Optimal Sequential Experimentation

Jorge D. Ramos-Mercado*

May 24, 2023

Abstract

I study sequential experimentation. An experimenter investigates an unknown state (of nature) by acquiring signals whose cost depend on their informativeness. Each signal has an imprecise (Gaussian) component that yields frequent information and a precise (Poisson) component yielding infrequent information. Optimally, experimenter only acquires precise information regarding a single state and may acquire some imprecise information to lower costs. This result is in contrast to Zhong (2022) in which experimentation costs depend on how much posterior beliefs are shifted by the experiment.

1 Introduction

Decision makers—e.g., policymakers, managers, etc.—rarely select among alternatives with known payoffs and ill-informed decisions are costly. For example, a manager hiring a criminal may be scolded, but he is fired if due diligence could have flagged the employee as a felon. Decision makers are thus incentivized to hire experimenters e.g., scientists, investigators, analysts, etc. Experimentation, however, is a time-consuming, stochastic process whose costs are mostly unrelated to how much uncertainty has been resolved. To

*Ramos-Mercado: University of Minnesota, 1925 Hanson Hall, Minneapolis, Minnesota 55455, ramos237@umn.edu.

this end, what is the way to sequentially experiment? In this paper, I characterize the optimal experiment when the state of the world is binary.[1]

Ever since Wald (1947), Blackwell (1951), and Blackwell and Girshick (1954), statisticians and economists have studied the sequential acquisition of information. This literature compares the relative informativeness of experiments and how long experimenters elicit information. Recently, however, Zhong (2022) considers the optimal signal acquisition strategy when an experimenter has access to a broad class of continuous-time signal processes. In particular, the paper allows the experimenter to elicit signals which allows beliefs to evolve given a time-dependent Jump-diffusion[2]. The paper in question provides an elegant and surprisingly simple characterization of the optimal experimentation process. Zhong finds that the optimal experimentation process is a pure jump-process with a time dependent arrival rate.

This result has a simple and intuitive interpretation. An experimenter can always pick precise signals that infrequently provides information (modeled as a Jump process) or imprecise signals that frequently generate information (Brownian motions with drift). Zhong finds that when the costs of running an experiment depend on how costly it is to shift the experimenter's beliefs, then costly, imprecise signals that only shift beliefs gradually are strictly suboptimal.

Costs associated with shifting the experimenter's beliefs come from the rational inattention literature (e.g., Sims 2003). Denti, Marinacci, and Rustichini (2023) find that the costs posited in the rational inattention literature are inconsistent with the costs of costly experimentation. For example, it is (almost surely) the case that the costs assumed from rational inattention cannot be interpreted as costs from running a laboratory, computational costs, or the effort exerted by policy analysts.

In this paper, I revisit the question of finding the optimal sequential experimentation strategy, but assume that the costs are associated with the informativeness of an explicit signal. In particular, a decision maker (DM) picks a continuous-time signal at times $t = 0, \mathrm{dt}$, for a fixed dt> 0, running from t to $t+$dt. The signal process converges to a jump-

[1] I make this assumption for tractability and ease of exposition. Most of the insights presented in this paper generalize when there is an arbitrary but finite set of signals.

[2] This assumption implies that the signal processes encompasses most Lévy processes; which only exclude pathological processes with few (if any) found applications to economics.

diffusion whose jumps and drift depend on the natural state of the world and the DM picks the drift, variance, and arrival rate of the possible jumps. My main result is that, intuitively, the experimenter seeks precise information confirming one state of nature; meanwhile, it also acquires some imprecise information. The experimenter seeks imprecise information, because such information can lower the overall cost of running an experiment.

The rest of the paper proceeds as follows. Section 2 presents the model and section 3 the results. Next, section 4 provides a numerical example for when it is optimal to only acquire Gaussian information. Section 5 discusses results and concludes. Proofs are delegated to the appendix A.

