The London School of Economics and Political Science

Essays in Information Economics

Evgeniya Kudinova

A thesis submitted to the Department of Economics
for the degree of Doctor of Philosophy

July 2024

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of approximately 29000 words.

Statement of inclusion of previous work

I confirm that Chapter 3 of this thesis is a revised version of the paper I submitted at the end of my Master of Research degree at the LSE in 2019.

Acknowledgements

This thesis is an outcome of a very long and intense journey, and I could not have managed to complete it without the help and support of many individuals.

First and foremost, I would never be able to express enough gratitude to Gilat Levy. She is simply the best supervisor I could ever wish for. Academically, I benefited infinitely many times from her creative, insightful and always deeply profound way of thinking about the economic models, and her almost miraculous ability to provide a speedy relief on the multiple occasions I felt completely stuck in my research path. On a more personal level, I gained from her incredibly wise, thoughtful and empathetic approach to supervising students, balancing between the research and life outside, and being a female in academia more generally. Thank you for always finding the right words to encourage and inspire me, especially when it is needed the most. I wish I become a supervisor and a mentor like you some day.

I am also indebted to my advisors Ronny Razin and Stephane Wolton for the many insightful discussions we shared and for always being available and ready to help. I am grateful to Ronny for being supportive at the earlier stages of my PhD journey and helping me find the way to the topics I enjoy. I also once learned your perspective of treating models like toys that theorists like to play with, which had a huge impact on how I treat them today. I thank Stephane for showing me what the true passion towards research means and how it spreads the waves of inspiration on everyone around. I believe I only truly learned to think critically because of you. Teaching EC453 alongside you is one of my favourite PhD experiences, so thank you for making this possible, as well as for making me feel welcomed in the Government department.

I am deeply grateful to the theory group at the LSE and those affiliated with it, Andrew Ellis, Matt Levy, Francesco Nava, Christopher Sandmann, Balazs Szentes, Inés Moreno de Barreda and Leeat Yariv, for many helpful comments and always challenging me to expand my knowledge and think out of the box. I also thank the rest of the faculty, especially Ethan Ilzetzki, Dimitra Petropoulou and Alwyn Young, for providing different perspectives and broadening my understanding of economics beyond micro-theoretical modeling. Thank you to my undergraduate supervisor Sergey Stepanov for sparking my interest towards economic research and making me believe that I am actually capable of doing it.

I wish to thank all the fellow, former and current, PhD colleagues, especially Thomas

Brzustowski, Alastair Langtry, Hongting Leng, Clement Minaudier, Chiara Sotis, Heidi Thysen, Lorenzo Vicari, Cecilia Wood and Alfred Zhang - you made my PhD journey brighter and easier through listening to my research aspirations and PhD life complaints, as well as being responsible for some of the good entertainments I had during the PhD years. I thank my friends outside academia, Kolya, Tanya, Anechka and Fuad, Nastya, Olya, Anya, Karina, Katya, Katya, Misha, Ksusha and many others - for not forgetting who I am and how I look after all the distance and years I keep disappearing and drowning in research - you are never forgotten, and I cherish our occasional encounters.

Lastly, but extremely importantly, I am blessed to have a supporting family by my side. I thank my sister Masha, my grandparents and especially my incredible parents, Ira and Misha: I am here, being the person I am, submitting my PhD thesis and starting the new chapter in my life, largely because of you and for you. For as long as I can remember myself, you did everything to ensure that I am happy and gain the best experiences. You were and are there for me whenever I stumble, always offering your warmth and wisdom. You are my role models and the reason I am excited and able to pursue the career in academia. Thank you for the unconditional love, support and encouragement you give me - I love you endlessly. I am also blessed to have met my partner Vadim, my found family. Thank you for the warmth, comfort and serenity you share and for always being there for me throughout this journey no matter the distance between us. Thank you for understanding me without words, for keeping me sane, for being my safe space and home. You make my life better every day, and I would not be here without you.

Abstract

This thesis consists of three essays in information economics. The first two chapters examine how economic agents adapt to risky new opportunities when they can invest in increasing the likelihood of a successful outcome, while also learning about its quality (i.e. experimenting). In the first chapter, I build on a single risky arm Poisson bandit environment with conclusive breakthroughs and explore how the ability to endogenously change the arm, by investing, affects experimentation (I assume that successful investment turns a bad arm into a good one). I find that the agent may behave according to one of the two regimes. She either acts non-monotonically with purely experimenting before and after the investment stage and eventually abandons the risky arm; or, due to converging to an interior belief, she never quits investing and experimenting until the risky option generates a success.

The second chapter studies a similar setting when the information arrives as conclusive breakdowns instead of breakthroughs. Despite the changed information structure, the agent still finds herself in one of the two regimes, but the mechanism behind those becomes very different: the ability to invest may make the agent willing to experiment and invest even after a breakdown, so she may never give up on the risky arm and remain in cycles of pure experimentation and investment stages.

The third chapter explores the idea that voters may communicate their protest against the government's silence on certain socially important issues (e.g. climate change) through elections. I use the common value elections framework to represent the traditional policies dimension and introduce an extra dimension that refers to the previously excluded policies. The voters are motivated by selecting a better fit along the salient dimension and signaling their views towards the previously excluded one. I show that the voters may signal their protesting views by abstaining from elections at a cost of increasing a risk of a wrong candidate selection, and identify conditions under which the citizens can achieve fully successful communication in large elections.

Contents

List of Figures

Chapter 1

Exploration, Exploitation, Amelioration: Experimentation with Endogenously Changing Arms

1.1 Introduction

Experimentation, e.g., with a new technology, is an inevitable step towards achieving progress. Yet, it involves a fundamental risk as the new technology may potentially turn out to be worse than the current one. Despite this opportunity cost, the possibility to learn that the new technology is indeed beneficial serves as a strong incentive to start and prolong experimentation at least for a while.

Due to the importance of experimentation for economic progress and the strategic nature of the problem, a large literature has analysed the optimal decision of experimenters. One aspect that the traditional models typically disregard, however, is the potential ability of decision makers to affect the outcome of experimentation instead of just learn about it. Indeed, in many applications the agents may actively engage in adapting to the new risky opportunities, and thus troubleshoot any issues on the way to ensure that the outcome is a success. A firm adopting a new technology may assign a team to monitor how it rolls out in the specific company environment and tailor it to become a better fit. Governments launching a new policy typically employ bureaucrats to achieve successful implementation.

A worker, switching to a new occupation and realizing that she may lack some necessary skills, may attend advanced training courses, while a team leader hiring a new trainee of uncertain quality can invest in training and supervising the apprentice. These examples raise the question of when investing in an adaptation process - an investment that is potentially costly in terms of time, effort or physical resources - is worthy to the agent. In this paper I show that being able to make such investments has a meaningful impact on experimentation and learning, stemming beyond the direct benefits thereof.

I analyse experimentation with the ability to actively change the risky option (I will refer to this process as training). Specifically, I augment the canonical Poisson bandit model, as formulated by Keller et al. (2005) (henceforth KRC). There is a single decision maker (or agent) who at each instance allocates her time between exploiting a safe arm that yields a fixed known payoff and experimenting using a risky one. The risky arm can be good or bad, where the good one achieves an observable breakthrough (also referred to as good news) at a Poisson arrival rate and is preferred to the safe one, while the bad one can never provide any positive payoff. On top of this traditional setup, I assume that, conditionally on using it, the agent can choose to endogenously change the risky arm: By paying some cost, she can turn a bad risky arm into a good one at some Poisson arrival rate; that is, if the agent invests in training, there is some probability that the bad risky arm becomes good, while a good arm remains unaffected. I further assume that once the change has occurred it is irreversible. Importantly, the event that training was successful is not directly observable to the agent; she can learn only through observing the breakthrough. [1]

The ability to invest in training the risky arm leads to significant qualitative changes compared to the case in which this is not possible. Specifically, training affects the evolution of beliefs about the probability that the risky arm is good. While the traditional Poisson bandits with good news imply that the agent necessarily gets more pessimistic in the absence of news arrival reading it as a signal of a bad arm, here the beliefs may evolve in a non-monotone manner. In fact, the positive effect of investing in training may counteract and even exceed the traditional negative effect on beliefs from receiving no news. I show that if the agent invests in improvements, she may become more optimistic even in the absence of news under relatively low priors, or still more pessimistic under high ones.

[1] I am considering the ability to improve the arm only conditional on using it. Such assumption fits the motivating examples and the story of investments in adaptation, when the agents learn more once they start the risky endeavor and can troubleshoot on the way. This is in contrast to making investments before the experimentation starts, which would be reflected through adjusting the prior belief on the arm being good, but become strictly dominated once experimentation starts, hence not affecting the analysis of this paper.

Such non-monotonicity in beliefs has a large impact on the agent's optimal strategies and the outcomes she can achieve.

My key result, derived from the non-monotonicity of belief evolution described above, is that the absence of good news does not necessarily lead to quitting experimentation. This is a sharp contrast to the standard models where the agent surely gives up on experimenting upon getting no news for a while, as then her beliefs gradually drop until they reach a stopping cutoff. In turn, when the ability to train guarantees a sufficiently high likelihood to improve the arm, once the agent starts training the arm, she will keep doing so 'forever' at least at some intensity even if no news arrives. This arises when the positive effect of improving the risky arm dominates the negative one from getting no news around the stopping cutoff (below which experimenting is no longer worthwhile), making the agent more optimistic irrespective of the informational flow. As such, the agent's belief diverges away from the cutoff deeper into the experimentation region, and the agent never quits experimenting: Her belief converges to the one where the two effects balance each other out, and the agent gets stuck training the arm until she finally observes the good news and resolves the uncertainty.

I also show that the overall optimal strategies structure may be non-monotone in actions. I find that the agent prefers to train the arm for some moderate beliefs, and may sometimes find it optimal to purely exploit the risky arm being both more optimistic and more pessimistic than in the training range. For lower productivity of training this leads to non-monotone actions path. Conditional on observing no news, the agent may first (for high enough priors) purely experiment for a bit, then train the risky arm as her beliefs go down and then again just experiment before switching to the safe option. This occurs because, as she anticipates quitting the risky arm soon, she no longer values its improvement enough to be willing to incur any extra cost.

For more efficient training, the non-monotone actions structure creates a sharp discontinuity in the agent's behaviour and the outcome of experimentation. Specifically, there exists a cutoff belief such that, when being slightly more pessimistic, the agent may give the risky arm a chance before exiting experimentation, but the path will necessarily lead her to switching to a safe option conditional on getting no news. In contrast, starting slightly above such belief, she will only become more optimistic and eventually get stuck on training the arm at certain intensity until news arrive. This ensures that she obtains a breakthrough almost surely in the limit, as opposed to much less optimistic outcome in the former scenario.

Overall, my results highlight that the ability to train impacts experimentation beyond its direct effect on the bad arms. In fact, training extends the experimenting horizon for the agent, which increases the likelihood of discovering initially good arms and reduces the chance of erroneous abandonment of those - an event occurring with strictly positive probability in traditional experimentation frameworks. The strength of these effects varies in a discontinuous manner given the variation in the agent's optimal actions paths.

My findings suggest important implications for the prior design of the arms and training mechanisms to anyone pursuing the goal of spreading risky arms adoption, or innovation more generally. I show that a small change in the productivity of training can cause a discontinuous switch from the pessimistic regime, where risky arms may be abandoned after a while, to the optimistic one, with all risky arms resulting in a breakthrough. Similarly, a marginal improvement of the original technology (arm) may lead to discontinuously higher investments in training and a better spread of innovation in the long run.

The change in action dynamics due to training also leads to novel technical implications. Firstly, the resulting value function does not have to be smooth in contrast to the standard good news experimentation models. This is due to the existence of a cutoff that the agent diverges away from being both above and below it. The threshold belief is then defined purely by *continuous pasting* as opposed to the smooth pasting argument, following the terminology introduced by Keller and Rady (2015). [2] Secondly, the cutoff belief where the agent gets stuck being indifferent between experimenting with or without training is absorbing on both sides: A more pessimistic agent will train and increase her belief until she converges to the cutoff, while a more optimistic one will exploit the risky arm and become more pessimistic. Such fully absorbing cutoffs are novel in the experimentation literature, and I establish an approach for defining these types of boundaries, which requires an extra optimality condition on top of traditional value matching and smooth pasting ones.

The rest of the Chapter is organized as follows. The next section discusses related literature and emphasizes the contribution of this paper. In Section 1.3 I present the model, analyse the evolution of beliefs and the value function, and provide the benchmark solution when the agent is not allowed to train the risky arm. I then establish the main results and qualitative implications of those in Section 1.4, including the discussion of the impact of the experimentation outcomes divergence in the long run in Subsection

[2] Their paper documents a similar diverging behaviour at the stopping cutoff in the experimentation with breakdowns, where the beliefs move away from the cutoff on one side and remain constant forever on the other. There, the benefit of learning the arm's type around the cutoff is null given that it leads to switching to the safe arm, an intuition that is very different from the one I identify in this paper.

1.4.4. Section 1.5 concludes by providing a broader perspective on how the ability to train changes the learning from experimenting.

1.2 Related literature

This paper contributes to the rich literature on the experimentation with Poisson bandits, pioneered by KRC, and explored further in Keller and Rady; Keller and Rady (2010; 2015), Klein and Rady (2011), Guo (2016), Lizzeri et al. (2024) and others. In all above, the risky arms types are fixed and stable. There are papers that analyse *restless* bandits, where the arms are allowed to randomly change in time irrespective of whether being used, as in Whittle (1988) or Keller and Rady (1999). Safronov (2021) considers agents who can learn by doing, that is, increase the lump sum payoff of the risky arms via becoming experienced in exploiting a certain option. Fryer and Harms (2018) analyse the environment where the risky arm's payoff increases if in use and decreases otherwise. In both of these papers, as in the classic restless arms models, the arms change exogenously once the agent decides to use them. In contrast, in the environment I study in this paper the agent actively chooses whether she wants the arm to change and this decision is separate from the exploitation one.

Quite a few papers consider experimentation subject to moral hazard. They typically focus on the principal-agent environments where the principal is providing incentives for the agent to exert effort for conducting experimentation. In such or similar setups, Bergemann and Hege; Bergemann and Hege (1998; 2005), Hörner and Samuelson (2013) and Hidir (2019) study various aspects of optimal contracting to incentivize experimentation by a single agent. Halac et al. (2016) introduce adverse selection on top of the moral hazard to the contracting problem. Halac et al. (2017) and Moroni (2022) extend the optimal incentives provision design further to contest-like environments where multiple agents compete in experimentation over the same tasks. Diverting from contractual design, Bonatti and Hörner (2011) analyse a model of team experimentation with binary type risky arm, where the arrival of breakthrough depends on the joint effort of the experimenting agents. In all these papers, if considered with a single agent, effort exertion transfers a safe arm into a risky one (determines the type of a good risky arm), or, in other words, gives rise to experimentation. This differs from the model discussed here, where improving the arm with training comes on top of experimentation, and the experimentation problem preserves for any training decision.

Finally, Fershtman and Pavan (2023) consider a decision maker who chooses between experimenting on some known arms and searching for the new ones at each instance. That is, they endogenize the set of arms available to the agent. An example of their framework can include discovering a new (and hence independent) exogenously changing arm in addition to the already available safe and fixed risky ones, and choosing one of the, now three, alternatives. Endogenizing the risky arm, as I propose in this paper, is very different from the three arms interpretation outlined above and thus is not nested there. It would require perfect correlation between the fixed and the restless risky arms, so any learning or pulling the changing arm affects them equally in both the belief and the underlying type.

1.3 Model

1.3.1 Baseline environment

Consider a single decision maker (agent) who, at each moment of time, chooses whether to use a safe or a risky arm. The risky arm is of a binary type $\lambda^\theta \in \{0, \lambda\}$, with $\lambda > 0$. If used in $[t, t+dt)$, such arm generates a payoff normalized to 1 with probability $\lambda^\theta dt$ conditional on its type. That is, the payoffs (breakthroughs) are generated at a Poisson arrival rate λ^θ. The arm's type is unobserved by the decision maker, and she has a prior that the arm is 'good' with some probability p_0, i.e. $\Pr(\lambda^\theta = \lambda) = p_0$. Alternatively, the agent can use a safe arm, which generates a constant payoff $s \in (0, \lambda)$, so that the good risky arm is preferred to the safe one, while the bad one is dominated.

Denote the share of time/effort allocated to risky arm in $[t, dt)$ as $\alpha_t \in [0,1]$ and the posterior belief about the risky arm's type conditional on the information available at time t as p_t, i.e. $\Pr(\lambda^\theta = \lambda | I_t) = p_t$. Then, the expected payoff generated in $[t, dt)$ is equal to $\big((1-\alpha_t)s + \alpha_t p_t \lambda\big)dt$, and the agent chooses $\alpha_t \in [0,1]_0^\infty$ to maximize the discounted present value of payoffs flow:

$$\max_{(\alpha_t \in [0,1])_0^\infty} E_{p_0}\left[\int_0^\infty e^{-rt}\big((1-\alpha_t)s + \alpha_t p_t \lambda\big)dt\right],$$

where $r > 0$ is a rate of exponential discounting.

1.3.2 Training/investing mechanism

I now add an additional ingredient to the model: suppose that the decision maker can also invest in improving the risky arm (that is, 'train') while exploiting it. Specifically, I assume that the training mechanism allows to turn a bad type of the arm into a good one (while it has no effect on the initially good arm). The bad arm becomes good at some Poisson arrival rate $\pi > 0$, i.e. with probability πdt if trained in $[t, t + dt)$. I refer to the event of the arm's type switch as 'successful training', and assume that this event is irreversible (that is, once the bad arm turns good, it remains good forever) and not directly observable (i.e., the agent cannot observe when the training succeeded directly, but only when the payoffs/breakthroughs are generated). The training mechanism is costly with a fixed cost $\kappa \geq 0$, which is payable no matter whether the training is successful or not; hence, if training occurs in $[t, t + dt)$, the cost κdt is paid.

I still denote by $\alpha_t \in [0, 1]$ the intensity of experimentation in $[t, t+dt)$ independently of whether training arises or not. However, the agent now also decides whether she wants to train the arm, and I introduce this decision through the intensity of training in $[t, t+dt)$, $\beta_t \in [0, 1]$. [3] Then, the decision maker's maximization problem becomes:

$$\max_{((\alpha_t,\beta_t)\in[0,1]^2)_0^\infty} E_{p_0}\left[\int_0^\infty e^{-rt}\left((1-\alpha_t)s + \alpha_t\left(p_t\lambda - \beta_t\kappa\right)\right)dt\right].$$

Note that the safe arm is used $1 - \alpha_t$ of the time, the risky arm (without training) $\alpha_t(1 - \beta_t)$, and the risky arm is trained and used $\alpha_t\beta_t$ share of the time. In addition, observe that the only direct effect of training on the payoffs is through its cost, while the benefit appears in the objective function only indirectly: It is built into the likelihood of the risky arm being good at time t, that is, p_t. [4]

1.3.3 Beliefs evolution

The way beliefs evolve depends on the intensities of experimentation and training, α_t, β_t. Under the exploitation of the safe arm, that is if $\alpha_t = 0$ in $[t, dt)$, there is no additional informational flow concerning the risky arm's type, meaning that beliefs remain unchanged,

[3] One way to think about β_t is as intensity of training in $[0, 1]$. Alternatively, one can imagine $\beta_t \in \{0, 1\}$. Such restricting assumption does not affect any of the results, just slightly changes the interpretation of some of them.

[4] Think about λ^θ being λ^θ_t instead. Indeed, with training, the risky arm's type is allowed to evolve endogenously rather than remain at λ^θ_0 forever under regular experimentation as in KRC. As such, under training, posterior belief $p_t \equiv \Pr(\lambda^\theta_t = \lambda)$ evolves because of the possible λ^θ_t evolution on top of the regular learning effect.

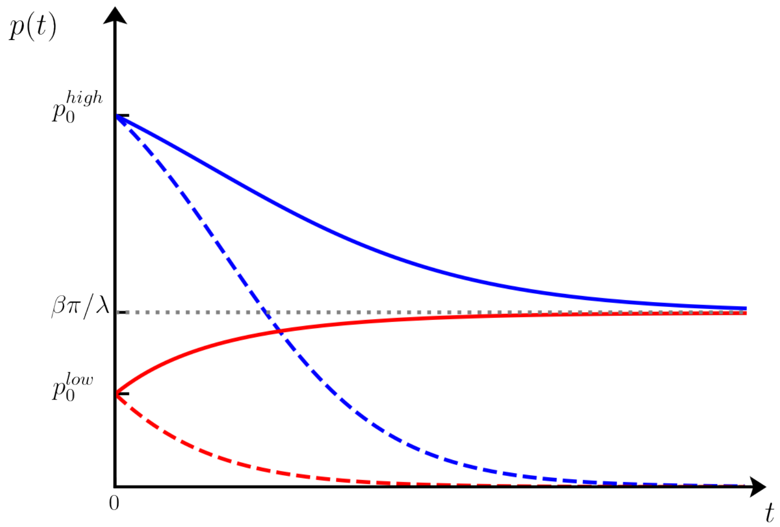

Figure 1.1: Evolution of beliefs conditional on receiving no news: for experimenting forever ($\alpha_t = 1, \beta_t = 0$) in dashed lines, and training forever at intensity $\beta_t = \beta > 0$ in solid ones; for high p_0 in blue, and low p_0 in red.

and so $p_{t+dt} = p_t$.

If experimentation occurs ($\alpha_t > 0$), the payoff flow from the risky arm signals the arm's type. Given the assumption of conclusive breakthroughs ($\lambda^\theta = 0$ for bad type), once a positive payoff is observed, the risky arm is certain to be good, that is belief discretely jumps upwards to $p_{t+dt} = 1$. No news instead serves as a negative 'signal' and leads to gradual belief updating downwards. In particular, following the classical Poisson bandits literature, $p_{t+dt} - p_t \equiv dp_t = -\alpha_t \lambda p_t(1-p_t)dt$. The belief is updated downwards more if the difference between the risky arm's types is greater ($\Delta\lambda^\theta = \lambda$ here), the intensity of experimentation is larger, and the more uncertain the agent is (for moderate beliefs p_t).

Now, on top of the standard learning effect, there is potentially another effect on belief evolution due to training. If the arm is good, training is redundant; if the arm is bad instead (with probability $1 - p_t$), there is a chance πdt that the training is successful and the arm type shifts to a good one. This creates a positive boost in the belief evolution of size $\beta_t \pi (1 - p_t)dt$, which counteracts the negative effect of getting no news. Note that this positive effect is stronger the stronger is the intensity of training.

The following lemma summarizes how beliefs change:

Lemma 1.1. *For any p_t and $\alpha_t > 0$, the risky arm is believed to be certainly good conditional on positive payoff of size 1 generated by it: $p_{t+dt}^{news} = 1$. Conditional on no payoff generated by the risky arm, beliefs follow: $dp_t = \alpha_t(\beta_t \pi - \lambda p_t)(1 - p_t)dt$.*

Notice that the ability to endogenously change the arm may result in non-monotone paths of p_t even conditional on no news observed, and the direction of belief evolution depends on the intensity of training β_t as well as the current belief p_t. Trivially, if the agent purely experiments with the risky arm ($\beta_t = 0$), having no news results in a gradual downward shift in beliefs, for any prior $p_0 \in (0,1)$ (as shown by dashed paths in the Figure 1.1). With training at some positive intensity (solid lines), the dynamics of belief varies depending on the prior optimism of an agent. If her belief p_0 is high ($p_0 > \beta\frac{\pi}{\lambda}$), she still becomes more pessimistic without the news, but training slows down the beliefs depreciation due to the boost of optimism it gives (plotted in blue), and the beliefs converge to $\beta\frac{\pi}{\lambda}$. In contrast, for relatively low beliefs $p_0 < \beta\frac{\pi}{\lambda}$ (red scenario), the informational effect of no news is weak enough that it is dominated by the positive influence of training. In fact, with continuous training the agent gets more optimistic, which gradually increases the informational content of getting no news and hence the strength of the negative effect on belief evolution up until the point where the two effects counteract each other exactly. At such belief ($p_t = \beta_t\frac{\pi}{\lambda}$) the agent gets 'stuck' for a while, and the question arises whether and when such a path can be optimal.

1.3.4 Bellman equation and the value function

In this subsection I derive the Hamilton-Jacobi-Bellman equation (HJB) that characterises the maximisation problem of interest, as well as the value function obtained from solving the equation. Note that all the extra dynamics occurring due to the ability to endogenously change the arm is represented through the belief about the arm's type, that is p_t. This means that a single state variable captures the entire dynamics of the problem, which allows for a tractable closed-form solution.

Denote by $V(p_t)$ the value function resulting from the optimal control problem. It consist of the monetary payoff obtained in $[t, t+dt)$ and the discounted continuation value of the problem. With the probability of observing good news from the risky arm, belief jumps to $p_{t+dt} = 1$ and otherwise changes according to the law of motion identified and discussed in section 1.3.3. Overall, it can be represented as:

$$V(p_t) = \max_{(\alpha_t,\beta_t)\in[0,1]^2} \big((1-\alpha_t)s + \alpha_t\,(p_t\lambda - \beta_t\kappa)\big)dt + e^{-rdt}\big(\alpha_t p_t\lambda dt V(1) + (1-\alpha_t p_t\lambda dt)V(p_t + dp_t)\big) \tag{1.1}$$

$$\text{s.t. } dp_t = \alpha_t(\beta_t\pi - \lambda p_t)(1-p_t)dt$$

Applying a first-order approximation on all the non-linear components, setting $(dt)^2$

terms to 0 and dividing the equation by dt (as well as dropping the time subscripts, since the time dimension is fully reflected through p_t) simplifies the problem further to:

$$rV(p) = \max_{(\alpha,\beta)\in[0,1]^2} \big((1-\alpha)s + \alpha\,(p\lambda - \beta\kappa)\big) + \alpha p\lambda\big(V(1) - V(p)\big) + V'(p)\alpha(\beta\pi - \lambda p)(1-p)$$

The equation shares a first-order differential-difference form as obtained in traditional Poisson bandit's models, and if the training option is shut down by setting $\beta = 0$, the equation is reduced to the KRC one. Even with $\beta > 0$, the interpretation of the HJB equation remains similar. The first part represents the current payoff obtained from picking a certain strategy. The latter parts describe the change in the value function due to the change in belief. If the good news is observed (at the rate $\alpha p\lambda$), there is a discrete jump in belief and value. Otherwise, at a rate approximated by 1, the belief shifts just marginally and incurs a marginal shift in the continuation value, $V'(p)$.

Yet, the ability to change the arm through β substantially changes the potential scope of the solutions as well as the resulting dynamics. Typically, the effects of acquiring new information move in opposite directions where the discrete jump from good news is positive, and the smooth shift in the absence of it is negative. In contrast, here the direction of the marginal shift is non-monotone and depends both on the state and the control variables value. That is, the agent endogenizes her learning from receiving no news, and it is possible that the continuation value increases with any information she gets (i.e. both informational effects in the equation are positive).

Importantly, the HJB equation is linear in both control variables, α and β. This implies that it is optimal for the agent to take pure actions: use a safe arm, purely experiment, or experiment with training the arm at full intensity. As such, the equation can be solved separately for each of the three modes resulting in the following value functions:

$$V(p) = \max_{\alpha,\beta} \begin{cases} \frac{s}{r} & \alpha = 0 \text{ (safe arm)} \\ \frac{\lambda}{r}p + C_{Exp}f(p) & \alpha = 1, \beta = 0 \text{ (experimentation)} \\ y(p) + C_{Tr}g(p) & \alpha = \beta = 1 \text{ (training)} \end{cases} \tag{1.2}$$

where $y(p) = p\left(\frac{\lambda}{r} - \frac{\kappa}{r+\lambda}\right) + (1-p)\left(\frac{\pi}{r+\pi}\left(\frac{\lambda}{r} - \frac{\kappa}{r+\lambda}\right) - \frac{\kappa}{r+\pi}\right)$, $g(p) = \frac{(1-p)^{1+\frac{r+\pi}{\lambda-\pi}}}{|p-\frac{\pi}{\lambda}|^{\frac{r+\pi}{\lambda-\pi}}}$, $f(p) = \frac{(1-p)^{1+\frac{r}{\lambda}}}{p^{\frac{r}{\lambda}}}$, and $\{C_{Exp}, C_{Tr}\} \geq 0$ are some arbitrary constants.

The value from pure experimentation ($\alpha = 1, \beta = 0$) and from using the safe arm

($\alpha = 0$) exactly replicate the value functions from KRC. The agent gets $\frac{s}{r}$ forever under using the safe arm (and never switches to any other action since there is no dynamic belief updating), and $\frac{\lambda}{r}p$ if she experiments forever. However, because the beliefs evolve, the agent may switch to other actions instead of experimenting forever, meaning that the actual value can be higher, which is captured by the adjustable non-negative extra component $C_{Exp}f(p)$.

The new part of the value function arises with the ability to train. The first linear and increasing part represents the average payoff from training the risky arm forever until the good news arrives and switching to purely using this arm immediately after. If the arm is already good (with probability p), the agent on average gets λ in each instance, hence the present value of it being $\frac{\lambda}{r}$, but pays the cost of training up until the first time T when she observes the good news and stops wasting resources, with $T \sim exp(\lambda)$. As such, the agent expects to pay $E_T[\int_0^T(e^{-rt}\kappa dt)] = \frac{\kappa}{r+\lambda}$. In turn, if the arm is bad (with probability $1-p$), the agent expects to start getting the good arm's payoff only once the training succeeds; denote this time as τ. Obtaining this payoff is discounted by $E_\tau[e^{-r\tau}] = \frac{\pi}{r+\pi}$, since $\tau \sim exp(\pi)$ by construction. Additionally, the cost of training has to be paid in each moment of time in $[0, \tau]$, with the present value of it being $E_\tau[\int_0^\tau(e^{-rt}\kappa dt)] = \frac{\kappa}{r+\pi}$. Overall, the first part highlights that the training mechanism provides a sort of insurance to the agent, where she now gets a positive payoff if the arm is bad but has to pay for it even in the good state, since the arm's type is not revealed immediately with training. In turn, the last part of the value function represents the extra value from switching between options given the potential belief evolution under training the arm, and is assumed to be non-negative by restricting solutions to $C_{Tr} \geq 0$. Having more options cannot hurt the agent, and if she strictly benefits from it, the free constant $C_{Tr} > 0$ will be determined as part of the optimal strategy.

1.3.5 Benchmark with no possibility to train

A useful benchmark is the case where the option of training the risky arm is unavailable. The following lemma reiterates the result previously established in KRC (see the paper for the proof).

Lemma 1.2 (Benchmark)**.** *In the absence of training, the optimal solution follows a cutoff structure with belief* $\hat{p} = \frac{sr}{\lambda(\lambda-s+r)} \in (0,1)$ *such that experimentation occurs for* $p > \hat{p}$*, and the safe arm is used otherwise. The solution is characterized by the continuous, smooth and globally convex value function:* $V(p) = max\{\frac{s}{r}, \frac{\lambda}{r}p + \left(\frac{s}{r} - \frac{\lambda}{r}\hat{p}\right)\frac{f(p)}{f(\hat{p})}\}$*, where* $f(p) = \frac{(1-p)^{1+r/\lambda}}{p^{r/\lambda}}$*.*

In the optimal solution, the agent experiments only if she is sufficiently optimistic. She engages with the risky arm more than her myopic counterpart (who only cares about the current payoff and quits for $p < \frac{s}{\lambda}$), because of the extra positive value from learning while experimenting and anticipating the future change in her actions that is captured by the $\left(\frac{s}{r} - \frac{\lambda}{r}\hat{p}\right)\frac{f(p)}{f(\hat{p})}$ term in the value function. The agent's belief gradually falls, as long as no news arrives, until she reaches a cutoff where she quits experimentation. At such cutoff $\hat{p}$, the agent's value from the safe and the risky arm match (*value matching* condition). There, she is also indifferent between experimenting for an extra instance and then returning to the optimal path immediately after and not experimenting at all. This is a marginal incentive which, given that the agent becomes more pessimistic with using the risky arm, translates to the *smooth pasting* condition ($V'(\hat{p}) = 0$).

In what follows I will refer to the cutoff and the value function from Lemma 1.2 as $\hat{p}, V_{bench}(p)$, respectively.

1.4 Main results

Being able to train interferes with the incentives to experiment and may significantly change the agent's optimal behaviour. Specifically, introducing the training mechanism of a certain efficiency can result in one of the two distinct regimes. The proposition below states these key results.

Proposition 1.1. *Assume sufficiently low cost of training.* [5] *Denote as $\underline{p}$ and $\overline{p}$ the lowest and highest belief, under which the agent trains the risky arm, respectively. Then, there exist $\underline{p}(\pi,\kappa)$ and $\overline{p}(\pi,\kappa)$, such that:*

i. If $\frac{\pi}{\lambda} < \underline{p}(\pi,\kappa)$, the agent's strategy is non-monotone in actions and leads to quitting experimentation for any p_0, unless news arrives: She uses a safe arm for $p < \hat{p}$, trains the risky arm in $(\underline{p},\overline{p}) \subset [\hat{p},1]$, and purely experiments otherwise.

ii. If $\frac{\pi}{\lambda} > \underline{p}(\pi,\kappa)$, the agent's strategy leads to experimenting 'forever' until news arrives for any $p_0 \geq \underline{p}$. She gets stuck at training the arm at full intensity until news arrives if $\frac{\pi}{\lambda} \in [\underline{p}(\pi,\kappa), \overline{p}(\pi,\kappa)]$, and at some reduced intensity $\beta^* \in (0,1)$ if $\frac{\pi}{\lambda} > \overline{p}(\pi,\kappa)$.

Under lower efficiency of training (when $\frac{\pi}{\lambda} < \underline{p}(\pi,\kappa)$), the agent trains only in some

[5] Specifically, $\kappa < \overline{\kappa}(\pi)$, where $\overline{\kappa}(\pi)$ is defined as a highest cost, at which the agent is willing to train the arm for at least a single belief. If $\kappa \geq \overline{\kappa}(\pi)$, the agent never trains the arm, and her solution matches the benchmark one established in Lemma 1.2. See precise definition and derivation of $\overline{\kappa}(\pi)$in Appendix A.2.

narrow region and gives up experimentation after not receiving a breakthrough for sufficiently long. That is, the agent mainly uses the risky arm to learn its type, and training enables her to increase the chances of observing success. Interestingly, her strategy is then non-monotone in actions: She gives up training and purely experiments before quitting experimentation, just as she does when being very optimistic.

In contrast, for $\frac{\pi}{\lambda} > \underline{p}(\pi, \kappa)$, the agent may get stuck on a certain belief and pursue training at some intensity until the news arrives. She converges to such belief both from above and from below: She trains the arm if she is more pessimistic, which rises her belief, while training the arm or purely experimenting in more optimistic states makes her more pessimistic and reduces her belief. Such strategy implies that she does not value learning per se that much and is mainly motivated by ensuring the arm succeeds when she uses it.

Importantly, the two regimes highlight that the ability to train the risky arm impacts experimentation through two distinct channels. First, training has a direct effect on the chance of successful outcome by turning a bad arm into a good one. Second, it increases the likelihood of learning that the good risky arm is good, an indirect and more subtle benefit. Training counteracts the negative impact of having no news, so the agent at least gets disappointed slower, or possibly even becomes more optimistic with training. This implies that the agent anticipates using the risky arm over a longer horizon than she would in the absence of training, which increases her chance of observing a breakthrough from the good arm before she stops. Clearly, the impact of these two channels varies across the regimes and leads to qualitatively different outcomes.

In the rest of the section I motivate the findings in more detail. Subsections 1.4.1 and 1.4.2 span the key results under low and high efficiency training, respectively. I then summarise the findings by discussing under which conditions each regime occurs and outlining the overall implications of training on experimentation in Subsections 1.4.3 and 1.4.4.

1.4.1 Low efficiency of training: Non-monotonicity in actions

The agent's optimal strategy under low training efficiency according to Proposition 1.1 part i can be represented by a scheme as in Figure 1.2. If the training mechanism is not too efficient (that is, if $\pi < \lambda\underline{p}(\pi, \kappa)$), training may improve the value of experimenting, but does not affect the overall strategy qualitatively. In the absence of news, the agent becomes more pessimistic with using the risky arm irrespective of whether she trains it,

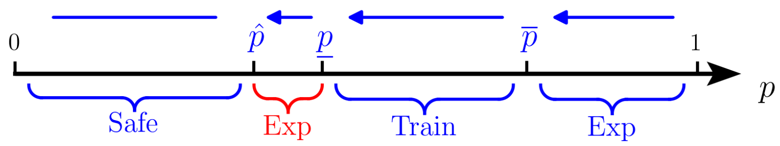

Figure 1.2: Agent's optimal strategy structure if $\frac{\pi}{\lambda} < \underline{p}(\pi, \kappa)$, as a function of belief p. The arrows above the real line indicate the direction of beliefs evolution in the absence of news.

since she is only willing to train for beliefs above $\frac{\pi}{\lambda}$, where the negative effect from no news dominates the positive influence of training. This implies that, no matter how optimistic the agent is initially, unless the news arrives her belief gradually decreases and converges to $\hat{p}$, where she quits experimentation, as she would without the ability to train.

