Dynamic Survival Bias in Optimal Stopping Problems[*]

Wanyi Chen

November 29, 2017

Abstract

In many economic scenarios, an analyst learns about a random variable by observing an ongoing rational experimentation. We assume an optimal stopping exercise with binary signals about a binary state of the world. The analyst observes an *public history* of experiments, but not an earlier experimentation *pre-history* of uncertain length.

In this setting, a *dynamic survivor bias* emerges: Bayes-updates ignoring the pre-history is more pessimistic than the sophisticated updates that accounts for all possible pre-histories consistent with an ongoing rational experimentation. We show that this bias is dynamic in the sense that the observation impacts the inference of the un-observed pre-history. Moreover, when the pre-history is stochastically larger, the downward bias is larger when the observed history is sufficiently encouraging for the experimenter. In general, we find that the analyst's Bayes-optimal inference critically depends on the ordering of the signal history and the combined knowledge of the signal realizations and the experimenter's actions.

My theory has implications for technology adoption in R& D settings, and formally subsumes a class of one armed bandits and the Wald experimentation problem, for instance.

Keywords: Optimal stopping; Bandit Problem; Survival Bias

[*] *I thank Lones Smith, William Sandholm, Marzena Rostek and Daniel Quint for conversations and feedback.*

1 Introduction

In many economic scenarios, an *analyst* (she) needs to learn about a random payoff. The distribution of the random payoff can be approached by Bayesian updating from trials of the random variable. However, the analyst cannot execute the trials herself and can only observe the trial results from an on-going experimentation. For instance, when a venture capitalist assesses a company researching on a risky project, she needs to learn the payoff of the project through the company's experiments. Before firms adopt new technologies or consumers adopt new products, they observe experiments from previous users. This paper studies the selection problem that arises when the observed information is partial. This selection problem is special since the available information to the analyst is shaped by an optimization problem of the *experimenter*. In all the examples mentioned before, the experimenter faces an optimal stopping problem as in Wald (1947). Executing trials is costly but rewards the experimenter by a payoff determined by the realizations of the random variable. Specifically, If the reward to the experimenter is the realization of the random payoff each time then it is a classical one-arm bandit problem[1] (consumers' trials on products), while if the reward is only the terminal realization of the random payoff, it is an optimal experimentation problem[2] (firms' trials on technologies). The experimenter will stop experimenting when sufficiently convinced that the payoff is low. Knowing the experimenter's optimization, the analyst updates her belief about the random payoff according to both the observed trial results and the inference of unobserved trial results that are consistent with the on-going experimention. We call this belief the *sophisticated posterior*.

If the analyst uses a simple Bayesian updating only from the observed trial results, her posterior of the random payoff will be biased down since she ignores the previous ongoing experimentation, which in itself is good news. We call the difference between this *naive posterior* and the sophisticated posterior the *dynamic survival bias* as the naive posterior doesn't account for the fact that the experimenter "survives" his stopping problem. Typically, a selection bias problem concerns static statistical inferences when the mere existence of the problem is itself informative. In my setting, the informational content in the ongoing experimentation changes period by period and thus the bias is dynamic.

The partial information problem may stem from many causes, for instance old historical data are lost, or the analyst has limited access to the complete data. In addition, there are cases where the analyst is required to only utilize partial data. For example the FDA conducts three phases of clinical trials before a new drug can be marketed. But the information in the physician labeling is only based on the phase 3 data[3]. In this

[1] As formulated in Gittins (1979)

[2] As formulated in Moscarini and Smith (2001)

[3] Drug Development and Review Definitions in the Investigational New Drug Application. `https://www.fda.gov/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/ApprovalApplications/InvestigationalNewDrugINDApplication/ucm176522.htm`

case the consumers may get a biased (conservative) information on the efficacy of the medicine.

The first goal of this paper is to study the properties of the dynamic survival bias. For simplicity of illustrating this situation, we assume the underlying random variable is Bernoulli distributed with an unknown probability. The experimenter's trial results are thus a sequence of successes and failures. We assume the available information to the analyst is a consecutive sub-sequence of the entire trial history and the experimenter's action of whether to continue or stop experimenting.

Essentially, the selection bias exists because the observation is a subsequence that's towards to the end of the entire trial history, while the trial results are not randomly distributed in the entire history. Since the experimenter will stop experimenting when sufficiently pessimistic about the payoff probability and earlier failures larger drop in the experimenter's incentive to continue experimenting. Therefore, it is more likely that the total trial history contains early successes and later failures. This logic also makes the pattern of the observed trial sequence important. In the basic Bayesian updating process, the number of successes and failures in the observation is a sufficient statistics. However, to derive the sophisticated posterior, the analyst needs to keep track of the ordering of signal realizations. Specially, a signal realizations sequence with earlier failures suggests better inference of pre-observation trial results, and thus implies a higher sophisticated posterior. Intuitively, suppose the pre-observation history already contains many failures, then observing earlier failures implies that the experimenter has to endure consecutive failures, which is a less likely situation. This also justifies the dynamics property of the survival bias, as different observation leads to different inference of the unobserved trial results. The ordering of successes and failures can be more important than the number of them. Signal realization sequences with more failures but earlier failures can imply higher posterior. In addition, the pattern of information also impacts whether the survival bias is larger when the survival time is (stochastically) longer. When the observed sequence of trial results contains early successes and large success numbers, the survival bias accumulates as the survival time gets (stochastically) longer. However, this is not necessarily true when the observation contains early failures and large failure numbers, since in this case, we are more certain that the pre-observation is good when it is shorter.

The other focus of this paper is to study the combined knowledge of the experimenter's signals and actions affects the analyst's learning. Observing the actions critically impacts the interpretation of the observed signal realizations. Naturally, different actions leads to different belief for the analyst, conditional on the same observed trial results. Calling this the first order effect, there's also the second order effect that different actions lead to opposite directions of the comparative statics of the sophisticated posterior on different observed trial results. Specifically, additional failures results in lower posterior when the experimenter is continuing experimenting after the analyst's observation. However, this is not true when the final action is stop, since more failures imply higher likelihood for longer pre-observation trials with more successes. In addition, the sophisticated posterior depends on the incentive of the experimenter, and thus his characteristics such as the cost of running the experiments. The combined knowledge of trial results and action reveals information about the experimenter and in turn affects the sophisticated posterior.

There has been lots of literatures that also consider the learning in stopping problems. For instance, the applications of the Bandit problem in a multi-agent setting. Bolton and Harris (1999) studies a model in which different agents play bandits with the same payoff distribution and can observe the other's trial results(signals). In this case, information is public good and their focus is the free riding problem. Rosenberg, Solan, and Vieille (2007) study a similar setting where the agents can only observe the actions of continuing or stopping playing the bandit, but not their trial results. On the other hand, there's also the applications of the Bandit problem in the contract problem framework. For example, Gomes, Gottlieb, and Maestri (2016) studies the screening and learning where the analyst sees the information but not the action of the experimenter. In all cases, the available information to the observer is either all of the signals, all of the actions, or both. Instead, I assume a natural situation where both action and signals are partially observable and emphasize the learning in a dynamic setting. Another branch of literature my work fits into is the social learning problem. In Smith, Sørensen, and Tian (2012), they argue that the information herding problem stems from the incomplete learning from a bandit player who forgets previous signals and ignores the information impact of his actions. Applying this idea to my model, I study the Bayesian optimal learning of a partially forgetful experimenter who can only keeps track on recent trial histories.

2 Model

A. The Experimenter

Consider an optimization of a rational Bayesian *experimenter* (he) at periods $-T, -T+1, \ldots, -1, 0, 1, \ldots$, for some nonnegative integer T. Let $X_t = 0, 1$ be a sequence of independent Bernoulli random variable, all with a fixed *success* chance (namely, $X_t = 1$) that is a random variable P on $[0, 1]$, and a *failure* chance ($X_t = 0$) of $1 - P$. We assume that the experimenter has a prior on P that follows a Beta distribution. We denote this prior by Υ and parameteriz it by a and b, such that its density on $p \in [0, 1]$ is a constant times $p^a(1-p)^b$.

The experimenter is engaged in an optimal stopping exercise with actions "stop" and "go" in periods $-T, -T+1, \ldots, -1, 0, 1, \ldots$. The game ends as soon as the experimenter stops. There are two focal optimal stopping problem applications. For simplicity, we focus in the main text on the one-armed bandit model (Gittins, 1979), in which the experimenter must forego a payoff of certain value $v \in (0, 1]$ each period that he draws X. This is his opportunity cost of experimentation. He collects the realization of X each time he draws it. He wishes to maximize the discounted expected payoffs, his discount factor is $\delta \in [0, 1)$. The Appendix considers the Wald sequential learning problem, namely, the optimal stopping problem that arises with costly sequential purchases of Bayesian signals prior to a terminal binary stopping decision.

B. The Analyst

The focus of this paper is instead on an *analyst*: She fully understands the optimization faced by the experimenter, and shares his prior[4] on P. In each period $n = 0, 1, \ldots$, if the experimentation is ongoing, then she sees the realized *public history* $H = \langle x_0, \ldots, x_n \rangle$ of trial results from the realizations of $\langle X_0, \ldots, X_n \rangle$ starting at time 0, as well as the action $a \in \{\text{go, stop}\}$ at time n. The public history H belongs to the space $\mathcal{H}$ of strings of 0's and 1's of any length. The analyst doesn't know the number of trials T in the *pre-history*, nor how they turned out. Her prior is that T obeys the geometric distribution with parameter $\gamma \in [0, 1)$, so that $\Pr(T = t) = \gamma^t (1 - \gamma)$ for all $t = 0, 1, \ldots$.

we explore the properties of the analyst's posterior[5] $F(p|S, a)$ on the probability $P = p$.