2 Model

I now present the model for a small, fixed time interval dt> 0. I study the limiting parametric choices as dt goes to 0. This approach ensures that the timing of play if clear. At time 0, nature draws a state of nature $\omega \in \{0, 1\}$ such that it draws $\omega = 1$ with an initial belief $\mu_0 \in (0, 1)$. At each time $t = 0, \mathrm{dt}, \ldots$, a Bayesian experimenter then observes the past experiments (to be defined below) and picks a new experiment or decides to stop. If the experimenter stops, the game ends. Otherwise, the experimenter picks a new experiment. An experiment is a collection of parameters $p_t \equiv (\sigma_t, (\lambda_{xt})_{x=0}^{1}) \in P \equiv \Re_+ \times [0, L]^2$, for constant $L \in (0, \infty]$, such that from time t to t+dt, it governs a signal process $(e_{t+s})_{s\in[0,\mathrm{dt}]}$ where e_t was the last signal observation observed at time t , $e_0 = 0$, and

$$\frac{de_{t+s}}{e_{t+s}} = \overbrace{\sigma_t[\omega \mathrm{dt} + dB_{t+s}]}^{\text{frequent and imprecise}} + \overbrace{dJ_{t+s}}^{\text{infrequent and precise}} \quad . \tag{1}$$

The process (B_{t+s}) is a standard Brownian motion and (J_{t+s}) is a jump process such that the processes are pairwise independent. The signal jumps from $e_{s^-} \equiv \lim_{\tau \nearrow s} e_\tau$ to $e^{-\eta_x} e_{s^-}$—for $\eta_x > 0, \eta_1 \neq \eta_0$—at an arrival rate of λ_{xt} if $x = \omega$ and 0 otherwise. For exposition, I assume that $0 < \eta_0 < \eta_1$. [3]

[3]One may assume the reverse order (i.e., $\eta_0 > \eta_1$ and will imply that the experimenter would try to verify that $\omega = 1$ rather than $\omega = 0$.

Next, I assume that the cost of an experiment is proportional to the informativeness of the signal. This means that given a belief $\mu_t \equiv Pr_t(\omega = 1)$, the cost of experiment e_t is

$$C(p_t) \equiv c(E_t[-d\ln(e_t)]/\text{dt})$$

for some convex, increasing, and smooth (i.e., twice continuously differentiable) function $c(\cdot) \geq 0$. The term in the inside of the cost function is the entropy of the experiment and the Ito's lemma's extension to a jump diffusion (diffusion from henceforth) implies that

$$\begin{aligned} E_t(d - \ln e) &= -E_t\left[\frac{de_t}{e_t} - \frac{1}{2}\frac{(de_t)^2}{e_t^2} - \mu_t\lambda_{1t}\text{dt}\eta_1 - (1-\mu_t)\lambda_{0t}\text{dt}\eta_0\right] \\ &= [\sigma_t(\sigma_t/2 - \mu_t) + \mu_t\lambda_{1t}\eta_1 + (1-\mu_t)\lambda_{0t}\eta_0]\text{dt}. \end{aligned}$$

Assuming that the cost of experimentation are proportional to the signal's entropy provides for analytic tractability. [4] Lastly, if the experimenter terminated the experiment at time $T = N\text{dt}$, for some $N = 0, 1, \ldots$, when his belief that $\omega = 1$ is $\mu_T \equiv Pr_T(\omega = 1)$ and after running experiments $p = (p_{n\text{dt}})_{n=0}^{N}$, his payoffs in at time $t \geq 0$ are

$$u_t(e, T) = E_t\left[e^{-r(T-t)}[1 - \kappa(\mu_T)] - \int_t^T re^{-r(s-t)}\text{dt}c(E_{m\text{dt}}[de_s]/\text{dt})\text{dt}\right] \tag{2}$$

such that $\kappa : [0, 1] \to \Re_+$ is a continuous differentiable, and single peaked function such that $\kappa(0) = \kappa(1) = 0$ and $r > 0$ is a discount rate. The function $\kappa(\cdot)$ models that the payoff to an experimenter crucially depends on what uncertainty remains.

