Thinking Ahead: The Decision Problem

by

Patrick Bolton* and Antoine Faure-Grimaud**

June 2005

Abstract: We propose a model of bounded rationality based on time-costs of deliberating current and future decisions. We model an individual decision maker's thinking process as a thought-experiment that takes time and let the decision maker 'think ahead' about future decision problems in yet unrealized states of nature. By formulating an intertemporal, state-contingent, planning problem, which may involve costly deliberation in every state of nature, and by letting the decision-maker deliberate ahead of the realization of a state, we attempt to capture the basic idea that individuals generally do not think through a complete action-plan. Instead, individuals prioritize their thinking and leave deliberations on less important decisions to the time or event when they arise.

JEL Classifications: D81, D84, C61

Key Words: Decision Problem, Thinking Costs, Bounded Rationality, Incomplete Planning

*David Zalaznick Professor of Business, Columbia Business School

**Professor of Finance, London School of Economics

We are grateful to Markus Brunnermeier, Luis Garicano, Oliver Hart, Margaret Meyer, John Moore, Nobuhiro Kiyotaki, Jean Tirole and Wei Xiong for helpful comments as well as the participants at several seminars for discussions and comments.

1 Introduction

This paper proposes a simple and tractable model of bounded rationality based on time-costs of deliberating current and future decisions. We introduce a deliberation technology based on the classical two-armed bandit problem (Gittins and Jones, 1974 and Rothschild, 1974) and model an individual decision maker's thinking process as a *thought-experiment* that takes time.

The basic situation we model is that of a boundedly rational decision-maker who thinks through a simple decision, such as which of two actions to take, by weighing in her mind the costs and benefits associated with each possible action through thought-experiments which take time. Eventually, following enough "thought-experimentation" the decision-maker (DM) becomes sufficiently confident about which action is best and takes a decision.

Although our model is built on the powerful multi-armed bandit framework, we depart from the classical bandit problem in a fundamental way by introducing the notion of 'thinking ahead' about future decision problems in yet unrealized states of nature. By formulating an intertemporal, state-contingent, planning problem, which may involve costly deliberation in every state of nature, and by letting the decision-maker deliberate ahead of the realization of a state, we attempt to capture the basic idea that individuals generally do not think through a complete action-plan. Instead, individuals *prioritize* their thinking and first think through the decisions that seem most important to them. They also generally leave deliberations on less important decisions to the time or event when they arise.

Such behavior is understandable if one has in mind deliberation costs but, as Rubinstein (1998) has noted, it is irreconcilable with the textbook model of the rational DM with no thinking costs:

"In situations in which the decision maker anticipates obtaining information before taking an action, one can distinguish between two timings of decision making: 1. Ex ante decision making. A decision is made before the information is revealed, and it is contingent on the content of the information to be received. 2. Ex post decision making. The decision maker waits until the information is received and then makes a decision. In standard decision problems, with fully rational decision makers, this distinction does not make any difference." [Rubinstein 1998, page 52]

There are at least three motivations for a model with costly deliberation such as ours. First, there is the obvious reason that the behavior captured by such a model is more descriptive of how individuals make decisions in reality[1]. Second, as we shall explain, such a model can provide new foundations for two popular behavioral hypotheses: "satisficing" behavior (Simon, 1955 and Radner, 1975) and decision-making under time pressure that takes the form of "putting out fires" (Radner and Rothschild, 1975). The main motivation of the current paper is to show how a model of decision-making with costly deliberation can explain both "satisficing" behavior and a prioritization of decision problems akin to "putting out fires". Third, a model with costly deliberation can also provide a tractable framework to analyze long-term incomplete contracting between boundedly rational agents. We analyze this contracting problem in our companion paper on "satisficing contracts" (Bolton and Faure-Grimaud, 2005).

Indeed, our initial objective was mainly to formulate a tractable framework of contracting between boundedly rational agents in response to Oliver Hart's observation that: "*In reality, a great deal of contractual incompleteness is undoubtedly linked to the inability of parties not only to contract very carefully about the future, but also to think very carefully about the utility consequences of their actions. It would therefore be highly desirable to relax the assumption that parties are unboundedly rational.*" [Hart, 1995, p. 81] However, having formulated our deliberation technology for boundedly rational agents we found that the decision problem is of sufficient independent interest to be discussed in a separate paper.

In our model the decision-maker starts with some prior estimate of the payoff associated with each possible action choice in every state of nature. She can either take her prior as her best guess of her final payoff and determine her optimal action-plan associated with that prior, or she can 'think' further and run an experiment on one action. This experiment will allow her to update her estimate of the payoff associated with that action and possibly to improve her action choice.

At each point in time, DM, thus, faces the basic problem whether to explore further the payoff associated with a particular action, search further other parts of her optimization problem, or make a decision based on what she has learnt so far. Since 'thinking ahead'

[1] Recent research in behavioral economics suggests the distinction between costly conscious deliberation and affective impulses, and that actual decisions are the outcome of complex interactions between affective and deliberative processes (see e.g. Metcalfe and Mischel, 1999, Bernheim and Rangel, 2002, Benabou and Pycia, 2002, and Loewenstein and O'Donoghue, 2004). A limitation of our model is that it entirely leaves out affective impulses. However, we think that this is not a critical limitation when our model is applied to business decisions and sophisiticated long-term contracting situations.

takes time, DM will generally decide to leave some future decisions that she is only likely to face in rare or distant states of nature to be thought through later.

As is well understood, thinking ahead and resolving future decisions allows a DM to make better current decisions only when current actions are partially or completely irreversible. Our decision problem, thus, involves irreversible actions, and thereby introduces a bias towards deliberation on future actions ahead of the realization of future states of nature.

To understand the underlying logic of our decision problem it is helpful to draw a parallel with the problem of irreversible investment under uncertainty involving a 'real option' (see Henry, 1974 and Dixit and Pyndick, 1996). In this problem the rational DM may choose to delay investment, even if it is known to generate an expected positive net present value, in an effort to maximize the value of the 'real option' of avoiding making investments that ex post turn out to have a negative net present value. If one interprets information acquisition through delayed investment as a form of time-consuming deliberation on future actions, one is able to infer from the 'real options' literature that a boundedly rational DM with time-costs of deliberation facing an irreversible action choice problem will behave like an investor facing a real option problem.

That is, the boundedly rational DM will postpone taking a current irreversible action until she is sufficiently confident that this action yields the highest payoff. The interesting observation, however, from our perspective is not so much that the boundedly rational DM will delay taking an action, but that she will eventually decide to act even if she has not fully resolved her entire future action-plan.

As helpful as the parallel with real options is in understanding the basic logic of our problem, it is not a perfect analogy. Also, to emphasize the differences with the real options problem we specialize our general framework to a decision problem with no real option value at all. That is, in our problem there is no doubt that a particular current action (investment) is clearly preferable. Nevertheless, in this problem a boundedly rational DM will generally choose to delay investment and think ahead about future decisions in some if not all future states of nature.

What is the reason for this delay, if no option value is present? The answer is that, by thinking ahead about future decisions the boundedly rational DM can reduce the time-lag between the realization of a state of nature and the date when DM takes a decision in that state. By reducing this time-lag the boundedly rational DM is able to reduce the overall expected lag between the time she makes a costly investment decision and the time when she recoups the returns from her investment. Because DM discounts future payoffs, reducing this time-lag raises her payoff. In other words, our framework models the general idea that *the benefit of 'thinking ahead' is to be able to react more promptly to new events, but the cost is delayed current decisions.* This is the main novel mechanism we study in this paper.

How does this framework provide a new foundation for satisficing behavior? In general it is optimal for the boundedly rational DM to engage in what we refer to as 'step-by-step' thinking. This involves singling out a subset of future decision problems and thinking these through first. If the thought-experiments on these problems reveal that the payoff from investing is appreciably higher than DM initially thought then DM will decide that she is *satisfied* with what she found and will choose to invest without engaging into further thinking or optimization on other future decisions she has not yet thought about. If, on the other hand, the thought-experiments reveal that the payoff from investing is no higher and possibly lower than initially thought then DM will continue thinking about other decision problems in yet unexplored future states of nature.

In other words, the boundedly rational DM will generally refrain from fully determining the optimal future action-plan and will settle on an incomplete plan which provides a satisfactory expected payoff. Note that, in our framework the satisficing threshold is determined endogenously, as the solution of an optimal stopping problem. Thus, our framework can address a basic criticism that has been voiced against the original satisficing hypothesis, namely that the satisficing threshold is imposed exogenously.

In what way does the boundedly rational DM behave as if she were 'putting out fires'? We show that quite generally the boundedly rational DM prioritizes her thinking by first choosing to think about the most important and urgent problems. It is in this sense that she behaves as if she were putting out fires. The original formulation of this behavioral hypothesis by Radner and Rothschild considered only very extreme situations, where DM had no choice but to put out fires. Our framework highlights that the general idea underlying the notion of putting out fires, that a boundedly rational DM prioritizes her thinking by focusing first on the most important problems or those most likely to arise, extends far beyond the extreme high stress situation considered by Radner and Rothschild.

Several other major insights emerge from our analysis. First, when the number of future decisions to think about is large so that the complexity of the overall planning problem is overwhelming then it is best to act immediately simply by guessing which action is best, and to postpone all thinking to the time when decision problems arise. This result is quite intuitive and provides one answer to the well known 'how-to-decide-how-to-decide' paradox one faces when one introduces deliberation costs into a rational decision-making problem (see Lipman, 1995)[2]. In the presence of deliberation costs DM faces a larger decision problem, as she has to decide how to economize on deliberation costs. Presumably this larger decision problem itself requires costly deliberation, which ought to be economized, etc. We suggest here that one way of resolving this paradox is to have DM simply act on a best guess without any deliberation when the problem becomes overwhelming.

In contrast, when the number of future decision problems (or states of nature) is more manageable then step-by-step thinking is generally optimal. For an even lower number of future problems complete planning is optimal (in particular, when there is only one state of nature, and therefore only one future problem to think about, then thinking ahead and complete planning is always optimal).

There is obviously a large and rapidly growing literature on bounded rationality and some of the ideas we have touched on have been explored by others. The literature on bounded rationality that is most closely related to ours is the one on decision procedures and costly deliberation (see Payne, Bettman, and Johnson, 1993, and Conlisk, 1996) and within this sub-literature our work is closest in spirit to Conlisk (1988). The main difference with Conlisk (1988) is that we formulate a different deliberation technology, with thinking modeled as thought experimentation and thinking costs taking the form of time thinking costs.

Another closely related paper is Gabaix and Laibson (2002). In this paper, a boundedly rational DM has to choose between a number of paths that provide different utility flows at a finite sequence of steps. Paths differ in transition probabilities from one step to another. DM can at some cost explore each step and evaluate the resulting utility flow before choosing one of these paths. Gabaix and Laibson assume that in doing so DM follows a simple heuristic that ignores the option value of exploring one step or another. In their model DM chooses which step to explore presuming that it would be the last time she will explore a step. DM stops exploring when the information value of all the unexplored steps is less than the exploration cost. Therefore there are two main differences between their approach and ours: 1) we introduce an exploration cost as the only source of bounded rationality while they also assume that DM follows a sub-optimal heuristic (in their model the actual choice may not converge to the fully rational solution when the deliberation cost goes to zero); 2) unlike

[2]A similar answer to this paradox is proposed by MacLeod (2004).

in our setup, Gabaix and Laibson do not consider the choice over thinking ahead or on the spot, as their DM has to make only one decision. Thus, in their model DM cannot explore a few steps first down a given path and possibly re-explore some other steps depending on what she has learned.

The literature on satisficing behavior and aspiration levels is also closely related. Indeed some research on satisficing behavior that builds on Simon (1955) has also attempted to tackle the question of the endogenous aspiration level but in a framework where costly deliberation by DM is not explicitly modeled (see Gilbert and Mosteller, 1966, Bruss, 2000, and Beckenkamp, 2004).

Finally, less related but still relevant, is the analysis of the design of organizations when boundedly rational economic agents can at some cost expand their knowledge (Garicano 2000). The optimal hirerachy has less knowledgeable agents at the bottom, dealing with simple and frequent tasks, and managers at the top to whom workers can refer when they encounter more difficult or less usual problems. Therefore, like in our set up, some agents (workers) choose not to think about complex and rare problems. In our model, this happens because those agents have the option to think about those later, if necessary.

The remainder of our paper is organized as follows. Section 2 presents a simplified version of our model and characterizes DM's optimal thinking and action plan in this setting. Section 3 derives a number of key comparative statics results for this simple model. Section 4 analyzes the general model with an arbitrary finite number of states. Section 5 concludes by summarizing our main findings and pointing to new directions of research. Finally two appendices contain the proofs of our main results.

2 A Simple Model of "Bandit" Rationality

The general dynamic decision problem we have in mind is a possibly infinite horizon (discrete time) problem involving an initial decision on which of n actions to take, as well as future decision problems that may be contingent on the initial action choice a_i and on a realized state of nature θ_j. The initial action may be taken at any time $t \geq 0$ and when an action is chosen at time t a state of nature is realized at some later time. When some action a_i has been chosen DM receives an immediate payoff $\omega(a_i)$. In addition, following the realization of the state of nature θ_j DM chooses another action a_{ijk} and obtains another payoff $\pi(a_i, a_{ijk}, \theta_j)$. Future payoffs are discounted and the discount factor is given by $\delta < 1$. Thus, the present discounted payoff when DM chooses action a_i in period t and action a_{ijk} in period $\tau \geq t$ is given by:

$$\delta^t \omega(a_i) + \delta^{t+\tau} \pi(a_i, a_{ijk}, \theta_j).$$

Although DM knows the true payoff $\omega(a_i)$ she does not know the true payoff $\pi(a_i, a_{ijk}, \theta_j)$. She starts out with a prior belief over those payoffs. Before taking any action, DM can learn more about the true payoff associated with that or any other action by engaging in *thought experimentation*. We model this thought experimentation in an exactly analogous way as in the multi-armed bandit literature. That is, in any given period t DM can 'think' about an action a_{ijk} in state θ_j and obtain a signal which is correlated with the true payoff. Upon obtaining this signal, DM revises her belief on the future payoff associated with the action a_{ijk}.

Thus, at $t = 0$ DM' s decision problem is to decide whether to pick an action a_i right away or whether to think ahead about one of the future decision problems. DM faces this same problem in subsequent periods, with possibly updated beliefs from earlier thought experimentation, as long as she has not picked an action a_i.

When she has chosen an action a_i some time elapses until a state of nature is realized. Upon realization of a state θ_j, DM' s decision problem is again to decide whether to pick an action a_{ijk} right away or whether to think about one of the actions she may take at that point. Should DM decide to think about the payoff associated with an action then she faces again this same decision problem in subsequent periods, with updated beliefs.

This general framework is clearly restrictive in some respects: we only allow for two rounds of action choice, the action sets are finite, the state-space is finite and learning through thought-experimentation can only be done for one action at a time. Yet, the framework is sufficiently general and versatile to be able to capture many dynamic decision problems boundedly rational DMs are likely to face in reality.