An interesting feature of this optimal strategy is that the agent's path in the absence of news is necessarily non-monotone in actions. If she starts sufficiently optimistic, she uses the risky arm, then starts training it once she is no longer convinced that the arm is good, and stops training a while before she gives up on the risky arm entirely. That is, she experiments purely for beliefs both higher and lower than the training region, even if the training cost is infinitesimally low. Intuitively, the benefit of training increases when p is lower, since it can only improve the bad arms. As such, its benefit almost disappears when the agent is very optimistic, and training is wasteful then. This rationalizes the agent's incentive to forgo training in $[\overline{p}, 1]$ for any positive cost.

The incentive to give up training but still experiment in $(\hat{p}, \underline{p})$ hinges on the belief dynamic and anticipated action switches in the future. The agent understands that the realization of benefit from training is delayed in time: If she trains the arm today and this training succeeds, it will take some time until the improved arm generates a breakthrough. However, she also anticipates that she will quit experimentation once her belief reaches the stopping cutoff $\hat{p}$ in some deterministic time, and she will never attain the benefit of training thereafter. Thus, when she is sufficiently close to $\hat{p}$, her marginal benefit from training gets very low, as she is very unlikely to learn about the newly occurring successes of training in the short time left before quitting the risky arm. This makes training no longer worthwhile and she stops training the arm. At the same time, she still has an incentive to experiment for a bit longer, since there is a positive chance of observing the breakthrough if the arm was originally good or the training succeeded in the past. She does so until her belief drops to $\hat{p}$, the benchmark cutoff: Once she quits training, her incentives fully coincide with the ones she would have in the absence of training option, and so she switches to the safe arm at the same belief.

Formally, the incentives to forgo the training are reflected in the marginal benefit of it, $V'(p)\pi(1-p)$, compared to the marginal cost κ. For high beliefs the direct benefit $\pi(1-p)$ is low, while for lower ones $V'(p)$ is low. At $\hat{p}$, $V'(\hat{p}) = 0$, by smooth pasting, as explained in Section 1.3.5. Therefore, the equivalence of the marginal cost and benefit of training must arise for $\underline{p} > \hat{p}$, since $V'(\underline{p}) > 0$ must hold for the benefit to cover even an infinitesimally small cost of training. The higher the cost is, the smaller the training incentives are, so $\underline{p}$ increases with the cost, while $\overline{p}$ declines.

The proposition below characterises the optimal agent's strategy formally, and the remainder of the subsection briefly motivates this technical characterisation.

Proposition 1.2. *If $\frac{\pi}{\lambda} < \underline{p}(\pi,\kappa)$, the agent's optimal strategy is fully characterized by a continuous, globally convex and smooth value function, with $\underline{p} \in (\frac{\pi}{\lambda}, \underline{p}(\pi,\kappa))$:*

$$V(p) = \begin{cases} \frac{s}{r} & p < \hat{p} \ \textit{(safe arm)} \\ \frac{\lambda}{r}p + (\frac{s}{r} - \frac{\lambda}{r}\hat{p})\frac{f(p)}{f(\hat{p})} & p \in [\hat{p}, \underline{p}] \ \textit{(pure exp)} \\ y(p) + \left(\frac{\lambda}{r}\underline{p} - y(\underline{p}) + (\frac{s}{r} - \frac{\lambda}{r}\hat{p})\frac{f(\underline{p})}{f(\hat{p})}\right)\frac{g(p)}{g(\underline{p})} & p \in [\underline{p}, \overline{p}] \ \textit{(training)} \\ \frac{\lambda}{r}p + \left(y(\overline{p}) - \frac{\lambda}{r}\overline{p} + \left(\frac{\lambda}{r}\underline{p} - y(\underline{p}) + (\frac{s}{r} - \frac{\lambda}{r}\hat{p})\frac{f(\underline{p})}{f(\hat{p})}\right)\frac{g(\overline{p})}{g(\underline{p})}\right)\frac{f(p)}{f(\overline{p})} & p > \overline{p} \ \textit{(pure exp)} \end{cases}$$

Notice that the value function that characterizes the low efficiency of training solution (also plotted in Figure A.1 in Appendix A.1) reflects the anticipated action dynamics. The agent benefits from switching to the other options for lower beliefs, and the non-linear components of $V(p)$ represent these options value. She gains extra value from starting to train the arm below $\overline{p}$, giving up on training at $\underline{p}$, and finally switching to the safe arm at $\hat{p}$ (or a subset of these depending on initial p), and the impact of these switches increases the closer the agent is to making a certain switch. That is, the latter components are decreasing in beliefs (as all the cutoffs are approached from the right) and convex.

Moreover, the value function shares the standard features as in the benchmark outlined in Section 1.3.5: $V(p)$ is continuous and smooth around all the cutoffs $(\hat{p}, \underline{p}, \overline{p})$. At any cutoff, the agent is indifferent only if she receives the same value from following any of the two actions and if she cannot benefit by marginally deviating to one of the actions for an instance and returning to the optimal path immediately after. The former condition, value matching, guarantees continuity of the value function. The latter translates to smooth pasting at all cutoffs, which hinges on the monotonically decreasing beliefs dynamics. For example, consider the lower boundary for training, $\underline{p}$. Training for an extra instance at $\underline{p}$ makes the agent more pessimistic and pushes her to the left of the cutoff, where she strictly prefers to experiment purely. Hence, she is indifferent in her marginal deviation to training as opposed to immediate pure experimentation only if its marginal cost and benefit conditional on experimenting are equal ($\lim_{p\to\underline{p}_-} V'(p)\pi(1-p) = \kappa$). This results in the smooth pasting requirement, $\lim_{p\to\underline{p}_-} V'(p) = \lim_{p\to\underline{p}_+} V'(p)$, once combined with value matching. A similar logic applies at the other cutoffs.

1.4.2 High efficiency of training: Convergence to training

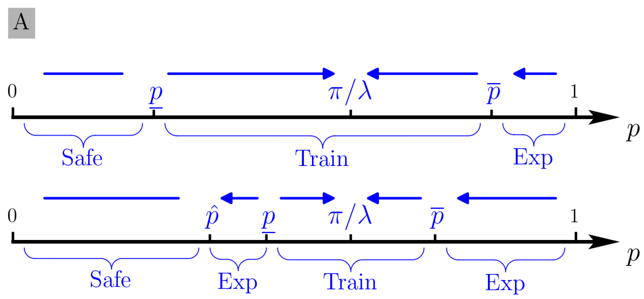

Figure 1.3: Agent's optimal strategy structure and beliefs evolution for moderately high training efficiency: for lower cost on the top, and higher cost at the bottom.

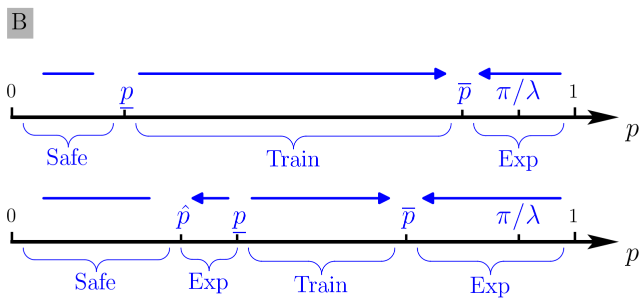

Figure 1.4: Agent's optimal strategy structure and beliefs evolution for very high training efficiency: for lower cost on the top, and higher cost at the bottom.

Figures 1.3 and 1.4 illustrate the possible optimal strategies according to Proposition 1.1 part *ii*. Strikingly different from traditional Poisson bandits models of experimentation, if the training mechanism is sufficiently efficient, the agent may never give up on the risky arm, even if no breakthrough arrives for a long time. This occurs whenever $\frac{\pi}{\lambda}$ is above $\underline{p}$. At $\underline{p}$ the agent becomes more optimistic with training, since its positive effect on belief dominates the negative one of receiving no news, and her belief will keep rising to $\frac{\pi}{\lambda}$ conditional on her pursuing training. If the cost of training is relatively low (when $\frac{\pi}{\lambda} \leq \overline{p}(\pi, \kappa)$), the agent finds it optimal to follow this path and get stuck at belief $\frac{\pi}{\lambda}$, where she keeps training the arm forever until it becomes good and generates a breakthrough. Starting at $p_0 > \frac{\pi}{\lambda}$, the agent gets more pessimistic with purely experimenting for beliefs above $\overline{p}$, switches to training at $\overline{p}$, which slows down her pessimistic learning but does not compensate it fully, so she gradually converges to $\frac{\pi}{\lambda}$ where she again gets stuck at training until the arm succeeds.

As training becomes more costly, her incentive to train at $\frac{\pi}{\lambda}$ disappears. Specifically, when $\frac{\pi}{\lambda} > \overline{p}(\pi, \kappa)$, the marginal benefit of training at $\frac{\pi}{\lambda}$ is lower than its marginal cost, either because the cost κ is too high, or the arm is very unlikely to be bad ($p = \frac{\pi}{\lambda}$ is close to 1), so the direct benefit of training almost vanishes. This implies that the agent quits training for beliefs around $\frac{\pi}{\lambda}$ and only experiments, which leads to her getting more pessimistic in the absence of news. Therefore, the agent can never achieve high enough belief through training and instead converges to the cutoff $\overline{p}$ both from below if she is more pessimistic and trains the arm, and from above if she is more optimistic and experiments purely. At such belief $\overline{p}$, she is exactly indifferent between training the arm and not, and there exists a unique stationary training intensity $\beta^* \equiv \frac{\lambda}{\pi}\overline{p} \in (0, 1)$ (given that $\overline{p} < \frac{\pi}{\lambda}$), such that she optimally remains at this boundary and trains the arm at reduced intensity until it succeeds.

In both of the scenarios outlined above, the agent ends up experimenting forever for any $p_0 \geq \underline{p}$, by converging to an interior belief and getting stuck there until the breakthrough arrives (due to training). Yet, if she is slightly more pessimistic than $\underline{p}$, she may still have some incentives to experiment: She may prefer to experiment without training for beliefs in $(\hat{p}, \underline{p})$, which is non-empty under high training costs. This creates a sharp discontinuity in the expected outcomes of experimentation around the cutoff $\underline{p}$, since experimenting below $\underline{p}$ eventually dies out unless the news arrives, in contrast to the experimentation above $\underline{p}$. I discuss the welfare implications of this observation further in Section 1.4.4.

Here, the motivation of forgoing training in $(\hat{p}, \underline{p})$ is different from the one in the low efficiency case, discussed in Section 1.4.1. Since the agent is willing to train at $\underline{p}$, the marginal benefit of training there (taking into account that she will keep training forever if she starts) should at least weakly exceed its cost. Starting training at $\underline{p} - \epsilon$ (for $\epsilon \to 0$) with continuing doing so forever after incurs a larger marginal benefit, as the arm is slightly more likely to be bad and thus gain from training. However, it also incurs the opportunity cost that training forever gives rise to, as the agent foregoes the safe arm's payoff forever the instance she starts training. This opportunity cost, relative to the benefit of training until news arrives, gets larger the more pessimistic the agent is.

If the cost of training is high, the relative opportunity cost becomes very large sooner (for higher belief), as the benefit of training the arm until the news arrives sharply decreases with the direct training cost. That is, the agent may prefer to quit training for some belief above $\hat{p}$, where she still values learning about the risky arm, but would rather get a safe payoff in the absence of news than train the arm over a long horizon. Hence, she purely experiments in $(\hat{p}, \underline{p})$. Formally, while the benefit of training must at least weakly exceed the cost for $p > \underline{p}$ ($\lim_{p \to \underline{p}_+} V'(p)\pi(1-p) \geq \kappa$), $\lim_{p \to \underline{p}_-} V'(p)\pi(1-p) \leq \kappa$ may hold, given that the agent anticipates different action dynamics on the different sides of the cutoff. This ultimately guarantees that for any belief below $\underline{p}$, the agent does not benefit from training for a short instance, so can never become optimistic enough to reach $\underline{p}$ and commence training the risky arm forever.

As the costs get lower, the value of training the risky arm until it succeeds increases for any belief, broadening the incentives to train (i.e. lowering $\underline{p}$). Once the cost is low enough, the value of training forever starting from belief $p_0 = \hat{p}$ strictly exceeds the value of marginally experimenting and giving up, implying that at the lower training cutoff the agent is indifferent between training forever and sticking to the safe arm immediately, $\underline{p} < \hat{p}$. Such cutoff decreases further with the cost, and if the reduction in cost is coupled with especially high training productivity π, it is possible that the agent is willing to train even a certainly bad risky arm, i.e. at $p = 0$, implying that any risky arm will be experimented upon forever.

The proposition below provides a technical characterization of the solution under high efficiency training, and the remainder of the subsection spans some formal aspects of it.

Proposition 1.3. *If $\frac{\pi}{\lambda} > \underline{p}(\pi,\kappa)$, the agent's optimal strategy is fully characterized by a continuous, globally convex and kinked at $\underline{p}$ (and smooth otherwise) value function.*

i. If $\frac{\pi}{\lambda} \in [\underline{p}(\pi,\kappa), \overline{p}(\pi,\kappa)]$, then $\{\underline{p}, \overline{p}\} = \{\underline{p}(\pi,\kappa), \overline{p}(\pi,\kappa)\}$, and

$$V(p) = \begin{cases} \max\{\frac{s}{r}, \frac{\lambda}{r}p + (\frac{s}{r} - \frac{\lambda}{r}\hat{p})\frac{f(p)}{f(\hat{p})}\} & p < \underline{p} \text{ (safe arm or pure exp)} \\ y(p) & p \in [\underline{p}, \overline{p}] \text{ (training)} \\ \frac{\lambda}{r}p + \left(y(\overline{p}) - \frac{\lambda}{r}\overline{p}\right)\frac{f(p)}{f(\overline{p})} & p > \overline{p} \text{ (pure exp)} \end{cases}$$

ii. If $\frac{\pi}{\lambda} > \overline{p}(\pi,\kappa)$, then $\overline{p} \in (\overline{p}(\pi,\kappa), \frac{\pi}{\lambda})$, and

$$V(p) = \begin{cases} \max\{\frac{s}{r}, \frac{\lambda}{r}p + (\frac{s}{r} - \frac{\lambda}{r}\hat{p})\frac{f(p)}{f(\hat{p})}\} & p < \underline{p} \text{ (safe arm or pure exp)} \\ y(p) + \left(\frac{\lambda}{r}\overline{p} - y(\overline{p}) + \frac{(\lambda\pi(1-\overline{p}) - \kappa r)\overline{p}}{r\pi(\overline{p} + \frac{r}{\lambda})}\right)\frac{g(p)}{g(\overline{p})} & p \in [\underline{p}, \overline{p}] \text{ (training)} \\ \frac{\lambda}{r}p + \left(\frac{(\lambda\pi(1-\overline{p}) - \kappa r)\overline{p}}{r\pi(\overline{p} + \frac{r}{\lambda})}\right)\frac{f(p)}{f(\overline{p})} & p > \overline{p} \text{ (pure exp)} \end{cases}$$

As in Section 1.4.1, the option value parts of the function reflect the anticipated action switches given the belief dynamics. Notably, whenever the agent converges to training at full intensity at $\frac{\pi}{\lambda}$ (for $\frac{\pi}{\lambda} \in [\underline{p}(\pi,\kappa), \overline{p}(\pi,\kappa)]$), once she starts training, she never switches to other options (until the news arrives), and so the value function $V(p)$ is linear in the training region and contains only the value of training the arm forever, $y(p)$ (see figure A.2 in Appendix A.1). Experimenting above $\overline{p}$ adds some value to the agent because she ends up stuck at training as well, which is reflected in the option value part.

As the cost rises to $\frac{\pi}{\lambda} > \overline{p}(\pi, \kappa)$, the agent converges to training the arm at a reduced intensity β^* and gets the extra value from it both when experimenting in $(\overline{p}, 1)$ and when training in $(\underline{p}, \overline{p})$. Indeed, while training for $p < \overline{p}$, she anticipates getting stuck at $\overline{p}$ and hence saving on the training costs when they would have exceeded the benefit - at $\overline{p}$, $V'(p)\pi(1-p) = \kappa$ exactly. Similarly, while using the arm for $p > \overline{p}$, she expects to start training at intensity β^* instead of becoming pessimistic with using the arm forever. Following the above, the options value components in the value function for $p > \underline{p}$ are increasing towards $\overline{p}$ (increasing and convex in $(\underline{p}, \overline{p})$, and decreasing and convex in $(\overline{p}, 1)$) (see figure A.3 in Appendix A.1).

Under both scenarios, the optimal action profile is sharply divided by the lower cutoff $\underline{p}$, where being above $\underline{p}$, the agent can never become more pessimistic than $\underline{p}$, and vice versa. As such, the agent's value above $\underline{p}$ does not depend on anything occurring below $\underline{p}$, including possibly experimenting there or using the safe arm, as well as the value of the cutoff itself. Similarly, being below $\underline{p}$, the agent cannot become more optimistic than $\underline{p}$, so her value for $p < \underline{p}$ is independent of $\underline{p}$ and anything above this cutoff.

The diverging behaviour around the lower training cutoff $\underline{p}$ implies that its value is determined fully by the value matching requirement, and it necessarily violates the smooth pasting as a result. Since training makes the agent more optimistic at $\underline{p}$, a short (marginal) deviation to training with returning to the optimal path immediately after pushes her deeper to the training region, implying that the value from such deviation fully matches the value to the right of the cutoff. Similarly, a short deviation to the action to the left of $\underline{p}$ (either using a safe arm or purely experimenting) either leaves the agent's belief unchanged or pushes her more to the left, thus matching $\lim_{p \to \underline{p}_-} V(p)$. Hence, any marginal indifference condition coincides with the value matching one, and smooth pasting that typically arises due to this marginal incentive need not hold.

When $\underline{p}$ arises from indifference between training the risky arm forever and using the safe arm (if $\underline{p} < \hat{p}$), the marginal cost of training, $s - \lambda \underline{p} + \kappa$, is larger than the benchmark one, $s - \lambda \hat{p}$, while the jump benefit from learning the news, $\lambda \underline{p} \frac{\lambda - s}{r}$, is smaller. So the marginal cost and benefit of training at $\underline{p}$ can be equal only if the smooth benefit under no news is strictly positive; i.e. $\lim_{p \to \underline{p}_+} V'(p) > 0$ must hold, violating smooth pasting. If the agent is indifferent between experimenting with and without training at $\underline{p}$, we can show that the marginal benefit of training exceeds the marginal cost to the right, $\lim_{p\to\underline{p}_+} V'(p)\pi(1-p) > \kappa$, while the cost dominates to the left, $\lim_{p\to\underline{p}_-} V'(p)\pi(1-p) \leq \kappa$, guaranteeing the convex kink and smooth pasting violation as well. [6]

Note that the value function remains smooth at the upper cutoff $\overline{p}$, where the agent switches from using the risky arm to training it. If the agent marginally forgoes training at $\overline{p}$, she becomes more pessimistic under pure experimentation and thus switches to training immediately. She is indifferent to such deviation only if the forgone cost of training matches the forgone training benefit, $V'(\overline{p})\pi(1-\overline{p}) = \kappa$, which, together with value matching, guarantees smooth pasting. Intuitively, since the threshold is absorbing at least from the right, if the agent approaches it from there, she will purely experiment for as long as the marginal cost of training exceeds its benefit and switch to improving the arm as soon as the benefit exceeds the cost; such strategy guarantees a smooth transition from one action to another.

If $\frac{\pi}{\lambda} > \overline{p}(\pi,\kappa)$, the upper cutoff $\overline{p}$ is absorbing on both sides - this is a novel boundary type, that typically does not occur in experimentation models. To define such cutoff, the traditional conditions of value matching and smooth pasting do not suffice, because they only identify two of the three free parameters in the triplet $\{\overline{p}, C_{Tr}, C_{Exp}\}$ (i.e. define $C_{Tr}(\overline{p}), C_{Exp}(\overline{p})$). The extra necessary condition guarantees that the agent cannot improve her value by shifting the cutoff of interest slightly, that is $\overline{p} \equiv \arg\max_q V(p;q)$, where $V(p;q) = \frac{\lambda}{r}p + C_{Exp}(q)f(p)$, implying that the boundary is defined through $C'_{Exp}(\overline{p}) = 0$ (or $C'_{Tr}(\overline{p}) = 0$, equivalently). If this extra optimality condition holds, the agent is indifferent between training only at the cutoff $\overline{p}$, strictly prefers training to the left of the cutoff and pure experimentation to the right, and cannot benefit from changing the intensity of training anywhere. [7]

[6] Such 'irregularity' of the cutoff $\underline{p}$ is similar to the one identified in Keller and Rady (2015) for the bad news information structure. However, the reason for the smooth pasting violation is different here. In KR it relies on the disappearance of the jump benefit at the cutoff (as learning the bad news result in quitting the experimentation, which coincides with the cutoff payoff exactly).

[7] To confirm that there are no further improvements, define $V_{indiff}(q) \equiv V(q;q)$ - such value function represents the value of being indifferent between experimenting with and without training and thus training at any intensity $\beta \in [0,1]$; such $V_{indiff}(q)$ is strictly concave, and plotted in Figure A.3. Then, the optimality condition $C'_{Exp}(\overline{p}) = 0$ implies $V'_{indiff}(\overline{p}) = V'(\overline{p})$, and so $V(p) \geq V_{indiff}(p)$ for any p, with equality occurring only at $p = \overline{p}$.

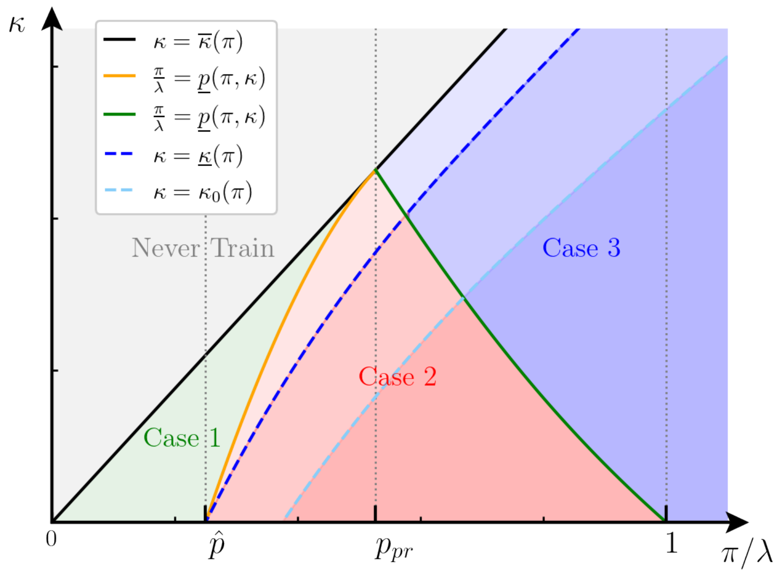

Figure 1.5: Training regimes for all possible values of π, κ, with π normalized to $\frac{\pi}{\lambda}$. $\frac{\pi}{\lambda} = \underline{p}(\pi, \kappa)$ and $\frac{\pi}{\lambda} = \overline{p}(\pi, \kappa)$ are defined in Proposition 1.1 and separate the space into three cases: case 1 - low efficiency training, case 2 - high efficiency with convergence to $\frac{\pi}{\lambda}$, case 3 - high efficiency with convergence to $\overline{p}$. The rest are informally defined and discussed in subsection 1.4.3.

1.4.3 Key findings summary and implications

Overall, there are two regimes arising from the ability to train: The experimentation incentives dominate in one, while the motive of improving the arm is prevailing in the other. Figure 1.5 illustrates when one or the other case arises for any training mechanism in the studied class, i.e. for any combination of π and κ, keeping the baseline parameters, $\{\lambda, s, r\}$, fixed. Training can only benefit the agent if it is not too costly, so she never trains the arm in the top left corner of the diagram. Then, $\frac{\pi}{\lambda}$ vs $\underline{p}(\pi, \kappa)$ constraint determines which of the two regimes occurs. The agent predominantly experiments with only marginal impact of training in the region in the bottom left (Case 1 region), which I previously referred to as 'low training efficiency' and studied in 1.4.1: That is, when the training productivity π is low, while its cost κ is moderately high. Otherwise, when training is either cheaper or more productive (to the right of the $\frac{\pi}{\lambda} = \underline{p}(\pi, \kappa)$ constraint), it has a more salient effect and changes the set of convergence outcomes, as described in section 1.4.2. Within this high efficiency region, the agent can either converge to training at full intensity at $p = \frac{\pi}{\lambda}$ (Case 2 on the diagram) or at reduced intensity at $p = \overline{p}$ (Case 3), which are separated along the $\frac{\pi}{\lambda} = \overline{p}(\pi, \kappa)$ constraint. Intuitively, the agent is willing to save some training costs by reducing intensity if the costs are high, so the latter scenario occurs above the constraint.

Being able to train not only improves the outcome of a bad risky arm, it also increases the likelihood of observing a breakthrough from initially good arms before quitting experimentation. The proposition below breaks down the qualitative impact of these effects on experimentation.

Proposition 1.4. *Having an ability to train:*

i. Improves the value of experimenting (for $p > \underline{p}$) if $\kappa < \overline{\kappa}(\pi)$, where $\overline{\kappa}(\pi)$ is a cost at which the agent is willing to train only at a single belief.

ii. Broadens the incentives to experiment beyond $\hat{p}$ if $\kappa < \underline{\kappa}(\pi)$, where $\underline{\kappa}(\pi)$ is a cost at which $\underline{p} = \hat{p}$. If $\frac{\pi}{\lambda} \leq \hat{p}$, this effect never occurs, even if training is costless.

iii. Guarantees a breakthrough almost surely for all prior beliefs above $\underline{p}$ if $\frac{\pi}{\lambda} > \underline{p}(\pi, \kappa)$.

The qualitative impact of training on experimentation varies across the training regimes. Trivially, if the agent trains the risky arm or anticipates doing so in the future, it improves her continuation value, otherwise she would not engage with the training mechanism. This occurs only if the cost is sufficiently low ($\kappa < \overline{\kappa}(\pi)$), as for higher costs training is fully dominated by pure experimentation, as indicated on Figure 1.5. Under lower efficiency (when $\frac{\pi}{\lambda} < \underline{p}(\pi, \kappa)$), this is the only impact training has on the agent. However, for $\frac{\pi}{\lambda} > \underline{p}(\pi, \kappa)$, the ability to train causes a change to the experimentation incentives and outcomes. Not only it enhances the value of experimenting, it also increases the incentives to experiment beyond $\hat{p}$ under low enough costs ($\kappa < \underline{\kappa}(\pi)$), and ensures that the breakthrough arises with probability 1 almost surely for priors above $\underline{p}$.

Intuitively, the effect of increasing the scope of experimentation to a wider range of beliefs should occur for any training productivity under low enough costs. This is, however, not so for low training productivity levels $\pi \leq \lambda\hat{p}$: The agent does not experiment more even if $\kappa = 0$ then. Under costless training, the only effect the training has as opposed to pure experimentation is in the marginal beliefs shift conditional on obtaining no news (this effect is weakly positive, so training dominates over pure experimentation). If the training efficiency is low, the agent does not benefit from this effect at the stopping cutoff, because she gets pessimistic there and smoothly transitions to the safe option (by Proposition 1, smooth pasting holds at $\underline{p}$). As a result, her incentive to train fully coincides with her incentive to experiment in the absence of training, and thus, $\underline{p} = \hat{p}$. Note that having a training option broadens the agent's incentives to experiment under low enough costs (when $\underline{p} < \hat{p}$), but can never reduce those (safe arm is never used for $p > \hat{p}$): If the agent does not want to train in the neighbourhood of $\hat{p}$, she still benefits from learning the risky arm's type there, and hence purely experiments at least.

Finally, the agent obtains a breakthrough almost surely whenever she gets stuck at training the arm with a least some intensity until the news arrives, i.e. if $\frac{\pi}{\lambda} > \underline{p}(\pi, \kappa)$ holds (implying $\frac{\pi}{\lambda} > \underline{p}$). Such dynamic ensures that she will keep experimenting 'forever', and if the arm is initially good, she will learn it eventually, $\lim_{\tau \to \infty} \Pr(\text{news by } \tau | \lambda^{\theta} = \lambda) = 1$. That is, training strengthens experimentation incentives so much that the agent never mistakenly abandons a good risky arm, which can occur with positive probability in the absence of training option. Similarly, if the arm is initially bad, the agent will train it 'forever' until it becomes good and generates a breakthrough, which also occurs almost surely in the limit.

There is a variation in the range of the priors which lead to a guaranteed breakthrough within the region of $\frac{\pi}{\lambda} > \underline{p}(\pi, \kappa)$, and Figure 1.5 reflects that. For low costs and high enough training productivity, the agent trains even a surely bad risky arm, so any prior belief of the agent results in a breakthrough. This region is constrained from above by $\kappa = \kappa_0(\pi)$, which indicates the highest cost at which $\underline{p} = 0$. The training incentive gradually narrows for costs in $(\kappa_0(\pi), \underline{\kappa}(\pi))$, so the agent never touches a risky arm if she is very pessimistic about it, but otherwise takes it and almost surely turns it to a breakthrough. If the cost gets even higher ($\kappa > \underline{\kappa}(\pi)$), an extra pure experimentation region, $(\hat{p}, \underline{p})$, arises, and the instances at which the agent necessarily obtains a breakthrough are only a subset of those where she is willing to experiment: Any experimentation below $\underline{p}$ will eventually be terminated in the absence of news.

1.4.4 Further welfare implications

In this section I build on the key results to draw further implications of training on experimentation. Specifically, I present a way to think about welfare in this setup.

Clearly, the agent's preference of one experimentation regime over another transitions smoothly. That is, the agent never has a discontinuous increase in her continuation value from a marginal change in any parameter. Yet, there is potential discontinuity in the expected outcome of experimentation that arises due to being able to train. Recall that under high efficiency of training and high enough cost the agent prefers experimenting both above and below the lower training cutoff $\underline{p}$. If she is above, she will keep training the arm until it generates a breakthrough, which occurs with probability 1 almost surely in the limit. In contrast, being below, she will use the risky arm for a short while to learn if it is good and abandon it in a deterministic time it takes to move from p_0 to $\hat{p}$ in the absence of news. Such dynamics results in a strictly positive likelihood of ending up with no breakthrough: The bad arms will never generate a success, and some of the good ones are abandoned before their true type is observed.

While in traditional models of experimentation any small change can only marginally impact the likelihood of learning the good news before switching to the safe arm, the above suggests that here it may lead to a discrete jump in the probability of observing success from strictly interior to 1. This discontinuity of the outcome is leveled out for the agent given her discounting, but may matter significantly more generally. An example of such is provided below.

Corollary 1.1. *Assume a social planner who values the future more than an experimenting agent, and can intervene in designing the prior information. Then, whenever $\frac{\pi}{\lambda} > \underline{p}(\pi, \kappa)$, the social planner:*

i. Can achieve a jump gain from marginal intervention in the prior p_0 around $\underline{p}$.

ii. May strictly prefer incomplete information environment to full information under sufficiently optimistic prior ($p_0 > \underline{p}$) and sufficiently high training cost $\kappa > \kappa_0(\pi)$.

To illustrate, imagine a social planner (he) who values only the long-term outcome of experimentation and does not care about how the steady state is achieved. Assume he can intervene in the prior information of the experimenting agent (she) through moving her prior from p_0 to some p_0' by paying a cost $c(|p_0' - p_0|)$, with some smoothly increasing $c(\cdot)$ such that $c(0) = 0$. Then, the first observation follows directly: If $\frac{\pi}{\lambda} > \underline{p}(\pi, \kappa)$, experimentation leads to a breakthrough almost surely for any prior above $\underline{p}$, and stops before achieving success with strictly positive probability otherwise. If the prior p_0 is just below the cutoff belief $\underline{p}$ (eg. $p_0 = \underline{p} - \epsilon$ with $\epsilon \to 0$), a slight manipulation that pushes the prior upwards is always attractive, as it gives the social planner a jump gain at a marginal cost of $c(\epsilon) \to 0$.

Interestingly, depending on the prior, the social planner may also strictly prefer to commit to not revealing the information about the arm's type to the experimenting agent. With full information, if $\kappa > \kappa_0(\pi)$, the agent would immediately adopt a good risky arm, but abandon a bad one in favor of the safe option, as training the bad risky arm is too costly for her. In contrast, without the information and with sufficiently optimistic prior ($p_0 > \underline{p}$), she would train the arm and achieve a success almost surely independently of the arm's initial type. This ensures a better outcome and is thus preferred by the social planner.

1.5 Concluding remarks

Overall, the ability to train risky arms enriches experimentation framework by allowing for various actions trajectories and qualitatively different outcome predictions. Specifically, training affects the possible belief dynamics and makes it non-monotone even in the absence of news. Such observation gives rise to two key results. First, the agent may behave in a non-monotone manner: She quits training the arm a while before giving up on experimentation. Second, the agent may optimally get stuck on a certain interior belief, with converging towards this belief in the absence of news both when being more optimistic by purely experimenting, and pessimistic, by training the arm. In such a strategy, the agent pursues training until the arm generates a success, and so obtains a breakthrough almost surely in the limit. Hence, the ability to train the arm may not just improve the bad arm's performance, but has a spillover effect on the rate of discovering the a priori good arms: Under sufficiently efficient training mechanism all the good risky arms will be discovered and used in the limit, while they are wrongly abandoned with some positive probability in traditional experimentation models.

Importantly, being able to improve the arms challenges our understanding of learning from experimentation. Specifically, I highlight that the agent can reach a success in experimentation by either discovering initially good arms or investing resources in the bad ones for so long that they generate an equal success. This suggests that observing experimentation outcomes of others conveys much less information than typically assumed. As such, this paper opens opportunities for future research on the issues of learning about the options while also improving those, and I hope that bringing this agenda to the models of strategic experimentation among multiple agents may unveil new insights in our understanding of these class of problems.

References

Bergemann, Dirk and Ulrich Hege (1998) "Venture capital financing, moral hazard, and learning," *Journal of Banking Finance*, 22 (6), 703–735.

——— (2005) "The financing of innovation: Learning and stopping," *RAND Journal of Economics*, 719–752.

Bonatti, Alessandro and Johannes Hörner (2011) "Collaborating," *American Economic Review*, 101 (2), 632–663.

Fershtman, Daniel and Alessandro Pavan (2023) "Searching for "Arms": Experimentation with Endogenous Consideration Sets," *Working paper*, `https://sites.google.com/site/danielfershtman`.

Fryer, Roland and Philipp Harms (2018) "Two-armed restless bandits with imperfect information: Stochastic control and indexability," *Mathematics of Operations Research*, 43 (2), 399–427.

Guo, Yingni (2016) "Dynamic delegation of experimentation," *American Economic Review*, 106 (8), 1969–2008.

Halac, Marina, Navin Kartik, and Qingmin Liu (2016) "Optimal contracts for experimentation," *The Review of Economic Studies*, 83 (3), 1040–1091.

——— (2017) "Contests for experimentation," *Journal of Political Economy*, 125 (5), 1523–1569.

Hidir, Sinem (2019) "Contracting for Experimentation and the Value of Bad News," Working paper.

Hörner, Johannes and Larry Samuelson (2013) "Incentives for experimenting agents," *The RAND Journal of Economics*, 44 (4), 632–663.

Keller, Godfrey and Sven Rady (1999) "Optimal Experimentation in a Changing Environment," *The Review of Economic Studies*, 66 (3), 475–507.

——— (2010) "Strategic experimentation with Poisson bandits," *Theoretical Economics*, 5 (2), 275–311.

——— (2015) "Breakdowns," *Theoretical Economics*, 10 (1), 175–202.

Keller, Godfrey, Sven Rady, and Martin Cripps (2005) "Strategic Experimentation with Exponential Bandits," *Econometrica*, 73 (1), 39–68.

Klein, Nicolas and Sven Rady (2011) "Negatively correlated bandits," *The Review of Economic Studies*, 78 (2), 693–732.

Lizzeri, Alessandro, Eran Shmaya, and Leeat Yariv (2024) "Disentangling Exploration from Exploitation,"Technical report, National Bureau of Economic Research.

Moroni, Sofia (2022) "Experimentation in organizations," *Theoretical Economics*, 17 (3), 1403–1450.

Safronov, Mikhail (2021) "Experimentation and Learning-by-Doing," Working paper.

Whittle, Peter (1988) "Restless bandits: activity allocation in a changing world," *Journal of Applied Probability*, 25 (A), 287–298.

Chapter 2

Experimentation and Amelioration with Breakdowns

2.1 Introduction

News do not always come positive. Indeed, in a lot of environments learning occurs through observing a series of breakdowns rather than good news. E.g., a company learns about a new computer software technology through the number of bugs it generates, a manager discovers their subordinate's lack of expertise through observing the project missing its targets, and a hospital evaluates its surgeons' competence based on the number and severity of complications after the surgeries they performed.

In this chapter I build on the framework established in Chapter 1 and study how the ability to invest in adaptation and improve its outcomes, through training, affects the incentives to learn and experiment, when the information arrives in a form of negative news. Specifically, I assume that the safe arm generates a moderate constant cost, and the bad risky arm causes a larger breakdown at some Poisson arrival rate, while the good arm never does. In turn, the training mechanism is fully preserved from a breakthroughs framework and allows the agent to switch the bad arm into a good one through costly investments, in a way that is not directly observed by the agent.

Some of the key findings from the breakthroughs environment are robust to the flipped signals structure. Specifically, the agent may still find herself never quitting experimentation. While the finding relies on the non-monotonicity in beliefs in the good news framework, the beliefs may no longer evolve non-monotonically with the breakdowns, as the absence of news and ability to train both make the agent more optimistic. Yet, training improves the agent's incentives to experiment when she is pessimistic, so she may keep experimenting and training the arm even after the breakdown, which generates the result.