I now justify several modeling choices of note. First, allowing σ_t to affect the variance *and* drift Brownian motion ensures that the experimenter is not able to set the variance in front of dB_t to 0 and the drift equal to (say) 1. Doing so would immediately reveal the true state of nature immediately and at arbitrarily small cost: thus, rendering the problem uninteresting. Secondly, this class of stochastic processes converges to a broad class of

[4] Otherwise, the optimal experimentation is (in general) proportional to the change in some cost function $dK(e_t)$. Such approach obfuscates the intuition why experimentation continues and how it is conducted. Alternatively, on may allow $C(p)$ to be a general cost function. Doing so, however, limits the predictions made by the model.

stochastic processes that are both broad and can be approximated with standard computational methods (e.g., Bruti-Liberati and Platen 2006).

3 Results

3.1 Auxiliary Results

I now present the key auxiliary result pinning down the paper's main insight: the optimal experiment elicits a positive amount of imprecise information (i.e., $\sigma_t \gg 0, \forall t \geq 0$). In particular, consider a smooth function of beliefs $F(\mu_t)$. How does one extend a notion of its derivative when the beliefs derived from a diffusion is also a diffusion? I define the right notion when the function $F(\cdot)$ is discounted by rate $r > 0$ as its generator[5]. Formally, at each time $t = 0, \mathrm{dt}, \ldots$, when the initial beliefs are $\mu_t = Pr_t(\omega = 1)$, the generator is defined as

$$\mathcal{L}(F(\mu_t), p_t) \equiv \lim_{\mathrm{s} \searrow 0} \frac{E_t[e^{-rs}[F(\mu_{t+s}) - F(\mu_t)|p_t]}{\mathrm{s}} = \lim_{\mathrm{s} \searrow 0} \frac{E_t[de^{-rs}F(\mu_{t+s})|p_t]}{\mathrm{s}}.$$

Deriving the generator is key for stating the value function which is solved by the optimal experimentation process. I now state the following result.

Lemma 3.1. *For each time $t = 0, dt, \ldots$, belief $\mu_t \in [0,1]$, parameters p_t, and a smooth function $F(\cdot)$ such that $F(0) = F(1) = 1$, it holds that*

$$\mathcal{L}(F(\mu_t), p_t) = \mu_t(1-\mu_t)[F''(\mu_t) - \sigma_t F'(\mu_t)]/2 + \bar{\lambda}_t(1 - F(\mu_t)) - rF(\mu_t) \quad (3)$$

where $\bar{\lambda}_t \equiv \mu_t \lambda_{1t} + (1-\mu_t)\lambda_{0t}$.

The proof of this theorem is tedious and uninteresting, so it is delegated to the appendix. Informally, it simply derives the evolution of beliefs given parameters chosen at time t and then derives how a function changes with changes in beliefs by using Ito's

[5] This is a sloppy use of notation as the generator (in general) is not required to refer to the discounted value of a function.

lemma. This result allows me to derive the value function below by simply imposing the principle of optimality.

3.2 Main Result

Given the auxiliary result stated above, I now state my main result. The main result is that the optimal experiment solves a rather standard value function. Define the experimenters maximum payoffs from either experimenting or terminating all experimentation when current beliefs are μ as $U(\mu)$. My main result is posing the optimal value function below.

Theorem 3.2. *At each belief $\mu \in [0,1]$, the optimal experiment solves*

$$rU(\mu) = \max\{1-\kappa(\mu), \max_{p=(\sigma,\lambda_0,\lambda_1)\geq 0} \mu(1-\mu)[U''(\mu) - \sigma U'(\mu)]/2 + \bar{\lambda}(1-U(\mu)) \\ - rc(\sigma(\sigma/2-\mu) + \lambda_1\eta_1\mu + \lambda_0\eta_0(1-\mu))\} \quad (4)$$

such that $\bar{\lambda} \equiv \lambda_1\mu + \lambda_0(1-\mu)$.