In this paper we specialize the framework described above to a problem of irreversible investment under uncertainty. In addition, we shall only allow DM to choose between two initial actions, which we label as *invest* and *don't invest*. Also, in the remainder of this section we will further specialize the model to the case where there are at most two states of nature and only two actions in each state. We describe this simpler model in greater detail below, while section 6 will consider the generalization to $N \geq 2$ states.

2.1 The model with two states of nature

In its simplest form our model has the following basic structure. It involves an initial investment decision with a set-up cost $I > 0$. If DM chooses to invest at date t, then at date $t+1$ the project ends up in one of two states: $\theta \in \{\theta_1, \theta_2\}$. We denote by μ the ex ante probability that state θ_1 occurs. When state θ_i is realized, investment returns are obtained only after DM has chosen one of two possible actions: a *risky* or a *safe* action. The return of the risky action, R, is unknown and may take two possible values, $R \in \{\underline{R}, \bar{R}\}$. The return of the safe action is known and equal to $S \in (\underline{R}, \bar{R})$, so that there is some prior uncertainty about which is the efficient action to take in state θ_i. We denote by ν_i the prior probability that the risky action is efficient in state θ_i. To begin with we shall assume that $\nu_i = \nu$ so that both states have the same payoff structure, but that payoffs are independently drawn across the two states.

As we have described above, DM may think about which decision to take in each state, which in our simple model is equivalent to experimenting (in DM's head) with the risky action. We formulate the simplest possible thought-experimentation problem and assume that when DM experiments with the risky action in any given period, there is a probability λ that she finds out the true payoff associated with the risky action, and a probability $(1-\lambda)$ that she learns nothing. In that case she must continue to think, or experiment, in subsequent periods until she gets lucky if she wants to find out the true payoff of the risky action[3].

A critical departure from the standard multi-armed bandit setup is that DM can choose to think through what should be done in some state θ_i before or after the realization of the state. She can think about what to do in one of these states or in both before acting. Of course, if she acts first, invests and only starts thinking after the realization of a state of nature, then there is no need to think through what to do in the state that is not realized.

Without much loss of generality we assume that at any date t DM can either think or act, but not do both at the same time. More precisely, we assume that each date t contains two subperiods: an early subperiod when DM has the choice between acting or thinking. If she thinks and her thinking is successful, she can in a later subperiod either pursue her

[3] Our model is related to Conlisk's approach to optimization costs. In Conlisk (1988) a decision maker learns about an optimal decision x^* by drawing a sequence of random variables $\tilde{x}$. This process is costly as it takes time to collect more draws, but the longer the person thinks, the better is the estimate of x^*. Our model can be viewed as a special case of Conlisk's, where the decision space is binary and where the draws are either fully informative or not informative at all.

thinking or make a decision on what action to take. Otherwise, time moves on to date $t+1$.

We shall make the following assumptions on the payoff structure:

Assumption: $A_1 : \nu\bar{R} + (1-\nu)\underline{R} > S$

$A_2 : \delta S - I > 0$

Assumption A_1 implies that the risky action is the best action for DM if she decides not to do any thinking. This is not a critical assumption for our main qualitative results. Assumption A_2, on the other hand, is more important. It ensures that the investment project has a positive net present value (NPV) when the safe action is always chosen. Under this assumption, any deliberation on future decisions, or knowledge of the payoff of the risky action, does not affect the initial decision whether to invest: there is always a way of generating a positive NPV, so that there is no ex-ante real option value to be maximized. As we explained earlier this assumption is imposed to highlight the difference of our simple model with the classical real options model.

2.2 Solving the Decision Problem

DM can decide to invest right away or to think first about what decision to take in future states of nature. If she were to invest right away then she would not be prepared to act immediately upon the realization of a state of nature θ_i, unless she decides to act only on the basis of her incomplete prior knowledge. But, if she prefers to first think through what to do in state θ_i she would delay the time when she would reap the returns from her investment. Thus, an important benefit of identifying the efficient decision in state θ_i ahead of time (before investing) is that DM will be able to act immediately following the realization of a state of nature and thus reduce the time gap between when she makes an investment outlay and when she reaps the returns from her investment. On the other hand, thinking before investing may not be that valuable and may unnecessarily delay investment. This is the basic deliberation tradeoff DM faces and that we now explore.

A lower bound for DM's payoff is the expected return obtained if DM acts immediately and does not think at all (we refer to this as the *no thinking* strategy). Under assumption A_1 DM then always selects the risky action and obtains an expected payoff:

$$V_{\emptyset} = -I + \delta\left[\nu\bar{R} + (1-\nu)\underline{R}\right].$$

Consider now the payoff DM could obtain by engaging in some thought experimentation. A first thinking strategy for DM is to invest right away at date 0, find out which state of nature prevails at date 1, and see whether some thinking is worthwhile once the state is realized.

Note that under our learning technology (where following each thought experiment, DM either learns the true payoff for sure or learns nothing) if it is optimal to undertake one experiment, then it is optimal to continue experimenting until the true payoff is found. Indeed, suppose that when state θ_i is realized DM prefers to experiment. Then, since the decision problem at that point is stationary, she will also prefer to continue experimenting should she gain no additional knowledge from previous rounds of experimentation. Our experimentation problem has been deliberately set up to obtain this particularly simple optimal stopping solution.

If, following successful experimentation DM learns that the true payoff of the risky action is $\underline{R}$ then she will optimally choose the safe action given that $\underline{R} < S$. If, on the other hand, she learns that the true payoff is $\bar{R}$ then she chooses the risky action. Thus, the expected present discounted payoff from thinking in state θ_i is given by

$$\underbrace{\lambda\left[\nu\bar{R}+(1-\nu)\,S\right]}_{\substack{\text{present discounted payoff}\\ \text{from learning the true}\\ \text{payoff in the first round}\\ \text{of experimentation}}} + \underbrace{(1-\lambda)\delta\lambda\left[\nu\bar{R}+(1-\nu)\,S\right]}_{\substack{\text{present discounted payoff}\\ \text{from learning the true}\\ \text{payoff in the second round}}}$$

$$+(1-\lambda)^2\delta^2\lambda\left[\nu\bar{R}+(1-\nu)\,S\right]+\sum_{t=3}^{\infty}(1-\lambda)^t\delta^t\lambda\left[\nu\bar{R}+(1-\nu)\,S\right].$$

Or, letting $\hat{\lambda}=\frac{\lambda}{1-(1-\lambda)\delta}$,the expected present discounted payoff from thinking in state θ_i can be written as:

$$\hat{\lambda}\left[\nu\bar{R}+(1-\nu)\,S\right].$$

Therefore once DM learns that the true state is θ_i she will prefer to think before acting if and only if:

$$\begin{aligned}\nu\bar{R}+(1-\nu)\underline{R} &\leq \hat{\lambda}\left[\nu\bar{R}+(1-\nu)\,S\right]\\ &\Leftrightarrow\\ \hat{\lambda} &\geq \hat{\lambda}_L\equiv\frac{\nu\bar{R}+(1-\nu)\underline{R}}{\nu\bar{R}+(1-\nu)S}.\end{aligned}$$

This condition essentially imposes a lower bound on DM' s thinking ability for her to be willing to engage in some thought experimentation. If DM were a very slow thinker (λ is very small) then it obviously makes no sense to waste a huge amount of time thinking. Thus, for sufficiently high values of λ, DM will choose to *think on the spot* if she has not done any thinking prior to the realization of the state of nature. In that case, when DM chooses to first invests and then think on the spot she gets an ex ante payoff V_L equal to:

$$V_L = -I + \delta\hat{\lambda}\left(\nu\bar{R} + (1-\nu)\,S\right)$$

Notice that under this strategy DM has to solve only one decision problem: the one she faces once the state of nature is realized.

To compare the payoffs of the two deliberation strategies considered so far it is convenient to introduce the following notation: Let

$$\rho^* \equiv \nu\bar{R} + (1-\nu)S$$

and

$$\rho \equiv \nu\bar{R} + (1-\nu)\underline{R}.$$

It is then immediate to see that the strategy of *no thinking* dominates the strategy of *thinking on the spot* if and only if:

$$V_L \geq V_\emptyset \Leftrightarrow \hat{\lambda} \geq \hat{\lambda}_L = \frac{\rho}{\rho^*}.$$

Consider next a third strategy available to DM, which is to *think ahead* about one or both states of nature. Suppose to begin with that $\mu = \frac{1}{2}$ and that DM, being indifferent between which state to think about first, starts with state θ_1. Again, if it is optimal to begin thinking about state θ_1 and DM does not gain any new knowledge from the first thought-experiment then it is optimal to continue thinking until the true payoff of the risky action in state θ_1 is found.

Suppose that DM has learned the true payoffs in state θ_1, under what conditions should she continue thinking about the other state before investing instead of investing right away and gambling on the realization of state θ_1? If she decides to think about state θ_2 she will again continue to think until she has found the true payoffs in that state. If she learns that the return on the risky action in state θ_2 is $\bar{R}$, her continuation payoff is

$$V_r = -I + \delta\left[\frac{\pi_1}{2} + \frac{\bar{R}}{2}\right],$$

where $\pi_1 \in \{S, \bar{R}\}$ is DM' s payoff in state θ_1. Similarly, if she finds that the best action in state θ_2 is the safe action, her continuation payoff is:

$$V_s = -I + \delta\left[\frac{\pi_1}{2} + \frac{S}{2}\right].$$

Therefore, DM' s expected continuation payoff from thinking ahead about state θ_2, given that she has already thought through her decision in state θ_1 is:

$$V_E^1 = \hat{\lambda}\left(-I + \delta\left(\frac{\pi_1}{2} + \frac{\rho^*}{2}\right)\right).$$

If instead of thinking ahead about state θ_2 DM immediately invests once she learns the true payoff in state θ_1 her continuation payoff is:

$$V_L^1 = -I + \delta\left(\frac{\pi_1}{2} + \frac{\max\{\rho, \hat{\lambda}\rho^*\}}{2}\right).$$

Thus, continuing to think about state θ_2 before investing, rather than investing right away is optimal if:

$$\Delta^1 \equiv V_E^1 - V_L^1 = -(1-\hat{\lambda})(\delta\frac{\pi_1}{2} - I) + \frac{\delta}{2}\left[\hat{\lambda}\rho^* - \max\{\rho, \hat{\lambda}\rho^*\}\right] \geq 0.$$

From this equation it is easy to characterize the solution of DM's continuation decision problem once she knows her true payoff in state θ_1. We state it in the following lemma:

Lemma 1: *Suppose that DM is thinking ahead and has already solved her decision problem in state θ_1, then it is better to invest right away and possibly think on the spot in state θ_2 rather than continue thinking ahead about state θ_2 if $\delta\frac{\pi_1}{2} - I \geq 0$.*
If, on the other hand, $\delta\frac{\pi_1}{2} - I < 0$, then there exists a cut-off $\hat{\lambda}_E^1$ such that thinking ahead about state θ_2 is preferred if and only if $\hat{\lambda} \geq \hat{\lambda}_E^1$.

Proof. see the appendix. ■

As one might expect, the decision to continue thinking ahead about state θ_2 depends on what DM has learned before. The higher is π_1 the less keen DM is to continue thinking. When DM finds a good outcome in state θ_1 she wants to reap the rewards from her discovery by accelerating investment. By thinking further about state θ_2 she delays investment and if she ends up in state θ_1 anyway her thinking will be wasted. The opportunity cost of these delays is captured by $\left(\delta\frac{\pi_1}{2} - I\right)$, the expected payoff in state θ_1. Note, in particular, that a thinking strategy such that DM stops thinking on a bad outcome where she learns that $\pi_1 = S$, but continues thinking on a good outcome, where $\pi_1 = \overline{R}$, is necessarily sub-optimal. This simple observation, we believe, highlights a basic mechanism behind *satisficing behavior*. Why do boundedly rational DMs settle with good but not necessarily optimal outcomes? Because they want to bring forward the time when they get the good reward, or as the saying goes, because *the best is the enemy of the good*.

Having characterized this key intermediate step we are now in a position to determine when DM should start to think ahead at all and when DM should defer until later all of her thinking. Depending on the value of π_1 several cases have to be considered.

First, suppose that $I \leq \delta\frac{S}{2}$. In that case, DM will *not* think about state θ_2 ahead of time, irrespective of the outcome of her optimization in state θ_1. DM will then think about at most one state. If she thinks ahead about one state her payoff is:

$$V_E = \hat{\lambda}\left[\nu\left(-I + \frac{\delta}{2}\overline{R} + \frac{\delta}{2}\max\{\rho, \hat{\lambda}\rho^*\}\right) + (1-\nu)\left(-I + \frac{\delta}{2}S + \frac{\delta}{2}\max\{\rho, \hat{\lambda}\rho^*\}\right)\right] =$$

$$\hat{\lambda}\left[-I + \frac{\delta}{2}\rho^* + \frac{\delta}{2}\max\{\rho, \hat{\lambda}\rho^*\}\right]$$

If she decides to invest without thinking her payoff is:

$$V_L = -I + \delta\max\{\rho, \hat{\lambda}\rho^*\}$$

Comparing the two payoffs V_E and V_L we immediately observe that:

Lemma 2: *When $I \leq \delta\frac{S}{2}$, thinking ahead is always dominated and DM chooses:*
- no thinking if and only if $\hat{\lambda} \leq \hat{\lambda}_L$, or
- thinking on the spot if and only if $\hat{\lambda} \geq \hat{\lambda}_L$.

Proof. see the appendix. ■

Lemma 2 establishes another important observation. If DM knows that she will stop thinking ahead irrespective of the outcome of her current thinking, then it is not worth thinking ahead about her current problem. In other words, it is only worth thinking ahead about some state θ_i if DM may want to continue thinking ahead about other states with positive probability; in particular, when the outcome of her current thinking is bad. In our simple model it is quite intuitive that DM would not want to do any thinking ahead if she can guarantee a high net return, which is the case when $\delta\frac{S}{2} \geq I$.

Second, suppose that $I \geq \delta\frac{\overline{R}}{2}$ and that $\hat{\lambda} \geq \hat{\lambda}_L$. In that case, DM wants to think about state θ_2 ahead of time, irrespective of the outcome of her optimization in state θ_1. In other words, if DM does any thinking ahead, she will want to work out a complete plan of action before investing. Thus, if she thinks ahead her payoff is:

$$V_E = \hat{\lambda}^2\left[\delta\rho^* - I\right]$$

while if she does not, she can expect to get:

$$V_L = -I + \delta\hat{\lambda}\rho^*$$

It is easy to check that in this case, $V_L < V_E$.