The agent still ends up in one of the two experimenting modes, where the more optimistic regime ensures the good arm is in use in the long-run almost surely (arises due to the agent's reluctancy to seize experimentation), whereas the agent terminates experimentation with a strictly positive probability in the more pessimistic one. The previous finding that the latter regime results in a non-monotonicity in actions conditional on the absence of news in the good news environment does not survive in the flipped information structure. This relies on the changed dynamics around the quitting cutoff. With the good news the benefit of training shrinks as the agent becomes more pessimistic and anticipates seizing experimentation soon (belief *converges* to the stopping cutoff). In contrast, the benefit of training strictly increases with agent's pessimism in the breakdowns framework, since she becomes more optimistic with no news and training, and so does not anticipate to quit (her belief *diverges* away from the cutoff).

Overall, while some results fail in the breakdowns specification, the ability to train the bad risky arms still has its bite: the positive influence of such training spans the a priori good arms in addition to the bad ones. More specifically, training expands the incentives to experiment for a wider range of beliefs. As such, the agent initiates experimenting under more priors, which increases the likelihood of the good risky arms being given a chance to be tested. This has a long term impact on the outcomes of experimentation since the good arms never fail a test - that is, never cause a breakdown.

The literature has established that flipping the signals structure in the models of experimentation can lead to very different experimenting dynamics and outcomes. Keller and Rady (2015) develop the machinery for working with the bad news model, when the stopping cutoff is irregular and violates the smooth pasting requirement that is traditionally present in the experimentation models with breakthroughs. Bonatti and Hörner (2017) and Wagner and Klein (2022) extend the breakdowns setting to allow for private actions and private information in strategic experimentation, respectively. Numerous recent studies exploit the differences generated by good and bad news information structures to address a range of more applied questions (see Halac and Prat (2016), Halac and Kremer (2020), as well as more recent Lizzeri et al. (2024), Ball and Knoepfle (2023), Bardhi et al. (2024) for examples). None of these papers endogenize the change in risky arms, which constitutes the novelty of the proposed framework, as argued in Chapter 1, and

I contribute to the above mentioned works by studying the effects of the information structure on experimentation when the agent can endogenously improve the risky option through investments. In line with the existing literature, my results highlight that different signals structures trigger different experimenting dynamics, while achieving similar experimentation outcomes.

The rest of the chapter is organized as follows. I specify the model including the beliefs evolution and value (cost) function derivation in Section 2.2 and conduct the main analysis in Section 2.3, which includes the optimal strategy characterisation, comparing it to the breakthroughs signal structure and discussing the qualitative implications of training. Section 2.4 concludes the analysis.

2.2 Model

2.2.1 Setup

Consider a model as in Keller and Rady (2015). There is a single agent who is choosing between a safe and a risky arm at each moment of time, where the time is treated continuously. The safe arm incurs a fixed cost of size $s > 0$ with certainty whenever in use. The risky arm, in turn, generates a cost normalized to size 1 at a Poisson arrival rate λ^b, where the rate depends on the arm's type: $\lambda^b \in \{\lambda, 0\}$, with $\Pr(\lambda^b = \lambda) = q_0$. That is, the arm is 'bad' if $\lambda^b = \lambda$, since it occasionally results in positive costs, while the 'good' arm never does given $\lambda^b = 0$. I further assume that $\lambda > s$, which implies that under complete information the agent would prefer the safe arm to the bad risky one, while the good arm dominates the safe one.

Similarly to Chapter 1, the agent can invest in ameliorating the risky arm, and the 'training' mechanism of doing so is exactly as in the breakthroughs model. At a cost κdt paid in dt, the bad risky arm becomes good with probability πdt (i.e. at the arrival rate π), while there is no effect on the a priori good arm. Moreover, once the first improvement, or success in training, occurred, the arm cannot be deteriorated or improved any further. Finally, the instance of the investment's success in not directly observable and can only be inferred from the absence of breakdowns. Given the breakdowns information structure, this implies that the agent is never perfectly sure whether the training succeeded, as she is never certain whether the arm is good.

The agent is looking to minimize the present value of the costs inflow. Following the notation from Chapter 1, denote by $\alpha_t \in [0,1]$ the intensity of experimentation in $[t, t+dt)$ and by $\beta_t \in [0,1]$ the intensity of investing in $[t, t+dt)$ conditional on experimenting. Then, in each moment in time the agent picks the optimal α_t and β_t that solve:

$$\min_{((\alpha_t,\beta_t)\in[0,1]^2)_0^\infty} E_{q_0}\left[\int_0^\infty e^{-rt}\left((1-\alpha_t)s + \alpha_t\left(q_t\lambda + \beta_t\kappa\right)\right)dt\right].$$

That is, the agent's costs value in each moment of time consists of the expected costs generated by the safe and the risky arms, weighted by intensities of being in use. She also exponentially discounts the future at rate $r > 0$. Finally, q_t denotes the agent's belief that the risky option is *bad* given the history available to her at time t.

2.2.2 Beliefs evolution

The following lemma states the evolution of beliefs, q_t, given the breakdowns information structure.

Lemma 2.1. *For any q_t and $\alpha_t > 0$, the risky arm is believed to be certainly bad conditional on a cost of size 1 generated by it: $q_{t+dt}^{news} = 1$. Conditional on no breakdown generated by the risky arm, beliefs follow: $dq_t = -\alpha_t q_t(\beta_t\pi + \lambda(1-q_t))dt$.*

Beliefs evolution depends on the information the agent observes in $[t, t+dt)$. Given some positive intensity of experimenting, the agent can observe a breakdown, which serves as a fully revealing signal that the risky arm is bad and that the training did not succeed, if any was done before time t ($q_t = 1$). Alternatively, the absence of a breakdown provides only very noisy information, and the effect of this can be decomposed into two. Firstly, the absence of news is a positive signal that the arm is slightly more likely to be good, which causes the belief to marginally decrease by $dq_t = -\alpha_t\lambda q_t(1-q_t))dt$. Secondly, due to training, the agent believes that the arm may have improved in the instant dt, which might have occurred if the arm was bad at time t (with probability q_t) and if the training succeeded in dt (with the probability $-\alpha_t\beta_t\pi dt$ - proportional to the intensity of training, β_t, and its productivity, π).

Notice that conditional on no news arriving the agent surely gets more optimistic (i.e. the belief that the arm is bad decreases). This happens because both the learning effect and the training one work in the same direction: no news is a good news, and improving the arm boosts the optimism even further. As such, training may be treated as a mechanism that speeds up learning, with the training intensity positively affecting the learning speed.

This is in contrast to slower learning with training in Chapter 1, where the two effects are counteracting each other. Overall, such dynamics ensures that conditional on no news beliefs evolve monotonically in the breakdowns model with ameliorating mechanism, as they would in the standard experimentation with Poisson bandits, so unless the agent exploits the safe arm, she converges to believing the risky arm is almost certainly good, $q_t \to 0$.

2.2.3 Bellman equation and the (costs) value function

The optimisation problem that the agent solves can be restated as an optimal control problem with the belief q_t as a single state variable. Given that the time dynamics is fully captured through the belief evolution path, from now on I omit the time subscripts. After standard manipulations, the Hamilton-Jacobi-Bellman equation (HJB) can be reduced to the following first-order differential-difference equation:

$$rV(q) = \min_{(\alpha,\beta)\in[0,1]^2} \big((1-\alpha)s + \alpha\,(q\lambda + \beta\kappa)\,\big) + \alpha q\lambda\big(V(1) - V(q)\big) - V'(q)\alpha q(\beta\pi + \lambda(1-q))$$

$V(q)$ is the costs value function that represents the present value of costs under optimally using the safe and the risky arms by the agent. [1] In its differential form, it consists of the expected costs that the agent bears today (i.e. in current dt) depending on the chosen strategy, and the change in value for the future (i.e. after dt passes) due to the new information obtained in dt. The future change in value can, in turn, be decomposed according to the possible signals structure. If the breakdown occurs (with probability $\alpha q\lambda$ in dt), the agent learns that the risky arm is bad, which causes a jump change in the belief from q to 1 and, as a result, a jump loss in terms of the continuation value, $V(1) - V(q)$. In the event of no bad news arriving (occurs with probability approaching 1), the agent smoothly updates her belief according to the law of motion described in Lemma 2.1, which marginally improves the continuation value through $V'(q)$. Overall, experimenting with the risky arm induces a jump loss from possibly discovering the breakdown and a smooth benefit from the absence of one, while a possibility to train the risky arm increases the smooth benefit relative to the jump loss.

Similarly to Chapter 1, the HJB equation is linear in its control variables, implying that for a given belief q the agent is ultimately choosing between the pure strategies of using the

[1] Henceforth, I will often refer to $V(q)$ simply as a value function. Yet, given that it represents the costs value and is thus in negative terms, I will say that the *value* to the agent increases as the *value function* decreases.

safe arm, purely experimenting without any training, and simultaneously experimenting and training the arm at a full intensity. The respective costs values following the solution of the Bellman equation are stated below:

$$V(q) = \min_{\alpha,\beta} \begin{cases} \frac{s}{r} & \alpha = 0 \text{ (safe arm)} \\ \frac{\lambda}{\lambda+r}\big(1+V(1)\big)q - C_{Exp}f(q) & \alpha = 1, \beta = 0 \text{ (experimentation)} \\ y(q,V(1)) - C_{Tr}g(q) & \alpha = \beta = 1 \text{ (training)} \end{cases} \tag{2.1}$$

where $y(q,V(1)) = q\frac{\lambda}{\pi+\lambda}\left((1+V(1))\frac{\pi+\lambda}{\pi+\lambda+r} + \frac{\kappa}{\pi+\lambda+r}\right) + \frac{\kappa}{r}\left(1 - q + q\frac{\pi}{\pi+\lambda}\right)$, $g(q) = \frac{(\frac{\pi}{\lambda}+1-q)^{1+\frac{r}{\lambda+\pi}}}{q^{\frac{r}{\lambda+\pi}}}$, $f(q) = \frac{(1-q)^{1+\frac{r}{\lambda}}}{q^{\frac{r}{\lambda}}}$, and $\{C_{Exp}, C_{Tr}\} \geq 0$ are some arbitrary constants.

The value function exhibits its traditional form and consists of two elements. The first component represents the costs value of using the same option forever. Under the safe arm, the agent pays the fixed cost s forever. With pure experimentation, she incurs the cost only if the risky arm is bad (with probability q), where the cost is delayed until $T \sim exp(\lambda)$ - the time of the first breakdown arrival - hence, discounted by $E_T[e^{-rT}] = \frac{\lambda}{\lambda+r}$, and is assumed to be of size 1 plus whatever continuation value the agent gets after learning the arm is bad, $V(1)$. Finally, the value of training the risky arm forever can be decomposed into three elements. If the risky arm is good (with probability $1-q$), it delivers no direct costs, but the agent trains it forever at a present value of cost, $\frac{\kappa}{r}$, since the information inflow never perfectly reveals that the arm is good and allows to seize costly training. If the arm is originally bad, the expected value it delivers depends on whether it is the breakdown or the success in training that arrives first. Denote by $\tau \sim exp(\pi)$ the moment when training succeeds, as opposed to $T \sim exp(\lambda)$ - the time of a first breakdown. With probability $\Pr(\tau < T) = \frac{\pi}{\pi+\lambda}$ the training is expected to succeed before the first bad news arrival, so the agent never gets any breakdown and keeps investing in the now good risky arm forever (pays $\frac{\kappa}{r}$). With the remaining probability, $\Pr(\tau > T) = \frac{\lambda}{\pi+\lambda}$, the breakdown is expected to occur before the training success. In this case, similarly to pure experimentation solution, the agent faces the cost 1 and continuation value of $V(1)$ discounted by time T in expectation, $E_\tau\left[E_T[e^{-rT}|T<\tau]\right] = \frac{\pi+\lambda}{\pi+\lambda+r}$, and pays the training cost during the time interval of T, $E_\tau\left[E_T[\int_0^T e^{-rt}\kappa dt|T<\tau]\right] = \frac{\kappa}{\pi+\lambda+r}$. Overall, as opposed to experimenting purely, training forever delays the occurrence of the breakdowns in expectation, but brings the necessity to pay for the training.

The second components in the value function represent the option value from the agent being able to dynamically adjust her behaviour as new information arrives and switch between the options instead of proceeding with the same action forever. Given the minimization nature of the problem and that the agent can opt out of switching between options, the option value cannot hurt the agent, and therefore enters negatively, with $g(q), f(q) > 0$ by construction and through imposing that $\{C_{Exp}, C_{Tr}\} \geq 0$. These free constants will be determined as part of the optimal strategy.

2.2.4 Benchmark with no possibility to train

A single agent solution from Keller and Rady (2015) provides a useful benchmark, as it states the optimal solution of the model from this Chapter in the absence of the possibility to train the risky arm. I outline their findings below (see the paper for the proof).

Lemma 2.2 (Benchmark)**.** *In the absence of training, the optimal solution follows a cutoff structure with the belief* $\hat{q} = \frac{s(\lambda+r)}{\lambda(s+r)} \in (0,1)$ *such that experimentation occurs for* $q < \hat{q}$*, and the safe arm is used otherwise. The solution is characterised by the piece-wise linear, weakly increasing value function with a concave kink at* $\hat{q}$*:* $V(q) = min\{\frac{s}{r}, \frac{\lambda}{\lambda+r}\left(1+\frac{s}{r}\right)q\}$*.*

The solution follows the familiar structure; however, its construction and the resulting dynamics are very different in comparison to the breakthroughs benchmark from Chapter 1. While the agent still reverts to the safe arm forever if the risky one is likely to be bad (i.e. for high values of belief q), under more optimistic beliefs experimenting with the risky arm with no news arriving makes the agent even more optimistic, and so she diverges away from the stopping cutoff $\hat{q}$ instead of converging to it. As a result of such dynamic, the agent seizes experimentation only once she observes a breakdown (as opposed to quitting due to becoming too pessimistic and gradually reaching the stopping cutoff after having no news for sufficiently long in the good news information environment). This implies that the only value she receives is from experimenting forever and switching to the safe arm after a first breakdown, so the value function is characterised by $V(1) = \frac{s}{r}$ and $C_{Exp} = 0$ - the option value component is zero since the agent never engages with any other options except for experimenting conditional on the absence of news.

At the cutoff belief, the agent is indifferent between experimenting and not, so the values from the safe and the risky arms should match there (the *value matching* condition). Together with normalizing the option value for experimenting to zero, this is sufficient to establish the cutoff $\hat{q}$. Notice that the *smooth pasting* requirement is necessarily violated. At the cutoff, the loss from experimenting, $-(s - \lambda\hat{q})$, is strictly positive, while the jump loss from learning the news is zero ($V(1) - V(\hat{q}) = \frac{s}{r} - \frac{s}{r} = 0$). Hence, the agent is only indifferent at the cutoff if the smooth benefit from experimentation covers the loss, implying that $\lim_{q \to \hat{q}_-} V'(q) > 0$.

I will use the solution from Lemma 2.2 as a benchmark and will preserve the notation of $\hat{q}$ and $V_{bench}(q)$ in what follows.

2.3 Main analysis

In this section I establish and analyse the solution to the model presented in Section 2.2. I start with summarising and discussing the optimal strategies in Subsection 2.3.1. I then provide the formal characterisation of the strategies, draw the comparison of the key results in the breakdowns model as opposed to the breakthroughs one, and discuss the qualitative impact of training on experimentation in Subsections 2.3.2, 2.3.3 and 2.3.4 respectively.

2.3.1 Key results

Similarly to the model with breakthroughs, the agent finds herself in one of the two regimes of experimentation. Yet the way how these are achieved is very different in the breakdowns model. The proposition below summarizes the main results.

Proposition 2.1. *Assume sufficiently low cost of training, $\kappa < \frac{s}{r}\pi$. Denote as $\underline{q}$ and $\overline{q}$ the lowest and highest belief, under which the agent trains the risky arm, respectively. Then, there exist $\pi^*(\kappa)$, such that:*

i. If $\pi < \pi^*(\kappa)$, the agent quits experimentation after a breakdown. She purely experiments for $q < \underline{q}$, trains the risky arm for $q \in [\underline{q}, \overline{q}]$, and uses the safe arm for $q > \overline{q}$.

ii. If $\pi > \pi^*(\kappa)$, the agent never quits experimentation even after a breakdown. She purely experiments for $q < \underline{q}$ and trains the risky arm for $q > \underline{q}$.

Under low training efficiency, the agent's behaviour follows the dynamic illustrated in Figure 2.1. She may start using and training the risky arm, which gradually makes her more optimistic, and then quit training at a certain deadline at $q = \underline{q}$ and experiment purely. If the arm is good originally, or if the training succeeds before the deadline at $\underline{q}$, she will keep experimenting forever (her belief will converge to $q \to 0$, but never reach 0 given no conclusive breakthrough signal is available to her). If the arm remains bad

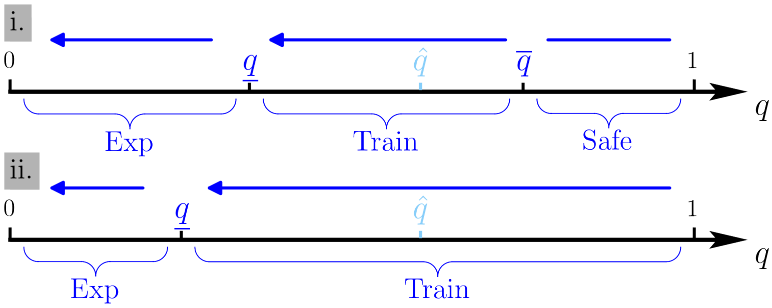

Figure 2.1: Agent's optimal strategy structure if *i.* $\pi < \pi^*(\kappa)$ and *ii.* $\pi > \pi^*(\kappa)$, as a function of belief q. The arrows indicate the direction of beliefs evolution in the absence of news.

by the deadline, as she keeps experimenting the breakdown will arrive with probability 1 almost surely in the limit and she will seize experimentation forever by switching to the safe option.

This solution structure suggests that the agent engages in training the risky arm only for some moderate beliefs. Given that the improvement mechanism affects only a priori bad arms, the benefit from training increases with the likelihood of the risky arm being bad, q. As such, whenever the agent is very optimistic, i.e. under low q, the benefit practically disappears, making training no longer worthwhile due to its costs. In fact, the agent purely experiments for the beliefs in a non-empty set $[0, \underline{q}]$ for any productivity π, even if the training costs are very low.

In contrast, whenever the agent is very pessimistic (under high q), the benefit of training is at its maximum. Yet, she understands that once she engages in training, she will become more attracted to using the risky arm and training it further for some time, since her belief q will monotonically decrease, and so she will do so until a breakdown arrives and makes her pessimistic again. Such path may be quite costly, especially if the cost of training is high and if its productivity is low (which increases a chance of getting a breakdown, making the paid costs meaningless). Instead, she may switch to using the safe arm forever immediately and save these costs. Overall, for high beliefs, the agent compares using the safe arm immediately against postponing it until the breakdown arrival. Under low productivity of training, the chance of the breakdown arrival is larger and keeps increasing with q, so she prefers to abandon the risky arm for high beliefs in $[\overline{q}, 1]$ and switch to the cheaper safe option forever.

As the training productivity gets larger (or its cost decreases), improving the arm becomes more worthwhile, and the training region expands until at a certain point the agent is willing to train even a surely bad risky arm (as illustrated in Figure 2.1 part ii.). This means that after each breakdown the agent prefers to start over with training the same risky arm rather than quit to the safe option. Hence, the agent truly *never gives up on experimenting.* While the arm remains bad, she finds herself in cycles between training and purely experimenting, where both activities make her more optimistic, and the breakdown arrival triggers a new cycle. Once the arm becomes good, the agent may train for a bit longer, as she cannot be sure if the arm is good given the information structure, and then purely experiment forever (as the breakdown would never arrive).

2.3.2 Formal characterisation

Proposition below gives formal characterisation of the solution.

Proposition 2.2. *Given sufficiently low cost of training, $\kappa < \frac{s}{r}\pi$, the agent's optimal strategy is fully characterized by a continuous, weakly increasing and globally concave value function that is smooth everywhere except for a single kink at $\overline{q}$ (whenever $\overline{q}$ exists).*

i. If $\pi < \pi^*(\kappa)$, then $\underline{q} \in (0, \hat{q})$, $\overline{q} \in (\hat{q}, 1)$, and

$$V(q) = \begin{cases} \frac{\lambda}{\lambda+r}\left(\frac{s}{r}+1\right)q & q < \underline{q} \\ y(q, \frac{s}{r}) - \left(y(\underline{q}, \frac{s}{r}) - \frac{\lambda}{\lambda+r}\left(\frac{s}{r}+1\right)\underline{q}\right)\frac{g(q)}{g(\underline{q})} & q \in [\underline{q}, \overline{q}] \\ \frac{s}{r} & q > \overline{q} \end{cases}$$

ii. If $\pi > \pi^*(\kappa)$, then $\underline{q} \in (0, \hat{q})$, and

$$V(q) = \begin{cases} \frac{\lambda}{\lambda+r}\left(V(1)+1\right)q & q < \underline{q} \\ y(q, V(1)) - \left(y(\underline{q}, V(1)) - \frac{\lambda}{\lambda+r}\left(V(1)+1\right)\underline{q}\right)\frac{g(q)}{g(\underline{q})} & q \geq \underline{q} \end{cases}$$

where $V(1)$ is defined by $V(1) = y(1, V(1)) - \left(y(\underline{q}, V(1)) - \frac{\lambda}{\lambda+r}\left(V(1)+1\right)\underline{q}\right)\frac{g(1)}{g(\underline{q})}$.

The value functions structure reflects the dynamics of the agent's optimal behaviour. When the agent purely experiments for optimistic beliefs in $[0, \underline{q}]$, she does not value the option of switching to other actions conditional on having no news, and so the value function contains only the linear component of experimenting 'forever'. The gradient of this linear component reflects the change in value conditional on observing a breakdown, $V(1)$, and so is determined as a part of the optimal solution: for lower training productivity the agent quits experimenting after a breakdown, so $V(1) = \frac{s}{r}$, while under high productivity,

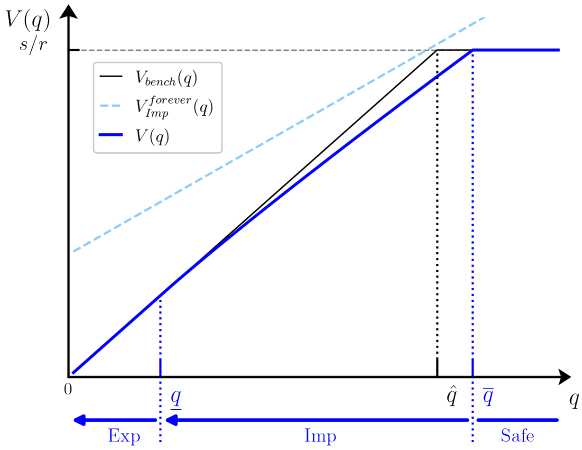

Figure 2.2: The optimal value function under low efficiency of training ($\pi < \pi^*(\kappa)$), as defined in Proposition 2.2, part i.

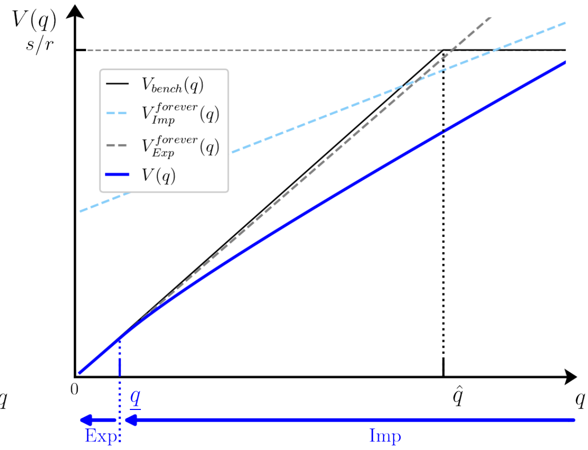

Figure 2.3: The optimal value function under high efficiency of training ($\pi > \pi^*(\kappa)$), as defined in Proposition 2.2, part ii.

she keeps experimenting and investing, and $V(1)$ is determined as the value from training the arm.

Once the agent trains the arm for beliefs above $\underline{q}$, she anticipates that as she gets more optimistic in the absence of news, she will quit costly training, which reduces the overall amount of costs she bears. The non-linear option value part in the function reflects the anticipated switch to pure experimentation and reduction in costs due to that. Notice that the value of training component, $y(q, V(1))$ is also endogenously determined through $V(1)$ - this is because the breakdown might occur during the training stage.

The lower cutoff $\underline{q}$ where the agent quits training is obtained by the indifference between training and experimenting (the value matching condition), as well as her marginal indifference, which results in the smooth pasting requirement. Intuitively, the agent becomes more optimistic around this boundary (q reduces), and thus quits training only once she does not want to train even for a single instance, which ensures the smooth transition. At the same time, $V(q)$ remains kinked at $\overline{q}$. This occurs due to the agent's diverging behaviour around the cutoff, similarly to the kink at $\hat{q}$ in the benchmark presented in Subsection 2.2.4. Such divergent behaviour necessarily results in the failure of the smooth pasting.

2.3.3 Breakdowns vs breakthroughs: Summary and comparison

Experimenting and training with breakdowns instead of breakthroughs leads to a very different dynamic, as suggested by Proposition 2.1 as opposed to Proposition 1.1 from

Chapter 1. Yet, some of the main results qualitatively hold for both information structures. I discuss whether and how the two main findings in the good news model - namely, the possibility to never quit experimenting and training until the news arrives and the potential non-monotonicity in actions conditional on the absence of news - survive in the bad news model.

The feature of experimenting 'forever' preserves for sufficiently efficient training mechanisms, in a sense that the agent is never attracted to switching to the safe option and finds herself with the good risky arm almost surely in the long-run. [2] Yet, the underlying mechanism and dynamic of achieving a similar qualitative outcome are very different. While in the breakthroughs model the result relies on the non-monotonicity in beliefs, so the agent converges to an interior one where she keeps experimenting and training at certain intensity until the good news arrives, this mechanism cannot occur with the breakdowns since the beliefs always decrease in the absence of news. Instead, with efficient training technology, learning the bad news does not discourage the agent enough to give up on the risky arm, as training it gives her a higher payoff than the safe option, and so she starts a new training-experimenting cycle.

Note that while both strategies achieve similar outcome of using the good risky arm in the limit, the dynamics of arriving there differs depending on the information structure. With the good news, depending on the prior belief, the agent either purely experiments and switches to investing at some intensity, or invests until the news arrives with potentially reducing training intensity at some point. The convergence to the reduced intensity of training can be interpreted as the agent alternating between short stages of training and experimenting purely. However, such dynamics is still very different from the cycles observed in the bad news environment, where the periods of training and purely experimenting last in time and are of randomly changing length each. Thus, the observed dynamics may be used as a predictor of the information structure in the environments where the type of receivable signals is unobserved.

The second key result, i.e. potential non-monotonicity in actions conditional on the absence of news, is not robust to the breakdowns model specification. Recall that non-monotonicity occurs under relatively low training efficiency, where in the good news model the agent purely experiments before and after the investment stage (i.e. in two disjoint

[2] As the good risky arms are never abandoned in the breakdowns model by construction (as the agent becomes more optimistic in the absence of news), a more insightful comparison is whether the agent is sufficiently incentivized to keep experimentation going if the arm may be bad and can generate a negative signal (either in the form of the absence of news with the breakthroughs, or as the bad news arrival with the breakdowns).

beliefs intervals). This is never possible in the bad news model, which relies on the belief dynamics under experimentation. While in the good news model the agent's incentives to train decrease as she becomes too pessimistic and anticipates terminating experimentation soon (captured through $V'(p) = 0$ at the stopping cutoff due to smooth pasting), with the bad news as the agent becomes more pessimistic her incentives for training grow stronger, since the training benefit keeps increasing in the belief q and there is no counter-acting force of the nearing experimentation termination. The latter is due to the agent's belief diverging away from the cutoff, and the violation of the smooth pasting there. [3] So, the agent is never incentivized to purely experiment for more pessimistic beliefs.

2.3.4 Implications of training on experimenting

Introducing the ability to ameliorate, or train, the risky arm undoubtedly increases the agent's value of experimenting, otherwise the agent would never engage in any. The proposition below breaks down the qualitative impact of training on experimentation.

Proposition 2.3. *An ability to train:*

i. Improves the value of experimenting for $q \in (\underline{q}, \overline{q})$ *if* $\pi \in (\underline{\pi}(\kappa), \pi^*(\kappa))$, *and for any* q *if* $\pi > \pi^*(\kappa)$, *where* $\underline{\pi}(\kappa)$ *is a training productivity level at which the agent is willing to train only at a single belief.*

ii. Necessarily broadens the incentives to experiment beyond $\hat{q}$ *for any* $\pi > \underline{\pi}(\kappa)$.

iii. Increases the probability of having a good arm in the long-run under at least some priors for any $\pi > \underline{\pi}(\kappa)$. *Specifically, this probability becomes 1 almost surely for any prior* q_0 *when* $\pi > \pi^*(\kappa)$.

Training has similar effects on the incentives to experiment no matter which of the two regimes the agent finds herself in. However, the strength of these effects differs. Trivially, whenever the agent engages in training, it reduces the costs of experimenting. Under relatively low productivity of training, this positive effect preserves only while the agent trains the arm (i.e. for $q \in (\underline{q}, \overline{q})$). Once she quits training, she never engages with it again, since she either keeps becoming more optimistic and thus experiments purely, or experiences a breakdown and quits the risky arm forever. As a result, the value is not affected by the presence of the improving mechanism. In contrast, for the high training

[3] Formally, the benefit of training over pure experimentation, $V'(q)\pi q$, can be shown to strictly increase for beliefs in the training range and above (while its cost is constant at κ). So if the agent trains at some belief q, training must strictly dominate pure experimentation for any beliefs above q.

productivity, the agent always believes that she might engage in training in the future: she does not train for low beliefs below $\underline{q}$, but as she is never certain that the risky arm is good, the probability of experiencing a breakdown is strictly positive for her, which will trigger a new training cycle. Anticipating this improves the value of purely experimenting, and so the overall costs decrease for any possible belief.

With the ability to train, the agent now experiments for a strictly larger set of beliefs. That is, no matter which of the regimes she finds herself in, she starts experimenting and training the arm while being more pessimistic than in the benchmark without training from Subsection 2.2.4, i.e. $\overline{q} > \hat{q}$ for any training mechanism (conditional on the mechanism being sufficiently attractive to be used). To understand this result, think about the incentives for the agent to train the arm under lowest attractive productivity level $\underline{\pi}(\kappa)$. She will then train the arm only for a single belief, which must coincide with the kink $\hat{q}$: for any higher belief marginal training has no impact as the agent is not willing to experiment in the next instance ($q + dq > \hat{q}$), while for any lower one, if the agent experiments at some q, she would also do so for higher beliefs in $(q, \hat{q})$, making lower productivity training also attractive. As such, the training region should expand around $\hat{q}$, implying that $\overline{q} > \hat{q}$.

Due to the information structure with breakdowns instead of breakthroughs, the agent can only discover whether the risky arm is bad, and can never confirm with certainty whether it is good. As such, introducing the ability to improve cannot directly affect the likelihood of discovering the good arms, but can increase the likelihood of the good arm being in use forever. Part iii of the Proposition 2.3 establishes that training necessarily increases such probability, trivially, by turning some of the bad arms into good before the first breakdown occurrence. Interestingly, under highly efficient training the probability of ending up with a good risky arm becomes 1 almost surely, since the agent has sufficient incentives to never terminate experimenting and occasional training, which ensures that any a priori bad arm will become good eventually.

The analysis above highlights the effects of training on experimenting through its ability to improve the a priori bad arms, i.e. its direct effect. Since any good arm in use will remain in use forever (due to never generating a breakdown), training cannot have any meaningful effect on the good arms already in use (e.g. affect the likelihood of their discovery as in the good news model in Chapter 1). Yet, there still exists a subtle indirect effect of training on the good arms.

Corollary 2.1. *Beyond improving the bad arms, training has a spillover effect on the a priori good arms through increasing the likelihood of them being used.*

Indeed, since the ability to train broadens the range of beliefs where the agent experiments, it also increases the range of beliefs for which the good risky arms start being in use. Specifically, for a non-empty set of prior beliefs in $(\hat{q}, \min\{\overline{q}, 1\})$, the risky arm may possibly be good (despite the probability of this being low), and training creates the incentive for the agent to start experimenting with it where the arm is not given a chance originally. Note that under the highly productive improvements, all the a priori good risky arms are in use, no matter what the prior likelihood of them being good was. To conclude, the positive influence of training the bad arms spreading beyond its direct effect is robust to the change in the information structure.

2.4 Concluding discussion

Overall, the ability to train the arm places the agent into one of the two experimenting regimes. In the more pessimistic one, training has a limited impact on the experimenting incentives and the optimal dynamics, where the agent engages with training only for some moderate beliefs, then quits her investments at a deadline as she gets more optimistic, but ultimately seizes experimenting if the breakdown arrives. As such, training improves the value to the agent while it lasts, as it increases the chances of the arm being good, but has no effect on the continuation value afterwards. This makes the outcome of the experimenting process qualitatively similar to the one achieved without the training opportunity. In contrast, training is highly impactful in the more optimistic regime where the agent becomes sufficiently incentivized to keep experimenting and training even after a breakdown, which ensures that she never quits experimenting and has a good arm in use in the limit almost surely.

Despite some of the results being qualitatively similar in the breakdowns and breakthroughs frameworks, the mechanisms of achieving those varies with the assumed information structure, as does the expected dynamics of actions given the training mechanism and signals structure. While the good news framework may generate as most one cycle of training-experimentation before the process termination (through non-monotonicity in actions under lower training efficiency), it is possible to generate infinitely many training-experimentation cycles with efficient training technologies and bad news signals. Due to the rich and non-trivial dynamics, the analysis presented in Chapters 1 and 2 may provide a rationale for observing different patterns of adaptation to new technologies in organisations, as well as draw predictions on the likelihood of successful adaptation depending on the organisational setup.

Note that considering the breakdowns information structure simplifies the optimal strategies and reduces the number of cases, while still exhibits the key qualitative features. Hence, such model specification may be appealing as a departing point to analyse various economically important extensions. One such question of interest may span the discussion of the impact of advance investments in developing the technologies, or improving them a priori, as opposed to the investments towards efficient implementation when the technology is already being adopted. [4]

The model presented in this Chapter can be easily adjusted to study such advance investments. Specifically, suppose that instead of training the arm conditionally on experimenting, the agent is only allowed to train it while using the safe one before the experimentation starts and cannot train the arm beyond this unique advance stage, with the training mechanism remaining exactly the same apart from its timing. Such modification decomposes the agent's decision problem into two stages, where she first decides under which beliefs she is willing to train the arm without experimenting and at which belief she quits such advance training. In the second stage, she then faces the standard experimentation problem as analysed in Section 2.2.4.

This version of the model allows to compare the effects of the pure advance training against the pure training while experimenting, and draw some conclusions on the importance of each type of investments. The proposition below summarizes these findings (without the proofs).

Proposition 2.4.

i. If the agent never trains conditionally on experimenting, then she is also never engages in advance training.

ii. If she is investing in implementation for non-empty $q \in [\underline{q}, \overline{q}]$, then the cutoff belief where the agent quits the prior investments is $q^* \in [\underline{q}, \overline{q}]$. Upper cutoff where she starts prior investments is above $\overline{q}$.

The result suggest that advance investments shift the training region towards more pessimistic beliefs, which has a twofold effect on the outcome of experimenting. On the one hand, expansion of the training region towards more pessimistic beliefs allows for even more of the a priori good arms to be used. On the other, the inability to learn with advance investments makes the agent more impatient to start experimenting sooner, so

[4] For instance, see Criscuolo and Narula (2007) and Popp (2006) for discussions about the two ivestment types in the contexts of multinational organisations and environmental policies development and execution.

she quits training while being more pessimistic, which increased the expected likelihood of the breakdown arrival. This simple result indicates the potential trade-off between the two investment types, suggesting that exploiting the baseline framework of this Chapter may yield economically interesting and non-trivial insights towards the suggested application.

References

Ball, Ian and Jan Knoepfle (2023) "Should the timing of inspections be predictable?" *Available at SSRN 4419716*.

Bardhi, Arjada, Yingni Guo, and Bruno Strulovici (2024) "Early-career discrimination: Spiraling or self-correcting?" *Working Paper*.

Bonatti, Alessandro and Johannes Hörner (2017) "Learning to Disagree in a Game of Experimentation," *Journal of Economic Theory*, 169, 234–269.

Criscuolo, Paola and Rajneesh Narula (2007) "Using multi-hub structures for international R&D: organisational inertia and the challenges of implementation," *Management International Review*, 47, 639–660.

Halac, Marina and Ilan Kremer (2020) "Experimenting with career concerns," *American Economic Journal: Microeconomics*, 12 (1), 260–288.

Halac, Marina and Andrea Prat (2016) "Managerial attention and worker performance," *American Economic Review*, 106 (10), 3104–3132.

Keller, Godfrey and Sven Rady (2015) "Breakdowns," *Theoretical Economics*, 10 (1), 175–202.

Lizzeri, Alessandro, Eran Shmaya, and Leeat Yariv (2024) "Disentangling Exploration from Exploitation,"Technical report, National Bureau of Economic Research.

