2-b)^2}}{e^{-0.5(r_1-c)^2} e^{-0.5(r_2-c)^2}} + \frac{e^{-0.5(r_1-b)^2} e^{-0.5(r_2-a)^2}}{e^{-0.5(r_1-c)^2} e^{-0.5(r_2-c)^2}}}. \tag{19}$$

Taking the derivative with respect to r_1 shows that the assumptions $a > c$ and $b > c$ imply that $p(\tau = 3 \mid r_1, r_2)$ is decreasing in r_1. This, in turn, implies that $EC_3(r_1, r_2)$ is increasing in r_1. It follows that $g_1(r_1)$, $g_2(r_1)$, $g_3(r_1)$ are increasing in r_1.

Next, the assumption that $b > c$ implies that $\frac{e^{-0.5(r_1-b)^2}}{e^{-0.5(r_1-c)^2}}$ is increasing in r_1 and the assumption that $a > c$ implies that $\frac{e^{-0.5(r_1-a)^2}}{e^{-0.5(r_1-c)^2}}$ is increasing in r_1. Overall, it follows that the LHS of equation 17 is increasing in r_1.

Moreover, $\frac{e^{-0.5(r_1-b)^2}}{e^{-0.5(r_1-c)^2}}$ and $\frac{e^{-0.5(r_1-a)^2}}{e^{-0.5(r_1-c)^2}}$ converge to zero as $r_1 \to -\infty$. In addition, $EC_3(r_1, r_2)$ converges to c as $r_1 \to -\infty$ implying that $g_3(r_1)$ converges to to zero as $r_1 \to -\infty$. Because $c < 0$, it follows that if r_1 is sufficiently low the LHS of equation 17 will be negative. Because $EC_3(r_1, r_2)$ converges to a as $r_1 \to +\infty$, $g_1(r_1)$, $g_2(r_1)$, $g_3(r_1)$ converge to a as $r_1 \to +\infty$. Because $a > 0$ and $a + b > 0$, and $\frac{e^{-0.5(r_1-b)^2}}{e^{-0.5(r_1-c)^2}}$ as well as $\frac{e^{-0.5(r_1-a)^2}}{e^{-0.5(r_1-c)^2}}$ are positive, it follows that if r_1 is sufficiently high, the LHS of equation 17 will be positive. It follows there is a unique value r_1^* that satisfies the desiderata stated above.

The optimal policy selects the unknown alternative in period one. It selects the unknown alternative in period two if $r_1 > r_1^*$. Finally, it selects the unknown alternative in period three, given that the unknown alternative has been chosen in periods one and two, whenever $EC_3(r_1, r_2) > 0$. Because $a > 0$, it follows that there exists a region such that $EC_3(r_1, r_2) > 0$.

Note, next, that $p(\tau = 3 \mid r_1, r_2)$ is symmetric in r_1 and r_2: $p(\tau = 3 \mid r_1, r_2) = p(\tau = 3 \mid r_2, r_1)$. It follows that the unknown alternatives of type I and type D are equally likely to satisfy the constraint $EC_3(r_1, r_2) > 0$. In other words, if the unknown alternative has been chosen in the first two periods, then the likelihood it is chosen in period 3 when it is of type I (increasing) is the same as when it is of type D (decreasing).

If the unknown alternative is of type I (increasing), its mean payoff is b. If it is of type D (decreasing) its mean payoff is a. Because we assumed that $a > b$, we have that r_1 is more likely to be above the threshold r_1^* if the alternative is of type D. Overall, the probability of selecting the unknown alternative in period 2 is lower if it is of the increasing type. This, and the symmetry argument about period 3 jointly imply that the unknown alternative is less likely to be chosen in period 3 if it is of the type I as compared to when it is of the decreasing type D.