Third, suppose that $I \geq \delta\frac{\overline{R}}{2}$ but $\hat{\lambda} \leq \hat{\lambda}_L$. In this case DM either thinks ahead and works out a complete plan of action before investing, or DM prefers not to do any thinking ever (thinking on the spot is always dominated by no thinking, since $\hat{\lambda} \leq \hat{\lambda}_L$). If she does not think at all she gets $V_\emptyset$ and if she thinks ahead her payoff is $V_E = \hat{\lambda}^2 [\delta\rho^* - I]$. Comparing $V_\emptyset$ and V_E, we find:

Lemma 3: *When $\hat{\lambda} \leq \hat{\lambda}_L$ thinking on the spot is always dominated and when $I \geq \delta\frac{\overline{R}}{2}$ DM chooses:*
- no thinking if $\hat{\lambda} \leq \hat{\lambda}_E = \sqrt{\frac{\delta\rho - I}{\delta\rho^* - I}}$, or
- thinking ahead (complete planning) if $\hat{\lambda} \geq \hat{\lambda}_E$.

Proof. see the appendix. ■

In this situation DM is a slow thinker but the opportunity cost of thinking ahead is also low as investment costs are high. Thinking on the spot is dominated because once investment costs are sunk DM wants to reap the returns from investment as quickly as possible. On the other hand, as long as investment costs have not been incurred, DM is less concerned about getting a low net expected return quickly.

Fourth, suppose that $\delta\frac{\bar{R}}{2} > I > \delta\frac{S}{2}$. In this intermediate case, DM wants to continue thinking ahead about state θ_2 only if the outcome of her thinking on state θ_1 is bad. Thus, if DM thinks ahead, then with probability ν she learns that the risky action has a payoff $\bar{R}$ in state θ_1 and stops thinking further about state θ_2. Later, of course, if state θ_2 is realized, DM may decide to think on the spot about what to do. With probability $1 - \nu$ instead, she learns that the return of the risky decision in state θ_1 is S and decides to continue thinking about state θ_2 before investing. Therefore, in this situation, her payoff if she thinks ahead is:

$$V_E = \hat{\lambda}\left[\nu\left(-I + \frac{\delta}{2}\overline{R} + \frac{\delta}{2}\max\{\rho, \hat{\lambda}\rho^*\}\right) + (1-\nu)\hat{\lambda}\left(-I + \frac{\delta}{2}S + \frac{\delta}{2}\rho^*\right)\right]$$

Having determined all the relevant payoffs we are in a position to derive the conditions under which DM prefers to think ahead. To limit the number of cases, and consistent with our focus on the choice between thinking ahead or deferring thinking, we narrow down our analysis to values of $\hat{\lambda}$ greater than $\hat{\lambda}_L$, so that no thinking is always dominated. We provide an analysis of the alternative situation in the Appendix.

Assumption: $\quad A_3 : \hat{\lambda} \geq \frac{\rho}{\rho^*}$

Proposition 1: *Under assumptions A_1, A_2, and A_3 the solution to DM's decision problem is as follows: DM prefers to think on the spot if and only if:*

$$I \leq \frac{\hat{\lambda}\delta\left[\rho^* - \nu\frac{\overline{R}}{2}\right]}{1 + \hat{\lambda} - \nu\hat{\lambda}}.$$

Otherwise DM prefers to think ahead, either adopting:

- *a* step-by-step *strategy where she first thinks about state θ_1 and continues her thinking about state θ_2 if she finds that the payoff in state θ_1 is S and if*

$$\frac{\hat{\lambda}\delta\left[\rho^* - \nu\frac{\overline{R}}{2}\right]}{1 + \hat{\lambda} - \nu\hat{\lambda}} \leq I \leq \frac{\delta\overline{R}}{2}.$$

If she uncovers that the payoff on state θ_1 is $\overline{R}$, she stops thinking ahead beyond state θ_1 and if state θ_2 is realized resumes her thinking then.

- *a* complete planning *strategy where she thinks ahead about both states before investing if $\frac{\delta\overline{R}}{2} \leq I$.*

Proof. see the appendix. ■

The following figure is helpful to summarize Proposition 1.

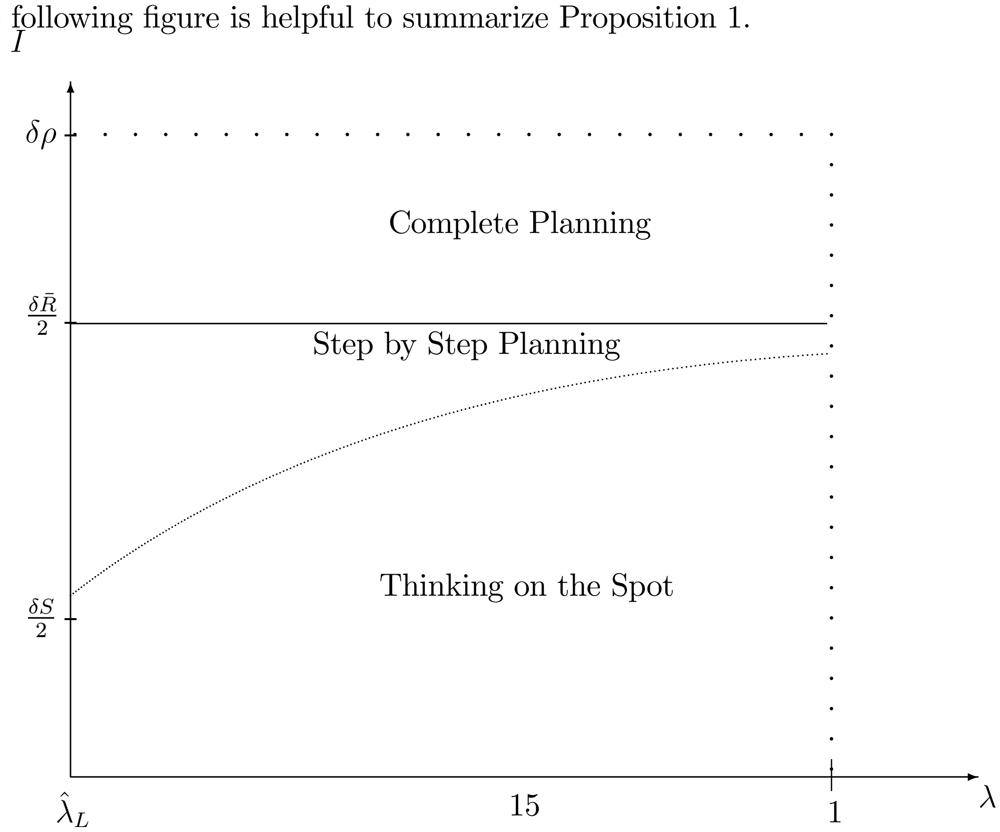

Figure 1

The figure maps out the three different regions: i) *complete planning*, where DM thinks everything through ahead of time; ii) *thinking on the spot*, where DM defers all the thinking to when the state of nature is realized, iii) *step-by-step planning*, where DM thinks first about state θ_1 before investing and maybe about state θ_2 as well. Recall that in this region, we could see DM thinking both before investing and after: for example, she thinks about what to do in state θ_1, learns some good news about that state and invests, but unfortunately state θ_2 is realized. She is then led to think again about what to do in that state before making a decision.

3 Comparative Statics in the simple model

Having characterized decision-making in the presence of positive deliberation costs in a simple symmetric example, we now explore how DM' s behavior changes with the underlying parameters of our problem, while continuing to retain our simple two-state structure.

3.1 Quick thinkers tend to 'think on their feet'

In our model quick thinkers have a higher λ. We show in this subsection that a DM with a higher λ is more likely to choose to think on the spot. To see this, consider a set of decision problems differing only in the size of I, with I distributed on the interval $(0, \bar{I}]$ according to some cumulative distribution function $F(.)$, with $\frac{\delta \overline{R}}{2} < \bar{I} < \delta S$. Then the following proposition holds:

Proposition 2: *The probability mass*

$$F(\frac{\hat{\lambda}\delta\left[\rho^* - \nu\frac{\overline{R}}{2}\right]}{1+\hat{\lambda}-\nu\hat{\lambda}})$$

is strictly increasing in λ. In other words, the set of problems for which DM is thinking on the spot is increasing in λ.

Proof. Immediate corollary of Proposition 1. ■

With high investment costs DM wants to plan ahead, whether she is a slow or fast thinker (as long as $\lambda \geq \hat{\lambda}_L$). As we have already explained, the reason is that this allows her to better align investment costs and monetary returns in time. When I is high the NPV of the investment is smaller and the opportunity cost of thinking ahead is low. In addition, the benefit of aligning cost and benefits goes up. Therefore both slow and fast thinkers then prefer to plan ahead.

In contrast, for intermediate values of investment costs, fast thinkers do more thinking on the spot than slow thinkers, who engage in step-by-step thinking. The reason is that, for fast thinkers the time gap between investment and the realization of returns is smaller and therefore matters less than for slow thinkers. As a result, they are more likely to prefer to bring forward the positive NPV by investing right away. Slow thinkers, on the other hand, are prepared to accept some delay in getting the positive NPV, in return for a higher NPV achieved by reducing the time lag between investment and recoupment.

Interestingly, as Figure 1 reveals, DM does not necessarily engage in complete planning as λ approaches 1. As long as $\lambda < 1$, what determines whether DM engages in complete planning is whether I is greater or smaller than $\frac{\delta \overline{R}}{2}$. Thus even for λ arbitrarily close to 1, DM will not engage in complete planning if $I < \frac{\delta \overline{R}}{2}$. It is only when there are no deliberation costs at all, so that $\lambda = 1$, that complete planning is always a rational behavior.

3.2 Think first about the most likely state

In the symmetric example we have studied DM faces maximum uncertainty about which state will occur, as each state is equally likely. We now explore how DM' s decisions change when μ, the probability that state θ_1 is realized is larger than $\frac{1}{2}$, while keeping the other characteristics of the model unchanged.

It is immediate to check that the payoff of the *thinking on the spot* strategy is unchanged when $\mu > \frac{1}{2}$. Similarly, the payoff associated with the *complete planning* strategy remains the same. On the other hand, the payoff associated with *step-by-step thinking* is affected in an important way when state θ_1 is more likely to arise.

When $\mu > \frac{1}{2}$ it is straightforward to see that the best step-by-step thinking strategy for DM is to think first about state θ_1, the most likely state. The costs of thinking ahead about one state only are the same whichever state DM thinks about, but the expected benefit in terms of DM's ability to act quickly following the realization of a state is higher for state θ_1, as this is the most likely state. Note also that DM is more likely to think ahead about one state only when $\mu > \frac{1}{2}$, as the marginal expected payoff of thinking ahead about the other state is reduced.

The continuation payoff obtained by thinking ahead about state θ_2, once DM knows that the payoff in state θ_1 is π_1 is given by:

$$V_E^1(\mu) = \hat{\lambda}\left(-I + \delta\left(\mu\pi_1 + (1-\mu)\rho^*\right)\right)$$

Compare this to the continuation payoff obtained by stopping to think ahead at that point and investing:

$$V_L^1 = -I + \delta\left(\mu\pi_1 + (1-\mu)\hat{\lambda}\rho^*\right),$$

and observe that the difference in continuation payoffs is given by:

$$\Delta^1 = V_L^1 - V_E^1(\mu) = (1-\hat{\lambda})(\mu\delta\pi_1 - I).$$

Thus, an increase in μ, the likelihood of state θ_1, results in an increase in Δ^1. As a result, there are now more values of I for which DM stops to think ahead once she knows the payoff in state θ_1. This has two implications:

First, DM is less likely to work out a complete action-plan before investing, so that the complete planning region in Figure 1 is now smaller.

Second, as the payoff of step-by-step thinking increases with μ, this strategy becomes more attractive relative to not thinking ahead at all.

Consequently we have:

Proposition 3: *A reduction in the uncertainty about states of nature reduces the attractiveness of* complete planning *and* thinking on the spot *and favors a* step-by-step *approach.*

Proof. Obvious. ■

As μ approaches 1, it is clear that thinking ahead about only state θ_1 is the best strategy: it allows DM to reduce the time between when she incurs investment costs and when she recoups her investment. In addition, thinking ahead is very unlikely to create unnecessary delays.

Vice-versa, as μ approaches $\frac{1}{2}$ and uncertainty about the state of nature increases, DM can respond by either working out a complete plan to deal with the greater uncertainty, or she can adopt a *wait-and-see* approach and defer all her thinking until after the uncertainty is resolved.

One may wonder whether this finding extends to changes in other forms of uncertainty. For instance, would DM change her behavior following a change in her prior belief ν_i in state θ_i that the risky action is efficient? It turns out that such a change in beliefs not only affects the perceived riskiness of the risky decision, but also changes the average payoff that DM can expect in state θ_i. We will show below that this change has several implications.

3.3 Solve easier problems first

So far we have interpreted the parameter λ as a measure of DM' s thinking ability. But λ can also be seen as a measure of the difficulty of the problem to be solved, with a higher λ denoting an easier problem. If we take that interpretation, we can let λ vary with the problem at hand and ask whether DM thinks first about harder or easier problems. Thus, suppose that the decision problem in state θ_1 is easier than the one in state θ_2. That is, suppose that $\lambda_1 > \lambda_2$. Holding everything else constant, we now explore how variations in the complexity of decision problems affects DM' s payoff and behavior. It is immediate that the payoff associated with the *complete planning* strategy is:

$$V_E = \hat{\lambda}_1 \hat{\lambda}_2 \left[\delta\rho^* - I\right]$$

and the payoff of *thinking on the spot* is:

$$V_L = -I + (\hat{\lambda}_2 + \hat{\lambda}_1)\frac{\delta\rho^*}{2}.$$

Comparing these two payoffs we can illustrate a first effect of variation in problem complexity. Take $\hat{\lambda}_1 = \hat{\lambda} + \varepsilon$ and $\hat{\lambda}_2 = \hat{\lambda} - \varepsilon$, where $\varepsilon > 0$[4]. Note, first, that the payoff under the *thinking on the spot* strategy is the same whether $\varepsilon = 0$, as we have assumed before, or $\varepsilon > 0$. On the other hand, the payoff under *complete planning*, $(\hat{\lambda}^2 - \varepsilon^2)\left[\delta\rho^* - I\right]$, is lower the higher is ε. Thus, increasing the variance in problem-complexity across states, while keeping average complexity constant, does not affect the payoff under the thinking-on-the-spot strategy as DM then only incurs average thinking costs. In contrast, when DM attempts to work out a complete action-plan, thinking costs compound. This is why DM is worse off under a complete planning strategy when the variance in problem-complexity increases.

How do differences in problem-complexity across states affect DM' s *step-by-step* thinking strategy? Intuitively, it seems to make sense to have a crack at the easier problem first. This intuition is, indeed, borne out in the formal analysis, but the reason why it makes sense to start first with the easier problem is somewhat subtle. Under the step-by-step thinking strategy, whenever DM ends up thinking about both states of nature, she compounds thinking costs in the same way as under complete planning. Under this scenario, whether she starts with state θ_1 or θ_2 is irrelevant.

[4]We assume that ε is small enough , so that even in state θ_2 DM will continue to think before acting if she does not know which decision is efficient. More precisely, we take ε to be small enough that assumption A_3 remains valid.