This value function implies that the previously defined generator plays a key role in the optimal value function. I conclude by this section by characterizing optimal choices. These choices can be derived via the standard, first order conditions

Corllary 3.3. *For each belief $\mu \in (0,1)$it is optimal to experiment, an optimal policy satisfies $\lambda_1(\mu) = 0, \lambda_0(\mu) \geq 0$. Moreover, if t $\lambda_0(\mu) > 0$ and $\tilde{C}(\mu) \equiv c'[\sigma(\mu)(\sigma(\mu)/2 - \mu) + \lambda_0(\mu)\eta_0]$, then it holds that*

$$1 - U(\mu) = \eta_0 r\tilde{C}(\mu).$$

If $\sigma(\mu) > 0$, then it further holds that

$$U'(\mu)\mu(1-\mu) + 2r(\sigma-\mu)\tilde{C}(\mu) = 0$$

These results simply follow from taking the first order condition. The fact that the experimenter at most tests one hypothesis follows from the observation that if both hypothesis are experimented at the same time, then either or $\tilde{C}(\mu) > 0$ and

$$r\tilde{C}(\mu)\eta_1 = r\tilde{C}(\mu)\eta_0 = 1 - U(\mu)$$

which would be a contradiction. Thus, the only possibility that this occurs is when $c'(0) = 0$ and $p = (0, 0, 0)$. Nevertheless, if the experiment finds it optimal to not elicit any kind of information, then posterior beliefs do not change and the experiment might as well terminate all experimentation. Next, for each policy $(\lambda_1, \lambda_0) > 0$, there policy $(0, \lambda_0 + \lambda_1[\mu/(1-\mu)])$ ensures that a jump arrives (in expectation) with the same rate and small costs.

Next, the first order conditions have informative implications. The experimenter first ensures that the marginal cost of eliciting a confirmatory signal that $\omega = 0$ equals to the expected gains from the signal arriving. A more interesting condition holds when $\sigma > 0$. The first order condition can be restated as

$$\overbrace{U'(\mu)\mu(1-\mu)}^{\Delta \text{ value of experimentation}} + \overbrace{2r(\sigma-\mu)\tilde{C}(\mu)}^{\Delta \text{ cost of experimentation}} = 0.$$

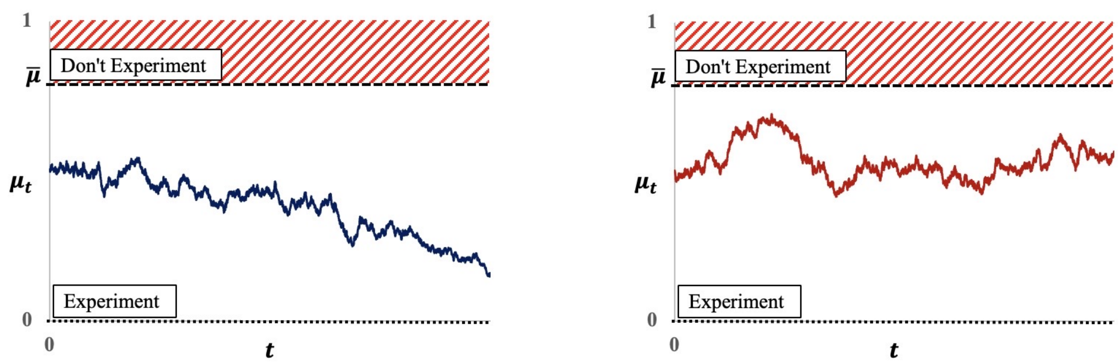
This implies that the variance can only be positive when $U'(\mu) < 0$ (i.e., the value of experimentation is falling) and the acquisition of imprecise information allows the experimenter to increase the value of experimentation and flow cost of experimentation. This second effect is a non-obvious role of acquiring imprecise information i.e., it may lower the data acquisition costs.

4 Stylized Example

In this section, I provide an example of when it is optimal for an experimenter to only acquire Gaussian signals. Suppose that $\kappa(x) = 4x(1-x)$, $\eta_0 = 1$, and that $c(x) = x$, then the value function $U(\cdot) \geq 0$ becomes

$$rU(\mu) = \max\{1 - 4\mu(1-\mu), \max_{p\geq 0} \mu(1-\mu)[U''(\mu) - \sigma U'(\mu)]/2 - \bar{\lambda}U(\mu) - \sigma(\sigma/2 - \mu)\}.$$

A simple first order condition clarifies that $\lambda_1 = \lambda_0 = 0$ is optimal. Solving for the value function, the optimal strategy characterized by a cutoff belief $\bar{\mu} \in (0,1)$. If $\mu > \bar{\mu}$, the experiment is terminated immediately. Otherwise, the experimenter picks $\sigma(\mu) = \mu[1 + \mu(1-\mu)]$. Below I simulate how beliefs evolve conditional on the true state of nature in figure 1.