Popp, David (2006) "Innovation in climate policy models: Implementing lessons from the economics of R&D," *Energy Economics*, 28 (5-6), 596–609.

Wagner, Peter A and Nicolas Klein (2022) "Strategic investment and learning with private information," *Journal of Economic Theory*, 204, 105523.

Chapter 3

Abstention as a Protest Vote

3.1 Introduction

Why do some people abstain from elections? This has been a remaining paradox in elections theory, with alternative reasons varying from bearing costs of participation or lack of information to base one's vote on to signaling or behavioral aspects like lack of ethical obligation. The paper aims at shedding the light on the issue and offers a novel explanation as to why people may strategically abstain from elections. In particular, in what follows I show that abstention can arise as a protest vote, which signals the society's misalignment with political status quo, that leaves out socially important and controversial matters.

To some extent social discontent with politics is present in many countries, independently of a country's political structure. For instance, consider the recent 'yellow vests' movement in France, where the core of the matter is claimed to lie in the citizens' dissatisfaction with policies orientation towards wealthy elites and ignoring the high hardly bearable costs of living for years. [1] Alternatively, think of the political situation in Russia, where the current establishment systematically neglects the lack of citizens' democratic rights and its actual convergence to autocracy, which evokes discontent in the society. The problem with these and many more potential examples is that despite the public sentiment existing politicians tend to exclude disturbing aspects from their political agenda and endorse the status quo silent political equilibrium at their benefit. Such neglecting behavior provokes discontent in the society and gives rise to citizens' desire to commu-

[1] See https://www.cnbc.com/2019/02/01/why-the-french-are-protesting.html and https://www.aljazeera.com/indepth/opinion/yellow-vests-protesting-france-181206083636240.html.

nicate their views and protest, so to say. In this paper I call this type of incentives a protesting motivation and explore how it affects one's decision to vote both positively and normatively.

My model builds upon the existing canonical voting models. In particular, I assume that the elections occur between two candidates, who are polarized in their attitude towards traditional salient political agenda and set their preferred policy upon winning the elections. The a priori centrist voters share an unobserved common shock about which candidate fits the environment better, and receive a private imperfect signal about the realization of this common preference. Such setup gives rise to so-called *election or pivotal motive* (see Feddersen, Pesendorfer (1996, 1997), Razin (2003), etc.), whereby the citizens participate in elections so as to contribute their private knowledge to select the better suiting candidate.

In additional to traditional policies, I assume that there are some non-salient ones that were historically excluded from political agenda. These could include any controversial topics like climate change response, policies towards various minorities (ethnical, LGBT, female rights, etc.), fight against corruption, inequality, insufficiency of political competition and other imperfections of existing political system, which societies actively debate about. To capture this intuition, I introduce another dimension of political response, broadly called an excluded or non-salient policy. I assume that the citizens are polarized in their attitudes towards this excluded policy, with each knowing their own preference, but being uncertain about which view is prevailing in the society overall. Hence, the citizens disagree in their preferences, but they all share the disagreement with the politicians' silence as a status quo. In response to such disagreement, I allow the candidates to be responsive to election outcome along this historically excluded dimension. Such extension gives rise to what I defined above as *protesting motive* among the citizens, which ascends from their desire to communicate the public debate to the politicians so that to shape the existing political agenda.

I study the citizens' behavior under the motives described above when they can support one of the candidates or abstain from the elections. In particular, I focus on a subset of pure perfect Bayesian equilibria, where all the citizens act symmetrically, and their decision on which candidate to support is independent from their views upon excluded policies. This class of equilibria is the most promising to achieve the full information and preferences aggregation and as such is of greater value to the society.

I show that under the assumptions above the citizens may use abstention as a signal of their views towards excluded policies – as a protest vote, so to say. In particular, under a simplifying assumption of the absence of any other motivations apart from the protesting one, I show that there always exists a separating equilibrium where one of the polarized groups votes and the other abstains, hence, effectively communicating the overall public sentiment in society and aggregating the preferences fully. This proves that a pure desire to protest suffices to explain imperfect (and non-zero) turnouts. Besides, this observation allows to justify why people vote in the environments where the election outcome is predetermined (in non-democracies, for instance).

I then extend the existence of separating equilibrium and, hence, efficient signaling about the public views to a general setup when the citizens face both election and protesting motives. I show that under the general setup, the supporters of one of the views face a trade-off between casting a vote for a better fitting candidate and thus enhancing information aggregation on a common value shock (election motive) at a cost of casting a wrong 'vote' along excluded policy dimension – that is, voting where the like-minded citizens pool on abstention and communicating wrong preference to the candidates. In other words, the election and protesting motives counter-act each other, which limits the existence of the separating equilibrium. I show that in large elections separation survives whenever the society is balanced enough (that is, the polarized views are supported by relatively equal shares of population) and when the private signals on common value shock are rather precise. The former implies that the tension in society and, thus, the protesting motive is the strongest, as well as the expected turnout is high enough no matter which view dominates. The latter, in turn, means that the information is well aggregated with less voters. Hence, both parameters contribute to relaxing the pressure on election incentive and make abstention (and separating equilibrium) more likely.

Finally, I conclude that in large elections it is possible to achieve full information aggregation and efficient preferences transmission simultaneously (under the existence of separating equilibrium). Whether this is desirable from the utilitarian standpoint (i.e. whether it maximizes the citizens' ex ante utility) depends on the citizens preferences: on their attitudes towards stability of political system in particular. If the citizens are strongly risk (instability) averse, they are more likely to tolerate the silent status quo at a cost of missing a chance to communicate their views on excluded policies. This is particularly so when the population is balanced enough in their views, which implies higher ex ante chances of policies moving in unfavorable direction. Hence, to overcome this inefficiency, the candidates should be moderately responsive on the excluded policy dimension. Note that in large elections the correct candidate selection occurs almost surely no matter what the turnout is, and given the citizens' common preferences, there is no information aggregation - efficiency tension along the traditional policies dimension.

Related literature This paper relates to several strands of literature. First, it contributes to the discussion on information aggregation in elections, which started from the seminal works of Austen-Smith and Banks (1996) and Feddersen and Pesendorfer (1996, 1997, 1999), who were the first to exploit the game theoretic approach in voting and consider the pivotal motives in how citizens vote rather than assume sincere voting throughout. These canonical models, mostly establishing possibility of full information aggregation in large electorates, were later enriched by bringing other incentives in consideration and often exhibiting the failure of information aggregation instead. These include adding various signaling motives (see more detailed discussion below) and, more recently, caring about the effect of a vote on victory margin (Herrera et al. (2019)). Instead of enriching incentives, Ekmekci and Lauermann (2020) consider the setup when turnout is correlated with the state of the world through the possibility of the elections organizer to manipulate the electorate and show the failure of information aggregation with pivotal motive solely.

The most relevant to this research extensions, however, are the ones that consider signaling in the elections setup. Many early models focus on signaling in sequential elections, where first elections help the voters to communicate their knowledge or preferences and through this affect consecutive ones. As such, Piketty (2000) explores how voters aggregate their information in common value environments to select better candidate in future, whereas in Shotts (2006) and Meirowitz and Tucker (2007) signaling of private preferences shapes the future politicians' behavior. In contrast with these models, Razin (2003) considers signaling in a single election. In particular, his signaling motive arises from candidates being responsive to the election outcome in their post-elections policies and ultimately interferes with the possibility to achieve the correct candidate selection and aggregate all the information simultaneously in common value environments. McMurray (2017a) extends such model to continuous states and signals and allows heterogenous precision of private information. In his model, citizens abstain if they are poorly informed or possess moderate beliefs about the true policy, so that not to make the policy response too extreme, and such abstention improves information aggregation in large elections. In this paper, I also consider signaling within a single election with the candidates' being responsive to election outcome, but my setup involves a combination of both private and common values and a two-dimensional policy response, which disentangles informational failure as in Razin (2003) and allows for efficient preferences communication through abstention.

This paper is also related to strategic abstention from elections. Originally, abstaining was thought to happen because of high voting costs (see Ledyard (1984) for an example). More advanced version of such theories introduces ethical voters, who apply an endogenously determined rule when deciding whether to vote (Feddersen, Sandroni (2006)). Alternatively, there are various studies about strategic reasons for abstention without any voting costs present. Pioneering in strategic abstention, Feddersen, Pesendorfer (1996) explore a model with a common value shock and show that abstention arises among uninformed independent voters who delegate their vote to better informed citizens. This result survives if a private value component (i.e. ideological preference) is added, given these ideological preferences are not too extreme (Feddersen, Pesendorfer (1999)), as well as in richer environments with continuous quality of information among the population or continuous states (McMurray (2013, 2017b)). Similarly, Herrera et al. (2019) emphasize that uninformed citizens can also abstain so that not to delude the victory margin by the marginal impact of their vote, which they claim to be a stronger motive than the swing voter's curse.

In addition to informational explanations for abstention, some studies focus on signaling through abstention. Shotts (2006) (and McMurray (2017a) similarly) shows that citizens with moderate views might choose to abstain so as to signal their centrist position and to drag post-elections policies to the less extreme ones. In contrast, I exploit a new interpretation of signaling with abstention, which is that abstaining communicates the polarized societal views and expresses the protest against the status quo silence on some socially valued issues. Kostadinova (2009) supports the similar intuition empirically on the example of concerns about corruption depressing the turnout in East European countries.

Last but not the least, this paper implies that the underlying political system is imperfect in a sense that it leaves some urgent issues out of salient political response, which is a feature prevailing in many non-democratic countries. As such, my model can be thought to represent elections in non-democracies and, thus, relates to the literature covering this topic. Broadly, such existing papers approach the question of why the elections are held in autocratic regimes despite the election outcome being often predetermined due to dictator's power to force it, and there are two prevailing explanations of this phenomenon. First, elections may work to aggregate the information on dictator's popularity and hence help in preventing costly revolutions (Cox (2009)). Alternatively, the format and results of elections can signal the strength of incumbent and their hold on power. Along these lines, Gehlbach and Simpser (2015) show that dictators manipulate elections to gain large victory margins in order to subjugate the bureaucrats. Somewhat similarly, Egorov and Sonin (2018) offer a plausible reason for choosing relatively fair versus non fair single candidate elections, according to which the incumbents use this or that type of election to signal their strength and non-fear of political rivals to the masses so as to prevent public protests.

All the papers above assume sincere voting throughout and do not account for electoral incentives. As such, I complement the existing literature by focusing on whether and why citizens choose to vote in non-democratic environments, study electoral incentives from the societal point of view and show that the protest and disagreement with politicians' silence can be costlessly expressed at the election stage through abstaining. This, in turn, may explain and support why the incumbents actively choose to hold elections: that is to aggregate the public sentiment and prevent costly protests in a timely manner, which vaguely resembles intuition in Cox (2009).

The paper proceeds as follows. Section 2 describes the model setup. Section 3 solves for a simplified benchmark where citizens' decision is driven purely by protesting motive, and then extends the results to general model allowing for both protesting and election motivations. In Section 4 I evaluate the existing equilibria from information aggregation and efficiency standpoints. I conclude the analysis in Section 5, present the proofs of main results in Section 6 and the references in Section 7.

3.2 Model

3.2.1 Political environment

There exists an established political agenda, which consists of a certain set of policies. For simplicity, all the traditional policies are pooled together to form a one-dimensional political response. Denote this as *a salient policy* $y \in \mathbb{R}$. However, the existing political system is imperfect in a sense that it remains silent on some socially important issues. This could include climate change response, corruption, domestic violence, authoritarianism of existing regime, insufficiency of political competition and other controversial topics. Assume that these excluded policies form another dimension of potential political response, and denote this dimension as *an excluded policy* $x \in \mathbb{R}$.

There are two candidates, L and R, who participate in a single election. In their views towards traditional policy y, the candidates are polarized symmetrically around zero, with $d \geq 0$ being a degree of polarization. That is, their policy responses upon being elected are equal to

$$y_L = -d, y_R = d.$$

However, both candidates have historically been silent about the excluded policy x, and this status quo of some political aspects being ignored is normalized to $x_0 = 0$. Given the silence was not broken before, it is assumed that the candidates' optimal outcome is to keep the status quo, i.e. candidates' bliss point is also $x_0 = 0$. However, the citizens may potentially misalign with the status quo silence, and prefer the politicians to introduce new policies either supporting or going against the controversial subjects. That is, if the citizens manage to communicate their sentiment to the politicians through elections, in order to prevent costly public protests and other negative consequences, the candidates might be forced to respond to the information transmitted to some degree. In particular, I assume that the candidates align in their response on excluded policies x and set them as

$$x_L(\Omega) = x_R(\Omega) = x_0 + \beta E[t|\Omega] = \beta E[t|\Omega],$$

With $t \in \mathbb{R}$ denoting the citizens' preference on the excluded policy (specified in subsection 2.2), and Ω corresponding to election outcome (discussed in subsection 2.3), the candidates' response is proportional to the expectation of citizens preferences based on observing election results. I assume that this response may be only partial, with $\beta \in [0, 1]$ representing the level of candidates' responsiveness to the information transmitted through the elections.

3.2.2 Populations and citizens

There are N (odd) citizens, who can potentially vote in elections. Each citizen has a certain two-dimensional preference over political system, which consists of an attitude towards traditional and excluded policies y and x.

All the citizens originally share the same bliss point $0 + s$ in terms of the traditional policy y, where s stands for a common value shock or state of the world, favoring one or the other candidate. The state space is symmetric with a shock taking values of $s \in \{-d,\ d\}$, each being equally likely. The exact realization of the state is unknown prior to the election for both candidates and electorate. However, each potential voter receives some imperfect information θ about the true state realization. These signals are distributed independently and identically across the electorate and take one of the two possible values $\theta \in \{l,\ r\}$, with a probability of a signal matching a state being $\Pr(l|s=-d) = \Pr(r|s=d) = p_\theta$, where parameter $p_\theta > \frac{1}{2}$ measures the precision of a signal.

In addition to this, the citizens are misaligned with the existing system in their silence towards excluded controversial aspects. As such, each citizen has a certain privately observed attitude towards these, which determines their type t. Each can either support, $t = t_f$ (for 'for'), or be 'against', $t = t_a$, a certain ignored aspect and would like the policies in line with their views to be introduced. I assume that the 'for' and 'against' types of citizens are polarized around zero, with $\gamma \geq 0$ representing the intensity of the views. Therefore, the t_f citizens share a common bliss point γ, while the bliss point of t_a type is $-\gamma$.

The distribution of preference across the society is unknown to both the members of the society and the candidates. I assume that the population can be of one of the two types, denoted by T – *mostly for* (F) or *mostly against* (A) excluded aspects, – with both types being equally likely a priori. In F population, each citizen is 'for' the excluded aspect with probability $\Pr(t_f|F) = p > \frac{1}{2}$ and 'against' otherwise, while in A population the majority of citizens is 'against' rather than 'for', and the likelihood of sharing the majoritarian view is also set as $\Pr(t_a|A) = p$. Hence, the population is equally biased in one of the directions in comparison to the status quo, however the direction of bias depends on the population type.

Given the preferences described above, each person's individual utility can be represented with a function as follows:

$$u(y,\ x,\ s,\ t) = u\left(|y-s| + |t-x|\right),$$

With $u(\cdot)$ strictly decreasing, concave, continuous and at least twice differentiable. $s \in \{-d,\ d\}$, $t \in \{-\gamma,\ \gamma\}$ are realizations of a state and a citizen's type respectively, while $y,\ x$ are two dimensions of responses set by a winning candidate after elections according to the rules described above. Overall, the citizens are interested in minimizing the traditional and excluded policies gaps $|y-s|$ and $|t-x|$, or, equivalently, in maximizing their expected utility, given uncertainty about the state realized and election outcome Ω at a time of casting a vote:

$$E\left[u(y,\ x,\ s,\ t)|\theta,\ t\right] = E\left[u\left(|y(\Omega)-s|+|t-x(\Omega)|\right)|\theta,\ t\right].$$

3.2.3 Game description and strategies

Let me now summarize the entire game. First, the nature determines the shock $s \in \{-d,\ d\}$ and the population type $T \in \{F, A\}$ (independently of each other). Each citizen perfectly learns their type $t \in \{t_f,\ t_a\}$ (which effectively signals the population type to them) and receives a signal $\theta \in \{l, r\}$, which is informative about the state realized. Based on her private information (θ, t), each citizen simultaneously decides whether she votes for a candidate L, R or abstains from elections, hence, their action space is $\mathcal{J} \in \{\mathcal{L}, \mathcal{R}, \mathcal{A}\}$. Once everyone made their decision, the election outcome $\Omega = (m, a)$ is publicly observed, with a out of N indicating the number of citizens who abstained from the election and m being the number of votes for candidate L. The candidate who gets higher vote share ($\frac{m}{N-a}$ for L vs. $\frac{N-a-m}{N-a} = 1 - \frac{m}{N-a}$ for R) wins the election, comes to power and sets the policies according to y_i and $x_i(\Omega)$, where $i = \{L, R\}$ is a winning candidate's index. In the event of a tie, the winner is picked randomly at 50:50 per cent chances. [2]

Similar to Razin (2003), in this paper I focus on symmetric perfect Bayesian equilibria (SPBE). By symmetry we mean that 1) citizens with identical information sets have identical strategies and 2) their strategies are symmetric with respect to candidates and signals θ. In particular, let $\Pr\left(\mathcal{A}|\theta,\ t_i\right) = \lambda_i(\theta)$ for $i = \{f, a\}$ be the probability that a type t_i citizen with a signal θ abstains, and $\Pr\left(\mathcal{L}|\theta,\ t_i\right) = \sigma_i(\theta)$ be a probability of voting for candidate L, with $\lambda_i(\theta) + \sigma_i(\theta) \leq 1$. Symmetry would imply that for each citizen $\Pr\left(\mathcal{L}|\theta, t_i\right) = \Pr\left(\mathcal{R}|\theta_-,\ t_i\right)$ for $\theta = \{l, r\}$, which is captured through the following:

Restriction 1 (R1. Symmetry with respect to candidates and signals θ)**.**

$$\forall i \in \{f, a\}\,,\quad \sigma_i(\theta) = 1 - \lambda_i\left(\theta_-\right) - \sigma_i\left(\theta_-\right).$$

Further to symmetry, I impose another condition to simplify the solution. In particular, I restrict attention to the equilibria where one's decision which candidate to vote for is independent of their excluded policies attitude, that is $\frac{\Pr(\mathcal{L}|\theta,\ t_i)}{\Pr(\mathcal{R}|\theta,\ t_i)} = c(\theta)$. Given Restriction 1 this can be written as follows:

[2] Note that here I assume that elections are declared legitimate no matter how low the turnout $(1-\frac{a}{N})$ is. The model can be extended to include the turnout barriers for elections legitimacy with the results remaining qualitatively similar to the original model.

Restriction 2 (R2. Irrelevance of citizen's type t).

$$\forall i \in \{f, a\}\,, \quad \frac{\sigma_i(\theta)}{1-\lambda_i(\theta)-\sigma_i(\theta)} = \frac{\sigma_i(\theta)}{\sigma_i\left(\theta_-\right)} = c(\theta).$$

Restriction 2 logically fits the model as the citizen's type t does not convey any additional information about the state of the world, whereas the candidates differ solely in their response to the state. Hence, it should be irrelevant to consider the type when making a choice on which candidate is a preferred one.

The following lemma investigates implications of the above mentioned restrictions for solving the model. Let $q_{\mathcal{J}}(s,\ T)$ for $\mathcal{J} \in \{\mathcal{L}, \mathcal{R}, \mathcal{A}\}$ be the probabilities that given a true state s and population type T a *randomly picked citizen* abstains from elections, votes for candidate L and votes for R respectively. Note that these are expressed by

$$q_{\mathcal{J}}(s,\ T) = \begin{cases} \sum_{\theta=l,r} \sum_{i=f,a} \Pr\left(\theta|s\right) \Pr\left(t_i|T\right)\lambda_i(\theta), & \text{if } \mathcal{J} = \mathcal{A} \\ \sum_{\theta=l,r} \sum_{i=f,a} \Pr\left(\theta|s\right) \Pr\left(t_i|T\right)\sigma_i(\theta), & \text{if } \mathcal{J} = \mathcal{L} \\ \sum_{\theta=l,r} \sum_{i=f,a} \Pr\left(\theta|s\right) \Pr\left(t_i|T\right)\left((1-\lambda_i(\theta)-\sigma_i(\theta))\right), & \text{if } \mathcal{J} = \mathcal{R} \end{cases}.$$

Lemma 3.1. *Assume individual strategies satisfy restrictions R1 and R2. Then,* $\forall\ T \in \{F, A\}$,

i. $\forall\ s \in \{-d, d\}\,,\ q_{\mathcal{A}}(s, T) = \overline{\lambda}_T$, *where* $\overline{\lambda}_T = \sum_{i=f,a} \Pr\left(t_i|T\right)\lambda_i$, *and* $\forall i \in \{f, a\}\,,\ \lambda_i(l) = \lambda_i(r) = \lambda_i$.

ii. $q_{\mathcal{L}}(-d, T) = q_{\mathcal{R}}(d, T) = q\left(1 - \overline{\lambda}_T\right)$ *and* $q_{\mathcal{L}}(d, T) = q_{\mathcal{R}}(-d, T) = (1-q)\left(1-\overline{\lambda}_T\right)$, *where* $q \in [1-p_\theta, p_\theta]$.

Hence, to specify any perfect Bayesian equilibrium that satisfies Restrictions R1 and R2 it is sufficient to determine three parameters, $\overline{\lambda}_F$, $\overline{\lambda}_A$ and q, which stand for the aggregate strategies. $\overline{\lambda}_F$ and $\overline{\lambda}_A$ indicate the average probability of abstention in F and A populations respectively. Note, that they are independent of the state realized (as well as of signals θ – see the proof), meaning that in any such equilibrium the number of abstainers a would be a signal of population type purely. At the same time, q represents the probability of voting for the state matching candidate given a randomly picked citizen votes, which is independent of her type. Thus, given the turnout $N-a$, the number of L votes m would signal the shock realization purely. To simplify the notation in further derivations for $\mathcal{J} \in \{\mathcal{L}, \mathcal{R}\}$ denote $\widetilde{q}_{\mathcal{J}}(s) \equiv \frac{q_{\mathcal{J}}(s,T)}{(1-\overline{\lambda}_T)}$, where $\widetilde{q}_{\mathcal{J}}(s) \in \{q,\ 1-q\}$ exactly highlights its independence from T.

Overall, the two imposed restrictions and Lemma 1, as a result, limit the scope of equilibria analysis to the ones where turnout and election results may convey information about both the state of the world and the population type. These equilibria are of interest since they potentially allow for the most efficient information transmission from citizens to political leaders.

3.3 Model analysis

3.3.1 Election outcome and politicians' response

To specify the citizens' behavior, we first need to consider how the election outcome affects the policies set by a winning candidate. Define the election outcome $\Omega = (m_N, a_N)$, where a_N and m_N are random variables with a_N being a number of abstainers out of N citizens and m_N indicating the number of votes for candidate L. Note that the probability of a certain outcome Ω given the model primitives is then a joint probability of (m_N, a_N), which follows trinomial distribution:

$$\Pr\left(m, a | s,\ T\right) = C_N^a C_{N-a}^m q_{\mathcal{A}}(s,T)^a q_{\mathcal{L}}(s,T)^m q_{\mathcal{R}}(s,T)^{N-a-m}.$$

In particular, using results established by Lemma 1, we can separate out the effects of population type and state between each other:

$$\Pr\left(m, a | s, T\right) = C_N^a \overline{\lambda}_T^a \left(1 - \overline{\lambda}_T\right)^{N-a} C_{N-a}^m \widetilde{q}_{\mathcal{L}}(s)^m \widetilde{q}_{\mathcal{R}}(s)^{N-a-m}.$$

The property of separability between probabilities of voting for a particular candidate ($\widetilde{q}_{\mathcal{L}}(s)^m \widetilde{q}_{\mathcal{R}}(s)^{N-a-m}$) and of voting versus abstaining ($\overline{\lambda}_T^a \left(1 - \overline{\lambda}_T\right)^{N-a}$) is of great convenience, since it allows to interpret the two-dimensional election outcome in the most informative way, which is summarized in the following lemma.

Lemma 3.2. [3] *Assume restrictions R1 and R2 are imposed.*

[3] The proof follows directly from Bayes rule (see Appendix).

i. When updating $\Pr(T|m,a)$*, only turnout is relevant, i.e.* $\forall\ T \in \{F, A\}$*,*

$$\Pr(T|m,a) = \Pr(T|a) = \frac{\overline{\lambda}_T^a \left(1-\overline{\lambda}_T\right)^{N-a}}{\sum_T \overline{\lambda}_T^a \left(1-\overline{\lambda}_T\right)^{N-a}}.$$

ii. When updating $\Pr(s|m,a)$*, probability of abstaining is irrelevant, i.e.* $\forall\ s \in \{-d, d\}$*,*

$$\Pr(s|m,a) = \frac{\widetilde{q}_{\mathcal{L}}(s)^m \widetilde{q}_{\mathcal{R}}(s)^{N-a-m}}{\sum_s \widetilde{q}_{\mathcal{L}}(s)^m \widetilde{q}_{\mathcal{R}}(s)^{N-a-m}}.$$

Notice that the lemma shows how the election outcome conveys two independent signals about the primitives of the model. In particular, the turnout signals population type purely (through number of abstainers a), while the voting outcome given turnout (m out of $N-a$) purely aggregates the individual messages about the realized state of the world. This efficient transmission of two-dimensional knowledge rises potential for full information aggregation through elections in equilibrium.

Let us consider the candidate's behavior upon winning the elections. Recall that being left or right partisan with certainty, the winning candidate is only responsive in terms of policies typically excluded from political agenda, aggregate preference for which is determined by population type. Therefore, he rationally updates his belief about the population type as in Lemma 2(i) and uses this belief to set an optimal response $x_i(\Omega), i \in \{L, R\}$. Corollary below summarizes how the policies are set.

Corollary 3.1. *Candidates shift excluded policies based on turnout only, i.e.*

$$x_L(\Omega) = x_R(\Omega) = x(a), \text{ and}$$

$$x(a) = \beta E[t|a] = \beta\overline{\gamma}\frac{\overline{\lambda}_F^a \left(1-\overline{\lambda}_F\right)^{N-a} - \overline{\lambda}_A^a \left(1-\overline{\lambda}_A\right)^{N-a}}{\sum_T \overline{\lambda}_T^a \left(1-\overline{\lambda}_T\right)^{N-a}},$$

where $\overline{\gamma} \equiv \gamma(2p-1)$.

The Corollary follows directly from Lemma 2, and unsurprisingly, confirms that the candidates are only responsive to turnout, rather than the voting outcome per se.

3.3.2 Benchmark case. Non-polarized candidates

In this subsection I consider a simplified version of the model described in Section 2. In particular, let me assume for now that there is no difference in the states of the world s, or equivalently the candidates are not polarized, that is $d = 0$. Hence, there is no uncertainty about traditional policies dimensions, and signals $\theta \in \{l, r\}$ become redundant as they neither affect the candidates' policy nor shape the voters preferences, as their bliss point is now $(0,\ t_i)$ depending on their attitude towards excluded policy solely. This means that the citizens are indifferent about the voting outcome per se, and are only motivated by signaling their type through turnout. In this subsection I show that this motivation alone suffices for the citizens to participate in elections, despite the voting per se being meaningless to them. In particular, citizens may use elections as a mechanism to signal their views on excluded aspects and costlessly indicate their protest against the existing political agenda. I call this type of incentive a *protesting motivation* (PM).

Looking at this simplified model may shed the light on why the citizens tend to participate in elections which are publicly known to be a 'pure farce' (following notation of Egorov, Sonin (2020)), where there is either a single candidate running unopposed, or only 'nominal' opposition is allowed to run, or all candidates are likely to set very similar policies, or the vote counting is commonly known to be unfair. All in all, given that the policy dimension plays no role under the simplified setup one can assume any type of political system where the election outcome is obvious in advance or where the policies are not really affected by who wins the election. For example, such scenarios are very likely to occur in non-democratic countries. Hence, this section can explain why some people may actively choose to participate in non-democratic elections, that is, they do so to signal their agreement (or acceptance) with the existing regime. On the contrary, by abstaining one signals their protest upon political agenda.

Let me theoretically characterize the protesting motive described above. First, define $\Delta(\theta,\ t_i)$ to be a difference in a citizen's expected utility from voting in election (denoted by $\mathcal{V}$) versus abstaining given her private information $(\theta,\ t_i)$, that is $\Delta(\theta,\ t_i) = E\left[u_{\mathcal{V}}(y,\ x,\ s,\ t)|\theta,\ t_i\right] - E\left[u_{\mathcal{A}}(y,\ x,\ s,\ t)|\theta,\ t_i\right]$. Note that, as discussed above, in this section the traditional policy gap $|y(\Omega) - s|$ is redundant irrespectively of election outcome, that is $|y(\Omega) - s| = 0$, as well as the signals θ are irrelevant, i.e. $\Delta(\theta,\ t_i) = \Delta(t_i)$. Given the candidates agree in their response towards the excluded policy shift x, voting for candidate L versus R does not affect citizens utility, hence $E\left[u_{\mathcal{L}}(x,\ t_i)\,|t_i\right] = E\left[u_{\mathcal{R}}(x, t_i)\,|t_i\right] =$

$E\left[u_{\mathcal{V}}\left(x,t_i\right)|t_i\right]$. As such, the difference is specified as:

$$\begin{aligned}\Delta\left(\ t_i\right) &= E\left[u_{\mathcal{V}}\left(x,t_i\right)-u_{\mathcal{A}}\left(x,t_i\right)|t_i\right] \\ &= \sum_{a=0}^{N-1}\sum_{m=0}^{N-a-1}\widetilde{\Pr}\left(m,a|t_i\right)\left[u\left(|t_i-x(a)|\right)-u\left(|t_i-x(a+1)|\right)\right] \\ &= \sum_{a=0}^{N-1}\widetilde{\Pr}\left(a|t_i\right)\left[u\left(|t_i-\beta E\left[t|a\right]|\right)-u\left(|t_i-\beta E\left[t|a+1\right]|\right)\right],\end{aligned}$$

or one can derive the linearized version as

$$\Delta\left(t_i\right)\cong u^{'}\sum_{a=0}^{N-1}\widetilde{\Pr}\left(a|t_i\right)\left(|t_i-\beta E\left[t|a\right]|-|t_i-\beta E\left[t|a+1\right]|\right),$$

where $u^{'}<0$.

In above $\widetilde{\Pr}\left(m,a|t_i\right)$ and $\widetilde{\Pr}\left(a|t_i\right)$ denote probabilities of a certain election outcome and turnout, respectively, as *perceived by a citizen*, i.e. without her own decision taken in account and given her private information. That is, in the form where both s and T matter, the perceived probabilities are:

$$\widetilde{\Pr}\left(m,a|\theta,\ t_i\right)=\sum_{T}\sum_{s}\widetilde{\Pr}\left(m,a|s,T\right)\Pr\left(s|\theta\right)\Pr\left(T|t_i\right),\ \text{and}$$

$$\widetilde{\Pr}\left(a|\theta,\ t_i\right)=\sum_{m=0}^{N-a-1}\widetilde{\Pr}\left(m,a|\theta,\ t_i\right)=\sum_{T}\widetilde{\Pr}\left(a|T\right)\Pr\left(T|t_i\right)=\widetilde{\Pr}\left(a|t_i\right),$$

with $\widetilde{\Pr}\left(m,a|s,T\right)$ following trinomial distribution of $N-1$ events (i.e. the joint probability of (m_{N-1},a_{N-1}), instead of (m_N,a_N)). Similarly, $\widetilde{\Pr}\left(a|T\right)$ obtained as a sum across all values of m follows binomial distribution: $C_{N-1}^{a}\overline{\lambda}_T^{a}\left(1-\overline{\lambda}_T\right)^{N-a-1}$. Note that in what follows I will always use tilde probabilities and expectations for the objects *perceived by a citizen*, that is out of $N-1$ without her vote counted, whereas all the objects without tilde denote the ones *perceived by candidates* – out of N observations.

Given the model specification for this section, when considering which decision to make, a citizen only cares about how she affects the candidates' beliefs about population type and therefore, the excluded policy shifts. That is, her whole utility change arises from *the protesting motivation (*PM*)*, as defined above:

$$\Delta\left(t_i\right)=\mathrm{PM}\left(t_i\right)=u^{'}\sum_{a=0}^{N-1}\widetilde{\Pr}\left(a|t_i\right)\left(|t_i-\beta E\left[t|a\right]|-|t_i-\beta E\left[t|a+1\right]|\right).$$

The protesting motivation aggregates the weighted change in excluded policies gap which voting against abstaining causes, for each possible abstention rate. Note that given each citizen perfectly knows her type, there is no uncertainty about the regime gaps on top of the uncertainty about the election outcome. The following lemma simplifies the expressions for protesting motivations for $i \in \{f, a\}$ and analyses how the motive prescribes to behave.

Lemma 3.3. *Protesting motivations are always of the opposing sign. In particular,*

i. $PM(t_f) = u'\beta \sum_{a=0}^{N-1} \widetilde{\Pr}(a|t_f)(E[t|a+1] - E[t|a]) = \begin{cases} < 0, & if\ \overline{\lambda}_F > \overline{\lambda}_A \\ = 0, & if\ \overline{\lambda}_F = \overline{\lambda}_A \\ > 0, & if\ \overline{\lambda}_F < \overline{\lambda}_A \end{cases}$;

ii. $PM(t_a) = -u'\beta \sum_{a=0}^{N-1} \widetilde{\Pr}(a|t_a)(E[t|a+1] - E[t|a]) = \begin{cases} > 0, & if\ \overline{\lambda}_F > \overline{\lambda}_A \\ = 0, & if\ \overline{\lambda}_F = \overline{\lambda}_A \\ < 0, & if\ \overline{\lambda}_F < \overline{\lambda}_A \end{cases}$.

To see the logic behind the lemma, observe that any citizen would have a more extreme view than how the candidates respond in their excluded policy update (unless $\beta = 1$ and $p_\gamma = 1$). Firstly, this is due to the fact that by setting the policy according to the population average views they aim to satisfy the preferences of the whole polarized population and would not go too extreme to prevent discontent of the other group. On top of that, candidates are prone to the status quo $x_0 = 0$ and shrink their response accordingly with $\beta \leq 1$. As such, the policy gap $t_i - \beta E[t|a]$ is always positive for the 'for' type (recall that $t_f = \gamma$) and negative for the 'against' one ($t_a = -\gamma$). Therefore, the protesting motivation is determined solely from the change in perceived population type from additional abstainer, and it works in opposing directions for the two types.

The exact sign of the motive, in turn, depends on the interpretation of turnout in equilibrium. If higher abstention transmits stronger 'for' signal, then 'for' citizens prefer to abstain and thus, $\text{PM}(t_f) < 0$, whereas 'against' citizens opt in to vote to signal their view ($\text{PM}(t_a) > 0$). Vice versa, if higher turnout signals 'for' views, then the 'for' type votes, $\text{PM}(t_f) > 0$, and 'against' one abstains, $\text{PM}(t_a) < 0$.

Whenever $\overline{\lambda}_F = \overline{\lambda}_A$, that is, both types behave identically, protesting motivations for both types are zero. This is explained by the fact that given identical behavior, the candidates cannot update their belief about the population preferences anyhow. Hence, they become unresponsive to election outcome, which makes the possibility of protesting through the elections impossible.

The equilibria existence in the presence of protesting motivation solely follows directly from Lemma 3 and is summarized in the proposition below. Given the choice who to vote for is redundant (can be anything without changing the remainder of the game), any equilibrium is fully characterized by the probabilities of abstention for each type of citizens, λ_f and λ_a.

Proposition 3.1 (equilibria existence)**.** *Assume $d = 0$. For any population size N, there always exist three equilibria:*

i. A pooling equilibrium where both types are equally likely to abstain, i.e. $\lambda_f = \lambda_a \in [0, 1]$.

ii. A pair of separating equilibria where one type always votes and the other abstains, i.e. $\lambda_f = 1$, $\lambda_a = 0$, *and* $\lambda_f = 0$, $\lambda_a = 1$.

First, there always exists a pooling equilibrium where both types behave identically, that is both abstain with the same probabilities, $\lambda_f = \lambda_a$. That means that the election results do not convey any information about the public sentiment, and the candidates are not able to update their beliefs about the population type. That being said, the citizens understand that their behavior is uninformative, hence, this kills the protesting motive they might potentially have, and they indeed randomize at any possible common probabilities. Needless to say, this outcome is highly inefficient as it fails to convey any message through elections and the question about why vote arises.

More importantly, there always[4] exists a pair of fully separating equilibria, in which one type of citizens always abstains, while the other one always votes. Given that the citizens are polarized at different extremes around candidate's response x, each type has an incentive to separate themselves from the other type and as such promote their group's interest. Hence, the turnout becomes a perfect signal of the aggregate views in the society, which the winning candidate optimally responds for (so as to prevent potentially costly protests later). Thus, the outcome is fully efficient in terms of transmitting information, and in such equilibria elections serve as an efficient mechanism which manages to communicate a message about the public sentiment.

[4]Existence of separating equilibria on the whole parameter space relies on the citizens' types bliss points being polarized around the status quo $x_0 = 0$ and on the symmetry and the absence of uncertainty about the bliss points. If the bliss points of both types were one sided from the status quo (say, represent moderate and extreme misalignment), then separating equilibrium would only survive if the candidates are responsive enough and the gap between the types is large enough. Under these conditions, the other results of the paper remain unchanged in one sided specification of the model.