Hence, the choice of which state to think about first only matters in the event when: i) the outcome of thinking is good news (that is, if DM learns that the true payoff is $\bar{R}$), so that DM engages in only partial planning before investment, and; ii) when the state of nature which DM has thought about is realized. Under this latter scenario DM is better off thinking first about the easier problem since she then gets to realize the net return from investment sooner. Formally, the payoffs under the two alternative step-by-step strategies are given by:

$$V_{E1} = \hat{\lambda}_1 \left[\nu \left(-I + \frac{\delta}{2}\overline{R} + \frac{\delta}{2}\hat{\lambda}_2 \rho^* \right) + (1-\nu)\hat{\lambda}_2 \left(-I + \frac{\delta}{2}S + \frac{\delta}{2}\rho^* \right) \right]$$

$$V_{E2} = \hat{\lambda}_2 \left[\nu \left(-I + \frac{\delta}{2}\overline{R} + \frac{\delta}{2}\hat{\lambda}_1 \rho^* \right) + (1-\nu)\hat{\lambda}_1 \left(-I + \frac{\delta}{2}S + \frac{\delta}{2}\rho^* \right) \right]$$

Therefore, as $\left(-I + \frac{\delta}{2}\overline{R}\right) > 0$, it is best to think first about the simple problem with the higher $\hat{\lambda}_i$. We summarize our discussion in the proposition below:

Proposition 4: *Let $\hat{\lambda}_1 = \hat{\lambda} + \varepsilon$ and $\hat{\lambda}_2 = \hat{\lambda} - \varepsilon$, with $\varepsilon > 0$ but small.*

Then, when DM chooses to think ahead, it is (weakly) optimal to think first about the easier problem, and the payoff associated with:
- *the* complete planning *strategy is decreasing in the difference in problem-complexity across states*,
- *the* thinking on the spot *strategy is unchanged,*
- *the* step-by-step *strategy, where DM thinks first about the simple problem (in state θ_1), rises for ε small enough.*

Proof. See the appendix. ∎

Thus, an important insight emerging from our analysis is that partial planning is more likely to take place in environments where problem-complexity varies across states. Moreover, the simpler problems are solved first, while the harder ones are deferred to later in the hope that they won't have to be solved. These conclusions are quite intuitive and reassuring.

3.4 Think first about the highest payoff state

If thinking first about more likely and easier problems makes sense, is it also desirable to think first about problems with a higher expected payoff? In our simple model expected returns may differ across states if either the probability of high returns on the risky action is different, or if returns themselves differ. We explore each variation in turn and find:

Proposition 5: *Suppose that either* $\nu_1 = \nu + \varepsilon$, and $\nu_2 = \nu - \varepsilon$ or that $S_1 = S + \varepsilon$, $S_2 = S - \varepsilon$, while $\overline{R}_1 = \overline{R} + \varepsilon$ and $\overline{R}_2 = \overline{R} - \varepsilon$, *for* $\varepsilon > 0$ *but small.*

- *Whenever some thinking ahead takes place, it is best to start thinking about the high payoff state* (θ_1).
- *The payoff associated with the* complete planning *or the* thinking on the spot *strategies is unaffected by changes in* ε, *however*
- *The payoff associated with the* step-by-step *strategy, where DM thinks first about the high return state, is increasing in* ε.

Proof. See the appendix. ■

This is again a reassuring and quite intuitive result. As DM seeks to get to the point where she will make a decision as quickly as possible it makes sense for her to first bring forward in time the highest expected payoffs, which are most likely to tip her decision whether to invest or continue thinking ahead.

3.5 Think about the most urgent problems first

We have argued that Radner and Rothschild's (1975) notion of *putting out fires* can be understood more generally as the idea that a boundedly rational DM prioritizes her thinking by focusing first on the most important problems or those most likely to arise. In this subsection we want to highlight a third aspect of this general idea: thinking first about the most urgent problems.

We model a problem as more or less urgent by varying the time lag between when DM decides to invest and when DM can make the decision on which action to take in a given state of nature. That is, we still assume that the state of nature is realized one period after DM invests, but now we only allow DM to act in either state θ_i at the earliest after a time-lag of $\Delta_i \geq 0$ after the realization of the state. Specifically, we consider an example here where $\Delta_1 = 0$ and $\Delta_2 > 0$, so that the problem in state θ_1 is in some sense more "urgent" than the problem in state θ_2.

Given this timing structure, we can show that it is optimal for DM to think ahead about the decision problem in state θ_2 only if it is also optimal to think ahead about the problem in state θ_1:

Proposition 6: *Suppose that DM can act following the realization of state θ_i at the earliest after a time-lag of $\Delta_i \geq 0$. Then if $\Delta_1 < \Delta_2$ it is never optimal to think ahead about state θ_2 before state θ_1.*

Proof. See the appendix. ■

Intuitively, there are two effects which favor thinking about the most urgent problem first: 1) should DM end up not facing this problem after all, she will then have time on her hands to think about the less urgent problem; 2) as more urgent problems tend to arise sooner they have a higher present discounted payoff. As we saw above, it is always optimal for DM to first think about the highest payoff states.

Consider, finally, the effect of discounting on DM's planning strategy. As can be seen immediately from Figure 1, as DM discounts the future more (that is, when δ is lower) she will, somewhat counter-intuitively, be led to do more planning other things equal. Indeed, she may either switch from a step-by-step policy to complete planning, or from thinking on the spot to a step-by-step planning policy. The only caveat is that at some point, when δ is sufficiently low, it does not pay to do any thinking on the spot or even any thinking ahead.

3.6 Exploiting Statistical Regularities

Suppose that DM has prior knowledge that payoffs of the risky action may be correlated across states of nature. Specifically, suppose that with prior probability 1/2 the payoff on the risky action is $\overline{R}$ in any given state θ_i . In addition, suppose that the probability that $R_i = \overline{R}$ conditional on $R_j = \overline{R}$ is given by $\varepsilon > 1/2$, so that payoffs on the risky action are positively correlated. How does this prior statistical regularity affect DM' s thinking strategy?

Note first that the expected ex-ante payoffs of *thinking on the spot* or *complete planning* are independent of the value of ε. Similarly the continuation payoffs remain the same as long as: i) following the discovery that $R_i = \overline{R}$ we have

$$\varepsilon\overline{R} + (1-\varepsilon)\underline{R} < \hat{\lambda}[\varepsilon\overline{R} + (1-\varepsilon)S],$$

and,

ii) following the discovery that $R_i = \underline{R}$ we have

$$S < \hat{\lambda}[\varepsilon S + (1-\varepsilon)\overline{R}].$$

Indeed, when these conditions hold DM's interim problem is still whether to do more thinking ahead or to postpone thinking until after the realization of the state, and the costs and benefits involved in this decision are unaffected by ε.

On the other hand, for high degrees of correlation the opposite inequalities hold:

$$\varepsilon\overline{R} + (1-\varepsilon)\underline{R} \geq \hat{\lambda}[\varepsilon\overline{R} + (1-\varepsilon)S],$$

or

$$\varepsilon \geq \varepsilon_1 \equiv \frac{\hat{\lambda}S - \underline{R}}{\overline{R} - \underline{R} - \hat{\lambda}(\overline{R} - S)},$$

and

$$S \geq \hat{\lambda}[\varepsilon S + (1-\varepsilon)\overline{R}],$$

or

$$\varepsilon \geq \varepsilon_2 \equiv \frac{\hat{\lambda}\overline{R} - S}{\hat{\lambda}(\overline{R} - S)}.$$

In this case DM would not want to think on the spot about R_j following the discovery of R_i. Therefore, DM would also not want to do any thinking ahead about state θ_j . Thus, for high degrees of correlation in payoffs DM would never want to pursue a complete planning strategy. As for the choice between thinking on the spot and thinking ahead about one state, it is determined as follows:

Proposition 7: *For $\varepsilon \geq \max\{\varepsilon_1, \varepsilon_2, \varepsilon_3\}$ thinking ahead about one state only is optimal, and for $\varepsilon_3 > \varepsilon \geq \max\{\varepsilon_1, \varepsilon_2\}$ thinking on the spot is optimal, where*

$$\varepsilon_3 = 1 - \frac{4I(1-\hat{\lambda})}{\hat{\lambda}\delta(\overline{R} - \underline{R})}.$$

Proof. Follows immediately by comparing the expected payoffs of the two strategies. ■

Thus, quite generally, whenever DM is aware of existing statistical regularities she will engage in more step-by-step thinking.

Interestingly, therefore, with correlated payoffs DM is more easily satisfied with partial optimization. In other words, there is more *satisficing* when payoffs are correlated. Note that the mechanism here is through future *satisficing* behavior feeding back into current *satisficing*. However, the flip side of this behavior is that DM will end up making more mistakes. Thus, we have the paradoxical situation where *although DM engages in more planning she ends up making more mistakes.*

Notice also that DM is more likely here to be satisfied with step-by-step thinking the lower is λ. In other words, slower thinkers are more likely to cut corners and therefore to make mistakes.

In sum, greater correlation in payoffs brings about a quicker response time by DM, but this comes at the expense of more decision errors. DM' s behavior in this environment, thus, gives the appearance of overconfidence.

4 The Model with N states

One may wonder to what extent our insights into optimal dynamic decision and deliberation strategies derived in our tractable two state example extend to more general settings. We attempt a limited exploration into this question in this section by partially analyzing the more general decision problem with a finite, arbitrary number, of states of nature $N \geq 2$.

We begin by looking at the case of N equiprobable states, each with the same structure: a safe action with known payoff S, and a risky action with unknown payoff $R \in \{\underline{R}, \overline{R}\}$. As before, we assume that payoffs of the risky action are identically and independently distributed across states and take DM' s prior belief to be $\Pr(R = \bar{R}) = \nu$. We also continue to assume that assumption A_3 holds.

As there is a finite number of states to explore, the optimal deliberation policy cannot be stationary. As and when DM discovers the solution to future decision problems in some states of nature by thinking ahead, she is more likely to stop thinking further, given that the remaining number of unexplored states diminishes. We also know from our previous analysis that she may be more or less willing to continue her thought experimentation before investing, depending on whether she learns 'good' or 'bad' news about her payoff in states she is thinking about. Therefore there is no hope in identifying a simple optimal stopping policy where, for instance, DM explores m^* out of N states and then stops. It is also unlikely that DM' s optimal policy would take the simple form where DM would stop thinking further once she discovers that she can obtain a minimum average return in the states she has successfully explored. To identify DM' s optimal policy we therefore proceed in steps and characterize basic properties the optimal policy must satisfy.

One property we have identified in our two state example is that if DM is sure to stop thinking and to invest after learning her payoff about the current state she thinks about, then she prefers not to think ahead about the current state. The next lemma establishes that this property holds in the more general model:

Lemma 4: *If DM prefers to stop thinking ahead and to invest, irrespective of what she learns about state θ_i, then she also prefers not to think ahead about state θ_i.*

Proof. Suppose that DM has already deliberated about m out of N states, and found that z_m of these states have a payoff for the risky action of $\overline{R}$. If she deliberates about the $(m+1)^{th}$ state and knows that she will then invest no matter what, she can expect to get:

$$\widehat{\lambda}\left[-I+\frac{\delta}{N}\left(z_m\overline{R}+(m-z_m)\,S+\rho^*+(N-(m+1))\,\widehat{\lambda}\rho^*\right)\right]$$

Thus, suppose by contradiction that DM decides to deliberate on this $(m+1)^{th}$ state. This is in her best interest if the payoff above is larger than what she could get by investing right away :

$$-I+\frac{\delta}{N}\left(z_m\overline{R}+(m-z_m)\,S+(N-m)\,\widehat{\lambda}\rho^*\right)$$

Therefore, it must be the case that

$$C_1: -I+\frac{\delta}{N}\left(z_m\overline{R}+(m-z_m)\,S+(N-(m+1))\,\widehat{\lambda}\rho^*\right)<0$$

Now, if DM were sure to stop deliberating after the $(m+1)^{th}$ state, she must be better off stopping further deliberations even when she learns bad news about the $m+1^{th}$ state. For that to be true, it must be the case that stopping even following bad news is better than continuing exploring just one more state $(m+2)$, before investing, or that:

$$-I+\frac{\delta}{N}\left(z_m\overline{R}+(m-z_m)\,S+S+(N-(m+1))\,\widehat{\lambda}\rho^*\right)\geq$$

$$\widehat{\lambda}\left[-I+\frac{\delta}{N}\left(z_m\overline{R}+(m-z_m)\,S+S+\rho^*+(N-(m+2))\,\widehat{\lambda}\rho^*\right)\right]$$

or:

$$-I+\frac{\delta}{N}\left(z_m\overline{R}+(m-z_m)\,S+S+(N-(m+2))\,\widehat{\lambda}\rho^*\right)\geq 0$$

or

$$-I+\frac{\delta}{N}\left(z_m\overline{R}+(m-z_m)\,S+(N-(m+1))\,\widehat{\lambda}\rho^*\right)-\frac{\delta}{N}\left(\widehat{\lambda}\rho^*-S\right)\geq 0$$

which implies that condition C_1 is violated, , as $\widehat{\lambda}\rho^*-S>0$. ■

One implication of this simple lemma is that if DM wants to do some thinking ahead, then with positive probability she may also want to work out a complete plan of action (e.g. in the event that she only learns bad news from her deliberations).

Recall that the reason why it may be in DM' s interest to do some thinking ahead is not to improve her current decision. In this respect, thinking ahead adds no value, as investment is guaranteed to yield a positive net present value. Rather, the reason why DM gains by thinking ahead is that she may be able to respond faster to the realization of a new state of nature. So the intuition behind lemma 4 is that if the gain from a quick response in state $(m+1)$ does not justify the cost of delaying investment, and this even after learning that the expected net present value of the investment will be lower, then it cannot possibly be the case that this gain exceeds deliberation costs prior to learning the bad news.

This intuition immediately suggests that the following property of the optimal deliberation policy must be satisfied.

Lemma 5: *It is never optimal to stop thinking ahead on learning bad news and to continue thinking ahead on learning good news.*

Proof. see Appendix B. ■

From these two simple observations we are able to infer that:

Theorem: *For N small enough, DM adopts a* step-by-step *strategy whereby she thinks ahead about some states, continues to do so upon learning bad news and invests only once she has accumulated enough good news.*

Proof. The two previous lemmata imply that DM never stops exploring upon receiving bad news. Indeed, if she did then she would not continue either when receiving good news. This in turn would imply that it would be best not to think ahead about that particular state. Therefore, when following a step-by-step strategy DM only stops on good news about the last state she explores.