(a) $\omega = 0$ *(b)* $\omega = 1$

Figure 1: ***Belief process simulation conditional on the true state*** ω**.**

5 Discussion

In this paper, I revisit the optimal theory of sequential experimentation and characterize the optimal, jump-diffusion signal process when the costs of experimentation explicitly depends on the informativeness of a signal. The optimal experiment seeks to only confirm one of the possible hypothesis with precise information that arrives infrequently. Meanwhile, the experimenter may find it optimal to acquire imprecise information that arrives frequently, because such experiments are informative and may mitigate costs.

There three parting remarks of note. First, if one allows for more general costs functions, then it is trivial that a general, jump-diffusion is optimal due to whichever cost function is specified. Such generality would (thus) not be interesting as it fails to restrict many processes. Secondly, it is possible to extend the setting herein by allowing for multiple states of nature. Doing so, however, increases the complexity of the problem without adding meaningful insights. Lastly, I make no claim that rational inattention models are ”wrong”, rather that its implications do not generalize.

References

[1] Blackwell, David. 1951. “Comparison of Experiments.” In Berkeley Symposium on Mathematical Statis- tics and Probability, edited by Jerzy Neyman, 93–102. Berkeley, CA: University of California Press.

[2] Blackwell, David, and M. Abraham Girshick. 1954.”Theory of Games and Statistical Decisions”. Hoboken, NJ: John Wiley & Sons.

[3] Bolton, Patrick, and Christopher Harris. 1999 “Strategic Experimentation,” Econometrica (ECTA) 67 (2), 349–374.

[4] Callander, Steven. 2011. “Searching and Learning by Trial and Error,” American Economic Review, 101 (6), 2277–2308.

[5] Caplin, Andrew, Mark Dean, and John Leahy. 2022. “Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy.” Journal of Political Economy 130 (6): 1676–1715.

[6] Christensen, Sören, Fabián Croce, Ernesto Mordecki, Paavo Salmien. *On Optimal stopping of Multidimensional Diffusions.* Journal of Stochastic Processes and their Applications, 2019 129(1): 2561-2581.

[7] Bloedel, Alexander W., and Weijie Zhong. 2021. “The Cost of Optimally Acquired Information.” Unpublished.

[8] Crandall, M., Ishii, H., and P.-L. Lions *User's Guide to Viscosity Solutions of Second Order Partial Differential Equations.* Bulletin of the American Mathematical Society 1992, 27: 1-67.

[9] Denti, Tommaso, Fabrizio Marinacci, and Aldo Rustichini. 2022. "Experimental Cost of Information". American Economic Review (2022) 12(9): 3106–3123.

[10] Denti, Tommaso. Forthcoming. "Posterior Separable Cost of Information." American Economic Review.

[11] Gollier, Christian. 2001. The Economics of Risk and Time. Cambridge, MA: MIT Press.

[12] JM Harrison. 2013. *Brownian Models of Performance and Control.* Cambridge University Press.

[13] Le Cam, L. 1996. "Comparison of Experiments: A Short Review." Lecture Notes–Monograph Series 30: 127–38.

[14] Macowiak, Bartosz, Filip Matejka, and Mirko Wiederholt. Forthcoming. "Rational Inattention: A Review." Journal of Economic Literature.

[15] Morris, Stephen, and Philipp Strack. 2019. "The Wald Problem and the Equivalence of Sequential Sampling and Static Information Costs." SSRN 2991567.

[16] Sims, Christopher A. 2003. "Implications of Rational Inattention." Journal of Monetary Economics 50(3): 665–90.

[17] Wald, A. 1947. "Sequential analysis". John Wiley.

[18] Zhong, Weijie. 2022 "Optimal Dynamic Information Acquisition". Econometrica 90(4):1537-1582.

A Proofs

A.1 Proof of Lemma 3.1.