3 Solution to the Experimenter's problem

Since the experiment observes a binomial experiment, his beliefs and thus behavior in any period only depends on the *state* of successes number σ and failure number ϕ, where $\sigma, \phi \in \mathbb{N}$. As is well-known (Gittins, 1979), his optimal behavior is described by an index rule $I(\sigma, \phi)$, where after seeing (σ, ϕ), he chooses to go iff his cost $v \leq I(\sigma, \phi)$. Note that this index also depends on the prior Υ and discount factor δ. We will specify $I(\sigma, \phi)$ in the Appendix. Let $H_0 = \langle x_{-T}, \ldots, x_{-1} \rangle$ be the pre-history of trial realizations. This belongs to the space $\mathcal{H}_0$ of strings of 0's and 1's of length T . Let $\sigma_t(H_0)$ and $\phi_t(H_0)$ be the respective numbers of successes and failures seen in periods $-T, \ldots - 1$. Let $|H|$ denote the length of a public history H and ' + " between sequences denote the operation of concatenation (E.g. $\langle 10 \rangle + \langle 01 \rangle = \langle 1001 \rangle$). We continue to use $\sigma_t(\cdot)$ and $\phi_t(\cdot)$ to denote the respective numbers of successes and failures in periods $0, 1, \ldots t$ seen in any total histories. With some abuse of notation, we later use $\sigma(\cdot)$ and $\phi(\cdot)$ without the subscript to denote the total number 1's and 0's in any sequence. Namely $\sigma(H_0 + H) = \sigma_{|H|}(H_0 + H)$, and $\phi(H_0 + H) = \phi_{|H|}(H_0 + H)$. Call $(\sigma(H_0 + H), \phi(H_0 + H))$ the ending state of $H_0 + H$. When the public history is H, let Γ_H denote the set of the pre-histories that contains the public history H. Since the experimenter uses the Gittins index rule every period, then his pre-history H_0 must belong to the *continuation set:*

$$\Gamma_H = \{H_0 | H_0 \in \mathcal{H}_0, I(\sigma_t(H_0 + H), \phi_t(H_0 + H)) \geq v, \text{ for all } t = -\text{T},\ldots\text{-1,-0,1}\ldots|\text{H}|\} \quad (1)$$

[4]This assumption is inessential and we only use it for convention. The main results in this study hold as long as the anlayst's prior is dominated by the experimenter's in the likelihood ratio order.

[5]The full expression of the analyst's posterior is $F(p|H, a; \gamma; \delta, v\Upsilon)$. For simplicity we supress this notation to $F(p|H, a)$ in most of the text .

Specifically, when $H = \varnothing$,

$$\Gamma_\varnothing = \{H_0 \in \mathcal{H}_0 : \ I(\sigma_t(H_0), \phi_t(H_0)) \geq v, \text{for all } t = -T, \ldots - 1\} \tag{2}$$

For any $H \in \mathcal{H}$, we have $\Gamma_H \subseteq \Gamma_\varnothing$. This property that Γ depends on H is the essential cause for the survival bias to be dynamic, which we will discuss in the next section.

4 Dynamic Survival Bias

Now let's turn to the analyst's problem. We first consider as a benchmark, the *naive posterior* based solely on the public history H and disregard the pre-history. Let $G(p|H)$ denote this cumulative distribution (cdf), namely the chance that the success chance $P \leq p$, given H alone. By contrast, the *sophisticated posterior* $F(p|H, a)$ is the correct Bayesian posterior belief of the analyst that accounts for all possible pre-histories consistent with the ongoing experimentation and the most recent action a, either go or stop.

Intuitively, the naive and sophisticated posteriors coincide if the experimenter's cost v is 0, or he is infinitely patient: $\delta = 1$. For in either case, the experimenter never stops, and not conditioning on the pre-history entails no loss of information. But since $v > 0$ and $\delta < 1$, the naive posterior is downwardly biased because it ignores the information from the event that the experimenter has not yet stopped. We call this difference between the naive and sophisticated posteriors the *dynamic survival bias*.

This is a selection bias because we only see the public sequence of trial results, which is not representative of the entire history. It is dynamic, since the inference will change over time, as we shall soon show.

Conversely, it is more likely for the observation to contain fewer success than in a sample from all the histories with the same size of the public history. In this sense, the observation is not representative of the total trial history.

The next result[6] exploits the likelihood ratio order $\succeq_{\text{lr}}$ on differentiable cdf's F_1, F_2, namely, $F_2 \succeq_{\text{lr}} F_1$ if $F_2'(p)/F_1'(p)$ increases in $p \in [0, 1]$.

Lemma 1. *The sophisticated posterior on the success chance P after "go" dominates the naive posterior in the likelihood ratio order: $F(\cdot|H, go) \succeq_{lr} G(\cdot|H)$ for all public histories H.*

In particular, $G(\cdot|\varnothing)$ is the assumed beta distribution prior belief on P, and is her sophisticated posterior belief on P conditional on the experimenter's survival until period 0.

Here we show $F(\cdot|\varnothing, \text{go}) \succeq_{\text{lr}} G(\cdot|\varnothing)$ with a particular example. The logic applies to general cases and a formal proof for Lemma 1 is in the appendix.

[6] The next result holds for all priors of P in $\Delta[0, 1]$ and not just the assumed beta prior.

Example. Suppose the experimenter is myopic (has a discount factor $\delta = 0$), has uniform prior on P and cost $v = 0.5$. In this case, the continuation set is

$$\Gamma_\varnothing = \{H_0 : \sigma_t(H_0) \geq \phi_t(H_0) \text{ for all } t = -T, \ldots -1\}$$

In addition, assume the analyst knows the pre-history length is 2. The only possible pre-histories are $\langle 10 \rangle$ and $\langle 11 \rangle$. The density $F'(p|\varnothing, \text{go})$ is thus proportional to $p^2 + p(1-p) = p$, the likelihood of the event that the experimenter is still going conditional on the realization of P being p. This dominates the uniform distribution in the likelihood ratio order as the likelihood ratio p is an increasing function of p.

This is not a coincident as the density of uniform distribution can be broken down into $1 = p^2 + 2p(1-p) + (1-p)^2$, corresponding to the pre-histories $\langle 11 \rangle$, $\langle 01 \rangle$, $\langle 10 \rangle$ and $\langle 00 \rangle$, unconditional on the experimenter's stopping problem. Note that the posterior of a certain pre-history of $\langle 00 \rangle$, $\langle 10 \rangle$ (or $\langle 01 \rangle$) and $\langle 11 \rangle$ are $\text{Beta}(1, 3)$, $\text{Beta}(2, 2)$ and $\text{Beta}(3, 1)$ respectively, increasing in the likelihood ratio order. Comparing the updating of the sophisticated posterior to the naive posterior, pre-histories with worse posteriors ($\langle 00 \rangle$ and $\langle 01 \rangle$) are eliminated.

Moreover, this bias is dynamic since the posterior on the pre-history changes with the public history, on contrast to usual survival bias where the bias is static. To illustrate this, let's consider the setting in the previous example again. Now instead of a null public history, let's assume the public history is $\langle 1 \rangle$, then the possible pre-histories are the same as before, $\langle 10 \rangle$ and $\langle 11 \rangle$. However if the public history is $\langle 0 \rangle$, the only possbile history is only $\langle 11 \rangle$, resulting in a posterior in period 0 that's proportional to p^2. We will address this later in section 5.

Hence, the public history is critical in estimating the unobserved pre-history with action "go". However when observing action "stop", the information conveyed by the public history is not so essential.

Let $B(\Gamma_H)$ denote the *boundary* of Γ_H:

$$B(\Gamma_H) = \{H_0 |\ H_0 \in \mathcal{H}_0 :\ H_0 \in \Gamma_H,\ H_0 + \langle 0 \rangle \notin \Gamma_H\} \tag{3}$$

When the action is "stop", the total histories must have the form $H_0 + H + \langle 0 \rangle$ where H_0 belongs to the boundary of the continuation set $B(\Gamma_H)$. The public history does affect the continuation set and thus the inference of the pre-histories . But the information from the action "stop" is much more important than that from the public history. Hence we don't draw conclusion about the comparison of $F(\cdot|H, \text{stop})$ and $G(\cdot|H)$.

5 Comparative Statics

Previously, we showed that the naive Bayesian updating results in a downward bias. Now we argue that it is also flawed in treating the number of successes and failures in the history as a sufficient statistics. In the sophisticated posterior, different ordering of the public information is essential. Now we introduce how to identify the relevant patterns in the public history.

We say sequence H' *lexicographically dominates* H if H' can be derive from H by moving a "0" to the left. For example, $\langle 011 \rangle \sqsupseteq_1 \langle 101 \rangle \sqsupseteq_1 \langle 110 \rangle$. Intuitively, H' has earlier failures when H' dominates H.

Next, we introduce the notion of comparison for the analyst's posteriors. We say a public history H' is more *encouraging* than H for the analyst after action a if H' leads to a sophisticated posterior that's higher in the second order stochastic dominance[7]. We denote this relation by $F(\cdot|H', a) \succeq_2 F(\cdot|H, a)$. The second order stochastic dominance implies a higher expectation of P with lower variance. This fits the analyst's goal of rendering a smart judgment of P.