Turning to the second part of the theorem, to see why it is best to do some thinking ahead when N is sufficiently small, it suffices to note that thinking ahead is obviously beneficial if $N = 1$. Also, the strategy of investing right away delivers a lower payoff than first thinking ahead about exactly one state (which, itself is a dominated strategy), if:

$$\underbrace{-I + \delta\widehat{\lambda}\rho^*}_{\text{invest immediately}} \qquad \leq \underbrace{\widehat{\lambda}\left[-I + \frac{\delta}{N}\rho^* + \frac{\delta}{N}(N-1)\widehat{\lambda}\rho^*\right]}_{\text{think ahead about one state}}$$

or,

$$\Leftrightarrow N \leq \frac{\delta\widehat{\lambda}\rho^*}{\delta\widehat{\lambda}\rho^* - I}$$

■

As stated, the theorem establishes that DM would want to do some thinking ahead if the number of possible states N is small enough. We also know from the analysis of the example with two states that even for $N = 2$, DM may prefer thinking ahead step-by-step rather than determine a complete action plan. However, what we cannot conclude from the theorem or our previous analysis is that DM would definitely not want to do any thinking ahead when N is sufficiently large. We establish this result in the next proposition.

Proposition 8: *Consider N equiprobable states. If $N \geq \frac{\delta S}{\delta S - I}$ the optimal thinking strategy is to invest right away and to think on the spot.*

Proof. Notice first that if $N \geq \frac{\delta S}{\delta S - I}$, DM will never think about the last state. That is, when $N \geq \frac{\delta S}{\delta S - I}$ then $m^* < N$. To see this, denote by Q_{N-1} the average payoff that DM expects from the $N-1$ other states. We then have:

$$\widehat{\lambda}\left[-I + \frac{\delta}{N}\rho^* + \frac{\delta}{N}(N-1)Q_{N-1}\right] \leq -I + \frac{\delta}{N}\widehat{\lambda}\rho^* + \frac{\delta}{N}(N-1)Q_{N-1}$$

$$\Leftrightarrow N(\delta Q_{N-1} - I) \geq \delta Q_{N-1}$$

as $Q_{N-1} \in (S, \overline{R})$ this is equivalent to $N \geq \frac{\delta Q_{N-1}}{\delta Q_{N-1} - I}$, which is true if $N \geq \frac{\delta S}{\delta S - I}$. Now from lemma 4 we know that if DM stops thinking ahead at the penultimate state irrespective of what she learns then she also prefers not to think ahead about the penultimate state, irrespective of what she learned before. By backward induction, it then follows that DM prefers no to do any thinking ahead. ■

Proposition 8 is quite intuitive. There is little to be gained by working out a highly incomplete action-plan and if any reasonable plan is complex and will take a long time to work out then the only reasonable course of action is to just hope for the best and not do any planning at all.

Another extreme situation where the best course of action is just to think on the spot is when the state of nature is highly transitory. Thus, for example, in the extreme case where there is a new *iid* draw of payoffs on the risky action every period there is no point in thinking ahead, as the knowledge obtained by DM on a particular state will already be obsolete by the time the state is realized.

In reality, people sometimes prefer to be faced with complex life situations where rational forethought makes little difference and the only course of action is to essentially *put their fate in God' s hands* so to speak. In such situations they are absolved of all responsibility for their fate and that makes them better off. In contrast, here our DM is always (weakly) worse off facing more uncertainty than less (as measured by the number of states N). Indeed, if one were to reduce the number of states below $\frac{\delta S}{\delta S - I}$, DM would want to do some thinking ahead and be better off as a result.

Just as some key properties of the optimal thinking strategy with time-deliberation costs derived in the two-state special case extend to an arbitrary number N of equiprobable states, some important comparative statics results also extend to this general setting. We report below two general results we have been able to prove on the optimal order in which DM should think ahead about future states.

The first result establishes that quite generally it is optimal for DM to think ahead first about the most likely states:

Proposition 9: *Consider N identical states with probabilities of realization $\mu_1 > \mu_2 > ... > \mu_N$. The optimal thinking strategy is to think ahead about the most likely states first.*

Proof. see Appendix B. ■

The second result establishes that it is also optimal for DM to think ahead first about the highest payoff states:

Proposition 10: *Consider N equiprobable states ranked in order of decreasing expected payoff, $\overline{R}_i = \overline{R} + \varepsilon_i$, $S_i = S + \varepsilon_i$, with $\varepsilon_1 > \varepsilon_2 > ... > \varepsilon_N$. Assume, in addition, that the ε_i are small enough that assumption A_3 holds in all N states: for all i, $\hat{\lambda} \geq \frac{x+\varepsilon_i}{y+\varepsilon_i}$. The optimal thinking strategy is then to think ahead about the highest payoff states first.*

Proof. see Appendix B. ■

The other comparative statics results we have derived in the two-state special case may also generalize, but we have not yet formally proved these generalizations.

5 Conclusion

The notion of bounded rationality has been with us for a long time now and few economists would dispute that the model of rational decision-making in most microeconomics textbooks is a poor description of how agents actually make decisions in reality. It is especially poor at describing decision making in dynamic decision problems, as normal human beings are not able to solve complex dynamic programming problems.

However, despite the descriptive limitations of the rational model, it continues to be the reference model in economics to describe individual behavior. Sometimes, justifications are offered for sticking with the rational model, such as evolutionary selection, unbiased errors which average out in the aggregate, or convergence to the rational choice through simple learning algorithms in stable environments. As pertinent as these justifications may be, they do not always apply in reality and often they are simply invoked as convenient excuses for staying within a well-understood paradigm that is easy to manipulate. Indeed, a basic difficulty in moving beyond the rational choice framework is that there are several alternative approaches one could take. Moreover, while bounded rationality models may fare better as descriptive models of decision-making in reality, they are also significantly more complex and unwieldy. That is a basic reason why these models have not been incorporated more readily into the mainstream.

This is why our primary concern in formulating our approach has been tractability. Although we have probably oversimplified the dynamic decision problems agents are likely to face in reality we nevertheless capture a basic aspect of most people's behavior when they face complex intertemporal decision problems: they think ahead only about the most important or salient aspects of the problem in their mind and leave the determination of less important aspects to a later time. By modeling the basic dynamic decision problem as one of optimization in the presence of time-deliberation costs, we have been able to characterize simple optimal incomplete planning strategies that resemble satisficing behavior. Thus, for example, we have been able to show that it is optimal for a decision-maker to think ahead first about the most likely future decisions she will face.

We have also been able to highlight a basic tradeoff that is likely to be present in most dynamic decision problems: the benefit of thinking ahead is that a decision-maker will be able to react more quickly to new events or challenges, but the cost is that she delays her current decision while she thinks through the most important future implications.

We have simplified our framework so much that our analysis may be seen to be only applicable to situations where information affects the timing of investment but not the decision whether to invest. In particular, a natural question is whether our analysis extends to situations where the investment under consideration could have a negative NPV? We think that all our analysis extends straightforwardly to situations where DM is uncertain to begin with whether the investment opportunity she faces has a positive NPV or not. Indeed, if under DM's prior beliefs the project has a negative NPV then there is no opportunity cost in thinking ahead. DM then finds it optimal to think before investing, and she will continue thinking until she has learned all the relevant information or up to the point when her revised beliefs are such that the project has a positive NPV. Should her beliefs evolve in that direction then we are back to the situation we have analyzed so far.

Perhaps a more important limitation of our framework may be that we exclude situations where DM has other irreversible options available besides investing in the project under consideration. For instance she may face a decision of selecting among several alternative irreversible investment projects. Adding this possibility to our framework can indeed lead to changes in the optimal thinking strategies we have characterized. In particular, for this more complex problem it may be optimal for DM to stop thinking ahead on learning bad news under a step-by-step thinking strategy, an outcome we could not obtain in our simple setup. We leave the characterization of optimal thinking strategies in this more general problem for future research.

References

Beckenkamp, Martin (2004) “Is there an optimization in bounded rationality? The ratio of aspiration levels”, Max Planck Institute for Research on Collective Goods working paper #2004/15

Benabou, Roland and Marek Pycia (2002) “Dynamic Inconsistency and Self-Control: A Planner-Doer Interpretation”, *Economics Letters*, 77, 419-424

Bernheim, Douglas and Antonio Rangel (2002), “Addiction and Cue-Conditioned Cognitive Processes”, NBER working paper #9329

Berry, D. A. and D. Fristedt (1985), *Bandit Problems: Sequential Allocation of Experiments*, London, Chapman Hall.

Bolton, Patrick and Antoine Faure-Grimaud (2004), “Thinking Ahead: Part II, Satisficing Contracts”, *mimeo.*

Bruss, T. (2000), “Mathematische Unterhaltungen: Der Ungewissheit ein Schnippchen schlagen”, *Spektrum der Wissenschaft*, 6, 106

Conlisk, John. “Costly Optimizers Versus Cheap Imitators” *Journal of Economic Behavior and Organization*, 1980, 275-293.

Conlisk, John. “Optimization Cost” *Journal of Economic Behavior and Organization*, 1988, 213-228.

Conlisk, John. “Why bounded rationality?” *Journal of Economic Literature*, June 1996, 34 (2), pp.669-700.

Dixit, A.K. and Pindyck, (1994) *Investment under Uncertainty*, Princeton University Press

Gabaix, Xavier and David Laibson, “A boundedly rational decision algorithm” *AER Papers and Proceedings*, May 200.

Gabaix, Xavier and David Laibson, “Bounded rationality and Directed Cognition” *mimeo, Harvard University,* 2002.

Gabaix, Xavier, David Laibson and Guillermo Moloche, “The allocation of Attention: Theory and Evidence” *mimeo Harvard University,* August 2003.

Garicano, Luis (2000). “Hierarchies and the Organization of Knowledge in Production”, *Journal of Political Economy* 108(5), 874-904.

Gilbert, J. and F. Mosteller, (1966) “Recognizing the Maximum of a Sequence”, *Journal of the American Statistical Association* 61, 35-73

Gittins, J.C. and D. M. Jones (1974), “A Dynamic Allocation Index for the Sequential

Design of Experiments.", in *Progress in Statistics*, Gani, Sarkadi and Vince (eds.), New York, North Holland

Gittins, J.C. (1979) "Bandit processes and dynamic allocation indices", *Journal of the Royal Statistical Society* B, 41, 2, p.148-77.

Henry, C. (1974a) "Option Values in the Economics of Irreplaceable Assets.", *Review of Economic Studies*, 41, 89-104

Henry, C. (1974b) "Investment Decisions under Uncertainty: The Irreversibility Effect", *American Economic Review*, 64, 1006–1012.

Jehiel, Philippe. "Limited horizon forecast in repeated alternate games," *Journal of Economic Theory*, 1995, 67: 497-519.

Kahneman, Daniel and Tversky, Amos. "Judgment Under Uncertainty: Heuristics and Biases", *Science*, 1974.

Loewenstein, George and Ted O'Donoghue, "Animal Spirits: Affective and Deliberative Processes in Economic Behavior", (2004), Carnegie Mellon Discussion Paper

Macleod, W. Bentley. (2004) "Towards a Quantum Theory of Learning" *mimeo, Columbia.*

Metcalfe, Janet and Walter Mischel (1999), "A Hot/Cool-System Analysis of Delay of Gratification: Dynamics of Willpower", *Psychological Review*, 106, 3-19

Payne, John W., James R. Bettman, and Eric J. Johnson, *The Adaptive Decision Maker*, 1993, Cambridge: Cambridge University Press.

Radner, R. (1975) "Satisficing", *Journal of Mathematical Economics*, 2 (2), 253-62.

Radner, R. and M. Rothschild (1975), "On the Allocation of Effort", *Journal of Economic Theory*, 10 (3), 358-76.

Rothschild, M. (1974) "A Two-Armed Bandit Theory of Market Pricing," *Journal of Economic Theory*, 9, 185-202.

Rubinstein, Ariel. (1998), *Modeling Bounded Rationality.* MIT Press: Cambridge.

Simon, Herbert. "A behavioral model of rational choice," *Quarterly Journal of Economics*, 1955, 69 (1).

Whittle, P. (1980) "Multi-armed Bandits and the Gittins Index", *Journal of the Royal Statistical Society*, 42, 143-149

Whittle, P. (1982), *Optimization over time: Dynamic Programming and Stochastic Control*, John Wiley, New York

Winston, G. (1989) "Imperfectly Rational Choice: Rationality as the Result of a Costly Activity", *Journal of Economic Behavior and organization*, 12 (1), 67-86.

APPENDIX A: The 2-state model

Proof of Lemma 1:

It is immediate from the previous equation that thinking on the spot dominates if

$$\delta\frac{\pi_1}{2} - I \geq 0.$$

Suppose now that $\delta\frac{\pi_1}{2} - I < 0$.

- If $\hat{\lambda} \geq \hat{\lambda}_L$, then $\max\{\rho, \hat{\lambda}\rho^*\} = \hat{\lambda}\rho^*$ and $\Delta = (1-\hat{\lambda})(\delta\frac{\pi_1}{2} - I) < 0$ so that thinking ahead dominates.

- If $\hat{\lambda} \leq \hat{\lambda}_L$, then $\max\{\rho, \hat{\lambda}\rho^*\} = \rho$ and

$$\Delta = (1-\hat{\lambda})(\delta\frac{\pi_1}{2} - I) + \frac{\delta}{2}\left[\rho - \hat{\lambda}\rho^*\right]$$

therefore,

$$\Delta \leq 0 \Leftrightarrow$$

$$-I + \delta\frac{\pi_1}{2} + \frac{\delta}{2}\rho \quad \leq \quad \hat{\lambda}\left[-I + \delta\frac{\pi_1}{2} + \frac{\delta}{2}\rho^*\right]$$

$$\Leftrightarrow$$

$$\hat{\lambda} \geq \hat{\lambda}_E \equiv \frac{-I + \frac{\delta}{2}(\pi_1 + \rho)}{-I + \frac{\delta}{2}(\pi_1 + \rho^*)} \blacksquare$$

Proof of Lemma 2:

The difference $\Delta = V_E - V_L$ is always negative as:

$$\begin{aligned}\Delta &= \hat{\lambda}\left[-I + \frac{\delta}{2}\rho^* + \frac{\delta}{2}\max\{\rho, \hat{\lambda}\rho^*\}\right] - \left[-I + \delta\max\{\rho, \hat{\lambda}\rho^*\}\right] \\ &= -\left[-I + \delta\max\{\rho, \hat{\lambda}\rho^*\}\right](1-\hat{\lambda}) + \hat{\lambda}\frac{\delta}{2}\left[\rho^* - \max\{\rho, \hat{\lambda}\rho^*\}\right]\end{aligned}$$

Case 1: $\hat{\lambda} \leq \hat{\lambda}_L$

$$\begin{aligned}\Delta &= -\left[-I + \delta\rho\right](1-\hat{\lambda}) + \hat{\lambda}\frac{\delta}{2}\left[\rho^* - \rho\right] \\ &= I - \delta\rho + \hat{\lambda}\left[-I + \delta\rho + \frac{\delta}{2}(\rho^* - \rho)\right]\end{aligned}$$

as the term in bracket is positive, Δ is at most equal to:

$$I - \delta\rho + \frac{\rho}{\rho^*}\left[-I + \delta\rho + \frac{\delta}{2}(\rho^* - \rho)\right] =$$

$$\frac{1}{\rho^*}(I\rho^* - \delta\rho\rho^* - I\rho + \delta\rho^2 + \frac{\delta}{2}(\rho^* - \rho)\rho) =$$

$$\frac{\rho^* - \rho}{\rho^*}(I - \frac{\delta}{2}\rho) < 0$$

and therefore thinking on the spot dominates thinking ahead. As $\hat{\lambda} \leq \hat{\lambda}_L$, no thinking dominates thinking on the spot.