Proof In this proof, I characterize the evolution of beliefs when the experimenter picks parameters $p_t = (\sigma_s, (\lambda_{\omega t})_{\omega=0}^{1})$ at each time $t = 0, \mathrm{dt}, \ldots$ for a fixed dt> 0. First, suppose that at time t, the experimenter expected $\omega = 1$ with probability μ_t and picks parameters p_t. I first characterize how beliefs evolve across each possible path that the experiment takes in the time interval $[t, t + \mathrm{dt}]$.

The first (and simplest) paths are when the experimenter observes a jump. Once, the experimenter observes a jump at some time $s \in [t, t + \mathrm{dt}]$ from e_{s^-} to $e_{s^-}e^{-\eta_0}$, the experimenter is certain that $\omega = 0$. Likewise, as soon as the experimenter observes a jump from e_{s^-} to $e_{s^-}e^{-\eta_1}$, he is certain that $\omega = 1$. The interesting case is when the process does not jump. There are no jumps from tome t to t+dt with a probability of $e^{-\lambda_{\omega t}\mathrm{dt}} = 1 - \lambda_{\omega t}\mathrm{dt} + o(\mathrm{dt})$ where $o(x)/x$ goes to 0 as $x \to 0$. In such case, beliefs as time $t + s$, for $s \in [0, \mathrm{dt}]$, only depend on the signal's realization at time t+s. This is because the processes considered satisfy the Strong Markov property. Given any change $x \equiv d\ln e_{t+s} \equiv \ln e_{t+s} - \ln e_t \in (-\infty, \infty)$ between time t and t+dt, the probability of observing such change is proportional to

$$Pr(d\ln e_t \equiv \ln e_{t+s} - \ln e_t = x|\omega, p_t) = \frac{e^{\frac{-(x-\sigma_t\omega s)^2}{2\sigma_t^2 s}}}{\sigma_t s\sqrt{2\pi}}. \tag{5}$$

By Bayes rule, the experimenters beliefs conditional prior μ_t, no jumps, and a change in the experiment signal of x is

$$\mu_{t+s} = \frac{\mu_t e^{-\frac{(x-\sigma_t s)^2}{2\sigma_t^2 s}-\lambda_{1t}s}}{\mu_t e^{-\frac{(x-\sigma_t s)^2}{2s\sigma_t^2}-\lambda_{1t}s} + (1-\mu_t)e^{-\frac{x^2}{2\sigma_t^2 s}-\lambda_{0t}s}} = \frac{\mu_t e^{\frac{x}{\sigma_t}+(\lambda_{0t}-\lambda_{1t}-1/2)s}}{\mu_t e^{\frac{x}{\sigma_t}+(\lambda_{0t}-\lambda_{1t}-1/2)s} + (1-\mu_t)}$$

The second expression expands the quadratic expression in the exponential, the groups all similar terms, and eliminates superfluous s terms.If one defines $y(e_s, s) \equiv e^{\frac{\ln e_s}{\sigma_t}+(\lambda_{0t}-\lambda_{1t}-1/2)s}$, then $y_e(e,s) = y(e,s)/(e\sigma_t)$ and $y_s(e,s) = y(e,s)(\lambda_{0t} - \lambda_{1t} - 1/2)$. Next, at each time $t+s$, for $s \in [0, \mathrm{dt}]$, current beliefs equal to $\mu_{t+s} = f[y(e_s,s)]$ such that $f(y) = \mu_t y/[\mu_t y + 1 - \mu_t]$ such that $f'(y) = f(y)(1-f(y))/y$. Consequently, it holds that

$$\frac{\mathrm{d}}{\mathrm{d}e_s}\mu_{t+s} = f'[y(e_s,s)]y_e(e,s) = (1-\mu_{t+s})\mu_{t+s}\frac{y(e_s,s)}{y(e_s,s)e_s\sigma_t} = \frac{(1-\mu_{t+s})\mu_{t+s}}{e_s\sigma_t},$$