Proposition 1. *After action "go", a public history is more encouraging if it is higher in the lexicographical order.*

$$H' \sqsupseteq H \Rightarrow F(\cdot|H', go) \succeq_2 F(\cdot|H, go)$$

Intuitively, starting from any state (σ, ϕ), a following trial results sequence that contains earlier failures always reaches states with lower Gittins index than a sequence that contains later failures. This can be seen in the exemplary table of Gittins indexes in Figures 1. Each of the grids represents a state. As the experiment goes on, states $(\sigma_t(H_0 + H), \phi_t(H_0 + H))$, $t = -T, \ldots, -1, 0, 1, \ldots$ forms a path in this table. A success leads the path to the right and a failure leads the path down. Thus earlier failures make the experimenter more likely to stop. Therefore, observing earlier failures while the experimenter continuing experimenting suggests that the pre-observation history must be good enough for the experimenter to endure the early failures.

Figures 1 shows the ending states $(\sigma_{|H|}(H_0 + H), \phi_{|H|}(H_0 + H))$ of all possible total histories $H_0 + H \in \Gamma_H$ for an experimenter with uniform prior, discount factor $\delta = 0$, and cost $v = 0.5$, where the public history are $H_1 = \langle 10 \rangle$ and $H_2 = \langle 01 \rangle$ respectively. Note that with $\langle 01 \rangle$, the pre-histories with the worst posteriors are eliminated. Essentially, the smaller is the continuation set Γ_H, the higher is the sophisticated posterior. The formal proof is in the appendix.

Next, we discuss the marginal effect of signals by showing how does the posteriors change when we append or prepend[8] one signal realization to a public history. We can regard prepending an observation as the analyst tracing back the trials for one period of time and appending an observation as the analyst continuing one period of observation. After action "go, very intuitively, concatenating successes (failures) to a public history makes

[7] $F_2 \succeq_2 F_1$ if $\int_0^x (F_1(t) - F_2(t))\, \mathrm{d}t \geq 0$ for all $x \in [0, 1]$

[8] We use "append" ("prepend") to refer to the operation of adding an element at the beginning (end) of a sequence, and use "concatenate" to refer to either of these two cases.

Changes in the Possible Total Histories with Different Public History

(a)

Number of Failures \ Number of Successes	0	1	2	3	4	5	6	7
0	0.50	0.67	0.75	0.80	0.83	0.86	0.88	0.89
1	0.33	0.50	0.60	0.67	0.71	0.75	0.78	0.80
2	0.25	0.40	0.50	0.57	0.63	0.67	0.70	0.73
3	0.20	0.33	0.43	0.50	0.56	0.60	0.64	0.67
4	0.17	0.29	0.38	0.44	0.50	0.55	0.58	0.62
5	0.14	0.25	0.33	0.40	0.45	0.50	0.54	0.57
6	0.13	0.22	0.30	0.36	0.42	0.46	0.50	0.53
7	0.11	0.20	0.27	0.33	0.38	0.43	0.47	0.50

(b)

Number of Failures \ Number of Successes	0	1	2	3	4	5	6	7
0	0.50	0.67	0.75	0.80	0.83	0.86	0.88	0.89
1	0.33	0.50	0.60	0.67	0.71	0.75	0.78	0.80
2	0.25	0.40	0.50	0.57	0.63	0.67	0.70	0.73
3	0.20	0.33	0.43	0.50	0.56	0.60	0.64	0.67
4	0.17	0.29	0.38	0.44	0.50	0.55	0.58	0.62
5	0.14	0.25	0.33	0.40	0.45	0.50	0.54	0.57
6	0.13	0.22	0.30	0.36	0.42	0.46	0.50	0.53
7	0.11	0.20	0.27	0.33	0.38	0.43	0.47	0.50

Figure 1:
These tables show the Gittins Index for each combination of success and failure numbers. The blue areas in (a) are the ending states corresponding to all possible total histories of the experimenter from whose trials the public history is $\langle 10 \rangle$. (b) The blue areas are the ending states corresponding to all possible total histories of the experimenter from whose trials the public history is $\langle 01 \rangle$. The red area highlights the states that are possible with public history $\langle 10 \rangle$ but not with $\langle 01 \rangle$.

it more (less) encouraging:

$$F(\cdot|H', \text{go}) \succeq_2 \; F(\cdot|H, \text{go}) \text{ if } \begin{cases} H' = \langle 1 \rangle + H & \text{or } H' = H + \langle 1 \rangle \\ H = \langle 0 \rangle + H' & \text{or } H = H' + \langle 0 \rangle \end{cases} \tag{4}$$

In particular,

$$F(\cdot|H', \text{go}) \succeq_{\text{lr}} (\cdot|H, \text{go}) \;\text{ if }\; H' = H + \langle 1 \rangle$$

Note that in this case, the notion of comparison is the likelihood ratio order, stronger than the second order stochastic dominance[9]. The reason is that in this case the inference about the pre-histories is not changed while in all other cases this inference evolves as additional signals pop out, namely $\Gamma_{H'} = \Gamma_H$ when $H' = H + \langle 1 \rangle$. Thus $F(\cdot|H', \text{go})$ can be computed from treating $F(\cdot|H, \text{go})$ as a prior and update it by one failure, while in the other cases, this property doesn't hold.

However, when the action is stop, the opposite holds if the experimenter is patient enough, namely his discount facot δ is sufficiently enough, and the public history contain enough failures, where the threshhold value of the number of failures depends on the discount factor and the prior Υ. We state this property formally in the next proposition.

Proposition 2. *After action "stop", concatenating failures (successes) to a public history makes it more(less) encouraging when the public history contains enough failures and the*

[9]The likelihood ratio order implies the first order stochastic dominance, and thus is stronger than the second order stochastic dominance.

The Sophisticated Posterior Rises with More Failures and Falls with More Successes

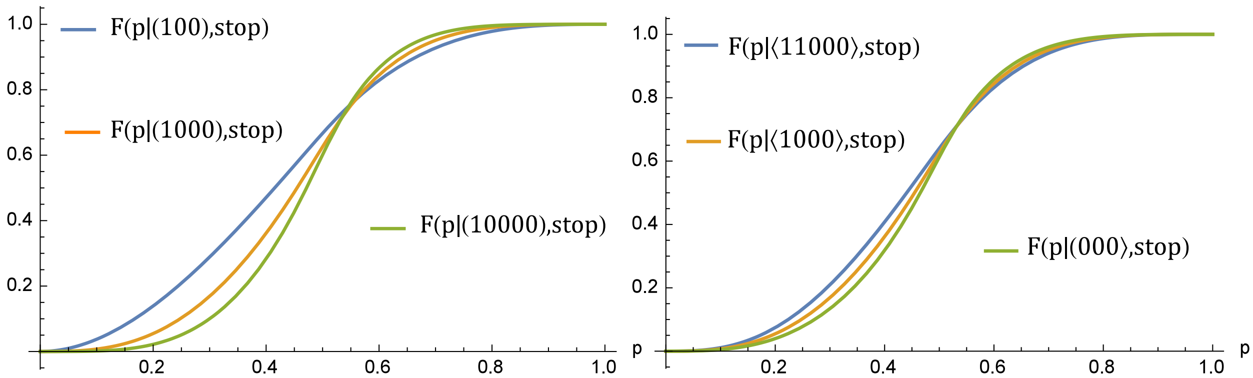

Figure 2:

These figures show the cdf and pdf of $F(\cdot|H, \text{stop})$ with different H's. Now we explain the second order stochastic dominance relationships shown in these figures. On the left panel, the orange line ($F(p|\langle 1000\rangle, \text{stop})$) single crosses the blue line ($F(p|\langle 100\rangle, \text{stop})$). Since the area between the blue line and the orange line before their intersection is larger than that after their intersection, the integral of $F(p|\langle 1000\rangle, \text{stop})$ for $p \in [0, q]$ is always lower than that of ($F(p|\langle 100\rangle, \text{stop})$), thus $F(p|\langle 1000\rangle, \text{stop}) \succeq_2 F(p|\langle 100\rangle, \text{stop})$. The other cases follow the same arguments.

experimenter is patient enough: $\exists \bar{\delta} > 0$ *and* $\bar{\phi}(\Upsilon, \delta)$ *such that when* $\delta > \bar{\delta}$ *and* $\phi(H) > \bar{\phi}$*,*

$$F(\cdot|H', stop) \succeq_2 \; F(\cdot|H, stop) \text{ if } H = \langle 1\rangle + H',\;\; H' = H + \langle 0\rangle \text{ or } H' = \langle 0\rangle + H$$

The intuition is that, when observing more failure, it is more likely that the experimenter has continued longer time with more successes such that he can endure the failures. This effect is stronger when the public history contain enough failures. Suppose one observed 10 failure and the experimenter stops, than an additional failure before it suggests that the pre-history must be good so that the experimenter can endure such consecutive failures. But when the observed failure number is few, this argument can not be supported.

Figure 2 shows the magnitude of the marginal effect of additional successes and failures after action "stop".