Case 2: $\hat{\lambda} \geq \hat{\lambda}_L$

$$\begin{aligned}\Delta &= -\left[-I+\delta\hat{\lambda}\rho^*\right](1-\hat{\lambda})+\hat{\lambda}\frac{\delta}{2}\left[\rho^*-\hat{\lambda}\rho^*\right] \\ &= (1-\hat{\lambda})\left[I-\frac{\delta}{2}\hat{\lambda}\rho^*\right]<0\end{aligned}$$

and therefore thinking on the spot dominates thinking ahead. As $\hat{\lambda} \geq \hat{\lambda}_L$, thinking on the spot also dominates no thinking.■

Proof of Lemma 3:

As in the proof of Lemma 2, there are two cases to consider:

Case 1: $\hat{\lambda} \leq \hat{\lambda}_L$

In that case, no thinking dominates thinking on the spot. Whether thinking ahead is best depends on the sign of

$$\Delta = \hat{\lambda}^2\left[\delta\rho^* - I\right] + I - \delta\rho$$

and this is positive for

$$I \geq \frac{\delta(\rho - \hat{\lambda}^2\rho^*)}{1-\hat{\lambda}^2}$$

The right hand side of this inequality is decreasing in $\hat{\lambda}$. The inequality cannot hold for $\hat{\lambda} = 0$, but it holds for $\hat{\lambda} = \frac{\rho}{\rho^*}$ as then this inequality becomes:

$$I \geq \frac{\delta(\rho - \frac{\rho^2}{\rho^{*2}}\rho^*)}{1-\frac{\rho^2}{\rho^{*2}}} = \frac{\delta\rho^*\rho}{\rho^*+\rho}$$

but as $I \geq \delta\frac{\overline{R}}{2}$, then $I > \frac{\delta\rho^*}{2}$ and we note that

$$\begin{aligned}\frac{\delta\rho^*}{2} &> \frac{\delta\rho^*\rho}{\rho^*+\rho} \\ &\Leftrightarrow \\ \rho^*+\rho &> 2\rho\end{aligned}$$

which is true since $\rho^* > \rho$. Therefore there exists a new threshold $\hat{\lambda}_E < \hat{\lambda}_L$ such that thinking ahead dominates no thinking if and only if $\hat{\lambda}$ exceeds that threshold.

Case 2: $\hat{\lambda} \geq \hat{\lambda}_L$

In that case, no thinking is dominated by thinking on the spot. Whether thinking ahead is best depends on the sign of

$$\Delta = \hat{\lambda}^2\left[\delta\rho^* - I\right] + I - \delta\hat{\lambda}\rho^*$$

and this is positive for

$$I \geq \frac{\delta\hat{\lambda}\rho^*}{1+\hat{\lambda}}$$

The right hand side of this inequality is increasing in $\hat{\lambda}$. It thus suffices to note that the inequality holds for $\hat{\lambda} = 1$. Indeed, we then have:

$$I > \frac{\delta \rho^*}{2}$$

which is true under our assumptions.■

Proof of Proposition 1:

First, lemmata 2 and 3 tell us that thinking ahead is dominated if $I \leq \frac{\delta S}{2}$ and that thinking on the spot is dominated if $I \geq \frac{\delta \overline{R}}{2}$.

Case 1: suppose first as we did in the text that $\hat{\lambda} \geq \hat{\lambda}_L$. Then surely for $I \geq \frac{\delta \overline{R}}{2}$, the best strategy is to think ahead as it dominates thinking on the spot, which itself dominates no thinking in that case. When I is so high DM thinks about both states (from lemma 1) before investing. Conversely for $I \leq \frac{\delta S}{2}$, deferring all thinking is best as it dominates both thinking ahead and no thinking.

We are left with the intermediate case where $\frac{\delta S}{2} \leq I \leq \frac{\delta \overline{R}}{2}$. In this case the best alternative to planning ahead is to follow the strategy of thinking on the spot. Indeed, we have:

$$\begin{aligned}
\Delta &\equiv V_E - V_L \\
&= \hat{\lambda}\left[\nu\left(-I + \frac{\delta}{2}\overline{R} + \frac{\delta}{2}\hat{\lambda}\rho^*\right) + (1-\nu)\hat{\lambda}\left(-I + \frac{\delta}{2}S + \frac{\delta}{2}\rho^*\right)\right] \\
&\quad - \left[-I + \delta\hat{\lambda}\rho^*\right] \\
&= \hat{\lambda}\left[\nu\left(-I + \frac{\delta}{2}\overline{R}\right) + \frac{\delta}{2}\hat{\lambda}\rho^* + (1-\nu)\hat{\lambda}\left(-I + \frac{\delta}{2}S\right)\right] \\
&\quad - \left[-I + \delta\hat{\lambda}\rho^*\right] \\
&= \hat{\lambda}\left[\hat{\lambda}\nu\left(-I + \frac{\delta}{2}\overline{R}\right) + (1-\hat{\lambda})\nu\left(-I + \frac{\delta}{2}\overline{R}\right) + \frac{\delta}{2}\hat{\lambda}\rho^* + (1-\nu)\hat{\lambda}\left(-I + \frac{\delta}{2}S\right)\right] \\
&\quad - \left[-I + \delta\hat{\lambda}\rho^*\right] \\
&= \hat{\lambda}\left[(1-\hat{\lambda})\nu\left(-I + \frac{\delta}{2}\overline{R}\right) + \hat{\lambda}\left(\delta\rho^* - I\right)\right] \\
&\quad - \left[-I + \delta\hat{\lambda}\rho^*\right] \\
&= \delta\rho^*\left(\hat{\lambda}^2 - \hat{\lambda}\right) - I\left(\hat{\lambda}^2 - 1\right) + \hat{\lambda}(1-\hat{\lambda})\nu\left(-I + \frac{\delta}{2}\overline{R}\right) \\
&= \left(\hat{\lambda} - 1\right)\left[\hat{\lambda}\delta\rho^* - I\left(1 + \hat{\lambda}\right) - \hat{\lambda}\nu\left(-I + \frac{\delta}{2}\overline{R}\right)\right],
\end{aligned}$$

and so we have $\Delta \geq 0$ if and only if:

$$\hat{\lambda}\delta\rho^* - I\left(1+\hat{\lambda}\right) - \hat{\lambda}\nu\left(-I + \frac{\delta}{2}\overline{R}\right) \leq 0,$$

or

$$I \geq \frac{\hat{\lambda}\delta\left[\rho^* - \nu\frac{\overline{R}}{2}\right]}{1+\hat{\lambda}-\nu\hat{\lambda}}.$$

Note that the right hand side is increasing and concave in $\hat{\lambda}$. Moreover, we have:

$$\frac{\delta S}{2} < \frac{\frac{\rho}{\rho^*}\delta\left[\rho^* - \nu\frac{\overline{R}}{2}\right]}{1+\frac{\rho}{\rho^*}(1-\nu)}$$

and,

$$\frac{\delta\overline{R}}{2} > \frac{\delta\left[\rho^* - \nu\frac{\overline{R}}{2}\right]}{2-\nu}$$

Case 2: suppose now that $\hat{\lambda} \leq \hat{\lambda}_L$. Then the best alternative to planning ahead is to follow the strategy of no thinking. From the previous lemmata, we can verify that for $I < \frac{\delta S}{2}$, the best strategy is to invest without any thinking taking place. For intermediate values, where $\frac{\delta S}{2} < I < \frac{\delta\overline{R}}{2}$, we have:

$$V_E = \hat{\lambda}\left[\nu\left(-I + \frac{\delta}{2}\overline{R} + \frac{\delta}{2}\rho\right) + (1-\nu)\hat{\lambda}\left(-I + \frac{\delta}{2}S + \frac{\delta}{2}\rho^*\right)\right]$$

and so

$$\begin{aligned}\Delta &= \hat{\lambda}\left[\nu\left(-I + \frac{\delta}{2}\overline{R} + \frac{\delta}{2}\rho\right) + (1-\nu)\hat{\lambda}\left(-I + \frac{\delta}{2}S + \frac{\delta}{2}\rho^*\right)\right] \\ &\quad - [-I + \delta\rho]\end{aligned}$$

Rearranging terms as before, we obtain that $\Delta \geq 0$ if and only if:

$$I \geq \frac{\hat{\lambda}\delta\left[\rho^* - \nu\frac{\overline{R}}{2}\right]}{1+\hat{\lambda}-\nu\hat{\lambda}} + \frac{\delta(\rho - \hat{\lambda}\rho^*)}{(1-\hat{\lambda})(1+\hat{\lambda}(1-\nu))} \tag{1}$$

which may or may not be feasible given that we need both $I \leq \frac{\delta}{2}\overline{R}$ and $\hat{\lambda} \leq \hat{\lambda}_L$ to hold. Inequality (1) in particular cannot be satisfied if $\overline{R} \leq \frac{2\rho^*\rho}{\rho^*+\rho}$, which is therefore a sufficient condition for no thinking to dominate thinking ahead. Otherwise thinking ahead may be best.

Similarly, when $\frac{\delta\overline{R}}{2} < I$, step-by-step thinking is dominated by complete planning. Whether DM starts investigating any state before investing depends on the sign of

$$\Delta = \hat{\lambda}^2\left[\delta\rho^* - I\right] - (\delta\rho - I)$$

and $\Delta > 0$ if:

$$\delta\rho^*\left[\hat{\lambda}^2 - \hat{\lambda}_L\right] + (1-\hat{\lambda}^2)I > 0$$

$$\Leftrightarrow I > \frac{\delta\rho^*}{(1-\hat{\lambda}^2)}(\hat{\lambda}_L - \hat{\lambda}^2)$$

which may or may not be true. Notice that it is more likely to be satisfied for $\hat{\lambda}$ close enough to $\hat{\lambda}_L$, as the right hand side of this last inequality is decreasing in $\hat{\lambda}$. A sufficient condition for thinking ahead to dominate in this parameter region is

$$\frac{\delta\overline{R}}{2} > \delta\rho$$

or $\overline{R} > 2\rho$.■

Proof of Proposition 4:

We first establish under what conditions the payoff of the step-by-step strategy is increasing in ε. Given that DM begins by exploring the easy state, she expects:

$$M_0 = (\hat{\lambda}+\varepsilon)\left[\nu\left(-I+\frac{\delta}{2}\overline{R}+\frac{\delta}{2}(\hat{\lambda}-\varepsilon)\rho^*\right)+(1-\nu)(\hat{\lambda}-\varepsilon)\left(-I+\frac{\delta}{2}S+\frac{\delta}{2}\rho^*\right)\right]$$

and so

$$\frac{\partial M_0}{\partial\varepsilon} = \nu\left(-I+\frac{\delta}{2}\overline{R}\right) - 2\varepsilon\left[\frac{\delta}{2}\rho^* + (1-\nu)\left(-I+\frac{\delta}{2}S\right)\right]$$

The value of this derivative is positive when ε goes to zero as $I \leq \frac{\delta}{2}\overline{R}$ in the region where step-by-step planning is optimal.■

Proof of Proposition 5:

Consider first the case where $\nu_1 = \nu+\varepsilon$, $\nu_2 = \nu-\varepsilon$, with $\varepsilon > 0$. In that case, the payoff of the *thinking on the spot*, or *complete planning* strategies are unchanged. In contrast, the payoff of the *step-by-step* strategy is affected by this average-belief-preserving spread as follows.

Whichever state θ_i DM thinks about first, DM will want to stop thinking further about the other state if she discovers that $\pi_i = \frac{\delta\overline{R}}{2}$, whenever $I \in [\frac{\delta S}{2}, \frac{\delta\overline{R}}{2}]$. It is then best for her to start thinking about state θ_1, the state with the higher prior belief ν_1. The reason is that she is then more likely to find that $\pi_1 = \frac{\delta\overline{R}}{2}$ and if state θ_1 arises, DM will be able to realize high returns relatively quickly. Note, therefore, that as more prior probability mass is shifted to the high return on the risky action in state θ_1, the step-by-step strategy also becomes relatively more attractive.

Second, suppose that returns themselves differ across the two states, and that returns in state θ_1 are higher than in state θ_2: $S_1 = S+\varepsilon$ while $S_2 = S-\varepsilon$, and $\overline{R}_1 = \overline{R}+\varepsilon$ while $\overline{R}_2 = \overline{R}-\varepsilon$. Again, we take ε to be small enough that assumption A_3 remains valid.

This redistribution of returns across states leaves DM' s expected payoff unaffected as long as $I < \frac{\delta(S-\varepsilon)}{2}$ or $I > \frac{\delta(\overline{R}+\varepsilon)}{2}$. Indeed, in that case she chooses to either defer all of her thinking until the uncertainty about the state is resolved, or to work out a complete plan before investing. In both cases, her expected payoff only depends on average returns $\frac{\rho_1^*+\rho_2^*}{2}$ and is therefore unaffected by changes in ε.

But if $\frac{\delta(S+\varepsilon)}{2} < I < \frac{\delta(\overline{R}-\varepsilon)}{2}$, DM will engage in step-by-step thinking and will stop thinking ahead if she learns that the risky decision is efficient in the state she thinks through first.

Once again, it is then best for her to think about the high payoff state first. The basic logic is the same as before: by thinking first about the high payoff state DM is able to bring forward in time the moment when she realizes the highest return.

Does this mean that when DM chooses to think ahead, she is always (weakly) better off thinking first about the high payoff state? The answer to this question is yes. There is one particular situation where conceivably this prescription might not hold. That is when $\frac{\delta(S-\varepsilon)}{2} < I < \frac{\delta(S+\varepsilon)}{2}$. In this case DM either thinks about the low return state first, or she defers all of her thinking to when the state of nature is realized. However, in the later case, her payoff is unaffected by changes in ε, while in the former it is decreasing in ε. What is more, for $\varepsilon = 0$ thinking on the spot dominates step-by-step thinking. Therefore, the latter strategy is then dominated by thinking on the spot for all $\varepsilon > 0$. Hence, it remains true that whenever thinking ahead pays, it is best to think first about the problem with the highest return.

Finally, for $\frac{\delta(\overline{R}-\varepsilon)}{2} < I < \frac{\delta(\overline{R}+\varepsilon)}{2}$, DM either thinks first about the high return state or works out a complete plan. In the latter case, her payoff is unaffected by ε. If she decides to think first about the high return state, her payoff goes up with ε so that eventually this strategy dominates complete planning. This completes the proof.