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}e_s^2} = \frac{\mathrm{d}}{\mathrm{d}e_s}\left[\frac{(1-\mu_{t+s})\mu_{t+s}}{e_s\sigma_t}\right] &= \frac{(1-2\mu_{t+s})(1-\mu_{t+s})\mu_{t+s}}{2e_s^2\sigma_t^2} - \frac{\sigma_t(1-\mu_{t+s})\mu_{t+s}}{2e_s^2\sigma_t^2}\\ &= \frac{(1/2-\mu_{t+s}-\sigma_t/2)(1-\mu_{t+s})\mu_{t+s}}{e_s^2\sigma_t^2},\end{aligned}$$

and

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\mu_{t+s} = f'[y(e_s,s)]y_e(e,s) &= (1-\mu_{t+s})\mu_{t+s}\frac{y(e_s,s)(\lambda_{0t}-\lambda_{1t}-1/2)}{y(e_s,s)}\\ &= (1-\mu_{t+s})\mu_{t+s}(\lambda_{0t}-\lambda_{1t}-1/2).\end{aligned}$$

I now derive the belief dynamics as a a function of the signal process (e_s) chosen by the experimenter at time t. Since beliefs converge to 1 when the signal jumps by $e^{-\eta_1}$ and to 0 when it jumps by $e^{-\eta_0}$, then one can denote $(J_{\mu t+s})$ as the beliefs corresponding jump process. Applying the Jump-diffusion version of Ito's lemma, it holds that

$$\begin{aligned}d\mu_{t+s} &= \frac{\mathrm{d}}{\mathrm{d}s}\mu_{t+s}ds + \frac{\mathrm{d}}{\mathrm{d}e_s}\mu_{t+s}de_s + \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}e_s^2}(de_s)^2\\ &= (1-\mu_{t+s})\mu_{t+s}(\lambda_{0t}-\lambda_{1t}-1/2)ds + (1-\mu_{t+s})\mu_{t+s}\frac{de_s}{e_s\sigma_t}\\ &\quad + (1/2-\mu_{t+s}-\sigma_t/2)(1-\mu_{t+s})\mu_{t+s}\frac{(de_s)^2}{(\sigma_t e_s)^2}\end{aligned}$$

$$= (1-\mu_{t+s})\mu_{t+s}(\lambda_{0t}-\lambda_{1t}-1/2)ds + (1-\mu_{t+s})\mu_{t+s}[\mu_{t+s}ds + dB_{t+s}]$$
$$+(1/2-\mu_{t+s}-\sigma_t/2)(1-\mu_{t+s})\mu_{t+s}ds+\mu_{t+s}(1-\mu_{t+s})\lambda_{1t}ds-\mu_{t+s}(1-\mu_{t+s})\lambda_{0t}ds+dJ_{\mu t+s}$$
$$= (1-\mu_{t+s})\mu_{t+s}(\lambda_{0t}-\lambda_{1t}-1/2)ds + (1-\mu_{t+s})\mu_{t+s}[\mu_{t+s}ds + dB_{t+s}]$$
$$+ (1/2-\mu_{t+s}-\sigma_t/2)(1-\mu_{t+s})\mu_{t+s}ds + \mu_{t+s}(1-\mu_{t+s})(\lambda_{1t}-\lambda_{0t})ds + dJ_{\mu t+s}$$

$$= (1-\mu_{t+s})\mu_{t+s}(\lambda_{0t}-\lambda_{1t}-1/2)ds + (1-\mu_{t+s})\mu_{t+s}dB_{t+s}$$
$$+ (1/2-\mu_{t+s}-\sigma_t/2)(1-\mu_{t+s})\mu_{t+s}ds + \mu_{t+s}(1-\mu_{t+s})(\mu_{t+s}+\lambda_{1t}-\lambda_{0t})ds + dJ_{\mu t+s}$$
$$= (1-\mu_{t+s})\mu_{t+s}(\lambda_{0t}-\lambda_{1t}-1/2)ds + (1-\mu_{t+s})\mu_{t+s}dB_{t+s}$$
$$+ (1/2+\lambda_{1t}-\lambda_{0t}-\sigma_t/2)(1-\mu_{t+s})\mu_{t+s}ds + dJ_{\mu t+s}$$
$$= -\frac{\sigma_t}{2}(1-\mu_{t+s})\mu_{t+s}ds + (1-\mu_{t+s})\mu_{t+s}dB_{t+s} + dJ_{\mu t+s}.$$