3.3.3 The general model analysis

For this section return to the original model where the candidates and the states of the world are polarized, that is $d > 0$. Now, the citizens care about the common shock and prefer a candidate that better fits the circumstances to be in power, since it minimizes the policy gap for them. As such, if one thinks that $s = -d$ is more likely, then they would have an incentive to cast a vote for candidate L, and vice versa. This creates additional motive for participation in election, which, following Razin (2003), I define as *election motivation (EM)*.

Denote each expected utility difference given a citizen's private information as $\Delta_{\text{kj}}(\theta, t_i)$, with a subscript kj standing for the difference in utilities from action k to action j, where $k, j \in \{\mathcal{A}, \mathcal{L}, \mathcal{R}\}$. Then, similar to section 3.2, the citizen's new linearized changes in utilities are

$$\Delta_{\text{kj}}(\theta, t_i) = E\left[u_k(y,\ x,\ s,\ t_i) - u_j(y,\ x,\ s,\ t_i)\,|\theta, t_i\right]$$
$$\cong u^{'} \sum_{a=0}^{N-1} \sum_{m=0}^{N-a-1} \widetilde{\Pr}(m, a|\theta, t_i)\, E\left[|t_i - x(\Omega_k)| + |y(\Omega_k) - s| - |t_i - x(\Omega_j)| - |y(\Omega_j) - s||m, a, \theta, t_i\right].$$

Decompose the change in utilities into protesting motivation $\mathrm{P}M_{\text{kj}}(\theta, t_i)$ and election motivation $EM_{\text{kj}}(\theta, t_i)$ as:

$$\Delta_{\text{kj}}(\theta, t_i) = PM_{\text{kj}}(\theta, t_i) + EM_{\text{kj}}(\theta, t_i)\,,\ \text{where}$$

$$PM_{\text{kj}}(\theta, t_i) = u^{'} \sum_{a=0}^{N-1} \sum_{m=0}^{N-a-1} \widetilde{\Pr}(m, a|\theta, t_i)\, E\left[|t_i - x(\Omega_k)| - |t_i - x(\Omega_j)||m, a, \theta, t_i\right]$$
$$= u^{'} \sum_{a=0}^{N-1} \widetilde{\Pr}(a|t_i)\left(|t_i - \beta E[t|a_k]| - |t_i - \beta E[t|a_j]|\right) = PM_{\text{kj}}(t_i)\,,\ \text{and}$$

$$EM_{\text{kj}}(\theta, t_i) = u^{'} \sum_{a=0}^{N-1} \sum_{m=0}^{N-a-1} \widetilde{\Pr}(m, a|\theta, t_i)\, E\left[|y(\Omega_k) - s| - |y(\Omega_j) - s||m, a, \theta, t_i\right],$$

with $\Omega_{\mathcal{A}} = (m, a+1)$, $\Omega_{\mathcal{L}} = (m+1, a)$ and $\Omega_{\mathcal{R}} = (m, a)$ denoting the election outcome given a abstainers and m L votes out of $N-1$ citizens and the remaining citizen choosing a subscript action. Observe that the protesting motivation, $PM_{\text{kj}}(\theta, t_i)$, is of a form identical to the one specified in section 3.2. This relies on the candidates updating their policy along the excluded dimension based on abstention only (according to Corollary 1), which makes the number of votes for this or that candidate irrelevant and enables to sum the probabilities over m for each given abstention rate. Hence, just as in the benchmark case, the protesting motivation is affected neither by the number of votes m, nor, therefore, by the state signals θ: that is, $PM_{\text{kj}}(\theta, t_i) = PM_{\text{kj}}(t_i)$.

In addition to the protesting motive, given candidates' polarization the exact winner of the election now matters. This affects the traditional policy gap, $|y(\Omega) - s|$, which gives rise to election motivation, $EM_{\text{kj}}(\theta, t_i)$, as specified above. Observe that whenever a citizen's decision does not affect the winner of the elections, the policy gap remains constant and the election motive shrinks. However, once a citizen believes that the elections are at a close margin (i.e. she is pivotal), she cares about who to cast the vote for, which ultimately affects her decision. Based on this logic, one can simplify the election motivation to

$$\begin{aligned}
EM_{\text{kj}}(\theta, t_i) &= u' \sum_{a=0}^{N-1} \sum_{m=0}^{N-a-1} \widetilde{\Pr}(m, a|\theta, t_i) \, E\left[|y(\Omega_k) - s| - |y(\Omega_j) - s||m, a, \theta, t_i\right] \\
&= u' \sum_{a=0, \text{ even}}^{N-1} \left[\widetilde{\Pr}\left(\frac{N-a-1}{2}, a|\theta, t_i\right) E\left[|y(\Omega_k) - s| - |y(\Omega_j) - s||m, a, \theta, t_i\right]\right] \\
&+ u' \sum_{a=1, \text{ odd}}^{N-2} \begin{bmatrix} \widetilde{\Pr}\left(\frac{N-a}{2} - 1, a|\theta, t_i\right) E\left[|y(\Omega_k) - s| - |y(\Omega_j) - s||m, a, \theta, t_i\right] \\ +\widetilde{\Pr}\left(\frac{N-a}{2}, a|\theta, t_i\right) E\left[|y(\Omega_k) - s| - |y(\Omega_j) - s||m, a, \theta, t_i\right] \end{bmatrix}.
\end{aligned}$$

As discussed above, $EM_{\text{kj}}(\theta, t_i)$ only takes into account the cases when a citizen is potentially pivotal. In particular, given the number of voters can be both odd and even, depending on a number of abstainers, the pivotal cases include the ones where the votes are spread evenly for the two candidates ($\frac{N-a-1}{2}$ vs. $\frac{N-a-1}{2}$, when a is even) and where one of the candidates is leading by one vote ($\frac{N-a}{2} - 1$ *vs.* $\frac{N-a}{2}$ or vice versa, if a is odd). Hence, there is a non-zero chance that the candidates will tie. For this case I assume that the winner is determined by a fair coin toss, each chosen to set a policy with probability $\frac{1}{2}$.

The following lemma explores which behavior each of the motivations promotes, for all the possible actions. For the remainder of the paper, let me assume that the 'for' citizens are at least as likely to abstain as the 'against' ones, i.e. $\overline{\lambda}_F \geq \overline{\lambda}_A$. This allows, without loss of generality, to get rid of duality of separating equilibria in Proposition 1 and instead to focus on the persistence of the equilibria in the presence of additional voting motives.

Lemma 3.4. *Assume* $\overline{\lambda}_F \geq \overline{\lambda}_A$.

i. When considering a choice to vote for candidate L *vs* R, $\forall t_i$:

$$PM_{LR}(t_i) = 0;$$

$$EM_{LR}(l, t_i) > 0 \;\; and \; EM_{LR}(r, t_i) < 0;$$

ii. When considering a choice to vote for candidate L vs abstain:

$$PM_{LA}(t_f) = \begin{cases} = 0 \; if \; \overline{\lambda}_F = \overline{\lambda}_A \\ < 0 \; if \; \overline{\lambda}_F > \overline{\lambda}_A \end{cases} \; and \; PM_{LA}(t_a) = \begin{cases} = 0 \; if \; \overline{\lambda}_F = \overline{\lambda}_A \\ > 0 \; if \; \overline{\lambda}_F > \overline{\lambda}_A \end{cases};$$

$$\forall t_i : \quad EM_{LA}(l, t_i) > 0 \;\; and \; EM_{LA}(r, t_i) < 0;$$

iii. Motivations are symmetric with respect to candidates and signals θ, i.e. $\forall t_i$:

$$PM_{LA}(t_i) = PM_{RA}(t_i);$$
$$EM_{LA}(\theta, t_i) = EM_{RA}(\theta_-, t_i).$$

As discussed above, the protesting motive is of a form identical to the one in section 3.2. It turns out that for the excluded policy gap considerations there is virtually no difference between which candidate to support even with polarized candidates. This is due to the politicians agreeing in their response, as well as updating their excluded policy based on turnout only, as discussed in section 3.1. Hence, the protesting motive between voting for one or the other candidate is redundant, $PM_{\text{LR}}(t_i) = 0$. For the same reason, the conclusion of the previous section holds: $PM_{\text{LA}}(t_i) = PM_{\text{RA}}(t_i) = PM_{\text{VA}}(t_i)$, and the sign analysis survives. In particular, when abstention signals 'for' views ($\overline{\lambda}_F > \overline{\lambda}_A$), 'for' type prefers to abstain and 'against' type – to vote, and when the signal is uninformative ($\overline{\lambda}_F = \overline{\lambda}_A$), both types are indifferent about their action.

As for the election motivation, first observe that the citizen's type does not affect the sign of the motive, but only the belief about the likelihood of a certain election outcome. Hence, only the signal θ would be crucial for the analysis below. When a citizen considers which candidate to vote for ($EM_{\text{LR}}(\theta, t_i)$) the result similar to Razin (2003) holds. In particular, in the event she is pivotal, she realizes that the decisions of others are uninformative to her since they are spread equally (or almost equally when $N - 1 - a$ is odd), and she has to decide based on her private signal θ, which makes one of the states more likely. In particular, whenever a citizen gets a signal $\theta = l$, she feels that the state $s = -d$ is more likely, which makes her better off if the candidate L wins, that is $EM_{\text{LR}}(l, t_i) > 0$. Vice versa, if $\theta = r$, then her belief about the state is closer to candidate R's policy, which makes candidate R more favorable and $EM_{\text{LR}}(r, t_i) < 0$.

Whenever a citizen considers a choice between voting for a particular candidate and abstaining, election motive is still in presence and prescribes her to behave in a pattern similar to when thinking which candidate to vote for. In particular, she prefers to vote for the left candidate if her $\theta = l$ and act against L if $\theta = r$, even if the 'act against' is abstention instead of supporting the right. This is because in case a citizen believes that the state is inclined to the right ($\theta = r$), abstention as opposed to supporting the left would either leave elections in a tie (if votes are equal so far) or affirm a favorable candidate R winning (if R leads by one vote so far). Both of these cases are preferred to their alternative of casting a vote for the left, since overall abstaining does not make a citizen worse off, while voting left does, which explains a negative sign of $EM_{\mathrm{LA}}(r, t_i)$. Given the analysis is restricted to symmetric equilibria (in lines with R1), motivations are symmetric with respect to candidates and signals, i.e. $EM_{\mathrm{LA}}(\theta, t_i) = EM_{\mathrm{RA}}(\theta_-, t_i)$; thus, same logic applies to voting right vs. abstaining selection.

To complete the above analysis, note that despite the election motive is still in presence when considering abstention, the magnitude of $EM_{\mathrm{LA}}(\theta, t_i)$ and $EM_{\mathrm{RA}}(\theta, t_i)$ is strictly lower than the strength of $EM_{\mathrm{LR}}(\theta, t_i)$. In particular, one may show that:

$$|EM_{\mathrm{LA}}(\theta, t_i)| + |EM_{\mathrm{RA}}(\theta, t_i)| = |EM_{\mathrm{LR}}(\theta, t_i)| .$$

To understand the logic behind this, consider a citizen is thinking about voting for the left against other alternatives. If she is more in favor of the candidate L, abstaining makes her at least not worse off in comparison to the world without her. In contrast, voting for R makes her worse off surely. Hence, her motivation to vote for L is stronger if considered against voting for R as a reference point rather than abstaining option, as she experiences a larger loss in this case. Technically speaking, voting for the left and opposed to the right affects all three pivotal cases, whereas by voting left as opposed to abstaining one makes a difference only in two out of three events. Same logic applies for other signals. To summarize, one may say that in some sense, when thinking about election motive solely, voting against makes a stronger statement than inaction.

Having analyzed what each motivation separately inclines citizens for, let us look at their cumulative effect. When deciding between which candidate to support, the motives never conflict each other, since the protesting one is just redundant, and the citizens act according to their election motive solely. However, in the choice between voting and abstaining, the conflict may arise. In particular, 'for' citizens have an incentive to abstain to communicate their views according to their desire to protest, but additionally are incentivized to support a favorable candidate according to their election motive. As such, the tradeoff arises for one of the citizens' types: the 'for' citizens would ideally want to cast a sincere vote for the candidate they perceive as a better fitting to the circumstances;

however, this would cast an insincere 'vote' along the voting-abstaining dimension – that is, by voting they will signal the opposite type instead of their own, which would distort the non-salient policy's shift and ultimately hurt them.

The following proposition characterizes the equilibria existence when the two motivations are in presence[5]. In particular, in what follows I will discuss the results assuming large elections, that is as the number of citizens $N \to \infty$. This simplifies the results derivation without losing much generality and approximates real world scenarios well enough. Recall that by Lemma 1 to formulate any equilibrium it is sufficient to determine q, $\overline{\lambda}_F$ and $\overline{\lambda}_A$, with the first being the probability to vote for a candidate that matches the state given a decision to vote, and the other two are the probabilities for a randomly picked citizen to abstain given the population type.

Proposition 3.2 (equilibria existence)**.** *Restrict attention to $\overline{\lambda}_F \geq \overline{\lambda}_A$. Then,*

i. In any equilibrium, if the citizens vote, they vote sincerely according to their signal θ, i.e. $q = p$;

ii. There always exists a pooling equilibrium with sincere voting where $\lambda_f = \lambda_a = 0$, i.e. all citizens vote irrespective of their type;

iii. In large elections[6], there exists a threshold $p^*(p_\theta) \in \left(\frac{1}{2}, 1\right)$, with $\frac{\partial p^*(p_\theta)}{\partial p_\theta} > 0$, such that $\forall p < p^*(p_\theta)$ there exists a separating equilibrium where 'for' type always abstains, $\lambda_f = 1$, and 'against' type always votes sincerely, $\lambda_a = 0$.

Firstly, in any equilibrium, if the citizens vote, they vote sincerely according to their private signal. This follows directly from Lemma 4 which states that whenever considering who to support the only incentive driving such decision is the election motivation, since both candidates would respond identically in terms of non-salient policy and update it based on turnout only. However, the vote can still affect the traditional policy implementation. Conditional on being pivotal a citizen correctly induces that the other's information about the state of the world aggregated is inconclusive and relies on her private signal instead, which makes her vote sincerely. Note that such strategy implies that whenever at least a share of citizens participates in election, they manage to fully aggre-

[5]The proposition focuses on pure strategy equilibria only. There additionally exists a partially separating equilibrium in which the 'for' type mixes between sincere voting and abstaining if the EM and PM exactly off-set each other. This serves as a moderate case between two extreme pure equilibria, and, hence, is omitted from the analysis as it does not contribute in terms of intuition.

[6]The large elections assumption (approximated with $N \to \infty$) is crucial for establishing a single threshold, but does not affect the equilibrium existence itself. If N is assumed finite and small instead, then different parameter space restrictions would be required to support the separating equilibrium.

gate the signals available to them, and as the turnout increases, the chances that their information suffices to select a correct candidate increases.

Given their desire to vote sincerely in any equilibrium, citizens choose only between voting for the candidate that matches their signal and abstaining. This observation allows to justify equilibria by comparing a single pair of strategies, that is, depending on the signal θ observed, $\Delta_{\text{LA}}(l, t_i)$ or $\Delta_{\text{RA}}(r, t_i)$, which are equal to each other according to symmetry results established in Lemma 4.

Proposition 2 states that there exist two types of pure strategy equilibria satisfying restrictions 1 and 2. First, there always exists a pooling equilibrium where both types of citizens vote. Similar to Proposition 1 result, when the types mimic each other's behavior, the candidates cannot update their beliefs about the population type. This enables them to stick to the status quo, which coincides with the average type according to the prior, $x = \beta Et = 0$, without any threat to their authority, since no protest was communicated to them. Such candidates' response undermines any protesting motivation and makes the citizens indifferent about their decision according to *PM*. However, now they additionally face election motive $EM_{\text{LA}}(l, t_i) = EM_{\text{RA}}(r, t_i)$, which is always positive by Lemma 4. Hence, *EM* pushes all citizens to vote, which is consistent with the candidates' beliefs. Note that in contrast to Section 3.2, now $\lambda_f = \lambda_a = 0$ is a single existing pooling equilibrium, which happens precisely due to the presence of election motive.

Additionally, there exists a separating equilibrium where the 'for' citizens abstain, while the 'against' ones vote. Observe that given the candidates' beliefs that the 'for' citizens are more likely to abstain, the 'against' ones are strictly inclined to vote due to both protesting and election motives. Indeed, sincere voting enables them to contribute to the state information aggregation and correct candidate selection on the one hand, and to separate themselves from the 'for' citizens and effectively communicate their views on the other. As opposed to this, the 'for' citizens' motives conflict each other. As discussed above, they face a tradeoff between contributing to the candidate selection according to election motivation and truthfully communicating their views about non-salient issues according to the protesting one. Obviously, the separating equilibrium exists only when the latter dominates the former, which turns out to occur if the opposing groups sizes are equal enough, that is if p is close enough to a half, and the private signal precision is high enough, i.e. p_θ is closer to 1.

To understand the logic behind the conditions for the 'for' type to abstain, consider how the key parameters affect the motivations. To start with, the composition of society matters, in particular, how the shares of two polarized groups relate. This has two effect is the model. First, when the majority is weak enough (that is, the prevailing views are supported by a share of citizens close enough to the half), then the tension between the groups is the strongest, and the chances of the previously non-salient policies to go in a less popular direction are the highest. This means that each 'for' citizen has a higher incentive to coordinate with her ideological fellows and sacrifice her private information in favor of abstaining and communicating the strength of their group. That is, weaker majority (lower p) enhances the protesting motivation. Secondly, the weaker majority implies that the turnout is high enough irrespectively of which views are prevailing in the society. In particular, having 'for' views, a citizen believes that her ideology is more likely to form a majority, hence, the electorate formed of the 'against' citizens is likely to consist of minority only, with the most likely turnout being $1-p$ share of population. The lower p is, the higher is the expected turnout, which implies more signals aggregated and, in turn, higher accuracy of voting outcome. This means that the event of being pivotal is less likely to occur, which reduces the election motive. On the contrary, with stronger majority low turnout seems more likely to the 'for' type, hence she feels stronger about casting a vote and by that preventing poor candidate selection. Therefore, the election motivation increases with the majority share p. Combining the two effects, observe that the shift in p changes the motives in the opposite directions: if p is lower, than PM increases, whereas EM decreases, which means that the desire to protest is more likely to dominate and incline the 'for' type to abstain. Hence, the separating equilibrium exists only when society is split equally enough.

Another key parameter is the precision of one's private information, p_θ. While it has no effect on the protesting motive, it shapes the election one greatly. As such, when the signals about the state of the world are precise enough, the information is more likely to be correctly aggregated with fewer signals. Therefore, the worry about low turnout and poor candidate selection is less salient, and the probability of being pivotal is lower. This makes the election motivation less pronounced, and the wish to protest is more likely to prevail over it. Therefore, the separating equilibrium with 'for' citizens abstaining is more sustainable under high precision of signals.

Notably, the other parameters of the model, including the candidate's polarization, d, their responsiveness, β, and citizens' polarization, γ, have only second-order effects on the equilibria existence (unless taken to extremes, obviously). This occurs because as elections grow large, both motives converge to infinitely small sums, where the dominance of one over the other is determined through rates of convergence rather than the constants affecting them. The parameters influencing the convergence rates include the ones that affect the probabilities of a certain election outcome, that is p and p_θ. Whereas the magnitude of the motives, determined through the remaining parameters, starts to matter only once both motives converge at the same speed. Alternatively, the effect of other parameters strengthens if the elections occur in the small enough societies, i.e. with N being finite and low enough. That being the case, higher candidates' responsiveness, as well as higher societal views polarization would enhance the protesting motive and make the separating equilibrium existence more likely, since the rise of both parameters makes the originally excluded policy update more pronounced. Higher candidates' polarization, in turn, would strengthen the election motive and therefore make separating equilibrium less likely to occur, because the loss from a wrong candidate selection gets larger with more polarized candidates. Hence, the above-mentioned parameters influence the magnitudes of candidates' policy responses, however, hardly affect the citizen's behavior in large elections.

3.4 Equilibria evaluation

Having analyzed the equilibria existence, let me now consider some properties of those. In particular, I will be interested in whether the equilibria manage to aggregate all the information available in the presence of uncertainty, as well as whether they are efficient from the citizens' perspective.

3.4.1 Correct candidate selection and majoritarian views implementation

The model specified features ex ante uncertainty over two dimensions. First, there is uncertainty about the state of the world realized. This matters for the candidate selection, since the polarized candidates are assumed to perfectly match the two opposing states in their salient policy. Recall that the citizens agree in their views over this dimension in a sense that they face a common value shock: there is one right candidate for all, but it is unknown which one a priori, with only an independent private signal available to each citizen. Therefore, the problem of the candidate selection boils down to the information aggregation task, with the key question being whether the citizens manage to select the correct candidate with the probability close to one. Define this equilibrium characteristic as the *correct candidate selection*.

Secondly, there is also uncertainty about the views prevailing in the population. In particular, it is a common knowledge that the society is split in two polarized groups (hence, the citizens disagree with each other over this dimension), however it is unknown which one forms a majority, with each citizen knowing only her own view, t_i, but not the number of like-minded people. The population type, in turn, affects how the previously excluded policies shift, which rises the problem of information equivalence (following notation of Prato, Wolton (2018)): the key question arising being whether the views are communicated correctly through elections, i.e. whether the non-salient policy shift reflects the true majoritarian views with probability close to one. In what follows I refer to this characteristic as the *majoritarian views implementation.*

The following corollary summarizes whether the equilibria discussed in section 3.3 manage to resolve both uncertainties effectively.

Corollary 3.2. *In large elections:*

i. Any equilibrium satisfies the correct candidate selection.

ii. The separating equilibrium where the 'for' supporters abstain and the 'against' ones vote satisfies the majoritarian views implementation. The pooling equilibrium where all citizens vote fails the criterion.

$\Longrightarrow$ *Both correct candidate selection and majoritarian views implementation are achieved simultaneously only if* $p < p^* (p_\theta)$.

To see the intuition behind the results, first recall that in any equilibrium, if the citizens vote, they cast their vote sincerely according to their private signal. This means that as elections grow large, if at least a proportion of citizens vote, by law of large numbers the share of votes for the state matching candidate converges to $p_\theta > \frac{1}{2}$, which ensures the victory of a correct candidate with probability one almost surely. Since both types of the equilibria feature the positive turnout (with the lowest possible converging to $1 - p$ share of population in the event when 'against' citizens comprise a minority and are the only ones to vote), the correct candidate is always selected irrespective of equilibrium realized.

In turn, for the originally excluded policy to overlap with the majoritarian preferences, the citizens need to coordinate within their ideological groups and separate themselves from the opposite type. In the separating equilibrium discussed in section 3.3, this is achieved through abstention of the 'for' citizens. As a result, in large elections, by law of large numbers the share of those who abstain converges to the share of 'for' citizens in the population almost surely: that is, the turnout is equal to $1 - p$ if 'for' is a majority view, and p if 'against' prevails instead. As such, the level of abstention allows to correctly deduce the population type and shift the non-salient policy accordingly, hence, making such equilibrium effective in terms of majoritarian views implementation in addition to correct candidate selection. As the citizens pool and vote irrespectively of their views, however, the equilibrium becomes ineffective since it fails to communicate public sentiment and the excluded policy no longer reflects majoritarian preferences.

Overall, both criteria hold only if the separating equilibrium is realized, which exists whenever $p < p^* (p_\theta)$, with $p^* (p_\theta)$ increasing in p_θ. Lower majoritarian power, p, and higher precision of private signals, p_θ, relax the pressure on candidate selection motive, since less votes suffice to aggregate the information well enough. Hence, this allows to focus on the population type transmission, which enables the majoritarian views implementation.

3.4.2 Equilibria efficiency

Another criterion to evaluate the equilibria is to consider their efficiency on the utilitarian grounds. In particular, define *efficiency* as an outcome where the citizens expected utility is maximized in the ex ante terms, with the efficient strategy not necessarily being a part of any equilibrium. That is efficiency is achieved when the traditional and excluded policy gaps are minimized.

First, observe that sincere voting is always efficient. This is due to the fact that such strategy ensures the correct candidate selection, therefore, in large elections, the traditional policy gap is trivially minimized, i.e.$\lim_{N\to\infty} |y(\Omega_N) - s| = 0$, whereas the voting strategy has no effect of the non-salient policy gap. Hence, sincere voting is always a part of an efficient strategy.

For the excluded policy gap minimization, consider how the ex ante expected utility looks like for each citizen. To start, recall that given citizens' polarization, the excluded policy update is always more moderate than the citizens' views, that is, $x(\Omega) \in [-\beta\overline{\gamma},\ \beta\overline{\gamma}] \subset (-\gamma,\ \gamma)$. Therefore, for any strategy (known to the candidates) that leads to the policy being updated to some x, the policy gap would be equal to $\gamma - |x|$ for the citizens comprising the majority, and $\gamma + |x|$ for each in the minority. Since ex ante probability of being in majority is p, the ex ante utility may be shown to converge to

$$U(x) = pu\left(\gamma - |x|\right) + (1-p)u\left(\gamma + |x|\right),$$

where $u(\cdot)$ is some decreasing concave function.

Observe that if the citizens pool at the same strategy, the candidates cannot update their beliefs and stick to the status quo, $x = 0$. Hence in the pooling equilibrium, the ex ante utility is equal to $u(\gamma)$. On the contrary, if the societal views are communicated effectively (that is, the majoritarian view implementation is satisfied), the policy update converges to $|x| = \beta\overline{\gamma}$, with the absolute sign depending on which group represents the majority. Note that whenever the majoritarian views implementation holds, that is, the citizens manage to separate themselves from each other fully, the policy update is the most extreme in the space of $x(\Omega) \in [-\beta\overline{\gamma},\ \beta\overline{\gamma}]$, which maximizes the ex post gap between the utility of majority and minority. Finally, any moderate policy response with $|x| \in (0,\ \beta\overline{\gamma})$ can be attained whenever the citizens transmit their signal imperfectly, i.e. mix between voting at abstaining at different rates for each group.

The corollary below summarizes the efficiency of the equilibria discussed in Section 3.3.

Corollary 3.3.

i. The pooling equilibrium where all the citizens vote is never efficient.

ii. The separating equilibrium where the 'for' type votes and the 'against' one abstains is efficient if and only if $x^* \geq \beta\overline{\gamma}$, where x^* is defined by $\frac{u'(\gamma - x^*)}{u'(\gamma + x^*)} = \frac{1-p}{p}$.

Given the shape of the utility function $u(\cdot)$, there are two effects taken into consideration by the citizens. The first is that the concavity implies that the citizens prefer certainty over the risky outcomes. In the context of elections, the citizens would rather tolerate stability in policies remaining at $x = 0$ despite their views are misaligned with the status quo. Not knowing which views comprise a majority, if they choose to protest against the status quo by abstaining, they risk revealing their minority and facing the policy moving in the direction unfavorable to them. This risk is especially profound in the separating equilibrium, since, as discussed above, the policy shifts are the most extreme in this case. On the other hand, however, being more likely to be a part of majority and thus reduce the policy gap may justify the risk. In particular, the most extreme policy shift minimizes the expected gap, which can ultimately increase the citizens' ex ante utility if their preference for stability is low enough.

The prevalence of one of the two effects determines the efficiency of the equilibria. To start, the pooling equilibrium is never efficient. This occurs because once the citizens persuade the candidates to a marginal policy shift, the instability risk remains low enough, while the expected policy gap decreases, making slight deviations optimal ex ante. In turn, the separating equilibrium can be efficient depending on the utility function the citizens face. If the function exhibits relatively low risk aversion, then the positive effect of decrease in average gap dominates the negative one of increased variance in possible outcomes, making the equilibrium strategies efficient. Alternatively, the candidates' responsiveness should be moderate enough in order to maintain the equilibrium efficiency. This is due to reduction of the variance of potential outcomes once β gets lower. [7]

Finally, to better estimate the likelihood of the separating equilibrium being efficient, consider a class of utilities $u(x) = -x^k,\ k \geq 1$, which are decreasing and concave in their argument. Then, the separating equilibrium is efficient whenever $x^* \geq \beta\overline{\gamma}$, with x^* defined by $\frac{\gamma - x^*}{\gamma + x^*} = \left(\frac{1-p}{p}\right)^{\frac{1}{k-1}}$. Observe that as $k \leq 2$ (that is, the function exhibits low enough risk aversion), $\frac{\gamma - x^*}{\gamma + x^*} = \left(\frac{1-p}{p}\right)^{\geq 1} \leq \frac{1-p}{p} \implies x^* \geq (2p-1)\gamma = \overline{\gamma} \geq \beta\overline{\gamma}$ – hence, the equilibrium is efficient. Alternatively, if $k > 2$, then $x^* < \overline{\gamma}$, and the condition required for efficiency is not guaranteed unless $\beta < \frac{x^*}{\overline{\gamma}} < 1$ – that is, additional restriction on candidates' responsiveness should be imposed. Note that if both conditions for efficiency fail, the citizens find communicating their views effectively suboptimal, and would rather blur their message so that not to experience too extreme risky policy amendments – hence, the full information equivalence may interfere with the efficiency of an outcome.

3.5 Discussion

Overall, I have rigorously shown that abstention can arise as a statement of protesting spirits in the society. In particular, the citizens are eager to separate each other through abstention if the majority is relatively weak and the signal precision is high enough – this ensures the highest protestive tension between the two polarized groups, as well as relaxes the election motive. In large elections, such equilibrium enables both full information aggregation, since the elections select the right candidate almost surely, and full information equivalence, due to the candidates' response in previously non-salient policies converging

[7] The effect of p on $x^* \geq \beta\overline{\gamma}$ likelihood is not considered, since both sides of inequality are affected similarly: both $\beta\overline{\gamma} = (2p-1)\beta\gamma$ and x^* are monotonously increasing in p and marginal effect on x^* depends on the curvature of utility function – this gives rise to incentives discussed above.

to the true majoritarian preferences. Moreover, under certain regularity conditions, the separating equilibrium is efficient in a sense that it maximizes citizens' ex ante expected utility.

Among other, the model sheds the light on why the societies may be interested in participating in elections despite its results predetermination and lack of political competition, which is often a feature of non-democratic elections, as well as on why such elections are held in general. I show that even in the absence of any election motivation, some citizens choose to vote so that to communicate the public sentiment and separate themselves from their opponents. This information aggregation may serve as a motive for holding elections that are meaningless per se.

On top of the work done in this paper, there are multiple directions to extend the results further. First, there are ways to generalize the model setup. This could include relaxing symmetry assumptions that I use, allowing for the two political dimensions to be correlated, introducing further heterogeneity in candidates' responses and citizens' views, extending dimensionality. The question remaining is whether the various generalized settings would preserve the full information aggregation along multiple dimensions.

In addition, the model discussed abstracts from strategic motives of the candidates by assuming their agreement in non-salient policy shift. Instead, one may study how electoral competition would interact with voters' motives and affect information aggregation results in a similar setup, as well as solve for the optimal candidates' behavior. Competition aspects may include the level of responsiveness to electoral results, as well as being responsive as opposed to committing to a certain platform more generally.

Moreover, the model is formulated in a static way where the elections urge a political action on policies previously excluded from political agenda. However, the whole idea of the change in salience of certain aspects rises the need for a more dynamic setup, where the views of society as well as the actuality of different aspects evolve through time. Hence, one may explore whether elections persistently prove to be an efficient mechanism for majoritarian views communication in a changing world, as well as which factors shape the policies.

Lastly, one may focus on the conceptual side of a matter, specifically on extending the reasons why people vote in non-democracies and why such elections are held in general. In case of elaborating on the protesting motives, which I focus on in this paper, one may study more thoroughly how the opposition can coordinate their forces to strengthen the protest in the presence of few different strategies for that. I find these concerns particularly actual and potentially fruitful and leave these questions for future research.

References

Austen-Smith, David, and Jeffrey S. Banks, 1996, Information aggregation, rationality, and the Condorcet jury theorem, *The American Political Science Review*, 90(1), 34 – 45.

Cox, Gary W., 2009, Authoritarian Elections and Leadership Succession, *APSA 2009 Toronto Meeting Paper, Available at SSRN:* `https://ssrn.com/abstract=1449034`.

Egorov, Georgy, and Konstantin Sonin, 2018, Election in Non-Democracies, *available at SSRN:* `https://ssrn.com/abstract=2497277`.

Ekmekci, Mehmet, and Stephan Lauermann, 2020, Manipulated electorates and information aggregation, *The Review of Economic Studies*, 87(2), 997 – 1033.

Feddersen, Timothy J., and Wolfgang Pesendorfer, 1996, The Swing Voter's Curse, *The American Economic Review*, 86(3), 408 – 424.

Feddersen, Timothy J., and Wolfgang Pesendorfer, 1997, Voting Behavior and Information Aggregation in Elections With Private Information, *Econometrica*, 65(5), 1029 – 1058.

Feddersen, Timothy J., and Wolfgang Pesendorfer, 1999, Abstention in Elections with Asymmetric Information and Diverse Preferences, *The American Political Science Review*, 93(2), 381 – 398.

Feddersen, Timothy J., and Alvaro Sandroni, 2006, A Theory of Participation in Elections, *The American Economic Review*, 96(4), 1271 – 1282.

Gehlbach, Scott, and Alberto Simpser, 2015, Electoral Manipulation as Bureaucratic Control, *American Journal of Political Science*, 59(1), 212 – 224.

Herrera, Helios, Aniol Llorente-Saguer, and Joseph McMurray, 2019, The Marginal Voter's Curse, *The Economic Journal*, 129(624), 3137 – 3153.

Kostadinova, Tatiana, 2009, Abstain or Rebel: Corruption Perceptions and Voting in East European Elections, *Politics and Policy*, 37(4), 691 – 714.

Ledyard, John O., 1984, The Pure Theory of Large Two-Candidate Elections, *Public Choice*, 44(1), 7 – 41.

McMurray, Joseph, 2013, Aggregating Information by Voting: The Wisdom of the

Experts Versus the Wisdom of the Masses, *Review of Economic Studies*, 80(1), 277 – 312.

McMurray, Joseph, 2017a, Voting as Communicating: Mandates, Multiple Candidates, and the Signaling Voter's Curse, *Games and Economic Behavior*, 102, 199 – 223.

McMurray, Joseph, 2017b, Ideology as Opinion: a Spatial Model of Common-value Elections, *American Economic Journal: Microeconomics*, 9(4), 108 – 140.

Meirowitz, Adam, and Joshua A. Tucker, 2007, Run Boris Run: Strategic Voting in Sequential Elections, *The Journal of Politics*, 69(1), 88 – 99.

Piketty, Thomas, 2000, Voting as Communicating, *The Review of Economic Studies*, 67, 169 – 191.

Prato, Carlo, and Stephane Wolton, 2018, Wisdom of the Crowd? Information Aggregation in Representative Democracy, *available at SSRN: https://ssrn.com/abstract=2995012.*

Razin, Ronny, 2003, Signaling and Election Motivations in a Voting Model with Common Values and Responsive Candidates, *Econometrica*, 71(4), 1083 – 1119.

Shotts, Kenneth W., 2006, A Signaling Model of Repeated Elections, *Social Choice and Welfare*, 27(2), 251 – 261.

Appendix A

Appendix to Chapter 1

A.1 Appendix. Figures

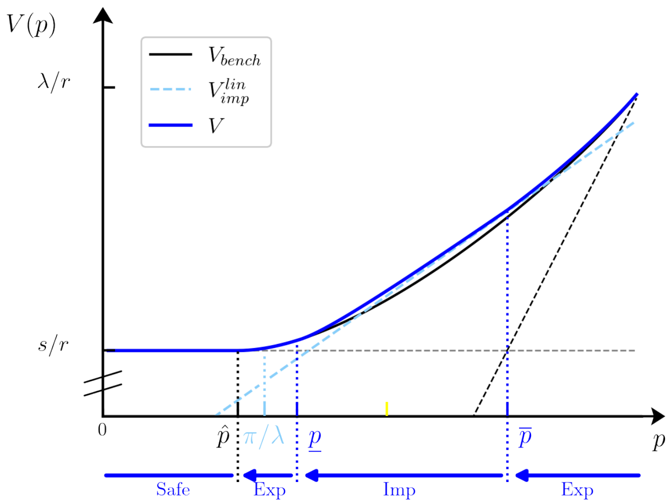

Figure A.1: The optimal value function under low efficiency of training ($\frac{\pi}{\lambda} < \underline{p}(\pi, \kappa)$), as defined in Proposition 1.2.

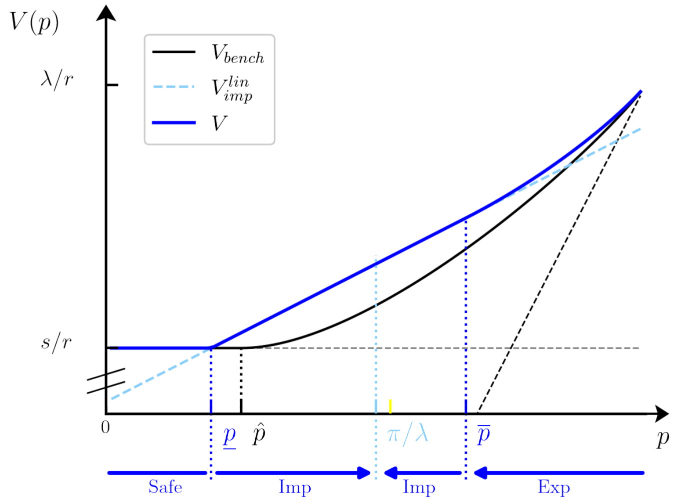

Figure A.2: The optimal value function under moderately high efficiency of training ($\frac{\pi}{\lambda} \in [\underline{p}(\pi,\kappa), \overline{p}(\pi,\kappa)]$), as defined in Proposition 1.3 part *i*. Plotted under relatively low costs, so illustrates the dynamics under top scheme on Figure 1.3.