In addition, adding successes (failures) to a public history at arbitrary positions doesn't necessarily makes it more encouraging. Here's an example:

Example. Suppose the experimenter has uniform prior on P, discount factor $\delta = 0.6$, and cost $v = 0.5$. We assume in this example the pre-history length is known to be 2. The tables in Figure 3 show the ending states of possible pre-histories with different public histories $\langle 10\rangle$ and $\langle 001\rangle$. When $H = \langle 10\rangle$, the possible pre-histories are $\langle 11\rangle$ and $\langle 10\rangle$,

		Number of Successes				
		0	1	2	3	4
Number of Failures	0	0.58	0.72	0.79	0.83	0.85
	1	0.39	0.55	0.64	0.7	0.74
	2	0.29	0.44	0.53	0.6	0.65
	3	0.23	0.37	0.46	0.53	0.58
	4	0.19	0.31	0.4	0.47	0.52
	5	0.16	0.27	0.36	0.42	0.48
	6	0.14	0.24	0.32	0.38	0.44

		Number of Successes				
		0	1	2	3	4
Number of Failures	0	0.58	0.72	0.79	0.83	0.85
	1	0.39	0.55	0.64	0.7	0.74
	2	0.29	0.44	0.53	0.6	0.65
	3	0.23	0.37	0.46	0.53	0.58
	4	0.19	0.31	0.4	0.47	0.52
	5	0.16	0.27	0.36	0.42	0.48
	6	0.14	0.24	0.32	0.38	0.44

Figure 3:
The gray areas are the states with the Gittins Indexes smaller than the cost $v = 0.5$. These area cannot be reached along the path of any history. The colored blocks show the possible paths for the public history $\langle 10\rangle$ and $\langle 001\rangle$.

while when $H = \langle 001\rangle$, the only possible pre-history is $\langle 11\rangle$. Thus, $F(p|\langle 001\rangle, \text{go})$ has a density proportional to $p^3(1-p)^2$, while $F(p|\langle 10\rangle, \text{go})$ has a density proportional to $p^3(1-p)^1 + 2p^2(1-p)^2$. We have[10] $F(p|\langle 001\rangle, \text{go}) \succeq_2 F(p|\langle 10\rangle, \text{go})$.

Next, we answer the question whether longer survival time T leads to larger bias. When certain conditions are met, the answer is yes.

Recall that $\Pr(T = t) = \gamma(1-\gamma)^t$. Higher γ implies that the experimenter expects the pre-history to be stochastically longer. Here we expand the expression of the sophisticated posterior to $F(\cdot|H, a; \gamma)$. With some abuse of notation, we use $F(\cdot|H, \text{go}; t)$ to denote the sophisticated posterior when we fixed the length of the pre-history at some t.

Proposition 3. **(a)** *After action "go", when the pre-observation history length rises stochastically, the sophisticated posterior is more encouraging: when $H \sqsupseteq \bar{H}$ or $H \sqsupseteq \bar{H} + \langle 1\ldots\rangle$ for some $\bar{H} \in B(\Gamma_H)$,*

$$1 > \gamma_2 > \gamma_1 > 0 \Rightarrow F(\cdot|H, go; \gamma_2) \succeq_{lr} F(\cdot|H, go; \gamma_1)$$

(b) *After action "stop", when the pre-observation history length rises stochastically, The sophisticated posterior is more encouraging when the experimenter is sufficiently patient: there $\exists\bar{\delta}$, when $\delta > \bar{\delta}$,*

$$1 > \gamma_2 > \gamma_1 > 0 \Rightarrow F(\cdot|H, \text{stop}; \gamma_2) \succeq_2 F(\cdot|H, \text{stop}; \gamma_1)$$

The action stop is irreversible and the experimenter makes the decision of whether to stop in each period of time. Continueing the experimentation for $t+1$ periods implies the continuation for t periods. The fact that the experimenter still pushing onward for

[10] $\int_0^q F(p|\langle 10\rangle, \text{go})\mathrm{d}p - \int_0^q F(p|\langle 001\rangle, \text{go})\mathrm{d}p = 2p^4(35 - 84q - 63q^2 + 25q^3)/35 \geq 0$ for $q \in [0, 1]$.

The Sophisticated Posterior Rises in the Pre-history Length

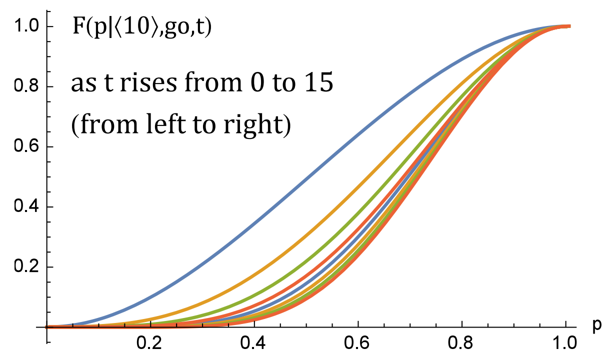

Figure 4:
This figrues plots $F(p|\langle 10\rangle, \text{go}; t)$ with various t when the experimenter is myopic, has uniform prior and outside option valua 0.5.

longer time is a stronger optimistic signal.. In other words, the survival bias accumulates as the survival time grows longer. We show the magnitude of the dynamic survival bias by plotting $F(p|H, \text{go}, t)$ varying the certain pre-history length t in Figure 4.

However, note that this property doesn't necessarily hold for public histories containing early failures or too many large failures compared to successes. In this case, we are certain about pre-history successes when the history length is short and this certainty is lost when the pre-history is long. For instance, if we see 10 failures and we know there's only one un-observed trial before that, we are pretty sure the un-observed trial is a success. However when we know there are 100 previous trials, we can hardly make judgements about them.

The intuition for (b) is very different from part (a). The motivation for the experimenter to continue experimenting comes from both the expectation of the next period payoff and future information gains. Namely there's a uncertainty bonus contained in the Gittins Index. With the Beta prior, the experimenter is more willing to continue with short histories when they yield close expectation of P. Hence, in the perspective of the analyst, the longer the experimenter had survived before his stop, the more encouraging are his pre-histories.

The sophisticated posterior depends on the incentive of the experimenter to continue the experimentation, which is determined by his prior distribution Υ, discount factor δ, and outside option value v. Thus we expand the expression of sophisticated posterior to $F(\cdot|H, a; \Upsilon, \delta, v)$. We complete the analysis of the benchmark model by discussing how does the characteristics of the experimenter affect the sophisticated posterior.

Proposition 4. *The sophisticated posterior falls in the experimenter's prior, discount factor and rises in his cost with both actions a=stop,go.*

$$\Upsilon \succeq_1 \Upsilon', \;\; \delta \geq \delta', \;\; \text{and } v' \geq v \Rightarrow F(\cdot|H, a; \Upsilon', \delta', v') \succeq_{lr} F(\cdot|H, a; \Upsilon, \delta, v), \;\; a = \textit{stop}, \textit{go}$$

When the experimenter is more optimistic in ex ante, more patient, or suffers less cost, he is more determined to push onward. Therefore the sophisticated analyst should be stochastically less optimistic and the survival bias is smaller.

6 Summary

I study a model in which an analyst learns the distribution of a Bernoulli random variable by observing actions and partial consecutive trial results of a Bayesian rational experimenter who conducts costly strategic experimentation on it to obtain a payoff determined by the realization of the random variable. This setting is different from existing related literatures. From the experimenter's incentive in choosing to continue or stop experimenting, the partial observation truncated in time is not a randomly selected sample from the total trial results, and the problem of missing historical information may result in a selection bias from naively Bayesian updating methods that causes a downward estimate of the payoff probability. This bias is dynamic in the sense that different observation affects the inference of the un-observed pre-history. In addition, the order that one learns conditionally i.i.d. signals now significantly matters for the sophisticated analyst's beliefs — a sequence of signal realizations with early failures lead to better estimation than a sequence with late failures, different from the basic principle of Bayesian updating. Moreover, upon observing stopping, more observed failures can be a optimistic signal.

7 Appendix

7.1 The Wald's Optimal Experimentation Problem

We use the formulation of the Wald's optimal experimentation problem as in Moscarini and Smith (2001). There are two stages for the experimenter. On the first stage, he must pay a cost c to draw X in periods $-T, -T+1, \ldots, -1, 0, 1, \ldots$. Unlike in the bandit problem, these trials only serve as signals, but do not yield payoff. After each trial, he chooses whether to continue the experiment or to strop. When he stops, he can take the risky action and obtain the realization of X or the safe action with a certain payoff v.

This setting is equivalent to the problem that the experimenter runs a sequential test for $E[X] = P < v$ and $P > v$ with a desired error determined by the experimentation cost c and the certain payoff value v.

By Huelsenbeck and Crandall (1997), given any total history $\bar{H}$, the experimenter will conduct a maximized sequential probability ratio test with the statistics

$$\Lambda(\bar{H}) = \frac{\sup\{p^{\sigma(\bar{H})}(1-p)^{\phi(\bar{H})} : p < v\}}{\sup\{p^{\sigma(\bar{H})}(1-p)^{\phi(\bar{H})} : p > v\}} \tag{5}$$

There exists $0 < \underline{\lambda} < \bar{\lambda}$ such that the experimenter will continue when$\Lambda(\bar{H}) \in [\underline{\lambda}, \bar{\lambda}]$, stop and take the safe action when $\Lambda(\bar{H}) < \underline{\lambda}$, stop and take the risky action when $\Lambda(\bar{H}) > \bar{\lambda}$. By the expression (5), for some $0 < k_1 < k_2$, the experimenter's continuation set can be expressed as

$$\Gamma_H = \{H_0 \in \mathcal{H}_0 | k_1 \phi_t(H_0 + H) \leq \sigma_t(H_0 + H)) \leq k_2 \phi_t(H_0 + H)) \text{ for all } t = -T, \dots\}$$

Now there are two boundaries for this set, the lower boundary is the same as (3) and we denote it as $\underline{B}(\Gamma_H)$ and the upper bound is

$$\bar{B}(\Gamma_H) = \{H_0 | \ H_0 \in \mathcal{H}_0 : \ H_0 \in \Gamma_H, \ H_0 + \langle 1 \rangle \notin \Gamma_H\}$$

Proposition 1 and 2(a) still hold in this setting since they only requires that the boundary of the continuation set is monotone, in the sense that if H_0, $H_0' \in B_i(\Gamma_H)$ and $\sigma(H_0') \geq \sigma(H_0')$, then $\phi(H_0') \geq \phi(H_0)$.