These are the beliefs dynamics given parameters p_t. I now characterize the how the discounted value of a smooth function of μ is expected to evolve over time. It is another application of Ito's lemma. Let $F(\cdot)$ be a smooth function such that $F(0) = F(1) = 1$, then Ito's lemmas implies that if $\bar{\lambda}_{t+s} \equiv \lambda_{1t}\mu_{t+s} + \lambda_{0t}(1-\mu_{t+s})$

$$E_t[de^{-rs}F(\mu_{t+s})] = -re^{-rs}F(\mu_t)ds+e^{-rt}F'(\mu_{t+s})E_t[d\mu_{t+s}]+e^{-rt}F''(\mu_{t+s})E_t[(d\mu_{t+s})^2/2]$$
$$= -re^{-rs}F(\mu_t)ds + e^{-rt}F'(\mu_{t+s})[-\sigma_t\mu_{t+s}(1-\mu_{t+s})/2 + \bar{\lambda}_{t+s}(1-F(\mu_t))]ds$$
$$+ e^{-rt}F''(\mu_{t+s})(1-\mu_{t+s})\mu_{t+s}/2ds$$
$$= [\mu_{t+s}(1-\mu_{t+s})[F''(\mu_{t+s})-\sigma_tF'(\mu_{t+s})]/2 + \bar{\lambda}_{t+s}(1-F(\mu_t)) - rF(\mu_{t+s})]e^{-rs}ds$$

Lastly, the diffusion's generator is defined as

$$\mathcal{L}e^{-rs}F(\mu_t) \equiv \lim_{\mathrm{s}\searrow 0}\frac{E_t[de^{-rs}F(\mu_{t+s})]}{\mathrm{s}}$$
$$= \lim_{\mathrm{s}\searrow 0} e^{-rs}\{[\gamma(\mu_{t+s},p_t)F'(\mu_{t+s}) + F''(\mu_{t+s})]\mu_{t+s}(1-\mu_{t+s}) - rF(\mu_{t+s}) - rF(\mu_{t+s})\}$$
$$= \mu_{t+s}(1-\mu_{t+s})[F''(\mu_{t+s})-\sigma_tF'(\mu_{t+s})]/2 + \bar{\lambda}_{t+s}(1-F(\mu_t)) - rF(\mu_{t+s})$$

This concludes the proof. ||||

A.2 Proof of Theorem 3.2.

Proof

Fix some small dt>0 and belief that $\omega = 1$ as $\mu \in [0,1]$. If $\mu = 0,1$, then beliefs will not change, any additional experimentation is costly and will not change beliefs, and the payoffs already reached 1. Consequently, the agent is strictly better off stopping immediately. The maximum payoffs are then $U(1) = U(0) = 1$. Otherwise, $\mu \in (0,1)$ and the agent either experiments or terminates all experimentation. If he decides to stop, then his terminal payoffs were given to be $1-\kappa(\mu)$. Otherwise, he pick an experiment $p = (\sigma, (\lambda_\omega)_{\omega=0}^1)$ and solves

$$U(\mu) = \max_{p\geq 0} e^{-r\text{dt}} - r\text{dt}C(p) + E[e^{-r\text{dt}}U(\mu')|p]$$

where μ' and the posterior beliefs given μ and the portion of the experiment run in the following time interval of length dt. Subtracting $U(\mu)$ from both sides of the equation, then implies that

$$\begin{aligned} 0 = \max_{p\geq 0} -r\text{dt}C(p) + E[e^{-r\text{dt}}[U(\mu') - U(\mu)|p] - (1-e^{-r\text{dt}})U(\mu) \\ = \max_{p\geq 0} -r\text{dt}[C(p) + U(\mu)] + E[e^{-r\text{dt}}[U(\mu') - U(\mu)|p] + o(\text{dt}). \end{aligned}$$

Next, I divide both sides of the expression by dt:

$$0 = \max_{p\geq 0} -r[C(p) + U(\mu)] + \frac{E[e^{-r\text{dt}}[U(\mu') - U(\mu)|p]}{\text{dt}} + \frac{o(\text{dt})}{\text{dt}}.$$

Taking the limit as dt goes to 0 and noting that hence implies the equation in the prompt. This concludes the proof.

||||