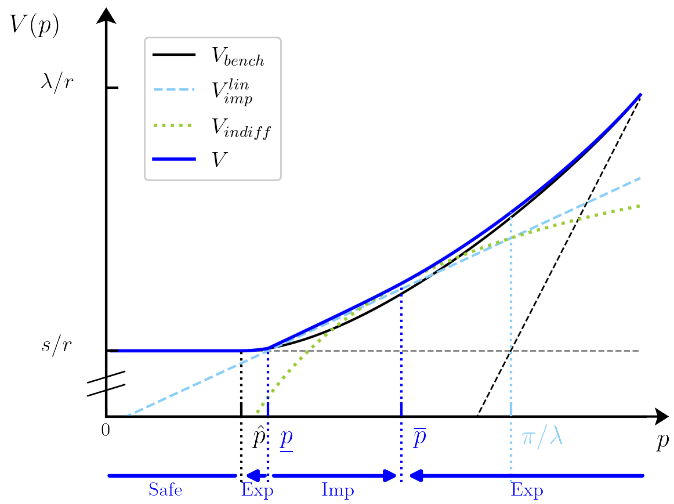

Figure A.3: The optimal value function under very high efficiency of training ($\frac{\pi}{\lambda} > \overline{p}(\pi,\kappa)$), as defined in Proposition 1.3 part *ii*. Plotted under relatively high costs, so illustrates the dynamics under bottom scheme on Figure 1.4.

A.2 Appendix. Proofs

To prove the main results I establish and prove some auxiliary lemmas.

Lemma A.2.1. *The ability to train the risky arm enhances the value iff $\kappa < \overline{\kappa}(\pi)$, where $\overline{\kappa}(\pi) \equiv \pi V'_{bench}(p')(1-p')$ and $p' \equiv \arg\max_{p\in(\hat{p},1)} V'_{bench}(p)(1-p)$.*

Proof. It is optimal to train the arm for some parameter space if it is optimal to train it for at least a single belief for at least a short instance dt. Denote this unique belief as p'. At this belief, the agent must be indifferent between training and some other action, so training is weakly dominated by experimentation for any p, and $V(p) = V_{bench}(p)$, as defined in Lemma 1.2.

A single indifference point cannot lie inside the stopping region $(0, \hat{p})$. There, experimentation is strictly dominated by safe arm and training is strictly dominated by experimentation: $s > p\lambda + p\lambda\big(V(1) - V(p)\big) - V'(p)\lambda p(1-p) > p\lambda - \kappa + p\lambda\big(V(1) - V(p)\big) + V'(p)(\pi - \lambda p)(1-p)$, where the latter inequality relies on $V'(p) = 0$ for $p \in (0, \hat{p})$. Hence, at p' the agent is indifferent between experimenting with and without training, which holds iff $V'_{bench}(p)\pi(1-p) = \kappa$.

As $V_{bench}(p)$ is strictly convex, $V'_{bench}(p)\pi(1-p)$ is quasi-concave for $p \in [\hat{p}, 1]$, with $V'_{bench}(p)\pi(1-p) = 0$ for $p = 1$ and $p \leq \hat{p}$ (as $V'_{bench}(\hat{p}) = 0$). As such, $V'_{bench}(p')\pi(1-p') = \kappa$ identifies p' uniquely only at $\max_{p\in[\hat{p},1]} V'_{bench}(p)(1-p)$ (solution is well defined and unique due to quasi-concavity of objective function). Moreover, the above guarantees that $p' \in (\hat{p}, 1)$ - strictly bounded away from $\hat{p}$ and 1. Finally, $\overline{\kappa}(\pi) \equiv \pi V'_{bench}(p')(1-p')$ defines the upper bound on the cost for training to remain attractive: any $\kappa < \overline{\kappa}(\pi)$ implies the training region under $V'(p)\pi(1-p) \geq \kappa$ is non-empty. Note that V_{bench} and p' are independent of π, κ by construction, so $\overline{\kappa}(\pi)$ is a linear function of π. ∎

Lemma A.2.2 (Conjecture on structure)**.** *If $\kappa < \overline{\kappa}(\pi)$, there always exist two cutoffs $\underline{p} \in [0, p')$ and $\overline{p} \in (p', 1]$, such that the agent trains the risky arm for beliefs in $[\underline{p}, \overline{p}]$, uses the safe arm for $p < min\{\underline{p}, \hat{p}\}$ and purely experiments otherwise.*

The pure experimentation range of beliefs $(\hat{p}, \underline{p})$ is non-empty for $\kappa \in (\underline{\kappa}(\pi), \overline{\kappa}(\pi))$, where $\underline{\kappa}(\pi)$ is a non-decreasing function, and $\forall \pi, \underline{\kappa}(\pi) < \overline{\kappa}(\pi)$.

Proof. Given the nature of conjecture, here I provide a more informal motivation for the proposed structure. I then fully prove that the conjecture is accurate in the proofs of main propositions.

$$rV(p) = \max_{(\alpha,\beta)\in[0,1]^2} \big((1-\alpha)s + \alpha(p\lambda - \beta\kappa)\big) + \alpha p\lambda\big(V(1) - V(p)\big) + V'(p)\alpha(\beta\pi - \lambda p)(1-p)$$

The optimal solution always involves a pure action ($\alpha = 0$; $\alpha = 1, \beta = 0$ or $\alpha = 1, \beta = 1$) due to linearity of $V(p)$ in α, β. Trivially, $V(p)$ is a non-decreasing function, implying that there exists a unique cutoff such that $\alpha = 0$ below this cutoff and $\alpha = 1$ above. Denote this cutoff as p^*.

By lemma A.2.1, under high training cost $\overline{\kappa}(\pi)$ the agent trains the arm at a unique belief $p' \in (\hat{p}, 1)$. This implies that she also trains the arm at p' for any $\kappa < \overline{\kappa}(\pi)$. By continuity of $V(p)$, for any $\kappa < \overline{\kappa}(\pi)$ there also must exist some neighbourhood around p' such that $V'(p)\pi(1-p) > \kappa$ holds there, implying that the agent is willing to train for at least a short instance dt within this neighbourhood, and never outside. Denote this region by $[\underline{p}, \overline{p}]$, with $\underline{p} < p'$ and $\overline{p} > p'$; it must be a single interval around p' by the logic above.

Note that $\overline{p}$ can go all the way up to $\overline{p} = 1$: If $\kappa = 0$, $V'(p)\pi(1-p) \geq 0$ holds even for $p = 1$, so training is weakly preferred to pure experimentation.

Similarly, under $\kappa = 0$, there always exists $\overline{\pi}$, such that $s < V'(0)\pi$ for $\pi > \overline{\pi}$ and the agent trains the arm even at $p = 0$, implying that $\underline{p} = 0$. Suppose the agent trains at $p = 0$. Then, $V(0) \geq y(0) = \frac{\pi}{r+\pi}\left(\frac{\lambda}{r} - \frac{\kappa}{r+\lambda}\right) - \frac{\kappa}{r+\pi}$ (as $C_{Tr}g(p) \geq 0$). With $\kappa = 0$, $y(0) \geq \frac{s}{r}$ iff $\pi \geq r\frac{s}{\lambda}/(1-\frac{s}{\lambda})$ - well defined cutoff for π. Note that $p^* = \underline{p}$ in this case.

Finally, consider very high cost: $\kappa = \overline{\kappa}(\pi) - \epsilon$, where $\epsilon \to 0$. Under such cost, the training region should be very narrow, $\underline{p} \to p'_-$ and $\overline{p} \to p'_+$. This implies that $\underline{p} > \hat{p}$ holds, as $p' > \hat{p}$ by lemma **??**. Hence, there exists some cost region $(\underline{\kappa}(\pi), \overline{\kappa}(\pi))$, such that $\underline{p} > \hat{p}$, and this region is non-empty for any value of π (as least $\kappa = \overline{\kappa}(\pi) - \epsilon$ belongs to the interval). $\underline{\kappa}(\pi)$ is defined via $\underline{p} = \hat{p}$.

Note that for $\kappa \in (\underline{\kappa}(\pi), \overline{\kappa}(\pi))$, $p^* = \hat{p}$: For $p < \underline{p}$, $dp \leq 0$, so the agent's incentives match the benchmark ones according to lemma 1.2 in full. Then, the stopping cutoff is defined by $\hat{p}$ exactly. Overall, $p^* = \min\{\underline{p}, \hat{p}\}$, and there may exist a non-empty pure experimentation region $(\hat{p}, \underline{p})$ for high enough costs. ■

Proposition 1.1 (part i) and Proposition 1.2:

Proof. Conjecture the structure according to lemma A.2.2, with non-empty $(\hat{p}, \underline{p})$, and $\frac{\pi}{\lambda} < \underline{p}$. This implies that $dp \leq 0$ for any p. Then, solve for the cutoffs consecutively from the lowest to highest.

At $\hat{p}$, value matching ensures $V(\hat{p}) = \frac{s}{r}$, which is rearranged to $C_{Exp}(\hat{p}) = (\frac{s}{r} - \frac{\lambda}{r}\hat{p})\frac{1}{f(\hat{p})}$. Note that it also implies that (1) $s = \lim_{p \to \hat{p}_+}[p\lambda + p\lambda\big(V(1) - V(p)\big) - V'(p)\lambda p(1-p)]$ - a value matching in differential form, with the right limit indicating the side where the agent experiments. Marginal indifference implies that the agent does not benefit (or lose) from a short deviation to experimentation at $\hat{p}$, which holds iff (2) $s = \lim_{p \to \hat{p}_-}[p\lambda + p\lambda\big(V(1) - V(p)\big) - V'(p)\lambda p(1-p)]$ - with the left limit due to $dp < 0$ at $\hat{p}$, and $\lim_{p \to \hat{p}_-} V'(p) = 0$ as $V(p) = \frac{s}{r}$ then. Combining (1) and (2) results in $\lim_{p \to \hat{p}_-} V'(p) = \lim_{p \to \hat{p}_+} V'(p)$ - that is, smooth pasting. So $V'(\hat{p})$ is well defined, and $V'(\hat{p}) = 0$ solves for $\hat{p} = \frac{sr}{\lambda(\lambda - s + r)}$.

By similar argument, smooth pasting holds at $\underline{p}$ and $\overline{p}$. It also directly translates to $V'(p)\pi(1-p) = \kappa$ at both cutoffs. Then, the value matching pins down the constants, $C_{Tr}(\hat{p}, \underline{p}) = \left(\frac{\lambda}{r}\underline{p} - y(\underline{p}) + C_{Exp}(\hat{p})f(\underline{p})\right)\frac{1}{g(\underline{p})}$ and $C_{Exp}(\hat{p}, \underline{p}, \overline{p}) = \left(y(\overline{p}) - \frac{\lambda}{r}\overline{p} + C_{Tr}(\hat{p}, \underline{p})g(\overline{p})\right)\frac{1}{f(\overline{p})}$, and smooth pasting solves for $\underline{p}$ and $\overline{p}$. Note that $\underline{p} > \hat{p}$, as conjectured, as $V'(\underline{p})\pi(1-\underline{p}) = \kappa$ holds only if $V'(\underline{p}) > V'(\hat{p}) = 0$. Also, $\underline{p} > \frac{\pi}{\lambda}$ is necessarily satisfied, as $V(\underline{p})$ is finite, while $\lim_{p \to \frac{\pi}{\lambda}}[y(p) + C_{Tr}(\hat{p}, \underline{p})g(p)] = +\infty$, confirming the initial conjecture. Hence, the solution is characterised by:

$$V(p) = \begin{cases} \frac{s}{r} & p < \hat{p} \ \textit{(safe arm)} \\ \frac{\lambda}{r}p + (\frac{s}{r} - \frac{\lambda}{r}\hat{p})\frac{f(p)}{f(\hat{p})} & p \in [\hat{p}, \underline{p}] \ \textit{(pure exp)} \\ y(p) + \left(\frac{\lambda}{r}\underline{p} - y(\underline{p}) + (\frac{s}{r} - \frac{\lambda}{r}\hat{p})\frac{f(\underline{p})}{f(\hat{p})}\right)\frac{g(p)}{g(\underline{p})} & p \in [\underline{p}, \overline{p}] \ \textit{(training)} \\ \frac{\lambda}{r}p + \left(y(\overline{p}) - \frac{\lambda}{r}\overline{p} + \left(\frac{\lambda}{r}\underline{p} - y(\underline{p}) + (\frac{s}{r} - \frac{\lambda}{r}\hat{p})\frac{f(\underline{p})}{f(\hat{p})}\right)\frac{g(\overline{p})}{g(\underline{p})}\right)\frac{f(p)}{f(\overline{p})} & p > \overline{p} \ \textit{(pure exp)} \end{cases}$$

Finally, continuity, smoothness and convexity of $V(p)$ follow directly from value matching and smooth pasting. ■

Proposition 1.1 (part ii) and Proposition 1.3:

Proof. Conjecture the structure according to lemma **??**, and $\frac{\pi}{\lambda} > \underline{p}$. Now, $dp \leq 0$ does not hold, so break the solution down into two cases.

i. Suppose $\frac{\pi}{\lambda} < \overline{p}$. This implies that when the agent trains, her beliefs converge to $\frac{\pi}{\lambda}$, which is also inside training region, and $dp = 0$ there; i.e. the agent never switches to other options in the absence of news after she starts training and $C_{Tr} = 0$ (check that continuation value from training until news arrives and switching to pure exploitation of risky arm immediately after is captured by $y(p)$ precisely).

At $\overline{p}$, value matching and smooth pasting must hold, by the argument as in the proof of Proposition 1 (part i) and Proposition 2. Thus, $C_{Exp}(\overline{p}) = \left(y(\overline{p}) - \frac{\lambda}{r}\overline{p}\right)\frac{1}{f(\overline{p})}$ and $\overline{p} = \frac{r\left(\frac{\pi}{r+\pi}\left(\frac{\lambda}{r}-\frac{\kappa}{r+\lambda}\right)-\frac{\kappa}{r+\pi}\right)}{\kappa+r\left(\frac{\pi}{r+\pi}\left(\frac{\lambda}{r}-\frac{\kappa}{r+\lambda}\right)-\frac{\kappa}{r+\pi}\right)}$, and $\overline{p}$ falls with κ (also, for $\kappa = 0$, $\overline{p} = 1$, as conjectured). Denote $\frac{r\left(\frac{\pi}{r+\pi}\left(\frac{\lambda}{r}-\frac{\kappa}{r+\lambda}\right)-\frac{\kappa}{r+\pi}\right)}{\kappa+r\left(\frac{\pi}{r+\pi}\left(\frac{\lambda}{r}-\frac{\kappa}{r+\lambda}\right)-\frac{\kappa}{r+\pi}\right)} \equiv \overline{p}(\pi,\kappa)$ - used in later proofs.

The lower bound $\underline{p}$ is determined purely by value matching. Specifically, define the solution to $\frac{s}{r} = y(p)$ as $p^*(\pi,\kappa)$. Solving the equation gives $p^*(\pi,\kappa) = \frac{\frac{s}{r}-\left(\frac{\pi}{r+\pi}\left(\frac{\lambda}{r}-\frac{\kappa}{r+\lambda}\right)-\frac{\kappa}{r+\pi}\right)}{\frac{\lambda}{r+\pi}\left(1+\frac{\kappa}{r+\lambda}\right)}$. If $p^*(\pi,\kappa) \leq \hat{p}$, this gives the solution to the lower bound, $\underline{p} = \max\{0, p^*(\pi,\kappa)\}$. $\frac{s}{r} = \left(\frac{\pi}{r+\pi}\left(\frac{\lambda}{r}-\frac{\kappa}{r+\lambda}\right)-\frac{\kappa}{r+\pi}\right) = y(0)$ defines $\kappa = \kappa_0(\pi)$, such that for $\kappa < \kappa_0(\pi)$ the following holds: $\underline{p} = 0$.

If $p^*(\pi,\kappa) > \hat{p}$, $\underline{p}$ solves for the lower root of $y(p) = V_{bench}(p)$ and it results in $\underline{p} > p^*(\pi,\kappa)$, implying that $\underline{p} > \hat{p}$. The agent purely experiments in $(\hat{p},\underline{p})$. Note that $\underline{p}$ is defined by $y(p) = V_{bench}(p)$, as the incentives of the agent for $p < \underline{p}$ coincide with the ones in the benchmark lemma **??** (she never trains there), so the value function below $\underline{p}$ coincides with $V_{bench}(p)$. The pure experimentation region $(\hat{p},\underline{p})$ is non-empty, whenever $p^*(\pi,\kappa) > \hat{p}$, so $\underline{\kappa}(\pi)$ in Lemma **??** is defined via $p^*(\pi,\kappa) = \hat{p}$.

Note that value matching at $\underline{p}$ also guarantees that the marginal indifference holds. If the agent marginally deviates to training at $\underline{p}$, her belief will increase as $dp > 0$, and so the continuation value of such marginal deviation is $\lim_{p\to\underline{p}_+}[p\lambda + p\lambda\big(V(1) - V(p)\big) - V'(p)\lambda p(1-p)]$ - coincides with the value to the right of the cutoff. That is, the marginal indifference fully coincides with value matching, and $\lim_{p\to\underline{p}_-} V'(p) = \lim_{p\to\underline{p}_+} V'(p)$ need not hold.

In fact, smooth pasting must be violated. For any $p < \underline{p}$, training hurts, so training in dt is unattractive as well, i.e. $V'(p)\pi(1-p) \leq \kappa$ for $p < \underline{p}$. This means that (1) $\lim_{p\to\underline{p}_-}[V'(p)\pi(1-p)] \leq \kappa$. For $p \in (\underline{p},\overline{p})$, $V'(p) = const$, as the value function is linear. Hence, $V'(p)\pi(1-p)$ is a linearly decreasing function, which is equal to κ at $\overline{p}$, given smooth pasting holds there. That is, (2) $\lim_{p\to\underline{p}_+}[V'(p)\pi(1-p)] > \kappa$, as $\underline{p} < \overline{p}$. (1) and (2) together imply that $\lim_{p\to\underline{p}_-} V'(p) < \lim_{p\to\underline{p}_+} V'(p)$ - so there must be a convex kink at $\underline{p}$.

Finally, define the solution to $\underline{p}$ in this case by $\underline{p}(\pi,\kappa)$. That is,

$$\underline{p}(\pi,\kappa) \equiv \begin{cases} \max\{0, p^*(\pi,\kappa)\} & p^*(\pi,\kappa) \leq \hat{p} \\ \arg\min\{y(p) = V_{bench}(p)\} & p^*(\pi,\kappa) > \hat{p} \end{cases}$$

The conjecture of solution i. holds iff $\frac{\pi}{\lambda} \in [\underline{p}(\pi,\kappa), \underline{p}(\pi,\kappa)]$. Summarizing the above, within this parametric restriction the solution is given by:

$$V(p) = \begin{cases} \max\{\frac{s}{r}, \frac{\lambda}{r}p + (\frac{s}{r} - \frac{\lambda}{r}\hat{p})\frac{f(p)}{f(\hat{p})}\} & p < \underline{p} \ \textit{(safe arm or pure exp)} \\ y(p) & p \in [\underline{p}, \overline{p}] \ \textit{(training)} \\ \frac{\lambda}{r}p + \big(y(\overline{p}) - \frac{\lambda}{r}\overline{p}\big)\frac{f(p)}{f(\overline{p})} & p > \overline{p} \ \textit{(pure exp)} \end{cases}$$

and $\{\underline{p}, \overline{p}\} = \{\underline{p}(\pi,\kappa), \overline{p}(\pi,\kappa)\}$.

ii. Suppose $\frac{\pi}{\lambda} > \overline{p}$. This implies that when the agent trains, $dp > 0$ and her beliefs converge to $\overline{p}$. Similarly, being above $\overline{p}$, she purely experiments and converges to $\overline{p}$ as well. As such, $\overline{p}$ is fully absorbing, and $V(p)$ depends solely on $\overline{p}$ (and not on $\underline{p}$ or $\hat{p}$) for $p > \underline{p}$. At $\overline{p}$ the agent is indifferent between training and not, so any β is optimal. However, the unique stationary solution occurs iff $\beta^* = \overline{p}/\frac{\pi}{\lambda} \in (0,1)$ as $\frac{\pi}{\lambda} > \overline{p}$ - this guarantees that $dp = 0$ at $\overline{p}$, so the agent remains at the cutoff once converges there.

At $\overline{p}$, value matching and smooth pasting must hold. Value matching, once using differential form, implies $\lim_{p\to\overline{p}_-}[p\lambda - \kappa + p\lambda\big(V(1) - V(p)\big) + V'(p)(\pi - \lambda p)(1-p)] = \lim_{p\to\overline{p}_+}[p\lambda + p\lambda\big(V(1) - V(p)\big) - V'(p)\lambda p(1-p)]$ (limits as specified as the agent trains for $p < \overline{p}$ and purely experiments for $p > \overline{p}$). This simplifies to (1) $\lim_{p\to\overline{p}_-}[V'(p)(\pi - \lambda p)(1-p)] - \kappa = \lim_{p\to\overline{p}_+}[-V'(p)\lambda p(1-p)]$. For the marginal incentive, at $\overline{p}$ the agent is indifferent between training and not for a marginal instance dt and following the optimal strategy right after. This implies that $\lim_{p\to\overline{p}_+}[p\lambda - \kappa + p\lambda\big(V(1) - V(p)\big) + V'(p)(\pi - \lambda p)(1-p)] = \lim_{p\to\overline{p}_-}[p\lambda + p\lambda\big(V(1) - V(p)\big) - V'(p)\lambda p(1-p)]$, with limits reflecting that marginal training increases p and marginal pure experimentation decreases it. The condition simplifies to (2) $\lim_{p\to\overline{p}_+}[V'(p)(\pi - \lambda p)(1-p)] - \kappa = \lim_{p\to\overline{p}_-}[-V'(p)\lambda p(1-p)]$. Combining (1) and (2) results in $\lim_{p\to\overline{p}_-} V'(p) = \lim_{p\to\overline{p}_+} V'(p)$ - smooth pasting, so $V'(\overline{p})$ is well defined.

Value matching and smooth pasting at $\overline{p}$ allow to define C_{Tr} and C_{Exp}. Specifically, $C_{Exp}(\overline{p}) = (y(\overline{p}) - \frac{\lambda}{r}\overline{p} + C_{Tr}(\overline{p})g(\overline{p}))\frac{1}{f(\overline{p})}$ and $C_{Tr}(\overline{p}) = \left(\frac{\lambda}{r}\overline{p} - y(\overline{p}) + \frac{(\lambda\pi(1-\overline{p}) - \kappa r)\overline{p}}{r\pi(\overline{p} + \frac{r}{\lambda})}\right)\frac{1}{g(\overline{p})}$. Once plugged in, this implies that $V_{Tr}(p;\overline{p}) = y(p) + \left(\frac{\lambda}{r}\overline{p} - y(\overline{p}) + \frac{(\lambda\pi(1-\overline{p}) - \kappa r)\overline{p}}{r\pi(\overline{p} + \frac{r}{\lambda})}\right)\frac{g(p)}{g(\overline{p})}$ and $V_{Exp}(p;\overline{p}) = \frac{\lambda}{r}p + \left(\frac{(\lambda\pi(1-\overline{p}) - \kappa r)\overline{p}}{r\pi(\overline{p} + \frac{r}{\lambda})}\right)\frac{f(p)}{f(\overline{p})}$. Maximizing these with respect to $\overline{p}$ translates to $\max_{\overline{p}} C_{Exp}(\overline{p})$ or $\max_{\overline{p}} C_{Tr}(\overline{p})$ equivalently. Solving maximization problem defines the upper cutoff implicitly with $\frac{\kappa}{\pi(1-\overline{p})} = \frac{\lambda}{(r+\lambda\overline{p})^2}(r + \lambda - \frac{\kappa}{\pi}r)$. Note that $\overline{p} < \frac{\pi}{\lambda}$ is necessarily satisfied, as $\lim_{p\to\frac{\pi}{\lambda}} V_{Tr}(p;\overline{p}) = +\infty$, while $V(\overline{p};\overline{p})$ is finite (and $V(p;\overline{p})$ is an increasing function). Hence, such solution satisfies the initial conjecture.

There are no improvements to the proposed solution via mixed strategies. Define $V_{Indiff}(q) \equiv V(q;q) = \frac{\lambda}{r}q + \left(\frac{(\lambda\pi(1-q) - \kappa r)q}{r\pi(q + \frac{r}{\lambda})}\right)$ - the value from mixing between experimentation with and without training (the agent mixes only if indifferent, so value matching and smooth pasting must be satisfied). $V_{Indiff}(q)$ is a concave function, while $V(p)$ is strictly convex around $\overline{p}$. $V_{Indiff}(q)$ and $V(p)$ have a single intersection at $\overline{p}$, by construction, implying that $V_{Indiff}(p) < V(p)$ for any p except $\overline{p}$ - guaranteeing no further improvements by mixing.

To conclude the solution, $\underline{p}$ is obtained solely by value matching, as in case i. I.e. $\underline{p} = \arg\min_p[V_{bench}(p) = V_{Tr}(p;\overline{p})]$. Define $\kappa_0(\pi)$ and $\underline{\kappa}(\pi)$ as costs satisfying $\underline{p} = 0$ and $\underline{p} = \hat{p}$, respectively. Note that the value function violates smooth pasting at $\underline{p}$ by the same argument as in case i.

Overall, the solution ii. is characterised by:

$$V(p) = \begin{cases} \max\{\frac{s}{r}, \frac{\lambda}{r}p + (\frac{s}{r} - \frac{\lambda}{r}\hat{p})\frac{f(p)}{f(\hat{p})}\} & p < \underline{p} \textit{ (safe arm or pure exp)} \\ y(p) + \left(\frac{\lambda}{r}\overline{p} - y(\overline{p}) + \frac{(\lambda\pi(1-\overline{p})-\kappa r)\overline{p}}{r\pi(\overline{p}+\frac{r}{\lambda})}\right)\frac{g(p)}{g(\overline{p})} & p \in [\underline{p}, \overline{p}] \textit{ (training)} \\ \frac{\lambda}{r}p + \left(\frac{(\lambda\pi(1-\overline{p})-\kappa r)\overline{p}}{r\pi(\overline{p}+\frac{r}{\lambda})}\right)\frac{f(p)}{f(\overline{p})} & p > \overline{p} \textit{ (pure exp)} \end{cases}$$

It holds whenever $\overline{p} < \frac{\pi}{\lambda}$. This is satisfied iff $\overline{p}(\pi,\kappa) < \frac{\pi}{\lambda}$ ($\overline{p}(\pi,\kappa)$ is defined in case i.). To see this, realize that at $\overline{p}(\pi,\kappa)$ the agent is indifferent by construction, so $\overline{p}(\pi,\kappa)$ lies on $V_{Indiff}(p)$, as it does on $y(p)$. Similarly, $\frac{\pi}{\lambda}$ is another intersection of $V_{Indiff}(p)$ and $y(p)$. $V_{Indiff}(\frac{\pi}{\lambda})$ is constant for any β by definition, including $\beta = 1$. At $p = \frac{\pi}{\lambda}$ and $\beta = 1$, $dp = 0$, so the agent is stuck and receives the value of training forever at full intensity until the news arrives - this is exactly $y(\frac{\pi}{\lambda})$. Given concavity of $V_{Indiff}(p)$, $\overline{p}(\pi,\kappa)$ and $\frac{\pi}{\lambda}$ are the only intersections of $V_{Indiff}(p)$ and $y(p)$, and $V_{Indiff}(p) > y(p)$ only for $p \in (\overline{p}(\pi,\kappa), \frac{\pi}{\lambda})$. At the same time, $V(\overline{p}) > y(\overline{p})$. Combining this with $V(\overline{p}) = V_{indiff}(\overline{p})$ ensures that $\overline{p} \in (\overline{p}(\pi,\kappa), \frac{\pi}{\lambda})$.

Cases i. and ii. are mutually exclusive: case i. holds if $\frac{\pi}{\lambda} \leq \overline{p}(\pi,\kappa)$, and case ii. holds if $\frac{\pi}{\lambda} > \overline{p}(\pi,\kappa)$. Finally, continuity and convexity of $V(p)$, as well as its smoothness everywhere but $\underline{p}$, follow directly in both cases from the proof provided. ■

Proposition 1.4:

Proof. Follows trivially from Lemmas A.2.1 and A.2.1 and the optimal strategies construction. ■

Appendix B

Appendix to Chapter 2

B.1 Appendix. Proofs

To prove the main results I establish and prove some auxiliary lemmas.

Lemma B.1.1. *The ability to train the risky arm enhances the value iff* $\kappa < \pi\frac{s}{r}$. *If* $\kappa = \pi\frac{s}{r}$*, the agent is indifferent between training the arm and not at a single belief* $\hat{q}$ *and strictly prefers pure experimentation or safe arm otherwise.*

Proof. Suppose that the agent is willing to train only at a single belief, q'. This implies that the optimal solution is characterized by $V_{bench}(q)$.

$q' \notin (\hat{q}, 1]$, as then $V'_{bench}(q) = 0$ and so $s < q\lambda + q\lambda\big(V(1) - V(q)\big) - V'(q)\lambda q(1-q) < q\lambda + \kappa + q\lambda\big(V(1) - V(q)\big) - V'(q)q(\pi + \lambda(1-q))$ - training is dominated by both experimenting and using the safe arm.

$q' \notin [0, \hat{q})$. Suppose $q' \in [0, \hat{q})$. Given the agent purely experiments in the benchmark in $[0, \hat{q})$, it must be that $\kappa = V'_{bench}(q')q'\pi$ for indifference at q', where $V'_{bench}(q) = \frac{\lambda}{\lambda+r}(1 + \frac{s}{r})$ $\forall q \in [0, \hat{q})$ (by Lemma 2.2). This implies that there exists a non-empty set of beliefs $(q', \hat{q})$, where $\kappa < V'_{bench}(q)q\pi$, and so the agent strictly prefers training there, which violates the premise of q' being a single belief where training is optimal.

As such, $q' = \hat{q}$, and to guarantee that such belief is unique, it must be that $\lim_{q \to \hat{q}^-} V'_{bench}(q)q\pi < \kappa$, or $\kappa \geq \frac{\lambda}{\lambda+r}(1 + \frac{s}{r})\hat{q}\pi \implies \kappa \geq \pi\frac{s}{r}$. So, training is used by the agent and strictly improves $V(q)$ iff $\kappa < \pi\frac{s}{r}$, and for $\kappa > \pi\frac{s}{r}$ it is strictly dominated by either pure experimentation or using the safe arm.

■

Lemma B.1.2 (Conjecture on structure)**.** *If* $\kappa < \pi\frac{s}{r}$*, there always exist two cutoffs* $\underline{q} \in [0, \hat{q})$ *and* $\overline{q} \in (\hat{q}, 1]$*, such that the agent trains the risky arm for beliefs in* $[\underline{q}, \overline{q}]$*, purely experiments for* $q < \underline{q}$ *and uses the safe arm for* $q > \overline{q}$.

Proof. Follows directly from Lemma B.1.1, as $q' = \hat{q}$ is the belief where training is the most beneficial. If $\kappa < \pi\frac{s}{r}$, by continuity of $V(q)$ there must exist a region $[\underline{q}, \overline{q}]$ such that $\hat{q} \in [\underline{q}, \overline{q}]$ where training is strictly beneficial. The agent then purely experiments for lower beliefs and switches to the safe arm for higher beliefs, according to the benchmark solution in Lemma 2.2.

Note that $\lim_{q\to 0} V'_{bench}(q)q\pi = 0 \leq \kappa$, so training is strictly dominated by pure experimentation for $q \to 0$ for any $\kappa > 0$, which implies that $\underline{q} > 0$ for any $\kappa > 0$, and $\underline{q} = 0$ iff $\kappa = 0$. In turn, $\overline{q} = 1$ is possible for some $\kappa > 0$, as proven below.

■

Proposition 2.1 and Proposition 2.2:

Proof. Conjecture the structure according to lemma B.1.2, and solve for the cutoffs consecutively from the lowest to highest.

Given the strictly decreasing belief dynamics and the conjectured strategy structure, $V(q)$ for $q < \underline{q}$ must be such that $C_{Exp} = 0$, since the agent purely experiments forever unless the breakdown arrives.

At $\underline{q}$, value matching ensures $V_{Exp}(\underline{q}) = V_{Tr}(\underline{q})$, which is rearranged to $C_{Tr}(\underline{q}) = (y(\underline{q}, V(1)) - \frac{\lambda}{\lambda+r}(1+V(1))\underline{q})\frac{1}{g(\underline{q})}$. Standard argument as presented in proof for Propositions 1.1 and 1.2 in Appendix A.2 establishes that the smooth pasting must hold at $\underline{q}$. Hence, the cutoff is uniquely defined via equation $V'_{Exp}(\underline{q}) = V'_{Tr}(\underline{q})$, or $\frac{\lambda}{\lambda+r}(1+V(1)) = \frac{\lambda}{\lambda+\pi+r}(1+V(1) - \frac{\kappa}{r}) - C_{Tr}(\underline{q}) * g'(\underline{q}) \implies \underline{q} = \frac{\kappa(\lambda+r)}{\lambda\pi(1+V(1))}$.

Note that the above characterizes the optimal solution $V(q)$ for any $q < \overline{q}$ up to $V(1)$, which is established in what follows. But before that, let us confirm that the agent is never willing to purely experiment for $q > \overline{q}$. At this instance, ignore the option of switching to the safe arm and consider the value of training as opposed to the value of purely experimenting for $q > \overline{q}$. The agent prefers to experiment for some belief q iff $\kappa > V'(q)\pi q$ (i.e. training cost exceeds its benefit), where $V'(q) = V'_{Tr}(q)$. Some algebraic manipulations allow to establish that $V'(q)\pi q$ is strictly increasing in q for any $q > \underline{q}$. This implies that given the agent trains the arm in $[\underline{q}, \overline{q}]$, $\kappa \leq V'(q)\pi q$ there, and so $\kappa < V'(q)\pi q$ holds for any $q > \overline{q}$, making the pure experimentation strictly dominated by training.

Finally, the value matching condition suffices to determine the upper cutoff $\overline{q}$. Assume q' solves $V_{Tr}(q) = y(q, V(1)) - \left(y(\underline{q}, V(1)) - \frac{\lambda}{\lambda+r}(V(1)+1)\,\underline{q}\right)\frac{g(q)}{g(\underline{q})} = \frac{s}{r} = V_{Safe}(q)$. The equation has a unique solution since its LHS monotonically increases in q. Whenever $q' \leq 1$ holds, $\overline{q} = q'$ implicitly defines the cutoff, and $V(1) = \frac{s}{r}$ complete the value function characterisation:

$$V(q) = \begin{cases} \frac{\lambda}{\lambda+r}\left(\frac{s}{r}+1\right)q & q < \underline{q} \\ y(q,\frac{s}{r}) - \left(y(\underline{q},\frac{s}{r}) - \frac{\lambda}{\lambda+r}\left(\frac{s}{r}+1\right)\underline{q}\right)\frac{g(q)}{g(\underline{q})} & q \in [\underline{q},\overline{q}] \\ \frac{s}{r} & q > \overline{q} \end{cases}$$

For $q' > 1$, the upper bound does not exist and the agent trains the arm for any $q > \underline{q}$ (as LHS of equation is less than the RHS). As such, the solution is characterised by:

$$V(q) = \begin{cases} \frac{\lambda}{\lambda+r}\left(V(1)+1\right)q & q < \underline{q} \\ y(q,V(1)) - \left(y(\underline{q},V(1)) - \frac{\lambda}{\lambda+r}\left(V(1)+1\right)\underline{q}\right)\frac{g(q)}{g(\underline{q})} & q \geq \underline{q} \end{cases}$$

where $V(1)$ is implicitly defined by $V(1) = y(1,V(1)) - \left(y(\underline{q},V(1)) - \frac{\lambda}{\lambda+r}\left(V(1)+1\right)\underline{q}\right)\frac{g(1)}{g(\underline{q})}$.

Finally, the condition of $q' = 1$ uniquely determines the boundary between the two distinct cases, and interpreted as $\pi = \pi^*(\kappa)$ in the main body of the Chapter.