Proof of Proposition 6:

Without loss of generality set $\Delta_1 = 0$. Suppose by contradiction that DM thinks ahead about state θ_2 first under a step-by-step thinking approach. Note that for this form of step-by-step thinking to be optimal we must have

$$\frac{\delta^{(1+\Delta_2)}}{2}S < I < \frac{\delta^{(1+\Delta_2)}}{2}\overline{R}.$$

Her payoff under this strategy is

$$V_2 = \hat{\lambda}\left[\nu\left(-I + \frac{\delta^{(1+\Delta_2)}}{2}\overline{R} + \frac{\delta}{2}\hat{\lambda}\rho^*\right) + (1-\nu)\hat{\lambda}\left(-I + \frac{\delta^{(1+\Delta_2)}}{2}S + \frac{\delta}{2}\rho^*\right)\right]$$

On the other hand if DM thinks ahead first about state θ_1 her payoff is

$$\begin{aligned} V_1 &= \hat{\lambda}\left[\nu\left(-I + \frac{\delta}{2}\overline{R} + \frac{\delta^{(1+\Delta_2)}}{2}\left[(1-(1-\lambda)^{\Delta_2})\rho^* + (1-\lambda)^{\Delta_2}\hat{\lambda}\rho^*\right]\right)\right. \\ &\quad \left. +(1-\nu)\max\left\{\begin{array}{c} -I + \frac{\delta}{2}S + \frac{\delta^{(1+\Delta_2)}}{2}\left[(1-(1-\lambda)^{\Delta_2})\rho^* + (1-\lambda)^{\Delta_2}\hat{\lambda}\rho^*\right], \\ \hat{\lambda}(-I + \frac{\delta}{2}S + \frac{\delta^{(1+\Delta_2)}}{2}\rho^*) \end{array}\right\}\right]. \end{aligned}$$

Thus,

$$V_1 \geq \hat{\lambda}\left[\nu\left(-I + \frac{\delta}{2}\overline{R} + \frac{\delta^{(1+\Delta_2)}}{2}\hat{\lambda}\rho^*\right) + (1-\nu)\hat{\lambda}\left(-I + \frac{\delta}{2}S + \frac{\delta^{(1+\Delta_2)}}{2}\rho^*\right)\right]$$

and

$$\hat{\lambda}\left[\nu\left(-I + \frac{\delta}{2}\overline{R} + \frac{\delta^{(1+\Delta_2)}}{2}\hat{\lambda}\rho^*\right) + (1-\nu)\hat{\lambda}\left(-I + \frac{\delta}{2}S + \frac{\delta^{(1+\Delta_2)}}{2}\rho^*\right)\right] \geq V_2.$$

APPENDIX B: The N-states model.

The derivation of the results of section 4 is made easier by formalizing DM's problem as follows. Denote by:

- $\pi_i \in \{\overline{R}, S\}$, the maximum payoff that DM can achieve in state θ_i,
- h_k, a history of payoff observations of length k; that is, $h_k \in \{\overline{R}, S\}^k$ is a sequence of payoffs uncovered in states $\theta_1 ... \theta_k$. For instance, $h_2 = (S, S)$ means that the optimal decision in the first two states, θ_1 and θ_2, has been found to be the safe action with payoff of S in both states.
- $\sigma(h_k)$, the probability that DM decides to explore state $k+1$ when she has already explored k states and has learned history h_k. That is, with probability $(1 - \sigma(h_k))$ she explores no more than k states and invests right away. Given these definitions we have: $\sigma(h_k) = 0 \Rightarrow \sigma(h_{k+1}) = 0$.
- Θ_k, the set of all possible continuation-histories following $k : \Theta_k = \{\overline{R}, S\}^{N-k}$.

Using these notations, we can represent the expected payoff of DM who decides to explore a first state before investing as:

$$\begin{aligned} M_0 &= E_{\Theta_0}\Big[\hat{\lambda}^{(1+\sigma(h_1)+\sigma(h_1)\sigma(h_2)+...)} \times (-I + \delta[\mu_1\pi_1 + \sigma(h_1)\mu_2\pi_2 + \sigma(h_1)\sigma(h_2)\mu_3\pi_3 + ...] \\ &\quad +\delta\hat{\lambda}[\rho_2^*\mu_2\left(1-\sigma(h_1)\right) + \rho_3^*\mu_3\left(1-\sigma(h_2)\right) + ...]\Big)\Big] \end{aligned}$$

Or,

$$M_0 = E_{\Theta_0}\left[\hat{\lambda}^{(1+\varphi_1)}\left(-I + \delta\left[\mu_1\pi_1 + \sum_{\tau=1}^{N-1}\chi_{1\tau} + \hat{\lambda}\sum_{\tau=1}^{N-1}\mu_{\tau+1}\rho_{\tau+1}^*(1-\sigma(h_\tau))\right]\right)\right],$$

where:

1. $\varphi_1 = \left(\sum_{\tau=1}^{N-1}\prod_{t=1}^{\tau}\sigma(h_t)\right)$ is the expected number of subsequent states that DM explores ahead given that DM decides to explore a first state, and
2. $\chi_{1\tau} = \left(\mu_{\tau+1}\pi_{\tau+1}\prod_{t=1}^{\tau}\sigma(h_t)\right)$ is the payoff in state $\tau+1$ times the likelihood of state $\tau+1$, multiplied by the probability of exploring state τ ahead under the exploration-plan $\{\sigma(h_t)\}$, given that DM decides to explore a first state.

For illustration purposes, this expression in the case where $N = 2$ becomes:

$$M_0 = E_{\Theta_0}\left[\hat{\lambda}^{1+\sigma(h_1)}\left(-I + \delta\left(\mu_1\pi_1 + \sigma(h_1)\mu_2\pi_2\right) + \delta\hat{\lambda}\rho_2^*\mu_2\left(1 - \sigma(h_1)\right)\right)\right]$$

and in the equiprobable symmetric state case, with parameter values satisfying $\frac{\delta S}{2} < I < \frac{\delta \overline{R}}{2}$, we have $\sigma(S) = 1$ and $\sigma(\overline{R}) = 0$. Thus, we obtain:

$$M_0 = \nu\hat{\lambda}\left(-I + \frac{\delta\overline{R}}{2} + \frac{\delta\hat{\lambda}\rho^*}{2}\right) + (1-\nu)\hat{\lambda}^2\left(-I + \frac{\delta S}{2} + \frac{\delta\rho^*}{2}\right).$$

Similarly, we can write the expected continuation value of exploring one more state when already $m-1$ states have been explored as:

$$E_{\Theta_{m-1}}\left[\hat{\lambda}^{(1+\varphi_m)}\left(-I + \delta\left[\sum_{\tau=1}^{m-1}\mu_\tau(z_\tau\overline{R} + (1-z_\tau)S)\right] + \right.\right.$$
$$\left.\left.\delta\left[\mu_m\pi_m + \sum_{\tau=m}^{N-1}\chi_{m\tau} + \hat{\lambda}\sum_{\tau=m}^{N-1}\mu_{\tau+1}\rho_{\tau+1}^*(1-\sigma(h_\tau))\right]\right)\right]$$

where again:

1. $\varphi_m = \left(\sum_{\tau=m}^{N-1}\prod_{t=m}^{\tau}\sigma(h_t)\right)$ is the expected number of subsequent states that DM explores ahead, given that DM decides to explore at least $m-1$ states,
2. $\chi_{m\tau} = \left(\mu_{\tau+1}\pi_{\tau+1}\prod_{t=m}^{\tau}\sigma(h_t)\right)$ is the payoff in state $\tau+1$ times the likelihood of state $\tau+1$, multiplied by the probability of exploring state τ ahead under the exploration-plan $\{\sigma(h_t)\}$, given that DM decides to explore at least $m-1$ states.

   And,
3. z_τ is an indicator variable that takes the value 1 if and only if DM finds out that the return on the risky action in state θ_τ is $\overline{R}$.

We can now proceed to prove the following results.

Lemma 5: *It is never optimal to stop thinking ahead on learning bad news and to continue thinking ahead on learning good news.*

Proof of Lemma 5*:*
(We prove this lemma without assuming that states are equiprobable, nor that their average payoff is identical). Suppose, by contradiction, that there exists an exploration strategy and history of length $m-1$ such that: $\sigma(h_{m-2}, \overline{R}_{m-1}) = 1$ and $\sigma(h_{m-2}, S_{m-1}) = 0$. For this exploration strategy to be optimal it must be the case that:

a) there exists a continuation strategy profile following good news in state $m-1$, $(\tilde{\sigma}(h_{m-2}, \overline{R}_{m-1}, \pi_m), \tilde{\sigma}(h_{m-2}, \overline{R}_{m-1}, \pi_m, \pi_{m+1}), ...\tilde{\sigma}(h_{m-2}, \overline{R}_{m-1}, \pi_m, ...\pi_N))$, such that:

$$E_{\Theta_{m-1}}\left[\hat{\lambda}^{1+\tilde{\varphi}_m}\left(-I+\delta\left(\sum_{\tau=1}^{m-2}\mu_\tau(z_\tau\overline{R}_\tau+(1-z_\tau)S_\tau)\right)+\right.\right.$$
$$\left.\left.\delta\mu_{m-1}\overline{R}_{m-1}+\delta\left(\mu_m\pi_m+\sum_{\tau=m}^{N-1}\tilde{\chi}_{m\tau}+\hat{\lambda}\sum_{\tau=m}^{N-1}\mu_{\tau+1}\rho^*_{\tau+1}(1-\tilde{\sigma}(h_\tau))\right)\right)\right]$$
$$\geq -I+\delta\left(\sum_{\tau=1}^{m-2}\mu_\tau(z_\tau\overline{R}_\tau+(1-z_\tau)S_\tau)\right)+\delta\mu_{m-1}\overline{R}_{m-1}+\delta\hat{\lambda}\sum_{\tau=m}^{N-1}\mu_{\tau+1}\rho^*_{\tau+1}$$

b) and, for any continuation strategy profile following bad news in state $m-1$, $(\sigma(h_{m-2}, S_{m-1}, \pi_m), \sigma(h_{m-2}, S_{m-1}, \pi_m, \pi_{m+1}), ...\sigma(h_{m-2}, S_{m-1}, \pi_m, ...\pi_N))$, we have:

$$E_{\Theta_{m-1}}\left[\hat{\lambda}^{1+\varphi_m}\left(-I+\delta\left(\sum_{\tau=1}^{m-2}\mu_\tau(z_\tau\overline{R}_\tau+(1-z_\tau)S_\tau)\right)+\right.\right.$$
$$\left.\left.\delta\mu_{m-1}S_{m-1}+\delta\left(\mu_m\pi_m+\sum_{\tau=m}^{N-1}\chi_{m\tau}+\hat{\lambda}\sum_{\tau=m}^{N-1}\mu_{\tau+1}\rho^*_{\tau+1}(1-\sigma(h_\tau))\right)\right)\right]$$
$$< -I+\delta\left(\sum_{\tau=1}^{m-2}\mu_\tau(z_\tau\overline{R}_\tau+(1-z_\tau)S_\tau)\right)+\delta\mu_{m-1}S_{m-1}+\delta\hat{\lambda}\sum_{\tau=m}^{N-1}\mu_{\tau+1}\rho^*_{\tau+1}$$

We claim this yields a contradiction: consider instead the reverse strategy where upon learning that state θ_{m-1} yields S_{m-1}, DM decides to follow the continuation strategy profile that she would have followed if she had learned instead that state θ_{m-1} yields $\overline{R}_{m-1}$, $(\tilde{\sigma}(h_{m-2}, \overline{R}_{m-1}, \pi_m), \tilde{\sigma}(h_{m-2}, \overline{R}_{m-1}, \pi_m, \pi_{m+1}), ...\tilde{\sigma}(h_{m-2}, \overline{R}_{m-1}, \pi_m, ...\pi_N))$.

This gives her:

$$M \equiv E_{\Theta_{m-1}}\left[\hat{\lambda}^{1+\tilde{\varphi}_m}\left(-I+\delta\left(\sum_{\tau=1}^{m-2}\mu_\tau(z_\tau\overline{R}_\tau+(1-z_\tau)S_\tau)\right)+\right.\right.$$
$$\left.\left.\delta\mu_{m-1}S_{m-1}+\delta\left(\mu_m\pi_m+\sum_{\tau=m}^{N-1}\tilde{\chi}_{m\tau}+\hat{\lambda}\sum_{\tau=m}^{N-1}\mu_{\tau+1}\rho^*_{\tau+1}(1-\tilde{\sigma}(h_\tau))\right)\right)\right]$$

Adding $\hat{\lambda}^{1+\tilde{\varphi}_m}\delta\mu_{m-1}S_{m-1}$ and subtracting $\hat{\lambda}^{1+\tilde{\varphi}_m}\delta\mu_{m-1}\overline{R}_{m-1}$ on both sides of condition a) we observe that this is larger than:

$$-I+\delta\left(\sum_{\tau=1}^{m-2}\mu_\tau(z_\tau\overline{R}_\tau+(1-z_\tau)S_\tau)\right)+\delta\mu_{m-1}\overline{R}_{m-1}+$$
$$\delta\hat{\lambda}\sum_{\tau=m}^{N-1}\mu_{\tau+1}\rho^*_{\tau+1}+\delta\mu_{m-1}E_{\Theta_{m-1}}\left[\hat{\lambda}^{(1+\tilde{\varphi}_m)}\left(S_{m-1}-\overline{R}_{m-1}\right)\right]=$$

$$-I+\delta\left(\sum_{\tau=1}^{m-2}\mu_\tau(z_\tau\overline{R}_\tau+(1-z_\tau)S_\tau)\right)+\delta\mu_{m-1}S_{m-1}+\delta\hat{\lambda}\sum_{\tau=m}^{N-1}\mu_{\tau+1}\rho^*_{\tau+1}$$
$$+\delta\mu_{m-1}E_{\Theta_{m-1}}\left[1-\hat{\lambda}^{(1+\tilde{\varphi}_m)}\right]\left(\overline{R}_{m-1}-S_{m-1}\right)$$

we observe that the last term is positive, and so we conclude that M is larger than the right hand side of b): there exists a continuation strategy that is preferred to the strategy of investing upon learning bad news in state θ_{m-1} .■

Proofs of Propositions 8 and 9:

We establish these propositions by showing that the optimal policy is to think ahead about the state with the highest expected payoff first. To simplify notations we suppose that states are ranked by decreasing expected payoff, such that $\mu_1\rho_1^*\geq\mu_2\rho_2^*\geq....\geq\mu_N\rho_N^*$.

Suppose by contradiction that the optimal exploration plan $\{\sigma^*(h_t)\}$ is such that the order in which states are explored is not by decreasing expected payoff. Then there exists at least one pair of states (θ_j,θ_k), adjacent in the order in which they are explored, such that the first of these two states to be explored, say state θ_j, has a lower expected payoff: $\mu_j\rho_j^*<\mu_k\rho_k^*$.

We now show that DM will then be better off interchanging the order in which these two states are explored. That is, DM is better off with exploration plan $\{\sigma^*_{k\vee j}(h_t)\}$ than with plan $\{\sigma^*(h_t)\}$.