More generally, Lemma1, proposition 1 through 3 holds for all optimal stopping problems where the experimenter's continuation set Γ_H can be expressed as

$$\{H_0 \in \mathcal{H}_0 | \sigma_t(H_0 + H) \geq k(t) \text{ for all } t = -T, \dots\}$$

with some integer $k(t)$ increasing in t. This applied to the setting of the bandit problems where the experimenter doesn't use the exact Gittins Index rules but approximate solutions or heuristic methods like the UCB algorithm[11].

7.2 Specify of the Gittins Index and the Sophisticated Posterior

We treat each pair of success and failure numbers (σ, ϕ) as a state for the experimenter's dynamic programming problem. Let $\mu(\sigma, \phi)$ denote the expected value of P given any history with σ successes and ϕ failures. We can express his value function as:

$$V(\sigma, \phi) = \max\{v + \delta V(\sigma, \phi), \ \mu(\sigma, \phi) + \delta\left(\mu(\sigma, \phi) V(\sigma + 1, \phi) + (1 - \mu(\sigma, \phi) V(\sigma, \phi + 1))\right)\}$$

The Gittins index in state (σ, ϕ) is the smallest fixed point of a recursively defined function[12]

$$W(\gamma|\sigma,\phi) = \max\{\gamma, \ \frac{\mu(\sigma,\phi)}{1/(1-\delta)} + \delta\left(\mu(\sigma,\phi) W(\gamma|\sigma+1, \phi) + (1 - \mu(\sigma,\phi)) W(\gamma|\sigma, \phi+1)\right)\} \quad (6)$$

where $\lim_{a\to\infty, b\to\infty} W(\gamma|\sigma, \phi) = \max\{\gamma/(1-\delta), \ \mu(\sigma, \phi)\}$.

[11] In the bandit problem, as in Bubeck, Cesa-Bianchi, et al. (2012), this algorithm uses the index $\mu(\bar{H}) + \sqrt{\alpha \ln |\bar{H}| / |\bar{H}|}$ where $\mu(\bar{H})$ denotes the mean of P given total history $\bar{H}$ for some $\alpha > 0$.

[12] See Gittins, Glazebrook, and Weber (2011)

Next, we show the formal expression of the sophisticated posterior.

$$F(p|H,\text{go};\gamma;\delta,\Upsilon,v)=\frac{\int_0^p\sum_{H_9\in\Gamma_H(\delta,\Upsilon,v)}z^{\sigma(H_9+H)}(1-z)^{\phi(H_9+H)}\gamma^{|H_0|}\mathrm{d}z}{\int_0^1\sum_{H_9\in\Gamma_H(\delta,\Upsilon,v)}z^{\sigma(H_9+H)}(1-z)^{\phi(H_9+H)}\gamma^{|H_0+H|}\mathrm{d}z} \tag{7}$$

$$F(p|H,\text{stop};\gamma;\delta,\Upsilon,v)=\frac{\int_0^p\sum_{H_0\in B(\Gamma_H(\delta,\Upsilon,v))}z^{\sigma(H_9+H)}(1-z)^{\phi(H_9+H)+1}\gamma^{|H_0|}\mathrm{d}z}{\int_0^1\sum_{H_0\in B(\Gamma_H(\delta,\Upsilon,v))}z^{\sigma(H_9+H)}(1-z)^{\phi(H_9+H)+1}\gamma^{|H_0|}\mathrm{d}z} \tag{8}$$

Let Ω_H denote the set of states corresponding to the histories in the continuation set Γ_H, namely

$$\Omega_H=\{(\sigma_{-1}(H_0),\phi_{-1}(H_0))|H_0\in\Gamma_H\}$$

Let Ω_H^B denote set of states corresponding to the histories in the boundary $B(\Gamma_H)$, namely

$$\Omega_H^B=\{(\sigma_{-1}(H_0),\phi_{-1}(H_0))|H_0\in B(\Gamma_H)\}$$

Assume state (σ,ϕ) is the σ,ϕth vertex in a 2-dimensional regular grid. Then Ω is a quasi-upper-triangular region in the square grid graph of (i,j), since $I(i,j)$ rises in i and falls in j [13]. We can partition it into $\Omega=\cup_{T=0}^{\infty}\cup_{i=b(T)}^{T}\{(i,T-i)\}$ where $b(T+1)=b(T)+1$ or $b(T+1)=b(T)$. Let $N(i,j|\Omega)$ be the number of staircase walks whose pathes are entirely inside Ω from $(0,0)$ to (i,j). Note that $N(i,j|\Omega)$ is the number of possible pre-histories with success number i and failure number j. Next, we show some important property of $N(i,j|\Omega)$ that's used in many of the proofs.

Claim 1. (a). $N(i,j|\Omega)=N(i-1,j|\Omega)+N(i,j-1|\Omega)$ when $(i,j)\in\Omega$ and $(i-1,j)\in\Omega$.

(b). $N(i,j|\Omega)=N(i,j-1|\Omega)$ when $\forall(i,j)\in\Omega$ and $(i-1,j)\notin\Omega$.

For example, when the experimenter has uniform prior is impatient ($\delta=0$), then $\Omega=\{(i,j):\ i\geq j\}$ and $N(i,j|\Omega)$ is the number of dyck path[14] from $(0,0)$ to (i,j).

Incorparating this, we can write (7) and (8) as

$$\begin{aligned}F(p|H,\text{go};\gamma;\delta,\Upsilon,v)&=\frac{\int_0^p\sum_{(i,j)\in\Omega_H(\delta,\Upsilon,v)}z^{i+\sigma(H)}(1-z)^{j+\phi(H)}\gamma^{i+j}\mathrm{d}z}{\int_0^1\sum_{(i,j)\in\Omega_H(\delta,\Upsilon,v)}z^{i+\sigma(H)}(1-z)^{j+\phi(H)}\gamma^{i+j}\mathrm{d}z}\\&=\frac{\int_0^p\sum_{T=0}^{\infty}\sum_{i=b(T)}^{T}N(i-\sigma(\mathrm{H}),j-\phi(\mathrm{H})|\Omega_H)z^i(1-z)^{T-j}\gamma^T\mathrm{d}z}{\int_0^1\sum_{T=0}^{\infty}\sum_{i=b(T)}^{T}N(i-\sigma(\mathrm{H}),j-\phi(\mathrm{H})|\Omega_H)z^i(1-z)^{T-j}\gamma^T\mathrm{d}z}\end{aligned} \tag{9}$$

[13] By Bellman (1956)

[14] The Catalan numbers.

$$F(p|H,\text{go};\gamma;\delta,\Upsilon,v)=\frac{\int_0^p\sum_{T=0}^{\infty}N(i-\sigma(H),j-\phi(H)|\Omega_H)z^i(1-z)^{T-j+1}\gamma^T\mathrm{d}}{\int_0^1\sum_{T=0}^{\infty}N(i-\sigma(H),j-\phi(H)|\Omega_H)z^i(1-z)^{T-j+1}\gamma^T\mathrm{d}} \tag{10}$$

When the experimenter is impatient enough and thus behaving in the myopic way, the sophisticated posterior can be expresses explicitly with certain cost values.

Let $c(m,\rho,r)=\frac{r}{m\rho+r}\binom{m\rho+r}{m}$ denote the mth fuss_catalan number.

E.g. When $\Upsilon\sim\text{Beta}(a,b)$ with $a\geq b$ and $a,b\in\mathbb{N}$, $v=0.5$ and[15] $\delta\leq 0.8$.

$$F(p|H,\text{go};T)=\frac{\int_0^p\sum_{s=\lceil T/2\rceil}^{T}\left(\sum_{m=0}^{T-s}c(m,2,r_1)c(T-s-m,2,r_2)\right)z^s(1-z)^f\mathrm{d}\Upsilon(z)}{\int_0^1\sum_{s=\lceil T/2\rceil}^{T}\left(\sum_{m=0}^{T-s}c(m,2,r_1)c(T-s-m,2,r_2)\right)z^s(1-z)^f\mathrm{d}\Upsilon(z)}$$

where $r_1(s)=|H|-2\sigma(H)+2s-T+1$, $r_2=a-b+1$.

When $\Upsilon\sim\text{Beta}(a,b)$ with $a\geq 2b+1$ and $a,b\in\mathbb{N}$, $v=2/3$.

$$F(p|H,\text{go};T)=\frac{\int_0^p\sum_{s=\lceil 2T/3\rceil}^{T}\left(\sum_{m=0}^{T-s}c(m,3,r_1)c(T-s-m,3,r_2)\right)z^s(1-z)^f\mathrm{d}\Upsilon(z)}{\int_0^1\sum_{s=\lceil 2T/3\rceil}^{T}\left(\sum_{m=0}^{T-s}c(m,3,r_1)c(T-s-m,3,r_2)\right)z^s(1-z)^f\mathrm{d}\Upsilon(z)}$$

where $r_1(s)=2|H|-3\sigma(H)+\text{mod}(T,3)+3s-2(T-1)$, $r_2=a-b+2$ and $\delta\leq 0.72$.

7.3 Proofs of the Main Results[16]

Proof of Lemma 1:

Step 1. we first show that $F(\cdot|\varnothing,\text{go})\succeq_{\text{lr}}G(\varnothing)$.

For simplicity we show the case when the prior has continuous density $f^0(p)$ and the analyst knows the pre-history length T. The arguments can easily be applied to general cases. Denote the pdf of $F(p|\varnothing,\text{go})$ as $f(p|\varnothing,\text{go})$,

$$f(p|\varnothing,\text{go})=\sum_{i=d}^{T}N(i,T-i|\Omega)p^i(1-p)^{T-i}f^0(z)/A \tag{11}$$

[15] The upperbound for the discount factor is from Banks and Sundaram (1992).