Continuity, concavity and smoothness everywhere except for a single kink at $\overline{q}$ (whenever $\overline{q}$ exists) of $V(q)$ follow directly from the optimal strategy construction. ■

Proposition 2.3:

Proof. i. The fact that $V(q) > V_{bench}(q)$ for $q \in [\underline{q},\overline{q}]$ follows trivially from the optimality of training in this beliefs range. Hence, it remains to prove that $V(q) > V_{bench}(q)$ for $q < \underline{q}$ whenever $\pi > \pi^*(\kappa)$. Under $\pi > \pi^*(\kappa)$, the solution is characterised by Proposition 2.1 part ii. Optimality of training at $q = 1$ implies that $V(1) < V_{Safe}(q) = \frac{s}{r}$. Hence, $q < \underline{q}$ $V(q) = \frac{\lambda}{\lambda+r}\left(V(1)+1\right)q < \frac{\lambda}{\lambda+r}\left(\frac{s}{r}+1\right)q = V_{bench}(q)$ (this also relies on the observation that $\underline{q} < \hat{q}$).

ii. $\pi > \underline{\pi}(\kappa)$ guarantees that the agent is willing to train the arm in the neighbourhood of at least one belief, which by Lemma B.1.1 is established to be $\hat{q}$. $\overline{q} > \hat{q}$ follows trivially from there.

iii. Assume the agent shares the belief $q \in (\underline{q},\overline{q})$ and faces the bad arm. Then, the probability of the training succeeding before the first breakdown arrival in a single training cycle is equal to $\frac{\pi}{\pi+\lambda}(1 - e^{-(\pi+\lambda)\tau^*(q)} > 0$, where $\tau^*(q)$ denotes the deterministic deadline of quitting training by reaching $\underline{q}$ conditional on the current belief q and the absence of news by then. As such, for any belief $q \in (\underline{q},\overline{q})$, the probability of having a good arm in the long-run strictly increases by $\frac{\pi}{\pi+\lambda}(1 - e^{-(\pi+\lambda)\tau^*(q)}$ if the arm is bad (with probability q), and weakly increases if the arm is already good (since once the good arm is in use, it generates no breakdowns and remains in use forever), with a strict increase from 0 to 1 for $q \in (\hat{q},\overline{q})$. Once the optimal strategy results in infinitely repeated training cycles conditional on the arm remaining bad (i.e. when $\pi > \pi^*(\kappa)$), the probability of the training eventually succeeding approaches 1 almost surely (a property of Poisson processes). Hence, as the agent never abandons the risky arm, she is guaranteed to end up with the good arm in the limit. ■

Appendix C

Appendix to Chapter 3

C.1 Proofs

Lemma 1

Proof. **Part (i)**: by R1 (symmetry): $\sigma_i(l) = 1 - \lambda_i(r) - \sigma_i(r) \;\; \Rightarrow \;\; \sigma_i(l) + \sigma_i(r) = 1 - \lambda_i(r)$ and $\sigma_i(r) = 1 - \lambda_i(l) - \sigma_i(l) \;\; \Rightarrow \;\; \sigma_i(l) + \sigma_i(r) = 1 - \lambda_i(l)$. This implies that $\lambda_i(l) = \lambda_i(r) = \lambda_i$ – independent of signal θ.

By definition of $q_{\mathcal{A}}(s,\ T)$: $q_{\mathcal{A}}(s,\ T) = \sum_{\theta=l,r} \sum_{i=f,a} \Pr(\theta|s) \Pr(t_i|T) \lambda_i(\theta) = \sum_{\theta=l,r} \Pr(\theta|s) \times \sum_{i=f,a} \Pr(t_i|T)\lambda_i = \sum_{i=f,a} \Pr(t_i|T)\lambda_i \equiv \overline{\lambda}_T$

Part (ii): by definition of $q_{\mathcal{J}}(s,\ T)$ and using R1: $q_{\mathcal{L}}(s,\ T) = \sum_{\theta=l,r} \sum_{i=f,a} \Pr(\theta|s) \Pr(t_i|T) \sigma_i(\theta)$

By construction, $\Pr(l|-d) = \Pr(r|d) = p_\theta$. Hence, $\Pr(\theta|s) = \Pr(\theta_-|-s) \quad \forall\, s = \{-d, d\}\,,\ \forall\, \theta = \{l, r\}$. Rearrange $q_{\mathcal{R}}(s,\ T)$ to obtain:

$$\begin{aligned} q_{\mathcal{R}}(s,\ T) &= \sum_{i=f,a} \Pr(t_i|T) \sum_{\theta=l,r} \Pr(\theta|s)\left(1 - \lambda_i(\theta) - \sigma_i(\theta)\right) \\ &= \sum_{i=f,a} \Pr(t_i|T) \sum_{\theta=l,r} \Pr(\theta|s) \sigma_i(\theta_-) = \sum_{i=f,a} \Pr(t_i|T) \sum_{\theta=l,r} \Pr(\theta_-|-s) \sigma_i(\theta_-) = q_{\mathcal{L}}(-s,\ T) \end{aligned}$$

Thus, $q_{\mathcal{L}}(-s,\ T) = q_{\mathcal{R}}(s,\ T)$, symmetry is proved. Using R2:

$$\begin{aligned}
q_{\mathcal{L}}(s,\ T) &= \sum_{i=f,a} \Pr(t_i|T) \sum_{\theta=l,r} \Pr(\theta|s)\sigma_i(\theta) = \sum_{i=f,a} \Pr(t_i|T) \left(\frac{\sum_{\theta=l,r} \Pr(\theta|s)\sigma_i(\theta)}{1-\lambda_i} \right)(1-\lambda_i) \\
&= \sum_{i=f,a} \Pr(t_i|T) \left(\frac{\Pr(l|s)\,\sigma_i(l) + \Pr(r|s)\,\sigma_i(r)}{\sigma_i(l)+\sigma_i(r)} \right)(1-\lambda_i) \\
&= \sum_{i=f,a} \Pr(t_i|T) \left(\frac{\Pr(l|s)\,\frac{\sigma_i(l)}{\sigma_i(r)} + \Pr(r|s)}{\frac{\sigma_i(l)}{\sigma_i(r)}+1} \right)(1-\lambda_i) \\
&= \sum_{i=f,a} \Pr(t_i|T) \left(\frac{\Pr(l|s)\,c(l) + \Pr(r|s)}{c(l)+1} \right)(1-\lambda_i) \\
&= \left(\frac{\Pr(l|s)\,c(l) + \Pr(r|s)}{c(l)+1} \right) \sum_{i=f,a} \Pr(t_i|T)(1-\lambda_i) = \left(\frac{\Pr(l|s)\,c(l) + \Pr(r|s)}{c(l)+1} \right)(1-\overline{\lambda}_T)
\end{aligned}$$

$$q_{\mathcal{L}}(-d,\ T) = q_{\mathcal{R}}(d,\ T) = q\left(1-\overline{\lambda}_T\right)$$

$$q_{\mathcal{L}}(d,\ T) = q_{\mathcal{R}}(-d,\ T) = (1-q)\left(1-\overline{\lambda}_T\right)$$

Note that $q = \frac{pc(l)+(1-p)}{c(l)+1} = \frac{c(l)}{c(l)+1}p + \left(1-\frac{c(l)}{c(l)+1}\right)(1-p) \in [1-p,p]$ by construction.

■

Lemma 2

Proof. **Part (i)** : by Lemma 1: $\Pr(m,a|s,T) = C_N^a C_{N-a}^m \overline{\lambda}_T^a \left(1-\overline{\lambda}_T\right)^{N-a} \widetilde{q}_{\mathcal{L}}(s)^m \widetilde{q}_{\mathcal{R}}(s)^{N-a-m}$. Using Bayes rule, update $\Pr(T|m,a)$:

$$\begin{aligned}
\Pr(T|m,a) &= \frac{\sum_s \Pr(m,a|s,T)}{\sum_T \sum_s \Pr(m,a|s,T)} = \frac{C_N^a C_{N-a}^m \overline{\lambda}_T^a \left(1-\overline{\lambda}_T\right)^{N-a} \sum_s \widetilde{q}_{\mathcal{L}}(s)^m \widetilde{q}_{\mathcal{R}}(s)^{N-a-m}}{\sum_T \left(C_N^a C_{N-a}^m \overline{\lambda}_T^a \left(1-\overline{\lambda}_T\right)^{N-a} \sum_s \widetilde{q}_{\mathcal{L}}(s)^m \widetilde{q}_{\mathcal{R}}(s)^{N-a-m} \right)} \\
&= \frac{\overline{\lambda}_T^a \left(1-\overline{\lambda}_T\right)^{N-a}}{\sum_T \overline{\lambda}_T^a \left(1-\overline{\lambda}_T\right)^{N-a}} = \Pr(T|a).
\end{aligned}$$

Part (ii): Using Bayes rule, update $\Pr(s|m,a)$:

$$\begin{aligned}
\Pr(s|m,a) &= \frac{\sum_T \Pr(m,a|s,T)}{\sum_T \sum_s \Pr(m,a|s,T)} = \frac{\sum_T C_N^a C_{N-a}^m \overline{\lambda}_T^a \left(1-\overline{\lambda}_T\right)^{N-a} \widetilde{q}_{\mathcal{L}}(s)^m \widetilde{q}_{\mathcal{R}}(s)^{N-a-m}}{\sum_s \left(\sum_T \left(C_N^a C_{N-a}^m \overline{\lambda}_T^a \left(1-\overline{\lambda}_T\right)^{N-a} \right) \widetilde{q}_{\mathcal{L}}(s)^m \widetilde{q}_{\mathcal{R}}(s)^{N-a-m} \right)} \\
&= \frac{\widetilde{q}_{\mathcal{L}}(s)^m \widetilde{q}_{\mathcal{R}}(s)^{N-a-m} \sum_T \overline{\lambda}_T^a \left(1-\overline{\lambda}_T\right)^{N-a}}{\left(\sum_T \overline{\lambda}_T^a \left(1-\overline{\lambda}_T\right)^{N-a} \right) \sum_s \widetilde{q}_{\mathcal{L}}(s)^m \widetilde{q}_{\mathcal{R}}(s)^{N-a-m}} = \frac{\widetilde{q}_{\mathcal{L}}(s)^m \widetilde{q}_{\mathcal{R}}(s)^{N-a-m}}{\sum_s \widetilde{q}_{\mathcal{L}}(s)^m \widetilde{q}_{\mathcal{R}}(s)^{N-a-m}}.
\end{aligned}$$

■

Corollary 1

Proof. By construction of the model, $r_L(\Omega) = r_R(\Omega) = \beta \mathrm{E}[t|\Omega]$. Using Lemma 2(i), calculate $E[t|\Omega]$: $E[t|\Omega] = E[t|m,a] = \sum_T \Pr(T|m,a) E[t|T] = \sum_T \Pr(T|a) E[t|T] = E[t|a]$. This proves independence from m.

$$
\begin{aligned}
E[x|a] &= \sum_T \Pr(T|a)\, E[t|T] \\
&= \frac{\overline{\lambda}_F^a\left(1-\overline{\lambda}_F\right)^{N-a}}{\sum_T \overline{\lambda}_T^a\left(1-\overline{\lambda}_T\right)^{N-a}}\left(p\gamma-(1-p)\gamma\right)+\frac{\overline{\lambda}_A^a\left(1-\overline{\lambda}_A\right)^{N-a}}{\sum_T \overline{\lambda}_T^a\left(1-\overline{\lambda}_T\right)^{N-a}}\left((1-p)\gamma-p\gamma\right) \\
&= \gamma(2p-1)\frac{\overline{\lambda}_F^a\left(1-\overline{\lambda}_F\right)^{N-a}-\overline{\lambda}_A^a\left(1-\overline{\lambda}_A\right)^{N-a}}{\sum_T \overline{\lambda}_T^a\left(1-\overline{\lambda}_T\right)^{N-a}}=\overline{\gamma}\frac{\overline{\lambda}_F^a\left(1-\overline{\lambda}_F\right)^{N-a}-\overline{\lambda}_A^a\left(1-\overline{\lambda}_A\right)^{N-a}}{\sum_T \overline{\lambda}_T^a\left(1-\overline{\lambda}_T\right)^{N-a}}.
\end{aligned}
$$

The last equation is obtained through defining $\overline{\gamma} \equiv \gamma(2p-1)$. The formula for $x(a)$ follows directly from above calculation. ■

Lemma 3

Proof. By definition, $\text{PM}(t_i) = u' \sum_{a=0}^{N-1} \widetilde{\Pr}(a|t_i)\left(|t_i - \beta E[t|a]| - |t_i - \beta E[t|t+1]|\right)$. Extend these for two misalignment types.

Part (i): For 'for' type (t_f), $\gamma \geq E[t|a]$ by construction, hence $\gamma \geq \beta E[t|a]$. Then:

$$
\begin{aligned}
\text{PM}(t_f) &= u' \sum_{a=0}^{N-1} \widetilde{\Pr}(a|t_f)\left(|\gamma-\beta E[t|a]|-|\gamma-\beta E[t|a+1]|\right) \\
&= u' \sum_{a=0}^{N-1} \widetilde{\Pr}(a|t_f)\left(\gamma-\beta E[t|a]-\gamma+\beta E[t|a+1]\right)=u'\beta\sum_{a=0}^{N-1} \widetilde{\Pr}(a|t_f)\left(E[t|a+1]-E[t|a]\right)
\end{aligned}
$$

Obviously, the sign of $\text{PM}(t_f)$ depends on $E[t|a+1]-E[t|a]$. Compute the difference as:

$$
\begin{aligned}
&E[t|a+1]-E[t|a] \\
&= \overline{\gamma}\frac{\overline{\lambda}_F^{a+1}\left(1-\overline{\lambda}_F\right)^{N-a-1}-\overline{\lambda}_A^{a+1}\left(1-\overline{\lambda}_A\right)^{N-a-1}}{\sum_T \overline{\lambda}_T^{a+1}\left(1-\overline{\lambda}_T\right)^{N-a-1}}-\overline{\gamma}\frac{\overline{\lambda}_F^a\left(1-\overline{\lambda}_F\right)^{N-a}-\overline{\lambda}_A^a\left(1-\overline{\lambda}_A\right)^{N-a}}{\sum_T \overline{\lambda}_T^a\left(1-\overline{\lambda}_T\right)^{N-a}} \\
&= \overline{\gamma}\left[\frac{\left(\overline{\lambda}_F^{a+1}\left(1-\overline{\lambda}_F\right)^{N-a-1}-\overline{\lambda}_A^{a+1}\left(1-\overline{\lambda}_A\right)^{N-a-1}\right)\left(\overline{\lambda}_F^a\left(1-\overline{\lambda}_F\right)^{N-a}+\overline{\lambda}_A^a\left(1-\overline{\lambda}_A\right)^{N-a}\right)}{\sum_T \overline{\lambda}_T^{a+1}\left(1-\overline{\lambda}_T\right)^{N-a-1}\sum_T \overline{\lambda}_T^a\left(1-\overline{\lambda}_T\right)^{N-a}}\right. \\
&\left.-\frac{\left(\overline{\lambda}_F^a\left(1-\overline{\lambda}_F\right)^{N-a}-\overline{\lambda}_A^a\left(1-\overline{\lambda}_A\right)^{N-a}\right)\left(\overline{\lambda}_F^{a+1}\left(1-\overline{\lambda}_F\right)^{N-a-1}+\overline{\lambda}_A^{a+1}\left(1-\overline{\lambda}_A\right)^{N-a-1}\right)}{\sum_T \overline{\lambda}_T^a\left(1-\overline{\lambda}_T\right)^{N-a}\sum_T \overline{\lambda}_T^{a+1}\left(1-\overline{\lambda}_T\right)^{N-a-1}}\right] \\
&= \frac{\overline{\gamma}\left(-2\overline{\lambda}_F^a\left(1-\overline{\lambda}_F\right)^{N-a}\overline{\lambda}_A^{a+1}\left(1-\overline{\lambda}_A\right)^{N-a-1}+2\overline{\lambda}_F^{a+1}\left(1-\overline{\lambda}_F\right)^{N-a-1}\overline{\lambda}_A^a\left(1-\overline{\lambda}_A\right)^{N-a}\right)}{\sum_T \overline{\lambda}_T^{a+1}\left(1-\overline{\lambda}_T\right)^{N-a-1}\sum_T \overline{\lambda}_T^a\left(1-\overline{\lambda}_T\right)^{N-a}} \\
&= \frac{2\overline{\gamma}\overline{\lambda}_F^a\left(1-\overline{\lambda}_F\right)^{N-a-1}\overline{\lambda}_A^a\left(1-\overline{\lambda}_A\right)^{N-a-1}\left(\overline{\lambda}_{\mathbf{F}}-\overline{\lambda}_{\mathbf{A}}\right)}{\sum_T \overline{\lambda}_T^{a+1}\left(1-\overline{\lambda}_T\right)^{N-a-1}\sum_T \overline{\lambda}_T^a\left(1-\overline{\lambda}_T\right)^{N-a}}
\end{aligned}
$$

Observe that the sign of $E[t|a+1]-E[t|a]$ depends purely on the sign of $\left(\overline{\lambda}_F-\overline{\lambda}_A\right)$ and is therefore independent of the value of a. Hence,

$$E[t|a+1]-E[t|a]=\begin{cases}>0, & if\ \overline{\lambda}_F>\overline{\lambda}_A\\ =0, & if\ \overline{\lambda}_F=\overline{\lambda}_A\\ <0, & if\ \overline{\lambda}_F<\overline{\lambda}_A\end{cases}$$

Given $u^{'}<0$,

$$\mathrm{PM}(t_f)=\begin{cases}<0, & if\ \overline{\lambda}_F>\overline{\lambda}_A\\ =0, & if\ \overline{\lambda}_F=\overline{\lambda}_A\\ >0, & if\ \overline{\lambda}_F<\overline{\lambda}_A\end{cases}$$

Part (ii): Repeat analysis for 'against' type (t_a). For t_a, $-\gamma\leq E[t|a]$ by construction, hence $-\gamma\leq\beta E[t|a]$. Then:

$$\begin{aligned}\mathrm{PM}(t_a)&=u^{'}\sum_{a=0}^{N-1}\widetilde{\Pr}(a|t_a)\left(|-\gamma-\beta E[t|a]|-|-\gamma-\beta E[\gamma|a+1]|\right)\\ &=u^{'}\sum_{a=0}^{N-1}\widetilde{\Pr}(a|t_a)\left(\gamma+\beta E[t|a]-\gamma-\beta E[t|a+1]\right)=u^{'}\beta\sum_{a=0}^{N-1}\widetilde{\Pr}(a|t_a)\left(E[t|a]-E[t|a+1]\right)\\ &=-u^{'}\beta\sum_{a=0}^{N-1}\widetilde{\Pr}(a|t_a)\left(E[t|a+1]-E[t|a]\right)\end{aligned}$$

Observe that $\mathrm{sign}(\mathrm{PM}(t_a))=-sign(\mathrm{PM}(t_f))$ given the construction. Hence, using analysis for part (i), we can directly conclude that:

$$\mathrm{PM}(t_a)=\begin{cases}>0, & if\ \overline{\lambda}_F>\overline{\lambda}_A\\ =0, & if\ \overline{\lambda}_F=\overline{\lambda}_A\\ <0, & if\ \overline{\lambda}_F<\overline{\lambda}_A\end{cases}$$

■

Proposition 1

Proof. Proof follows directly from Lemma 3.

Case 1. Suppose $\overline{\lambda}_F=\overline{\lambda}_A$. Then, $\mathrm{PM}(t_f)=PM(t_a)=0$, which means that both types are indifferent between voting and abstaining as long as $\overline{\lambda}_F=\overline{\lambda}_A$. To guarantee $\overline{\lambda}_F=\overline{\lambda}_A$, it must be that $p\lambda_f+(1-p)\lambda_a=(1-p)\lambda_f+p\lambda_a\Rightarrow\lambda_f=\lambda_a$ (assuming $p>\frac{1}{2}$). Hence, there always exists an equilibrium, where both types mix at $\lambda_f=\lambda_a\in[0,1]$.

[Note that if $\lambda_f=\lambda_a=1$ or $\lambda_f=\lambda_a=0$, then we face pure strategies. If we set off equilibrium beliefs to be $E[t|a]=0$ for any level of a, then these will replicate the logic above exactly].

Case 2. Suppose $\overline{\lambda}_F>\overline{\lambda}_A$. Then, by Lemma 3 $\mathrm{PM}(t_f)<0\Rightarrow\lambda_f=1$ ('for' type strictly prefers abstaining), and $\mathrm{PM}(t_a)>0\Rightarrow\lambda_a=0$ ('against' type best responds by voting). Hence, $\overline{\lambda}_F=p\lambda_f+(1-p)\lambda_a=p$ and $\overline{\lambda}_A=1-p$. Since, $p>\frac{1}{2}$, then $p>1-p$,

and $\overline{\lambda}_F > \overline{\lambda}_A$ holds indeed. Hence, $\lambda_f = 1$ and $\lambda_a = 0$ forms an equilibrium.
Assuming $\overline{\lambda}_F < \overline{\lambda}_A$ gives rise to symmetric equilibrium, where $\lambda_f = 0$ and $\lambda_a = 1$. ■

Lemma 4

Proof. Consider $PM_{\text{kj}}(t_i) = u' \sum_{a=0}^{N-1} \widetilde{\Pr}(a|t_i)(|t_i - \beta E[t|a_k]| - |t_i - \beta E[t|a_j]|)$, as derived in the paper. The excluded policy gap depends on a_j, which is abstention given a abstainers out of N-1 plus remaining citizen playing j. That is, $a_L = a_R = a$ and $a_A = a+1$. Then,

$$\begin{aligned}\mathbf{PM}_{\text{LR}}(\mathbf{t_i}) &= u' \sum_{a=0}^{N-1} \widetilde{\Pr}(a|t_i)(|t_i - \beta E[t|a_L]| - |t_i - \beta E[t|a_R]|) \\ &= u' \sum_{a=0}^{N-1} \widetilde{\Pr}(a|t_i)(|t_i - \beta E[t|a]| - |t_i - \beta E[t|a]|) = \mathbf{0}\end{aligned}$$

$$\begin{aligned}\text{PM}_{\text{LA}}(\mathbf{t_i}) &= u' \sum_{a=0}^{N-1} \widetilde{\Pr}(a|t_i)(|t_i - \beta E[t|a_L]| - |t_i - \beta E[t|a_A]|) \\ &= u' \sum_{a=0}^{N-1} \widetilde{\Pr}(a|t_i)(|t_i - \beta E[t|a]| - |t_i - \beta E[t|a+1]|)\end{aligned}$$

$$\begin{aligned}\text{PM}_{\text{RA}}(\mathbf{t_i}) &= u' \sum_{a=0}^{N-1} \widetilde{\Pr}(a|t_i)(|t_i - \beta E[t|a_R]| - |t_i - \beta E[t|a_A]|) \\ &= u' \sum_{a=0}^{N-1} \widetilde{\Pr}(a|t_i)(|t_i - \beta E[t|a]| - |t_i - \beta E[t|a+1]|) = \text{PM}_{\text{LA}}(\mathbf{t_i})\end{aligned}$$

Observe that $\text{PM}_{\text{LA}}(t_i) = u' \sum_{a=0}^{N-1} \widetilde{\Pr}(a|t_i)(|t_i - \beta E[t|a]| - |t_i - \beta E[t|a+1]|)$ exactly replicates the definition of PM used in Lemma 3. Hence, the conclusions of Lemma 3 hold here for $\text{PM}_{\text{LA}}(t_i) = \text{PM}_{\text{RA}}(t_i)$:

$$\text{PM}_{\text{LA}}(\mathbf{t_f}) = \begin{cases} = \mathbf{0} \textbf{ if } \overline{\lambda}_\mathbf{F} = \overline{\lambda}_\mathbf{A} \\ < \mathbf{0} \textbf{ if } \overline{\lambda}_\mathbf{F} > \overline{\lambda}_\mathbf{A} \end{cases} \text{ and } \text{P}\mathbf{M}_{\text{LA}}(\mathbf{t_a}) = \begin{cases} = \mathbf{0} \textbf{ if } \overline{\lambda}_\mathbf{F} = \overline{\lambda}_\mathbf{A} \\ > \mathbf{0} \textbf{ if } \overline{\lambda}_\mathbf{F} > \overline{\lambda}_\mathbf{A} \end{cases}$$

as shown in Lemma 3 proof. This completes PM sign analysis.

Now consider EM as derived in the paper:

$$\begin{aligned}EM_{\text{kj}}(\theta, t_i) &= u' \sum_{a=0,\text{ even}}^{N-1} \left[\widetilde{\Pr}\left(\frac{N-a-1}{2}, a|\theta, t_i\right) E[|y(\Omega_k) - s| - |y(\Omega_j) - s||m, a, \theta, t_i] \right] \\ &+ u' \sum_{a=1,\text{ odd}}^{N-2} \begin{bmatrix} \widetilde{\Pr}\left(\frac{N-a}{2} - 1, a|\theta, t_i\right) E[|y(\Omega_k) - s| - |y(\Omega_j) - s||m, a, \theta, t_i] \\ +\widetilde{\Pr}\left(\frac{N-a}{2}, a|\theta, t_i\right) E[|y(\Omega_k) - s| - |y(\Omega_j) - s||m, a, \theta, t_i] \end{bmatrix}\end{aligned}$$

Consider how $y(\Omega_j)$ depends on the action j in three pivotal cases (the first brackets indicate votes distribution of N-1 citizens, i.e. $(L\ votes,\ R\ votes) = (m,\ N-1-a-m)$):

$$\left(\frac{N-a-1}{2},\ \frac{N-a-1}{2}\right):\quad y\left(\Omega_j\right) = \begin{cases} -d, & if\ j = L \\ d, & if\ j = R \\ Tie, & if\ j = A \end{cases}$$

$$\left(\frac{N-a}{2}-1,\ \frac{N-a}{2}\right):\quad y\left(\Omega_j\right) = \begin{cases} Tie, & if\ j = L \\ d, & if\ j = R \\ d, & if\ j = A \end{cases}$$

$$\left(\frac{N-a}{2},\ \frac{N-a}{2}-1\right):\quad y\left(\Omega_j\right) = \begin{cases} -d, & if\ j = L \\ Tie, & if\ j = R \\ -d, & if\ j = A \end{cases}$$

Note that I assume that in case of a tie a winning candidate is chosen by tossing a fair coin, i.e. $\Pr(-d) = \Pr(d) = \frac{1}{2}$. Hence, the expected policy gap in the event of a tie is:

$$\begin{aligned} E\left[|y\left(\Omega_j\right) - s||\mathbf{\Omega_j}\mathbf{=}\ \mathbf{Tie}, \theta, t_i\right] &= \frac{1}{2}E\left[|-d-s||\theta, t_i\right] + \frac{1}{2}E\left[|d-s||\theta, t_i\right] \\ &= \frac{1}{2}E\left[d+s|\theta, t_i\right] + \frac{1}{2}E\left[d-s|\theta, t_i\right] = \frac{1}{2}E\left[2d|\theta, t_i\right] = \mathbf{d}, \end{aligned}$$

where the signs of the moduli are fully determined by $y\left(\Omega_j\right) \in \{-d, d\}$, since $s \in \{-d, d\} \Longrightarrow -d \leq s \leq d$.

Now, given $y\left(\Omega_j\right)$ updates, calculate EM for each action pair (again, moduli are determined by $y\left(\Omega_j\right)$, since $\Longrightarrow -d \leq s \leq d$).

<u>For L vs R:</u>

$$\begin{aligned} EM_{\text{LR}}\left(\theta, t_i\right) &= u^{'} \sum_{a=0,\text{ even}}^{N-1} \left[\widetilde{\Pr}\left(\frac{N-a-1}{2}, a|\theta, t_i\right) E\left[|-d-s| - |d-s||m, a, \theta, t_i\right]\right] \\ &+ u^{'} \sum_{a=1,\text{ odd}}^{N-2} \left[\begin{array}{l} \widetilde{\Pr}\left(\frac{N-a}{2}-1, a|\theta, t_i\right) E\left[Tie - |d-s||m, a, \theta, t_i\right] \\ +\widetilde{\Pr}\left(\frac{N-a}{2}, a|\theta, t_i\right) E\left[|-d-s| - Tie|m, a, \theta, t_i\right] \end{array}\right] \\ &= u^{'} \sum_{a=0,\text{ even}}^{N-1} \left[\widetilde{\Pr}\left(\frac{N-a-1}{2}, a|\theta, t_i\right) E\left[d+s-d+s|m, a, \theta, t_i\right]\right] \\ &+ u^{'} \sum_{a=1,\text{ odd}}^{N-2} \left[\begin{array}{l} \widetilde{\Pr}\left(\frac{N-a}{2}-1, a|\theta, t_i\right) E\left[d-d+s|m, a, \theta, t_i\right] \\ +\widetilde{\Pr}\left(\frac{N-a}{2}, a|\theta, t_i\right) E\left[d+s-d|m, a, \theta, t_i\right] \end{array}\right] \\ &= 2u^{'} \sum_{a=0,\text{ even}}^{N-1} \left[\widetilde{\Pr}\left(\frac{N-a-1}{2}, a|\theta, t_i\right) E\left[s|m, a, \theta, t_i\right]\right] \\ &+ u^{'} \sum_{a=1,\text{ odd}}^{N-2} \left[\begin{array}{l} \widetilde{\Pr}\left(\frac{N-a}{2}-1, a|\theta, t_i\right) E\left[s|m, a, \theta, t_i\right] \\ +\widetilde{\Pr}\left(\frac{N-a}{2}, a|\theta, t_i\right) E\left[s|m, a, \theta, t_i\right] \end{array}\right] \end{aligned}$$

Observe that the sign of $E\left[s|m, a, \theta, t_i\right]$ for 3 pivotal cases would determine the sign of $EM_{\text{LR}}\left(\theta, t_i\right)$.

Thus, consider what $E\left[s|m,a,\theta,t_i\right]$ is equal to:

$$\begin{aligned}
\widetilde{\Pr}\left(s|m,a,\theta,t_i\right) &= \frac{\sum_T \widetilde{\Pr}\left(m,a|s,T\right)\Pr\left(\theta|s\right)\Pr\left(t_i|T\right)}{\sum_T\sum_s \widetilde{\Pr}\left(m,a|s,T\right)\Pr\left(\theta|s\right)\Pr\left(t_i|T\right)} \\
&= \frac{\sum_T C^a_{N-1}C^m_{N-1-a}\overline{\lambda}^a_T\left(1-\overline{\lambda}_T\right)^{N-1-a}\widetilde{q}_{\mathcal{L}}(s)^m\widetilde{q}_{\mathcal{R}}(s)^{N-1-a-m}\Pr\left(\theta|s\right)\Pr\left(t_i|T\right)}{\sum_s\left(\sum_T\left(C^a_{N-1}C^m_{N-1-a}\overline{\lambda}^a_T\left(1-\overline{\lambda}_T\right)^{N-1-a}\Pr\left(t_i|T\right)\right)\widetilde{q}_{\mathcal{L}}(s)^m\widetilde{q}_{\mathcal{R}}(s)^{N-1-a-m}\Pr\left(\theta|s\right)\right)} \\
&= \frac{\widetilde{q}_{\mathcal{L}}(s)^m\widetilde{q}_{\mathcal{R}}(s)^{N-1-a-m}\Pr\left(\theta|s\right)\sum_T\overline{\lambda}^a_T\left(1-\overline{\lambda}_T\right)^{N-1-a}\Pr\left(t_i|T\right)}{\sum_T\left(\overline{\lambda}^a_T\left(1-\overline{\lambda}_T\right)^{N-1-a}\Pr\left(t_i|T\right)\right)\sum_s\left(\widetilde{q}_{\mathcal{L}}(s)^m\widetilde{q}_{\mathcal{R}}(s)^{N-1-a-m}\Pr\left(\theta|s\right)\right)} \\
&= \frac{\widetilde{q}_{\mathcal{L}}(s)^m\widetilde{q}_{\mathcal{R}}(s)^{N-1-a-m}\Pr\left(\theta|s\right)}{\sum_s\left(\widetilde{q}_{\mathcal{L}}(s)^m\widetilde{q}_{\mathcal{R}}(s)^{N-1-a-m}\Pr\left(\theta|s\right)\right)} = \widetilde{\Pr}\left(s|m,a,\theta\right)
\end{aligned}$$

$$\begin{aligned}
E\left[s|m,a,\theta,t_i\right] = E\left[s|m,a,\theta\right] &= d\left[\widetilde{\Pr}\left(d|m,a,\theta\right)-\widetilde{\Pr}\left(-d|m,a,\theta\right)\right] \\
&= d\frac{\widetilde{q}_{\mathcal{L}}(d)^m\widetilde{q}_{\mathcal{R}}(d)^{N-1-a-m}\Pr\left(\theta|d\right)-\widetilde{q}_{\mathcal{L}}(-d)^m\widetilde{q}_{\mathcal{R}}(-d)^{N-1-a-m}\Pr\left(\theta|-d\right)}{\sum_s\left(\widetilde{q}_{\mathcal{L}}(s)^m\widetilde{q}_{\mathcal{R}}(s)^{N-1-a-m}\Pr\left(\theta|s\right)\right)} \\
&= d\frac{(1-q)^m q^{N-1-a-m}\Pr\left(\theta|d\right)-q^m(1-q)^{N-1-a-m}\Pr\left(\theta|-d\right)}{\sum_s\left(\widetilde{q}_{\mathcal{L}}(s)^m\widetilde{q}_{\mathcal{R}}(s)^{N-1-a-m}\Pr\left(\theta|s\right)\right)}
\end{aligned}$$

Thus, for pivotal values of m:

$$\begin{aligned}
E\left[s|m=\frac{\mathbf{N-a-1}}{\mathbf{2}},a,\theta\right] &= d\frac{(1-q)^{\frac{N-a-1}{2}}q^{\frac{N-a-1}{2}}\Pr\left(\theta|d\right)-q^{\frac{N-a-1}{2}}(1-q)^{\frac{N-a-1}{2}}\Pr\left(\theta|-d\right)}{q^{\frac{N-a-1}{2}}(1-q)^{\frac{N-a-1}{2}}\sum_s\Pr\left(\theta|s\right)} \\
&= d\left[\Pr\left(\theta|\mathbf{d}\right)-\Pr\left(\theta|-\mathbf{d}\right)\right]
\end{aligned}$$

$$\begin{aligned}
E\left[s|m=\frac{\mathbf{N-a}}{\mathbf{2}}-\mathbf{1},a,\theta\right] &= d\frac{q^{\frac{N-a}{2}-1}(1-q)^{\frac{N-a}{2}-1}\left[q\Pr\left(\theta|d\right)-(1-q)\Pr\left(\theta|-d\right)\right]}{q^{\frac{N-a}{2}-1}(1-q)^{\frac{N-a}{2}-1}\left[q\Pr\left(\theta|d\right)+(1-q)\Pr\left(\theta|-d\right)\right]} \\
&= d\frac{\left[q\Pr\left(\theta|d\right)-(1-q)\Pr\left(\theta|-d\right)\right]}{\left[q\Pr\left(\theta|d\right)+(1-q)\Pr\left(\theta|-d\right)\right]} = d\frac{\mathbf{q}-\Pr\left(\theta|-\mathbf{d}\right)}{q\Pr\left(\theta|d\right)+(1-q)\Pr\left(\theta|-d\right)} \\
\widetilde{E}\left[s|m=\frac{\mathbf{N-a}}{\mathbf{2}},a,\theta\right] &= d\frac{(1-q)\Pr\left(\theta|d\right)-q\Pr\left(\theta|-d\right)}{(1-q)\Pr\left(\theta|d\right)+q\Pr\left(\theta|-d\right)} = d\frac{\Pr\left(\theta|\mathbf{d}\right)-\mathbf{q}}{(1-q)\Pr\left(\theta|d\right)+q\Pr\left(\theta|-d\right)}
\end{aligned}$$

The sign of those depends on the brackets in bold. Hence, given $q\in[1-p_\theta,p_\theta]$:

$$\theta=l\Rightarrow\begin{cases}\Pr\left(\theta|d\right)-\Pr\left(\theta|-d\right)=1-2p_\theta<0\\ q-\Pr\left(\theta|-d\right)=q-p_\theta\leq 0\\ \Pr\left(\theta|d\right)-q=(1-p_\theta)-q\leq 0\end{cases}$$

$$\theta=r\Rightarrow\begin{cases}\Pr\left(\theta|d\right)-\Pr\left(\theta|-d\right)=2p_\theta-1>0\\ q-\Pr\left(\theta|-d\right)=q-(1-p_\theta)\geq 0\\ \Pr\left(\theta|d\right)-q=p_\theta-q\geq 0\end{cases}$$

Thus, as $u^{'}<0$, $\forall t_i$:

$$\mathbf{EM}_{\mathrm{LR}}\left(\mathbf{l},\mathbf{t_i}\right)>\mathbf{0}\ \ \textbf{and}\ \mathbf{EM}_{\mathrm{LR}}\left(\mathbf{r},\mathbf{t_i}\right)<\mathbf{0}$$

For L vs A:

$$
\begin{aligned}
EM_{\mathrm{LA}}(\theta,t_i) &= u^{'}\sum_{a=0,\ \text{even}}^{N-1}\left[\widetilde{\Pr}\left(\frac{N-a-1}{2},a|\theta,t_i\right)E\left[|-d-s|-Tie|m,a,\theta,t_i\right]\right]\\
&+u^{'}\sum_{a=1,\ \text{odd}}^{N-2}\left[\begin{array}{c}\widetilde{\Pr}\left(\frac{N-a}{2}-1,a|\theta,t_i\right)E\left[Tie-|d-s||m,a,\theta,t_i\right]\\+\widetilde{\Pr}\left(\frac{N-a}{2},a|\theta,t_i\right)E\left[|-d-s|-|-d-s||m,a,\theta,t_i\right]\end{array}\right]\\
&=u^{'}\sum_{a=0,\ \text{even}}^{N-1}\left[\widetilde{\Pr}\left(\frac{N-a-1}{2},a|\theta,t_i\right)E\left[s|m,a,\theta,t_i\right]\right]\\
&+u^{'}\sum_{a=1,\ \text{odd}}^{N-2}\left[\widetilde{\Pr}\left(\frac{N-a}{2}-1,a|\theta,t_i\right)E\left[s|m,a,\theta,t_i\right]\right]
\end{aligned}
$$