There are four different types of histories h_t under the optimal exploration plan $\{\sigma^*(h_t)\}$ to consider:

1. under h_t neither of the states (θ_j,θ_k) is explored ahead,
2. both states (θ_j,θ_k) are explored ahead,
3. exactly one of the states (θ_j) is explored ahead,
4. DM invests after exploring θ_j upon finding $\overline{R_j}$ and continues exploring θ_k upon finding $\underline{R}_j$.

Observe that in the first two cases the expected payoffs under respectively $\{\sigma^*_{k\vee j}(h_t)\}$ and $\{\sigma^*(h_t)\}$ are the same. In the third case, if DM inverts the order of θ_j and θ_k and otherwise leaves her exploration plan unchanged, she will be strictly better off. Indeed, under $\{\sigma^*(h_t)\}$ her expected payoff just before exploring state θ_j and having explored all states θ_i, $i\in E$, is:

$$\hat{\lambda}\left[-I+\delta\sum_{i\in E}\mu_i\pi_i+\delta\mu_j\rho_j^*+\delta\hat{\lambda}\mu_k\rho_k^*+\hat{\lambda}\sum_{i\in U}\delta\mu_i\rho_i^*\right],$$

where U denotes the subset of states that remain unexplored. Under $\{\sigma^*_{k\vee j}(h_t)\}$ her expected payoff just before exploring state θ_k and having explored all states θ_i, $i\in E$, is:

$$\hat{\lambda}\left[-I+\delta\sum_{i\in E}\mu_i\pi_i+\delta\mu_k\rho_k^*+\delta\hat{\lambda}\mu_j\rho_j^*+\hat{\lambda}\sum_{i\in U}\delta\mu_i\rho_i^*\right].$$

Thus, the incremental payoff obtained from inverting the order of these two states is

$$\delta\hat{\lambda}(1-\hat{\lambda})(\mu_k\rho_k^* - \mu_j\rho_j^*) > 0.$$

In the fourth case the strategy of proof is very similar. DM invests after exploring θ_j upon finding $\overline{R}$ and continues exploring θ_k upon finding $\underline{R}$. We show that if DM instead explored state θ_k first and stuck to exactly the same policy after that exploration, she will be made better off.

In a first step, suppose that when starting with state θ_j DM invests when finding $\overline{R}$ while she explores state θ_k only upon finding $\underline{R}$, and then stops exploring: exploration will either stop at state θ_j or will cover states θ_j and θ_k but no more. Starting with state θ_j provides DM with the payoff:

$$\begin{aligned} M_j &= \hat{\lambda}\nu_j\left[-I+\delta\sum_{i\in E}\mu_i\pi_i+\delta\mu_j\overline{R}_j+\delta\hat{\lambda}\mu_k\rho_k^*+\hat{\lambda}\sum_{i\in U}\delta\mu_i\rho_i^*\right]+ \\ &\quad \hat{\lambda}^2(1-\nu_j)\left[-I+\delta\sum_{i\in E}\mu_i\pi_i+\delta\mu_jS_j+\delta\mu_k\rho_k^*+\hat{\lambda}\sum_{i\in U}\delta\mu_i\rho_i^*\right]\end{aligned}$$

where U denotes the subset of states other than θ_j and θ_k that remain unexplored. Under $\{\sigma_{k\vee j}^*(h_t)\}$ where DM starts with state θ_k and at most explores another state θ_j she can expect to get:

$$\begin{aligned} M_k &= \hat{\lambda}\nu_k\left[-I+\delta\sum_{i\in E}\mu_i\pi_i+\delta\mu_k\overline{R}_k+\delta\hat{\lambda}\mu_j\rho_j^*+\hat{\lambda}\sum_{i\in U}\delta\mu_i\rho_i^*\right]+ \\ &\quad \hat{\lambda}^2(1-\nu_k)\left[-I+\delta\sum_{i\in E}\mu_i\pi_i+\delta\mu_kS_k+\delta\mu_j\rho_j^*+\hat{\lambda}\sum_{i\in U}\delta\mu_i\rho_i^*\right]\end{aligned}$$

Thus, when $\nu_k=\nu_j$ the incremental payoff obtained from inverting the order of these two states is

$$\delta\hat{\lambda}(1-\hat{\lambda}\delta)(\mu_k\rho_k^* - \mu_j\rho_j^*) > 0.$$

If now, $\nu_k>\nu_j$ but $\mu_i=\frac{1}{N},\overline{R}_i=\overline{R}$ and $S_i=S$ for all i, then the previous expressions become:

$$\begin{aligned} M_j = \hat{\lambda}\nu_j&\left[-I+\frac{\delta}{N}\sum_{i\in E}\pi_i+\frac{\delta}{N}\overline{R}+\frac{\delta}{N}\hat{\lambda}\rho_k^*+\hat{\lambda}\frac{\delta}{N}\sum_{i\in U}\rho_i^*\right]+ \\ \hat{\lambda}^2(1-\nu_j)&\left[-I+\frac{\delta}{N}\sum_{i\in E}\pi_i+\frac{\delta}{N}S+\frac{\delta}{N}\rho_k^*+\hat{\lambda}\frac{\delta}{N}\sum_{i\in U}\rho_i^*\right]\end{aligned}$$

and:

$$M_k = \hat{\lambda}\nu_k \left[-I + \frac{\delta}{N}\sum_{i\in E}\pi_i + \frac{\delta}{N}\overline{R} + \frac{\delta}{N}\hat{\lambda}\rho_j^* + \hat{\lambda}\frac{\delta}{N}\sum_{i\in U}\rho_i^* \right] +$$

$$\hat{\lambda}^2(1-\nu_k)\left[-I + \frac{\delta}{N}\sum_{i\in E}\pi_i + \frac{\delta}{N}S + \frac{\delta}{N}\rho_j^* + \hat{\lambda}\frac{\delta}{N}\sum_{i\in U}\rho_i^* \right]$$

we then have $M_k - M_j =$

$$\hat{\lambda}(1-\hat{\lambda})(\nu_k - \nu_j)\left[-I + \frac{\delta}{N}\sum_{i\in E}\pi_i + \hat{\lambda}\frac{\delta}{N}\sum_{i\in U}\rho_i^* \right]$$

$$+\hat{\lambda}^2\frac{\delta}{N}(\rho_j^* - \rho_k^*) + (v_k - v_j)\hat{\lambda}\frac{\delta}{N}\left[\overline{R} - \hat{\lambda}S\right] =$$

$$\hat{\lambda}(1-\hat{\lambda})(\nu_k - \nu_j)\left[-I + \frac{\delta}{N}\sum_{i\in E}\pi_i + \hat{\lambda}\frac{\delta}{N}\sum_{i\in U}\rho_i^* \right]$$

$$+(\nu_k - \nu_j)\hat{\lambda}\frac{\delta}{N}(1-\hat{\lambda})\overline{R} > 0$$

We have thus shown that if DM starts exploring state θ_j and explores at most one other state θ_k (upon finding that $R = \underline{R}$), she is better off inverting the order of θ_j and θ_k.

In a second step, consider now the possibility that DM explores θ_j, stops exploring further if she finds that the return in that state is $\overline{R}_j$, continues exploring if S_j, and then either stops exploring upon finding that state θ_k returns $\overline{R}_k$, or explores exactly one more state if she finds that state θ_k returns S_k. This strategy yields an expected payoff of:

$$\begin{aligned} M_j &= \hat{\lambda}\nu_j \left[-I + \delta\sum_{i\in E}\mu_i\pi_i + \delta\mu_j\overline{R}_j + \delta\hat{\lambda}\mu_k\rho_k^* + \hat{\lambda}\sum_{i\in U}\delta\mu_i\rho_i^* \right] + \\ & \hat{\lambda}^2(1-\nu_j)\nu_k \left[-I + \delta\sum_{i\in E}\mu_i\pi_i + \delta\mu_j S_j + \delta\mu_k\overline{R}_k + \hat{\lambda}\sum_{i\in U}\delta\mu_i\rho_i^* \right] + \\ & \hat{\lambda}^3(1-\nu_j)(1-\nu_k)\left[-I + \delta\sum_{i\in E}\mu_i\pi_i + \delta\mu_j S_j + \delta\mu_k S_k + \hat{\lambda}\sum_{i\in U}\delta\mu_i\rho_i^* \right] \end{aligned}$$

If now DM started instead with state θ_k and otherwise stuck with the same strategy (i.e. invests as soon as she gets one piece of good news and continues exploring for a maximum of 3 states otherwise), she would get:

$$\begin{aligned} M_k &= \hat{\lambda}\nu_k \left[-I + \delta \sum_{i \in E} \mu_i \pi_i + \delta \mu_k \overline{R}_k + \delta \hat{\lambda} \mu_j \rho_j^* + \hat{\lambda} \sum_{i \in U} \delta \mu_i \rho_i^* \right] + \\ & \hat{\lambda}^2 (1-\nu_k)\nu_j \left[-I + \delta \sum_{i \in E} \mu_i \pi_i + \delta \mu_k S_k + \delta \mu_j \overline{R}_j + \hat{\lambda} \sum_{i \in U} \delta \mu_i \rho_i^* \right] + \\ & \hat{\lambda}^3 (1-\nu_j)(1-\nu_k) \left[-I + \delta \sum_{i \in E} \mu_i \pi_i + \delta \mu_j S_j + \delta \mu_k S_k + \hat{\lambda} \sum_{i \in U} \delta \mu_i \rho_i^* \right] \end{aligned}$$

Observe that the difference in payoffs does not depend on what happens if DM goes on exploring more than states θ_k and θ_j. This therefore establishes that if $M_k \geq M_j$ for the strategies under consideration then this inequality also holds for any continuation strategy that involves exploring more than states θ_k and θ_j.

We have:

- when $\nu_j = \nu_k$, $M_j - M_k$ equals:

$$\hat{\lambda}\nu\delta \left[\mu_j \overline{R}_j + \hat{\lambda}\mu_k \rho_k^* - \delta \mu_k \overline{R}_k - \hat{\lambda} \mu_j \rho_j^* \right] +$$

$$\hat{\lambda}^2 (1-\nu)\nu\delta \left[\mu_j S_j + \mu_k \overline{R}_k - \mu_k S_k - \mu_j \overline{R}_j \right]$$

or rearranging,

$$\hat{\lambda}\nu\delta \left[\mu_j \overline{R}_j + \hat{\lambda}\mu_k \rho_k^* - \mu_k \overline{R}_k - \hat{\lambda} \mu_j \rho_j^* + \right.$$

$$\left. (1-\nu)\hat{\lambda} \left[\mu_j S_j + \mu_k \overline{R}_k - \mu_k S_k - \mu_j \overline{R}_j \right] \right]$$

$$= \hat{\lambda}\nu\delta \left[\mu_j \overline{R}_j + \hat{\lambda}\mu_k \nu \overline{R}_k - \mu_k \overline{R}_k - \hat{\lambda} \mu_j \nu \overline{R}_j + (1-\nu)\hat{\lambda} \left[\mu_k \overline{R}_k - \mu_j \overline{R}_j \right] \right]$$

$$= \hat{\lambda}\nu\delta (\mu_j \overline{R}_j - \mu_k \overline{R}_k)(1-\hat{\lambda}) \leq 0$$

if either $\mu_j < \mu_k$ or $\overline{R}_j < \overline{R}_k$.

- If now, $\nu_k > \nu_j$ but $\mu_i = \frac{1}{N}, \overline{R}_i = \overline{R}$ and $S_i = S$ for all i, then the previous expressions become:

$$\begin{aligned} M_j - M_k &= \hat{\lambda}(\nu_j - \nu_k) \left[-I + \frac{\delta}{N} \sum_{i \in E} \pi_i + \frac{\delta}{N}\overline{R} + \frac{\delta}{N}\hat{\lambda}\rho^* + \hat{\lambda} \sum_{i \in U} \frac{\delta}{N}\rho^* \right] + \\ & \hat{\lambda}^2 \left[(1-\nu_j)\nu_k - (1-\nu_k)\nu_j \right] \left[-I + \frac{\delta}{N} \sum_{i \in E} \pi_i + \frac{\delta}{N} S + \frac{\delta}{N}\overline{R} + \hat{\lambda} \sum_{i \in U} \frac{\delta}{N}\rho^* \right] \end{aligned}$$

Simplifying,

$$\begin{aligned} M_j - M_k &= \hat{\lambda}(v_j - v_k) \left[-I + \frac{\delta}{N} \sum_{i \in E} \pi_i + \frac{\delta}{N}\overline{R} + \frac{\delta}{N}\hat{\lambda}\rho^* + \hat{\lambda} \sum_{i \in U} \frac{\delta}{N}\rho^* \right] + \\ & \hat{\lambda}^2 \left[\nu_k - \nu_j \right] \left[-I + \frac{\delta}{N} \sum_{i \in E} \pi_i + \frac{\delta}{N} S + \frac{\delta}{N}\overline{R} + \hat{\lambda} \sum_{i \in U} \frac{\delta}{N}\rho^* \right] \end{aligned}$$

Or,

$$
\begin{aligned}
M_j - M_k &= \hat{\lambda}(\nu_j - \nu_k)\left[-I + \frac{\delta}{N}\sum_{i\in E}\pi_i + \frac{\delta}{N}\overline{R} + \frac{\delta}{N}\hat{\lambda}\rho^* + \hat{\lambda}\sum_{i\in U}\frac{\delta}{N}\rho^*\right] + \\
&\quad \hat{\lambda}^2\left[\nu_k - \nu_j\right]\left[-I + \frac{\delta}{N}\sum_{i\in E}\pi_i + \frac{\delta}{N}S + \frac{\delta}{N}\hat{\lambda}\rho^* - \frac{\delta}{N}\hat{\lambda}\rho^* + \frac{\delta}{N}\overline{R} + \hat{\lambda}\sum_{i\in U}\frac{\delta}{N}\rho^*\right]
\end{aligned}
$$

And,

$$
\begin{aligned}
M_j - M_k &= \hat{\lambda}(\nu_j - \nu_k)(1-\hat{\lambda})\left[-I + \frac{\delta}{N}\sum_{i\in E}\pi_i + \frac{\delta}{N}\overline{R} + \frac{\delta}{N}\hat{\lambda}\rho^* + \hat{\lambda}\sum_{i\in U}\frac{\delta}{N}\rho^*\right] + \\
&\quad \hat{\lambda}^2(\nu_k - \nu_j)\left[\frac{\delta}{N}S - \frac{\delta}{N}\hat{\lambda}\rho^*\right] \\
&\leq 0
\end{aligned}
$$

Therefore, DM is better off starting thinking ahead about state θ_k. This establishes that when DM's strategy involves investing when she finds one piece of good news and continuing exploring ahead otherwise, she is better off thinking ahead first about state θ_k. As we have also established that the same is true when DM's strategy involves investing no matter what she finds in exploring states θ_j or θ_k. The same is true if DM explores both states no matter what she finds. We can therefore conclude that DM is always better off exploring first state θ_k.■