[16] The order of proofs are slightly different from the order of the propositions presented in the main text. Many results follow similar arguments. we organized the proofs in a way that the mathematical inferences flow more naturally.

where $A = \int_0^p \sum_{i=d}^T N(i, T{-}i|\Omega) z^i (1{-}z)^{T-i} \mathrm{d}F^0(z)$ and $(d, T{-}d)$ is on the lower boundary of Ω. Note that $1 = (1-z+z)^T = \sum_{i=0}^T \binom{T}{i} z^i (1-z)^{T-i}$ and $G(\varnothing) = F^0(p)$, thus we can rewrite the density of the naive posterior with null public history as

$$G(\varnothing) = \mu^0(p) = \sum_{i=0}^{T} \binom{T}{i} p^i (1-p)^{T-i} \mu^0(p)$$

Claim 2. Suppose a sequence of differentiable probability density functions $f_1, f_2 ... f_n$ satisfy $f_j'/f_j > f_i'/f_i$ for $\forall i < j$. Let $g^k = \sum_{i=1}^n a_i^k f_i / \sum_{i=1}^n a_i^k$, $k = 1, 2$. When $a_j^1/a_i^1 > a_j^2/a_i^2$ for all $j > i$, we have $g^2 \succeq_{\mathrm{lr}} g^1$. This follows easily from the fact that

$$\left(\frac{g^2}{g^1}\right)' = \frac{\sum_{i=1}^n a_i^1 f_i' \sum_{j=1}^n a_j^2 f_j - \sum_{i=1}^n a_i^1 f_i \sum_{j=1}^n a_j^2 f_j'}{(\sum_{j=1}^n a_j^2 f_i)^2}$$

Step 2. From Claim 2, we prove $F(H, \mathrm{go}) \succeq_{\mathrm{lr}} G(H)$ by showing that for $\forall j > i$,

$$N(j, T-i|\Omega)/N(i, T-i|\Omega) \geq \binom{T}{j} / \binom{T}{i} \tag{12}$$

Obviously when $0 \leq i < i' \leq b(T)$, the desired inequality holds as $N(i, T-i|\Omega) = 0$.

Note that

$$N(i, j|\mathbb{N}^2) = \binom{i+j}{i} \text{ and } \frac{N(i', T-i'|\mathbb{N}^2)}{N(i, T-i|\mathbb{N}^2)} = \binom{T}{i'} / \binom{T}{i}$$

We show (7.3) by showing that $N(j, T-j|\Omega)/N(i, T-i|\Omega)$ falls in the "size" of Ω:

Claim 3. If $\Omega' = \Omega \cup \{(a, b)\}$ for $\forall c \in \mathbb{N}$ and $i < i'$, we have

$$N(i', c-i'|\Omega')N(i, c-i|\Omega) \leq N(i, c-i|\Omega')N(i', c-i'|\Omega) \tag{13}$$

For non-triviality, assume $b \geq 1$ and (a, b) is on the lower boundary of Ω.

Firstly, as $N(a, b|\Omega) = 0 < N(a, b|\Omega')$ and $N(a{-}k, b{+}k|\Omega) = N(a{-}k, b{+}k|\Omega')$, for all $k \neq 0$, we have (13) for all $c = a + b$.

Next, since for arbitrary non-negative numbers A, B, C, D, E, F satisfying $AD \leq BC$ and $CF < DE$, we have $(A+C)(D+F) \leq (B+D)(C+E)$, we have (13) for $c = a{+}b{+}1$ by Claim 1.

Finally, by induction, (13) holds for $c = T$ as desired.

Step 3. $G(H)$ is the posterior with the Bayesian updating of the public history H based on $G(\varepsilon)$, let $\hat{F}(H, \mathrm{go})$ be the posterior with the Bayesian updating of the public history H based on $F(\varepsilon, \mathrm{go})$.

Since $F(\cdot|\varepsilon, \mathrm{go}) \succeq_{\mathrm{lr}} G(\cdot|\varepsilon)$, obviously $\hat{F}(\cdot|H, \mathrm{go}) \succeq_{\mathrm{lr}} G(\cdot|H)$

Obviously $\Omega_H \in \Omega_0$ and for some $d' > d$,

$$F(p|H, \text{go}) = \int_0^p \sum_{i=d'}^{T} N(i, j|\Omega_H) z^i (1-z)^{T-i} \mathrm{d}F^0(z)$$

Thus $F(\cdot|H, \text{go}) \succeq_{\text{lr}} \hat{F}(\cdot|H, \text{go}) \succeq_{\text{lr}} G(\cdot|H)$.

Proof of Proposition 3(a):

Step 1. For simplicity we show the case where the analyst has a prior with continuous density function $f^0(z)$ and the pre-observation history length is known to be T. The arguments can easily be applied to general cases.

When $H \succeq_1 \bar{H}$ or $H \succeq_1 \bar{H} + \langle 1 \ldots \rangle$ for some $\bar{H} \in B(\Gamma_H)$, every sequence in $\Gamma_\varnothing$ is a possible pre-history, namely $\Gamma_H = \{H_0 + H|\ H_0 \in \Gamma_\varnothing\}$. Thus we only need to show $F(\cdot|\varnothing, \text{go}; T+1) \sqsupseteq_1 F(\cdot|\varnothing, \text{go}; T)$

Following Step 1 of the proof for Lemma 1, the pdf of the sophisticated posterior is

$$\begin{aligned} f(p|\varepsilon, \text{go}; T) &= \sum_{i=b(T)}^{T} N(i, T-i|\Omega_\varnothing) p^i (1-p)^{T-i} f^0(p)/A & (14) \\ &= \sum_{i=b(T)}^{T} N(i, T-i|\Omega_\varnothing) p^i (1-p)^{T-i} (p+1-p) f^0(p)/A & (15) \\ &= \sum_{i=b(T)}^{T+1} \left(N(i, T-i|\Omega_\varnothing) + N(i-1, T-i+1|\Omega_\varnothing)\right) p^i (1-p)^{T+1-i} f^0(p)/A \end{aligned}$$

On the other hand,

$$f(p|\varepsilon, \text{go}; T+1) = \sum_{i=b(T+1)}^{T+1} N(i, T+1-i|\Omega_\varnothing) p^i (1-p)^{T+1-i} f^0(p)/A$$

From the quasi-upper-triangular shape of Ω, either $\bar{i}(T+1) = \bar{i}(T)+1$ or $\bar{i}(T+1) = \bar{i}(T)$.

Step 2. Recall (1) that $N(i, T+1-i|\Omega) = N(i, T-i|\Omega) + N(i-1, T+1-i|\Omega)$. When $\bar{i}(T+1) = \bar{i}(T)$, we have $f(p|\varepsilon, \text{go}; T+1) = f(p|\varepsilon, \text{go}; T)$ for any p and thus $F(\varepsilon, \text{go}; T+1) = F(\varepsilon, \text{go}; T)$. When $b(T+1) = b(T)+1$, by Claim 2, $F(\cdot|\varnothing, \text{go}; T+1) \sqsupseteq_1 F(\cdot|\varnothing, \text{go}; T)$. Thus when $H \succeq_1 \bar{H}$ or $H \succeq_1 \bar{H} + \langle 1 \ldots \rangle$ for some $\bar{H} \in B(\Gamma_H)$, the sophisticated posterior $F(\cdot|\varnothing, \text{go}; T)$ rises in T in the likelihood ratio order. Still by Claim 2, $F(\cdot|\varnothing, \text{go}; \gamma)$ rises in γ.

Renarks. The condition for this result important. When the public history contain early failures, Γ_H is only a sub set of $\{H_0 + H|\ H_0 \in \Gamma_\varnothing\}$. From the arguments in

Proposition 1, pre-histories in $\{H_0+H|\ H_0 \in \Gamma_\varnothing\}$ with most failures and fewest successes will be excluded. Treat the pre-histories in $\{H_0+H|\ H_0 \in \Gamma_\varnothing\}$ as a benchmark where the posterior at period 0 rises in t. With each t, the posterior conditional on Γ_H is higher than that conditional on $\{H_0+H|\ H_0 \in \Gamma_\varnothing\}$. Since longer pre-history implies more combinations of possible pre-histories. When the pre-history is short, the exclusion of several bad pre-histories makes the posterior much higher, but when the history length is long, this effect is not so strong. Thus increasing of pre-history length doesn't necessarily increase the posterior.

Formally, when $H \succeq_1 \bar{H}$ or $H \succeq_1 \bar{H} + \langle 1 \ldots \rangle$ for some $\bar{H} \in B(\Gamma_H)$, we can express (15) alternatively as

$$f(p|\varepsilon, \text{go}; T) = \sum_{(i,j)\in\Omega_\varnothing} N(i,j|\Omega_\varnothing)p^i(1-p)^j f^0(p)/A$$

However when the condition does not hold,

$$f(p|\varepsilon, \text{go}; T) = \sum_{(i,j)\in\Omega_H} N(i,j|\Omega_\varnothing)p^i(1-p)^j f^0(p)/A$$

Note the difference in the set for summation. The arguments in Step 2 fails.

Proof of Proposition 3(b):

Step 1. From Kelly et al. (1981), for any $k > 0$, there exists a sufficiently large $\bar{\delta}$ such that when $\delta > \bar{\delta}$, $I(a+k, b+1) < I(a,b)$. Hence for any Υ and v, there exists a sufficiently large $\bar{\delta}$ such that when $\delta > \bar{\delta}$, if $I(a,b) < v$, we have $I(a+k_1, b+k_2) < v$ and thus we can arrange the boundary set of any Γ_H as $\{H_1, H_2 \ldots\}$ such that $\sigma(H_j) \geq \sigma(H_i)$ and $\phi(H_j) \geq \phi(H_i)$ when $j > i$ and $\mu(\sigma(H_j), \phi(H_j)) > \mu(\sigma(H_i), \phi(H_i))$.

Claim 4. If $\mu(\sigma,\phi) \leq (\sigma', \phi')$ for $\sigma' > \sigma,\ \phi' > \phi$, then the posterior distribution $F_B \succeq_2 F_A$ where F_i has pdf $f_i(p) = p^\sigma(1-p)^\phi / \int_0^1 q^\sigma(1-q)^\phi \mathrm{d}\Upsilon(q)$ for $i = 1, B$.

Proof. When $p \in (0,1)$,

$$f_1(p)/f_2(p) = \left(\int_0^1 q^{\sigma'}(1-q)^{\phi'}\mathrm{d}\Upsilon(q) / \int_0^1 q^{\sigma}(1-q)^{\phi}\mathrm{d}\Upsilon(q)\right) p^{\sigma'-\sigma}(1-p)^{\phi'-\phi}$$

this is a bump shaped function and thus will cross any horizontal line at most twice. Since $\int_0^1 f_1(p)\mathrm{d}p = \int_0^1 f_2(p)\mathrm{d}p = 1$ and $f_1(0) = f_2(0) = f_1(1) = f_2(1) = 0$, f_1 will cross f_2 twice in $p \in (0,1)$ first from below at p_1 and then from above at p_2.

Let $K(q,p) = \begin{cases} 1 & q \leq p \\ 0 & q > p \end{cases}$. K is totally positive in degree 3.

$F_1(p) - F_2(p) = \int_0^p (f_1(q) - f_2(q))\,\mathrm{d}q = \int_0^1 K(q,p)\,(f_1(q) - f_2(q))\,\mathrm{d}q$. Since $f_1 - f_2$ have two sign changes in $(0,1)$, so does $F_1(p) - F_2(p)$ by Choi and Smith (2017).

Suppose F_1 and F_2 do have two sign changes, again by Choi and Smith (2017), F_2 will cross F_1 first from below and then above at p_1' and p_2' respectively. Then $F_2(p) < F_1(p)$ in $p \in (p_2', 1)$. Since $F_1(1) = F_2(1) = 1$, there must $\exists p_3 < 1$ s.t. when $p \geq p_3$, $f_1(p) < f_2(p)$. However this is not possible as $f_1(p) \geq f_2(p)$ when $p \in [p_2, 1]$. Thus F_2 can only cross F_1 once at some $\hat{p} \in (0,1)$ from below. Then either $F_2 \succeq_2 F_1$ or $F_1 \succ_{\text{convex}} F_2$ holds depending on

$$\int_{p=0}^{\hat{p}} (f_1(p) - f_2(p))\,\mathrm{d}p \geq \int_{p=\hat{p}}^{1} (f_2(p) - f_1(p))\,\mathrm{d}p$$

or

$$\int_{p=0}^{\hat{p}} (f_1(p) - f_2(p))\,\mathrm{d}p < \int_{p=\hat{p}}^{1} (f_2(p) - f_1(p))\,\mathrm{d}p$$

respectively. However $F_1 \succ_{\text{convex}} F_2$ would imply $\mu(\sigma,\phi) > \mu(\sigma',\phi')$, contradicting our assumption, thus $F_2 \succeq_2 F_1$.

Step 2. By Step 1 and Claim 2, when $\delta > \bar{\delta}$, we have $G(\cdot|H_j + H) \succ_2 G(\cdot|H_i + H)$ for any $j > i$ and $H_i, H_j \in \Gamma_H$. Thus $F(\cdot|H, \text{stop}, \gamma_2) \succ_2 F(\cdot|H, \text{stop}, \gamma_1)$ for $\gamma_2 > \gamma_1$.

Proof of Proposition 1:

We first show this result fixing the pre-history length, namely

$$H' \sqsupseteq_1 H \Rightarrow F(\cdot|H', \text{go}; T) \succeq_2 F(\cdot|H, \text{go}; T)$$

From the semi-upper-triangular shape of Ω, obviously $\Omega_H \subset \Omega_{H'} \cup B(\Gamma_{H'})$. Let Ω_H^T denote the set of states corresponding to the histories in the continuation set with length T, then $\Omega_H \cup \{(i,j)|i+j = T\}$. Either $\Omega_H^T = \Omega_{H'}^T$ or $\Omega_{H'}^T = \Omega_{H'}^T \backslash \{(b(T), T - b(T))\}$. In the second case, $\Omega_{H'}^T$ can be derived from Ω_H^T by exclude the state with the most failures.

From (9),

$$F(p|H, \text{go}; T) = \frac{\int_0^p \sum_{i=b(T)}^T z^{i+\sigma(H)}(1-z)^{T-i+\phi(H)}\gamma^T \mathrm{d}z}{\int_0^1 \sum_{i=b(T)}^T z^{i+\sigma(H)}(1-z)^{T-i+\phi(H)}\gamma^T \mathrm{d}z}$$

and

$$F(p|H', \text{go}; T) = \frac{\int_0^p \sum_{i=\tilde{b}(T)}^T z^{i+\sigma(H)}(1-z)^{T-i+\phi(H)}\gamma^T \mathrm{d}z}{\int_0^1 \sum_{i=\tilde{b}(T)}^T z^{i+\sigma(H)}(1-z)^{T-i+\phi(H)}\gamma^T \mathrm{d}z}$$

where $\tilde{b}(T) = b(T)$ or $b(T) + 1$.

By Claim 2, $F(p|H, \text{go}; T) \succeq_{\text{lr}} F(p|H', \text{go}; T)$ for any T.

Next, we consider states in the set $\Omega_H^T \backslash \Omega_{H'}^T$. From similar arguments as in the proof of Proposition 3(b), the cdf function

$$\tilde{F}(p) = \frac{\int_0^p \sum_{H_0 \in \Omega_H^T \backslash \Omega_{H'}^T} z^{\sigma(H_0)+\sigma(H)}(1-z)^{\phi(H_0)+\phi(H)}\gamma^{\sigma(H_0)+\phi(H_0)}\mathrm{d}z}{\int_0^1 \sum_{H_0 \in \Omega_H^T \backslash \Omega_{H'}^T} z^{\sigma(H_0)+\sigma(H)}(1-z)^{\phi(H_0)+\phi(H)}\gamma^{\sigma(H_0)+\phi(H_0)}\mathrm{d}}$$

single crosses $G(p|H)$ at some point $\tilde{p}$ while $E[p|\tilde{F}] \leq E[p|G(\cdot|H)]$. We have

$$\int_{\tilde{p}}^1 \Big(\tilde{F}(z) - G(z|H)\Big)\,\mathrm{d}z \geq \int_0^{\tilde{p}} \Big(G(z|H) - \tilde{F}(z)\Big)\,\mathrm{d}z$$

Since $F(\cdot|H,\text{go}) \succeq_{\text{lr}} G(\cdot|H)$, thus $F(\cdot|H,\text{go}) \succeq_1 G(\cdot|H)$. Therefore there exists some $\hat{p} < \tilde{p}$ such that $\int_{\hat{p}}^1 \Big(\tilde{F}(z) - F(z|H,\text{go})\Big)\,\mathrm{d}z \geq \int_0^{\hat{p}} \Big(F(z|H,\text{go}) - \tilde{F}(z)\Big)\,\mathrm{d}z$, namely $F(\cdot|H,\text{go}) \succeq_2 \tilde{F}(\cdot)$. Since $F(p|H,\text{go})$ can be regarded as the average from $F(p|H',\text{go})$ and $\tilde{F}(p)$, we must have $F(\cdot|H',\text{go}) \succeq_2 F(\cdot|H',\text{go})$.

Proof of Proposition 2:

(a) When $H' = H + \langle 1 \rangle$, it is easy since $\Gamma_H = \Gamma_{H'}$. For the other cases, we first show this result fixing the total history length, namely

$$F(\cdot|H',\text{go};T) \succeq_{\text{lr}} F(\cdot|H,\text{go};T+1) \text{ if } H' = \langle 1 \rangle + H \text{ or } H' = H + \langle 1 \rangle$$

and

$$F(\cdot|H',\text{go};T+1) \succeq_{\text{lr}} F(\cdot|H,\text{go};T) \text{ if } H = \langle 0 \rangle + H' \text{ or } H = H' + \langle 0 \rangle$$

From Claim 2, we will get the desired result by showing that the weight of the good state is higher(lower) when we concatenate 1 (0) to a public history, namely

$$N(a+1,b|\Omega)/N(a,b|\Omega) < N(a,b+1|\Omega)/N(a-1,b+1|\Omega)$$

and

$$N(a,b+1|\Omega)/N(a,b|\Omega) > N(a,b+2|\Omega)/N(a-1,b+1|\Omega)$$

By Claim 1, in both cases it suffices to show that for $\forall a, b$ and Ω,

$$N(a,b|\Omega)^2 \leq N(a-1,b+1|\Omega)N(a+1,b-2|\Omega) \tag{16}$$

Step 1, Let $n_1, n_2, \cdots n_k$ be a sequence of number such that $n_i n_j \geq n_{i-t}n_{j+t}$ for $\forall j \geq i$ and $t \leq \min\{i, k-j\}$. compose $m_1, m_2 \cdots m_{k-1}$ with some number K such that $m_i = n_i + n_{i+1}$ for $i > K$ and $m_i = 0$ for $i \leq K$. now we show $m_i m_j \geq m_{i-t}m_{j+t}$ for $\forall j \geq i$ and $t \leq \min\{i, k-1-j\}$.

If $i \leq K$, then $i - t \leq K$ and thus $m_i m_j \geq m_{i-t}m_{j+t}$ for $\forall j \geq i$ and $t \leq \min\{i, k-j\}$;

If $i > K$, then for $\forall j \geq i$ and $t \leq \min\{i, k-j\}$,

$$\begin{aligned} m_i m_j &= (n_i + n_{i+1})(n_j + n_{j+1}) = n_i n_j + n_{i+1} n_j + n_i n_{j+1} + n_{i+1} n_{j+1} \\ &\geq n_{i-t} n_{j+t} + n_{i-t} n_{j+t+1} + n_{i-t+1} n_{j+t} + n_{i+1-t} n_{j+1+t} \\ &= (n_{i-t} + n_{i-t+1})(n_{j+t} + n_{j+t+1}) \geq m_{i-t} m_{j+t} \end{aligned}$$

Step 2, for arbitrary a, b, Define

$$\Omega' = \bigcup_{T \leq a+b+1} \bigcup_{t \leq T} \{(a-t, b+t)\}$$

If $(i,j) \in \Omega$, Let $\tilde{N}_{i,j} = N(i,j|\Omega)$ and $N_{i,j} = N_{i-1,j} + N_{i,j-1}$; if $(i,j) \in \Omega' \backslash \Omega$, let $\tilde{N}_{i,j} = 0$. By Step 1, let $n_i = \tilde{N}_{-b-1+i,b+1-i}$ and $k = a+b+1$, since $N_{00} = 1$ and $N_{ij} = 0$ for $i \neq 0$ and $j \neq 0$, we have $n_i n_j \geq n_{i-t} n_{j+t}$ for $\forall j \geq i$ and $t \leq \min\{i, k-j\}$. By induction, (16) holds for $\forall a, b$ and Ω.

(b) From proposition 3(b), we will ge the desired result by showing that the weight of states corresponding to longer histories are higher when the public history contain more failures or less successes. By the expression (10), this means for any (i,j), $(i',j') \in \Omega_H^B$ where $i \leq i'$, $j \leq j'$.

$$\frac{N(i'+1-\sigma(H), j'-\phi(H)|\Omega)}{N(i+1-\sigma(H), j-\phi(H)|\Omega)} > \frac{N(i'-\sigma(H), j'-\phi(H)|\Omega)}{N(i-\sigma(H), j-\phi(H)|\Omega)} \tag{17}$$

and

$$\frac{N(i'-\sigma(H), j'-\phi(H)+1|\Omega)}{N(i-\sigma(H), j-\phi(H)+1|\Omega)} < \frac{N(i'-\sigma(H), j'-\phi(H)|\Omega)}{N(i-\sigma(H), j-\phi(H)|\Omega)} \tag{18}$$

Step 1. We first show the case when the experimenter is impatient, the cost $v = k/(k+1)$ for some positive integer k and the prior is Beta$(k \cdot n, n)$ for any positive integer n. In these cases, the continuation set of states $\Omega_0 = \bigcup_{t=1,2,3\ldots} \{(i,j)|i \geq k \cdot j\}$ and $\Omega_H = \bigcup_{t=1,2,3\ldots} \{(i,j)|i + \sigma(H) \geq k(j + \phi(H))\}$. By Aval (2008). $N(i,j|\Omega_0)$ can be expressed as the jth fuss-catalan number with parameter $\rho = k+1$ and $r = i - k \cdot j + 1$, namely

$$N(i,j|\Omega_0) = c(j, k+1, i) = \frac{i - k \cdot j + 1}{j+i+1} \binom{j+i+1}{j} \tag{19}$$

Now we show (17) for any $i' = i + a$, $j' = j+1$ and $i \geq k \cdot j$. where[17] $a \leq k$.

The left hand side of (17) is

$$\frac{N(i+1, j|\Omega_0)}{N(i,j|\Omega_0)} = \frac{j+i+2}{i+2} \cdot \frac{i - k \cdot j + 2}{i - k \cdot j + 1} \cdot \frac{j+i+1}{j+i+2} \tag{20}$$

[17]For the special example we consider in this step we only need to show the case $a = k$, but we show the general case with $a \leq k$ for step 2 where the boundary set is not so regular.

The right hand side of (17) is

$$\frac{N(i+1+k,j+1,|\Omega_0)}{N(i+k,j+1|\Omega_0)} = \frac{j+i+a+3}{i+a+2}\cdot\frac{i+a-k-k\cdot j+3}{i+a-k-k\cdot j+2}\cdot\frac{j+i+a+2}{j+i+a+3} \tag{21}$$

Obviously the third term in (20) is smaller than (21). Then let's consider the second term. Since $i \geq k\cdot j \geq a\cdot j$, $(j+i+2)/(i+2) < (a+1)/a$ and thus

$$\frac{j+i+2}{i+2} < \frac{j+i+2+a+1}{i+2+a} = \frac{j+i+a+3}{i+a+2}$$

Similarly the first term in (20) is smaller than (21). Thus (17) holds.

Step 2. In the general case, there exists some $0 < k_1 < k_2$, so that the continuation set

$$\underset{t=1,2,3\dots}{\cup}\{(i,j)|i \geq k_2\cdot j\} \subset \Omega_0 \subset \underset{t=1,2,3\dots}{\cup}\{(i,j)|i \geq k_1\cdot j\}$$

Fixing j, when i is sufficiently large, the four terms in (17) approaches (19) for some equal $k \in (K_1, K_2)$. Since tha ratio of (20) to (21) is strictly less than 1, (17) holds. To be more specific, there exists $\bar{i}(i,j)$ such that (17) holds when $i > \bar{i}(i,j)$. Let $\bar{\phi} = \sup\{\arg\min_{\phi}(\bar{i}(i,j), j+\phi) \notin \Omega_0\}$ and then when $\phi(H) > \bar{\phi}$, for each $(i,j) \in \Omega_H$, $i > \bar{i}(i,j)$ and (17) holds.

Proof of Proposition 4:

We first show that $\Omega(\Upsilon,\delta,v) \subseteq \Omega(\Upsilon',\delta',v')$ if $\Upsilon' \succeq_1 \Upsilon$, $\delta' \geq \delta_2$ and $v' \leq v$.

Since $\Omega(\Upsilon,\delta,v) = \{(i,j) : I(i,j|\Upsilon,\delta,) \geq v\}$, thus the fall of v and increase of Υ and δ expands Ω. We only need to show that the $I(i,j|\Upsilon',\delta') \geq I(s,f|\Upsilon,\delta)$ for$\forall i,j$.

Recall the specification of the Gittins index from (6), we expand the notation of the value function to $W(\gamma|\sigma,\phi;\Upsilon,\delta)$. since W rises in δ for all γ, the fixed point $I(\sigma,\phi|\Upsilon,\delta)$ also rises in δ by Milgrom and Shannon (1994).

On the other hand, when $\Upsilon' \sqsupseteq_1 \Upsilon$, we have $\mu(i,j|\Upsilon') \geq \mu(i,j|\Upsilon)$. Thus $W(\gamma|\sigma,\phi;\Upsilon',\delta) \geq W(\gamma|\sigma,\phi;\Upsilon,\delta)$ for all γ, then $I(\sigma,\phi|\Upsilon',\delta) \geq I(\sigma,\phi|\Upsilon,\delta)$ for $\forall\sigma,\phi$ and δ.

The rest of the proof follows from Claim 2 and the same argument in the proof of Proposition 1.

References

Aval, J.-C. (2008): "Multivariate Fuss–Catalan numbers," *Discrete Mathematics*, 308(20), 4660–4669. 7.3

Banks, J. S., and R. K. Sundaram (1992): "A class of bandit problems yielding myopic optimal strategies," *Journal of Applied Probability*, 29(3), 625–632. 15

Bellman, R. (1956): "A problem in the sequential design of experiments," *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 16(3/4), 221–229. 13

Bolton, P., and C. Harris (1999): "Strategic experimentation," *Econometrica*, 67(2), 349–374. 1

Bubeck, S., N. Cesa-Bianchi, et al. (2012): "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Foundations and Trends® in Machine Learning*, 5(1), 1–122. 11

Choi, M., and L. Smith (2017): "Ordinal aggregation results via Karlin's variation diminishing property," *Journal of Economic Theory*, 168, 1–11. 7.3

Gittins, J., K. Glazebrook, and R. Weber (2011): *Multi-armed bandit allocation indices*. John Wiley & Sons. 12

Gittins, J. C. (1979): "Bandit processes and dynamic allocation indices," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 148–177. 1, 2, 3

Gomes, R., D. Gottlieb, and L. Maestri (2016): "Experimentation and project selection: Screening and learning," *Games and Economic Behavior*, 96, 145–169. 1

Huelsenbeck, J. P., and K. A. Crandall (1997): "Phylogeny estimation and hypothesis testing using maximum likelihood," *Annual Review of Ecology and Systematics*, 28(1), 437–466. 7.1

Kelly, F., et al. (1981): "Multi-armed bandits with discount factor near one: The Bernoulli case," *The Annals of Statistics*, 9(5), 987–1001. 7.3

Milgrom, P., and C. Shannon (1994): "Monotone comparative statics," *Econometrica: Journal of the Econometric Society*, pp. 157–180. 7.3

Moscarini, G., and L. Smith (2001): "The Optimal Level of Experimentation," *Econometrica*, 69(6), 1629–1644. 2, 7.1

Rosenberg, D., E. Solan, and N. Vieille (2007): "Social Learning in One-Arm Bandit Problems," *Econometrica*, 75(6), 1591–1611. 1

Smith, L., P. Sørensen, and J. Tian (2012): "Informational herding, optimal experimentation, and contrarianism," *Revised and resubmitted for Review of Economic Studies*. 1

Wald, A. (1947): "Foundations of a general theory of sequential decision functions," *Econometrica, Journal of the Econometric Society*, pp. 279–313. 1