$EM_{\mathrm{LA}}(\theta,t_i)$ depends on the signs of $E[s|m,a,\theta,t_i]$ just as $EM_{\mathrm{LR}}(\theta,t_i)$ did, and $E[s|m,a,\theta,t_i]$ sign analysis holds in the same way. Hence, analysis for the sign of EM remains the same and $\forall t_i$:

$$\mathbf{EM}_{\mathrm{LA}}(\mathbf{l},\mathbf{t_i})>\mathbf{0}\ \ \textbf{and}\ \mathbf{EM}_{\mathrm{LA}}(\mathbf{r},\mathbf{t_i})<\mathbf{0}$$

For R vs A symmetry:

$$
\begin{aligned}
EM_{\mathrm{RA}}(\theta,t_i) &= u^{'}\sum_{a=0,\ \text{even}}^{N-1}\left[\widetilde{\Pr}\left(\frac{N-a-1}{2},a|\theta,t_i\right)E\left[|d-s|-Tie|m,a,\theta,t_i\right]\right]\\
&+u^{'}\sum_{a=1,\ \text{odd}}^{N-2}\left[\begin{array}{c}\widetilde{\Pr}\left(\frac{N-a}{2}-1,a|\theta,t_i\right)E\left[|d-s|-|d-s||m,a,\theta,t_i\right]\\+\widetilde{\Pr}\left(\frac{N-a}{2},a|\theta,t_i\right)E\left[Tie-|-d-s||m,a,\theta,t_i\right]\end{array}\right]\\
&=u^{'}\sum_{a=0,\ \text{even}}^{N-1}\left[\widetilde{\Pr}\left(\frac{N-a-1}{2},a|\theta,t_i\right)E\left[-s|m,a,\theta,t_i\right]\right]\\
&+u^{'}\sum_{a=1,\ \text{odd}}^{N-2}\left[\widetilde{\Pr}\left(\frac{N-a}{2},a|\theta,t_i\right)E\left[-s|m,a,\theta,t_i\right]\right]\\
&=u^{'}\sum_{a=0,\ \text{even}}^{N-1}\left[\widetilde{\Pr}\left(\frac{N-a-1}{2},a|\theta,t_i\right)E\left[s|m,a,\theta_-,t_i\right]\right]\\
&+u^{'}\sum_{a=1,\ \text{odd}}^{N-2}\left[\widetilde{\Pr}\left(\frac{N-a}{2}-1,a|\theta_-,t_i\right)E\left[s|m,a,\theta_-,t_i\right]\right]=EM_{\mathrm{LA}}(\theta_-,t_i)
\end{aligned}
$$

Where the symmetry results used are:

- $-E\left[s\middle|m=\frac{N-a-1}{2},a,\theta,t_i\right]=-d\left[\Pr(\theta|d)-\Pr(\theta|-d)\right]=d\left[\Pr(\theta|-d)-\Pr(\theta|d)\right]=d\left[\Pr(\theta_-|d)-\Pr(\theta_-|-d)\right]=E\left[s\middle|m=\frac{N-a-1}{2},a,\theta_-,t_i\right]$
- $-E\left[s\middle|m=\frac{N-a}{2},a,\theta,t_i\right]=-d\frac{\Pr(\theta|d)-q}{(1-q)\Pr(\theta|d)+q\Pr(\theta|-d)}=d\frac{q-\Pr(\theta_-|-d)}{(1-q)\Pr(\theta_-|-d)+q\Pr(\theta_-|d)}$ $=E\left[s\middle|m=\frac{N-a}{2}-1,a,\theta_-,t_i\right]$
- $\widetilde{\Pr}\left(\frac{N-a}{2},a|\theta,t_i\right)=\sum_T\sum_s\widetilde{\Pr}\left(\frac{N-a}{2},a|s,T\right)\Pr(s|\theta)\Pr(T|t_i)=$ $\sum_T\sum_s C_{N-1}^{a}C_{N-1-a}^{\frac{N-a}{2}}\overline{\lambda}_T^{a}\left(1-\overline{\lambda}_T\right)^{N-1-a}\widetilde{q}_{\mathcal{L}}(s)^{\frac{N-a}{2}}\widetilde{q}_{\mathcal{R}}(s)^{\frac{N-a}{2}-1}\Pr(s|\theta)\Pr(T|t_i)$

$= \sum_T \sum_s C^a_{N-1} C^{\frac{N-a}{2}-1}_{N-1-a} \overline{\lambda}^a_T \left(1-\overline{\lambda}_T\right)^{N-1-a} \widetilde{q}_{\mathcal{L}}(-s)^{\frac{N-a}{2}-1} \widetilde{q}_{\mathcal{R}}(-s)^{\frac{N-a}{2}} \Pr\left(-s|\theta_-\right)\Pr\left(T|t_i\right) =$
$\sum_T \sum_s \widetilde{\Pr}\left(\frac{N-a}{2}-1, a|-s, T\right) \Pr\left(-s|\theta_-\right)\Pr\left(T|t_i\right) = \widetilde{\Pr}\left(\frac{N-a}{2}-1, a|\theta_-, t_i\right)$

■

Proposition 2

Proof.

$$PM_{\text{kj}}\left(t_i\right) = u^{'} \sum_{a=0}^{N-1} \widetilde{\Pr}\left(a|t_i\right)\left(\left|t_i - \beta E\left[t|a_k\right]\right| - \left|t_i - \beta E\left[t|a_j\right]\right|\right)$$

Part 1: Suppose a citizen decides to vote. Then, the sign of $\Delta_{\text{LR}}\left(\theta, t_i\right)$ determines who to cast a vote for. By Lemma 4, $\forall t_i$:

$$\Delta_{\text{LR}}\left(\theta, t_i\right) = EM_{\text{LR}}\left(\theta, t_i\right) = \begin{cases} > 0, & if\ \theta = l \\ < 0, & if\ \theta = r \end{cases}$$

Thus, if a citizen votes, she strictly prefers voting for L if $\theta = l$ and for R if $\theta = r$ – that is, votes sincerely; hence $\Pr\left(\mathcal{L}|l,\ t_i\right) = \sigma_i(l) = 1 - \lambda_i$ and $\Pr\left(\mathcal{L}|r,\ t_i\right) = \sigma_i(r) = 0$. This means that by definition:

$$q_{\mathcal{J}}(s,\ T) = \begin{cases} \sum_{\theta=l,r} \sum_{i=f,a} \Pr\left(\theta|s\right)\Pr\left(t_i|T\right)\sigma_i(\theta), \text{ if } \mathcal{J}=\mathcal{L} \\ \sum_{\theta=l,r} \sum_{i=f,a} \Pr\left(\theta|s\right)\Pr\left(t_i|T\right)\left(\left(1-\lambda_i-\sigma_i(\theta)\right)\right), \text{ if } \mathcal{J}=\mathcal{R} \end{cases}$$
$$= \begin{cases} \Pr\left(l|s\right)\sum_{i=f,a}\Pr\left(t_i|T\right)\left(1-\lambda_i\right), \text{ if } \mathcal{J}=\mathcal{L} \\ \Pr\left(r|s\right)\sum_{i=f,a}\Pr\left(t_i|T\right)\left(1-\lambda_i\right), \text{ if } \mathcal{J}=\mathcal{R} \end{cases} = \begin{cases} \Pr\left(l|s\right)\left(1-\overline{\lambda}_T\right), \text{ if } \mathcal{J}=\mathcal{L} \\ \Pr\left(r|s\right)\left(1-\overline{\lambda}_T\right), \text{ if } \mathcal{J}=\mathcal{R} \end{cases}$$

Hence, $\widetilde{q}_{\mathcal{L}}(s) = \Pr\left(l|s\right)$ and $\widetilde{q}_{\mathcal{R}}(s) = \Pr\left(r|s\right)$. Or, linking to q, $\widetilde{q}_{\mathcal{L}}(-d) = \widetilde{q}_{\mathcal{R}}(d) = \mathbf{p}_\theta = \mathbf{q}$, and $\widetilde{q}_{\mathcal{L}}(d) = \widetilde{q}_{\mathcal{R}}(-d) = 1 - p_\theta = 1 - q$.

Part 2: Given part 1, the citizen's equilibrium actions are reduced to either voting sincerely or abstaining. Hence, to support any equilibrium one needs to consider $\Delta_{\text{LA}}\left(l, t_i\right)$ and $\Delta_{\text{RA}}\left(r, t_i\right)$ only. By symmetry of the motivations (Lemma 4 part 3):

$$\Delta_{\text{LA}}\left(l, t_i\right) = PM_{\text{LA}}\left(t_i\right) + EM_{\text{LA}}\left(l, t_i\right) = PM_{\text{RA}}\left(t_i\right) + EM_{\text{RA}}\left(r, t_i\right) = \Delta_{\text{RA}}\left(r, t_i\right).$$

Hence, it is sufficient to restrict attention to $\Delta_{\text{LA}}\left(l, t_i\right)\ \forall t_i$.

Suppose $\overline{\lambda}_F = \overline{\lambda}_A$. Then, by Lemma 4, $PM_{\text{LA}}\left(t_f\right) = PM_{\text{LA}}\left(t_a\right) = 0$, and $\forall t_i,\ EM_{\text{LA}}\left(l, t_i\right) > 0$. Hence, $\Delta_{\text{LA}}\left(l, t_i\right) = EM_{\text{LA}}\left(l, t_i\right) > 0 \implies \mathcal{L} \succ \mathcal{A}$. That is, both types vote sincerely and never abstain: $\lambda_f = \lambda_a = 0$.

By definition, $\overline{\lambda}_T = \sum_{i=f,a} \Pr\left(t_i|T\right)\lambda_i \implies \overline{\lambda}_F = \overline{\lambda}_A = 0$. $\overline{\lambda}_F = \overline{\lambda}_A$ holds, hence, $\lambda_f = \lambda_a = 0$ is an equilibrium.

Part 3: Suppose $\overline{\lambda}_F > \overline{\lambda}_A$. Then, by Lemma 4, $PM_{\text{LA}}\left(t_f\right) < 0,\quad PM_{\text{LA}}\left(t_a\right) > 0$, and

$\forall t_i, \quad EM_{\text{LA}}(l, t_i) > 0.$

For t_a, $PM_{\text{LA}}(t_a) > 0$, $EM_{\text{LA}}(l, t_a) > 0 \implies \Delta_{\text{LA}}(l, t_a) > 0 \implies \mathcal{L} \succ \mathcal{A}$ and $\lambda_a = 0$.

For t_f, $PM_{\text{LA}}(t_f) < 0$, $EM_{\text{LA}}(l, t_f) > 0$ – hence, determining the sign of $\Delta_{\text{LA}}(l, t_f)$ requires further investigation.

If $\Delta_{\text{LA}}(l, t_f) > 0$, then $\lambda_f = 0$, and $\overline{\lambda}_F = \overline{\lambda}_A = 0$ – contradicts $\overline{\lambda}_F > \overline{\lambda}_A \implies$ not an equilibrium.

If $\Delta_{\text{LA}}(l, t_f) < 0$, then $\lambda_f = 1$. Then, $\overline{\lambda}_F = p$, $\overline{\lambda}_A = 1 - p \implies \overline{\lambda}_F > \overline{\lambda}_A$ – equilibrium would hold. Hence, it remains to check, when $\Delta_{\text{LA}}(l, t_f) < 0$ holds. Alternatively, check when $|PM_{\text{LA}}(t_f)| > |EM_{\text{LA}}(l, t_f)| = EM_{\text{LA}}(l, t_f)$ (last equality follows from Lemma 4).

First, consider $|PM_{\text{LA}}(t_f)|$. As derived in Lemma 3:

$$|PM_{\text{LA}}(t_f)| = \left|u'\right| \beta \sum_{a=0}^{N-1} \widetilde{\Pr}(a|t_f)\left(E[t|a+1] - E[t|a]\right)$$

Using

$$\widetilde{\Pr}(a|t_f) = \sum_T \widetilde{\Pr}(a|T)\Pr(T|t_f) = C_{N-1}^a \left(\overline{\lambda}_F^a \left(1 - \overline{\lambda}_F\right)^{N-1-a} p + \overline{\lambda}_A^a \left(1 - \overline{\lambda}_A\right)^{N-1-a} (1-p)\right),$$

$$E[t|a+1] - E[t|a] = \frac{2\overline{\gamma}\,\overline{\lambda}_F^a \left(1 - \overline{\lambda}_F\right)^{N-a-1} \overline{\lambda}_A^a \left(1 - \overline{\lambda}_A\right)^{N-a-1} \left(\overline{\lambda}_F - \overline{\lambda}_A\right)}{\sum_T \overline{\lambda}_T^{a+1} \left(1 - \overline{\lambda}_T\right)^{N-a-1} \sum_T \overline{\lambda}_T^a \left(1 - \overline{\lambda}_T\right)^{N-a}},$$

Obtain that $|PM_{\text{LA}}(t_f)| =$
$\left|u'\right| \beta \sum_{a=0}^{N-1} C_{N-1}^a \left(\overline{\lambda}_F^a \left(1 - \overline{\lambda}_F\right)^{N-1-a} p + \overline{\lambda}_A^a \left(1 - \overline{\lambda}_A\right)^{N-1-a} (1-p)\right) \frac{2\overline{\gamma}\,\overline{\lambda}_F^a\left(1-\overline{\lambda}_F\right)^{N-a-1}\overline{\lambda}_A^a\left(1-\overline{\lambda}_A\right)^{N-a-1}\left(\overline{\lambda}_F-\overline{\lambda}_A\right)}{\sum_T \overline{\lambda}_T^{a+1}\left(1-\overline{\lambda}_T\right)^{N-a-1} \sum_T \overline{\lambda}_T^a\left(1-\overline{\lambda}_T\right)^{N-a}}.$

In equilibrium $\overline{\lambda}_F = p$, $\overline{\lambda}_A = 1 - p$. Hence, simplify further to:

$$\begin{aligned}
|PM_{\text{LA}}(t_f)| &= \left|u'\right| 2\overline{\gamma}\beta(2p-1) \times \\
&\times \sum_{a=0}^{N-1} C_{N-1}^a \frac{\left(p^{a+1}(1-p)^{N-1-a} + (1-p)^{a+1}p^{N-1-a}\right) p^a(1-p)^{N-1-a}(1-p)^a p^{N-1-a}}{\left(p^{a+1}(1-p)^{N-1-a} + (1-p)^{a+1}p^{N-1-a}\right)\left(p^a(1-p)^{N-a} + (1-p)^a p^{N-a}\right)} \\
&= \left|u'\right| 2\overline{\gamma}\beta(2p-1) \sum_{a=0}^{N-1} \frac{C_{N-1}^a p^{N-1}(1-p)^{N-1}}{\left(p^a(1-p)^{N-a} + (1-p)^a p^{N-a}\right)}
\end{aligned}$$

Claim 1. As $N \to \infty$, $|PM_{\text{LA}}(t_f)|$ is monotonously decreasing in p.

Proof:

$$\left|PM_{\text{LA}}\left(t_f\right)\right|_p^{'} \propto 2\sum_{a=0}^{N-1}\frac{C_{N-1}^a p^{N-1}(1-p)^{N-1}}{(p^a(1-p)^{N-a}+(1-p)^a p^{N-a})} + (2p-1)\times\sum_{a=0}^{N-1} C_{N-1}^a\frac{(p(1-p))^{N-1}}{(p^a(1-p)^{N-a}+(1-p)^a p^{N-a})}\times$$
$$\left(\frac{-(N-1)(2p-1)}{p(1-p)} - \frac{\left(ap^{a-1}(1-p)^{N-a}-(N-a)p^a(1-p)^{N-1-a}-a(1-p)^{a-1}p^{N-a}+(N-a)(1-p)^a p^{N-1-a}\right)}{(p^a(1-p)^{N-a}+(1-p)^a p^{N-a})}\right)$$
$$=\sum_{a=0}^{N-1}\frac{C_{N-1}^a p^{N-1}(1-p)^{N-1}}{(p^a(1-p)^{N-a}+(1-p)^a p^{N-a})}$$
$$\left(2-\frac{(N-1)(2p-1)^2}{p(1-p)} - \frac{(2p-1)\left(ap^{a-1}(1-p)^{N-a}-(N-a)p^a(1-p)^{N-1-a}-a(1-p)^{a-1}p^{N-a}+(N-a)(1-p)^a p^{N-1-a}\right)}{(p^a(1-p)^{N-a}+(1-p)^a p^{N-a})}\right)$$
$$=\sum_{a=0}^{N-1}\frac{C_{N-1}^a p^{N-1}(1-p)^{N-1}}{(p^a(1-p)^{N-a}+(1-p)^a p^{N-a})}\times$$
$$\times\left(2-\frac{(2p-1)}{p(1-p)}\left((N-1)(2p-1)+\frac{(a-Np)}{\left(1+\left(\frac{p}{1-p}\right)^{N-2a}\right)}+\frac{(N(1-p)-a)}{\left(1+\left(\frac{1-p}{p}\right)^{N-2a}\right)}\right)\right)$$

One may check that $\arg\min_a \frac{(a-Np)}{\left(1+\left(\frac{p}{1-p}\right)^{N-2a}\right)}+\frac{(N(1-p)-a)}{\left(1+\left(\frac{1-p}{p}\right)^{N-2a}\right)}=\frac{N}{2}$, hence

$$\begin{aligned}(N-1)(2p-1)+\frac{(a-Np)}{\left(1+\left(\frac{p}{1-p}\right)^{N-2a}\right)}+\frac{(N(1-p)-a)}{\left(1+\left(\frac{1-p}{p}\right)^{N-2a}\right)} &\geq (N-1)(2p-1)+\frac{N(1-2p)}{2}\\ &=\left(\frac{1}{2}N-1\right)(2p-1)\end{aligned}$$

This implies that as $N\to\infty$

$$\begin{aligned}\left|PM_{\text{LA}}\left(t_f\right)\right|_p^{'} &\leq \sum_{a=0}^{N-1}\frac{C_{N-1}^a p^{N-1}(1-p)^{N-1}}{(p^a(1-p)^{N-a}+(1-p)^a p^{N-a})}\left(2-\frac{(2p-1)}{p(1-p)}\left(\frac{1}{2}N-1\right)(2p-1)\right)\\ &=\left(2-\frac{(2p-1)^2}{2p(1-p)}(N-2)\right)\sum_{a=0}^{N-1}\frac{C_{N-1}^a p^{N-1}(1-p)^{N-1}}{(p^a(1-p)^{N-a}+(1-p)^a p^{N-a})}<0\end{aligned}$$

Where the last inequality follows from the sum being positive, and $\lim_{N\to\infty}\left(2-\frac{(2p-1)^2}{2p(1-p)}(N-2)\right)=-\infty<0$. Hence, for large enough N, $\left|PM_{\text{LA}}\left(t_f\right)\right|_p^{'}<0 \implies \left|PM_{\text{LA}}\left(t_f\right)\right|$ is decreasing in p. ■(*Claim* 1)

Now, consider $\left|EM_{\text{LA}}\left(l,t_f\right)\right|$. As derived in Lemma 4 proof:

$$\begin{aligned}EM_{\text{LA}}\left(\theta,t_f\right) &= u^{'}d\left[\Pr\left(\theta|d\right)-\Pr\left(\theta|-d\right)\right]\sum_{a=0,\text{ even}}^{N-1}\widetilde{\Pr}\left(\frac{N-a-1}{2},a|\theta,t_f\right)\\ &+u^{'}d\frac{q-\Pr\left(\theta|-d\right)}{q\Pr\left(\theta|d\right)+(1-q)\Pr\left(\theta|-d\right)}\sum_{a=1,\text{ odd}}^{N-2}\widetilde{\Pr}\left(\frac{N-a}{2}-1,a|\theta,t_f\right)\end{aligned}$$

Then, for $\theta = l$:

$$EM_{\text{LA}}(l,t_f) = u^{'}d\left[1-2p_\theta\right]\sum_{a=0,\text{ even}}^{N-1}\widetilde{\Pr}\left(\frac{N-a-1}{2},a|l,t_f\right)$$
$$+u^{'}d\frac{q-p_\theta}{q(1-p_\theta)+(1-q)p_\theta}\sum_{a=1,\text{ odd}}^{N-2}\widetilde{\Pr}\left(\frac{N-a}{2}-1,a|l,t_f\right)$$
$$=u^{'}d\left[1-2p_\theta\right]\sum_{a=0,\text{ even}}^{N-1}\widetilde{\Pr}\left(\frac{N-a-1}{2},a|l,t_f\right)$$

Where the last step follows from the fact that $q = p_\theta$ in any equilibrium (see part 1 proof).

Using

$$\widetilde{\Pr}\left(\frac{N-a-1}{2},a|l,t_f\right) = \sum_T\sum_s\widetilde{\Pr}\left(\frac{N-a-1}{2},a|s,T\right)\Pr(s|l)\Pr(T|t_f)$$
$$=\sum_T C_{N-1}^a\overline{\lambda}_T^a\left(1-\overline{\lambda}_T\right)^{N-1-a}\Pr(T|t_f)\,C_{N-1-a}^{\frac{N-a-1}{2}}\left(q(1-q)\right)^{\frac{N-a-1}{2}}$$
$$=C_{N-1}^a\left(p^{a+1}(1-p)^{N-1-a}+(1-p)^{a+1}p^{N-1-a}\right)C_{N-1-a}^{\frac{N-a-1}{2}}\left(p_\theta(1-p_\theta)\right)^{\frac{N-a-1}{2}},$$

obtain

$$EM_{\text{LA}}(l,t_f) = \left|u^{'}\right|d\left[2p_\theta-1\right]\times$$
$$\times\sum_{a=0,\text{ even}}^{N-1}C_{N-1}^a\left(p^{a+1}(1-p)^{N-1-a}+(1-p)^{a+1}p^{N-1-a}\right)C_{N-1-a}^{\frac{N-a-1}{2}}\left(p_\theta(1-p_\theta)\right)^{\frac{N-a-1}{2}}$$

Claim 2. As $N\to\infty$, $EM_{\text{LA}}(l,t_f)$ is monotonously increasing in p and monotonously decreasing in p_θ.

Proof:

$$EM_{\text{LA}}(l,t_f)^{'}_{p_\theta}\propto$$
$$2\sum_{a=0,\text{ even}}^{N-1}C_{N-1}^a\left(p^{a+1}(1-p)^{N-1-a}+(1-p)^{a+1}p^{N-1-a}\right)C_{N-1-a}^{\frac{N-a-1}{2}}\left(p_\theta(1-p_\theta)\right)^{\frac{N-a-1}{2}}+(2p_\theta-1)\times$$
$$\times\sum_{a=0,\text{ even}}^{N-1}C_{N-1}^a\left(p^{a+1}(1-p)^{N-1-a}+(1-p)^{a+1}p^{N-1-a}\right)C_{N-1-a}^{\frac{N-a-1}{2}}\frac{N-a-1}{2}\left(p_\theta(1-p_\theta)\right)^{\frac{N-a-1}{2}-1}(1-2p_\theta)$$
$$=\sum_{a=0,\text{ even}}^{N-1}C_{N-1}^a\left(p^{a+1}(1-p)^{N-1-a}+(1-p)^{a+1}p^{N-1-a}\right)C_{N-1-a}^{\frac{N-a-1}{2}}\left(p_\theta(1-p_\theta)\right)^{\frac{N-a-1}{2}}\times$$
$$\times\left(2-\frac{(2p_\theta-1)^2}{p_\theta(1-p_\theta)}\frac{N-1-a}{2}\right)$$

Observe that $2-\frac{(2p_\theta-1)^2}{p_\theta(1-p_\theta)}\frac{N-1-a}{2}<0$ iff $a<N-1-\frac{4p_\theta(1-p_\theta)}{(2p_\theta-1)^2}=N-\frac{1}{(2p_\theta-1)^2}$. As $N\to\infty$, $2-\frac{(2p_\theta-1)^2}{p_\theta(1-p_\theta)}\frac{N-1-a}{2}<0$ iff $\frac{a}{N}<1-\frac{1}{(2p_\theta-1)^2N}\to 1$, that is, almost everywhere ($\Pr\left(2-\frac{(2p_\theta-1)^2}{p_\theta(1-p_\theta)}\frac{N-1-a}{2}<0\right)=$

$\Pr\left(\frac{a}{N} < 1 - \frac{1}{(2p_\theta-1)^2 N}\right) \to \Pr(a < N) = 1)$. Hence, $EM_{\text{LA}}\left(l, t_f\right)'_{p_\theta} < 0 \implies EM_{\text{LA}}\left(l, t_f\right)$ decreases in p_θ. $\blacksquare$(*Claim* 2)

Since the two motives move in the opposing directions in p, check which of the two dominate at the extreme values of p.

Claim 3. Assume large elections, i.e. $N \to \infty$. Then, as $p \to 1$, $EM_{\text{LA}}\left(l, t_f\right) > \left|PM_{\text{LA}}\left(t_f\right)\right|$, and, as $p \to \frac{1}{2}$ and $p_\theta \to 1$, $EM_{\text{LA}}\left(l, t_f\right) < \left|PM_{\text{LA}}\left(t_f\right)\right|$.

Proof:

$$\begin{aligned}
\left|PM_{\text{LA}}\left(t_f\right)\right| &= \left|u'\right| 2\overline{\gamma}\beta(2p-1) \sum_{a=0}^{N-1} \frac{C_{N-1}^{a} p^{N-1}(1-p)^{N-1}}{\left(p^a(1-p)^{N-a} + (1-p)^a p^{N-a}\right)} \\
&= \left|u'\right| 2\overline{\gamma}\beta(2p-1) \sum_{a=0}^{N-1} \frac{C_{N-1}^{a}}{\left(p^{a-(N-1)}(1-p)^{1-a} + (1-p)^{a-(N-1)}p^{1-a}\right)}
\end{aligned}$$

$$\begin{aligned}
EM_{\text{LA}}\left(l, t_f\right) = &\left|u'\right| d\left(2p_\theta - 1\right) \times \\
&\times \sum_{a=0,\text{ even}}^{N-1} C_{N-1}^{a}\left(p^{a+1}(1-p)^{N-1-a} + (1-p)^{a+1}p^{N-1-a}\right) C_{N-1-a}^{\frac{N-a-1}{2}}\left(p_\theta\left(1-p_\theta\right)\right)^{\frac{N-a-1}{2}}
\end{aligned}$$

<u>Part 1.</u> Take $p = 1 - e$, with $e \to 0$. Then,

$$\begin{aligned}
\lim_{e\to 0}\left|PM_{\text{LA}}\left(t_f\right)\right| &= \lim_{e\to 0}\left|u'\right| 2\overline{\gamma}\beta(1-2e) \sum_{a=0}^{N-1} \frac{C_{N-1}^{a}}{\left((1-e)^{a-(N-1)}e^{1-a} + e^{a-(N-1)}(1-e)^{1-a}\right)} \\
&= \lim_{e\to 0}\left|u'\right| 2\overline{\gamma}\beta \sum_{a=0}^{N-1} \frac{C_{N-1}^{a}}{\left(\frac{1}{e^{a-1}} + \frac{1}{e^{(N-1)-a}}\right)} = 0
\end{aligned}$$

Where last equation follows from the fact that $\frac{1}{e^{a-1}} + \frac{1}{e^{(N-1)-a}} \to \infty$ for any value of a, since either $a - 1 > 0$ or $(N-1) - a > 0$.

$$\begin{aligned}
\lim_{e\to 0} EM_{\text{LA}}\left(l, t_f\right) = &\lim_{e\to 0}\left|u'\right| d\left(2p_\theta - 1\right) \times \\
&\times \sum_{a=0,\text{ even}}^{N-1} C_{N-1}^{a}\left((1-e)^{a+1}e^{N-1-a} + e^{a+1}(1-e)^{N-1-a}\right) C_{N-1-a}^{\frac{N-a-1}{2}}\left(p_\theta\left(1-p_\theta\right)\right)^{\frac{N-a-1}{2}} \\
&= \left|u'\right| d\left(2p_\theta - 1\right) \times \lim_{e\to 0} C_{N-1}^{N-1} e^{N-1-(N-1)} C_{N-1-(N-1)}^{\frac{N-1-(N-1)}{2}}\left(p_\theta\left(1-p_\theta\right)\right)^{\frac{N-1-(N-1)}{2}} \\
&= \left|u'\right| d\left(2p_\theta - 1\right) \times \lim_{e\to 0} e^0 = \left|u'\right| d\left(2p_\theta - 1\right) > 0
\end{aligned}$$

Hence, $\lim_{e\to 0} EM_{\text{LA}}\left(l, t_f\right) > \lim_{e\to 0}\left|PM_{\text{LA}}\left(t_f\right)\right|$.

<u>Part 2.</u> Take $p_\theta = 1 - e$, with $e \to 0$. This does not affect $\left|PM_{\text{LA}}\left(t_f\right)\right|$, and by Claim 2

guarantees the lowest value of $EM_{\text{LA}}(l,t_f)$ in p_θ.

$$\begin{aligned}\text{EM}_{\text{LA}}(l,t_f) &= \left|u^{'}\right| d(1-2e)\times \\ &\times \sum_{a=0,\text{ even}}^{N-1} C_{N-1}^{a}\left(p^{a+1}(1-p)^{N-1-a}+(1-p)^{a+1}p^{N-1-a}\right) C_{N-1-a}^{\frac{N-a-1}{2}}\left((1-e)e\right)^{\frac{N-a-1}{2}} \\ &\overset{e\to 0}{\to} \left|u^{'}\right| d\left(p^N+(1-p)^N\right)\end{aligned}$$

Since $\lim_{e\to 0}((1-e)e)^{\frac{N-a-1}{2}}=0$ unless $\frac{N-a-1}{2}=0 \iff a=N-1$.

Consider a lower bound of $|PM_{\text{LA}}(t_f)|$ as

$$\begin{aligned}|PM_{\text{LA}}(t_f)| &> \left|u^{'}\right| 2\overline{\gamma}\beta(2p-1)\sum_{a=0}^{N-1}\frac{C_{N-1}^{a}p^{N-1}(1-p)^{N-1}}{(1-p)^N+p^N} \\ &= \left|u^{'}\right| 2\overline{\gamma}\beta(2p-1)\frac{p^{N-1}(1-p)^{N-1}}{(1-p)^N+p^N}\sum_{a=0}^{N-1}C_{N-1}^{a} = \left|u^{'}\right| 2\overline{\gamma}\beta(2p-1)\frac{2^{N-1}p^{N-1}(1-p)^{N-1}}{(1-p)^N+p^N}\end{aligned}$$

Where the bound is obtained by replacing the denominator with its highest value: $\arg\max_{a\in\{0,1,\ldots,N-1\}}\left(p^a(1-p)^{N-a}+(1-p)^a p^{N-a}\right)=0$ with maximum value equal to $(1-p)^N+p^N$.

Finally compute the ratio of motivations:

$$\frac{|PM_{\text{LA}}(t_f)|}{\text{EM}_{\text{LA}}(l,t_f)} > \frac{2\overline{\gamma}\beta}{d}(2p-1)\frac{2^{N-1}p^{N-1}(1-p)^{N-1}}{\left(p^N+(1-p)^N\right)^2}$$

Take $p=\frac{1}{2}+\frac{1}{N}\overset{N\to\infty}{\to}\frac{1}{2}$.

$$\frac{|PM_{\text{LA}}(t_f)|}{\text{EM}_{\text{LA}}(l,t_f)} > \frac{2\overline{\gamma}\beta}{d}\left(\frac{2}{N}\right)\frac{2^{N-1}\left(\frac{1}{2}+\frac{1}{N}\right)^{N-1}\left(\frac{1}{2}-\frac{1}{N}\right)^{N-1}}{\left(\left(\frac{1}{2}+\frac{1}{N}\right)^N+\left(\frac{1}{2}-\frac{1}{N}\right)^N\right)^2} = \frac{8\overline{\gamma}\beta}{d}\frac{2^N\left(1+\frac{2}{N}\right)^{N-1}\left(1-\frac{2}{N}\right)^{N-1}}{N\left(\left(1+\frac{2}{N}\right)^N+\left(1-\frac{2}{N}\right)^N\right)^2}\overset{N\to\infty}{\to}\infty$$

Where the limit follows from $\lim_{N\to\infty}\frac{\left(1+\frac{2}{N}\right)^{N-1}\left(1-\frac{2}{N}\right)^{N-1}}{\left(\left(1+\frac{2}{N}\right)^N+\left(1-\frac{2}{N}\right)^N\right)^2}<\infty$ and $\lim_{N\to\infty}\frac{2^N}{N}=\infty$. Hence, $|PM_{\text{LA}}(t_f)|>\text{EM}_{\text{LA}}(l,t_f)$. ■(*Claim* 3)

Combine results of Claims 1-3. In large elections, as $p\to 1$, which ensures the largest $EM_{\text{LA}}(l,t_f)$ and the lowest $|PM_{\text{LA}}(t_f)|$, $EM_{\text{LA}}(l,t_f)>|PM_{\text{LA}}(t_f)| \implies \Delta_{\text{LA}}(l,t_f)>0$. Vice versa, as $p\to\frac{1}{2}$ and $p_\theta\to 1$, which ensures the lowest $EM_{\text{LA}}(l,t_f)$ and the largest $|PM_{\text{LA}}(t_f)|$, $EM_{\text{LA}}(l,t_f)<|PM_{\text{LA}}(t_f)| \implies \Delta_{\text{LA}}(l,t_f)<0$. Hence, given continuity and monotonicity, there must exist a single threshold $p^*(p_\theta)$, such that $\Delta_{\text{LA}}(l,t_f)=0$, and for $p<p^*(p_\theta)$, $\Delta_{\text{LA}}(l,t_f)<0$ holds $\implies \lambda_f=1 \implies \overline{\lambda}_F>\overline{\lambda}_A$ – in large elections separating equilibrium holds as long as $p<p^*(p_\theta)$.

Finally, by Claim 2, $EM_{\text{LA}}(l,t_f)$ decreases with $p_\theta \implies$ for any given p, p_θ, s.t. $\Delta_{\text{LA}}(l,t_f)^{p,\,p_\theta}=0$, increasing p_θ to $p_\theta^{'}>p_\theta$ implies $\Delta_{\text{LA}}(l,t_f)^{p,\,p_\theta^{'}}<0$. Then, there must exist $p^{'}>p$, s.t. $\Delta_{\text{LA}}(l,t_f)^{p^{'},p_\theta^{'}}=0$. By definition, $p=p^*(p_\theta)$ and $p^{'}=p^*\left(p_\theta^{'}\right)$. $p_\theta^{'}>p_\theta$

$\implies p^*\left(p_\theta^{'}\right) = p^{'} > p = p^*\left(p_\theta\right)$. Hence, $p^*\left(p_\theta\right)$ is increasing with p_θ. ■

Corollary 3

Proof.

$$E\left[u|\theta, t_i\right] = \sum_{a=0}^{N-1}\sum_{m=0}^{N-a-1} \widetilde{\Pr}\left(m, a|\theta, t_i\right) E\left[u\left(|t_i - x(a)| + |y(m,a) - s|\right)|m, a, \theta\right]$$

Observe that $\widetilde{\Pr}\left(m, a|\theta, t_i\right)$ can be decomposed to

$$\widetilde{\Pr}\left(m, a|\theta, t_i\right) = \sum_{T\in\{F,A\}} \Pr\left(T|t_i\right)\widetilde{\Pr}\left(a|T\right)\widetilde{\Pr}\left(m|a,\theta\right).$$

Consider $\widetilde{\Pr}\left(a|T\right),\ \ \widetilde{\Pr}\left(m|a,\theta\right)$ as $N\to\infty$. By law of large numbers, given sincere voting:

$$\widetilde{\Pr}\left(m|a,\theta\right) \to \begin{cases} 1,\ if\ m = \Pr\left(l|s\right)\left(N-1-a\right) \equiv \overline{m}(s) \\ 0,\ otw \end{cases}$$

With $p_\theta > \frac{1}{2}$, as $N\to\infty$, $y\left(\overline{m}(s), a\right) = y\left(\overline{m}(s)+1, a\right) = \begin{cases} -d,\ if\ s=-d \\ d,\ if\ s=d \end{cases} \implies$ $|y-s| = 0,\ \forall s$. That is, the salient policy gap is always 0 irrespective of state, signal, or one's action chosen.

Again, by law of large numbers:

$$\widetilde{\Pr}\left(a|F\right) \to \begin{cases} 1,\ if\ a = p(N-1) \\ 0,\ otw \end{cases}$$

$$\widetilde{\Pr}\left(a|A\right) \to \begin{cases} 1,\ if\ a = (1-p)(N-1) \\ 0,\ otw \end{cases}$$

Combining all the findings, obtain that

$$E\left[u|\theta, t_i\right] \to \Pr\left(F|t_i\right) u\left(\left|t_i - x\left(a_i^F\right)\right|\right) + \Pr\left(A|t_i\right) u\left(\left|t_i - x\left(a_i^A\right)\right|\right),$$

where a_i^T denotes the abstention given population T and i's action. That is, in the equilibrium where 'for' type abstains and 'against' one votes:

$$\begin{aligned} E\left[u|\theta, t_i\right] &\to \begin{cases} \mathrm{pu}\left(\left|\gamma - x\left(p(N-1)+1\right)\right|\right) + (1-p)u\left(\left|\gamma - x\left((1-p)(N-1)+1\right)\right|\right),\ if\ i=f \\ (1-p)u\left(\left|-\gamma - x\left(p(N-1)\right)\right|\right) + pu\left(\left|-\gamma - x\left((1-p)(N-1)\right)\right|\right),\ if\ i=a \end{cases} \\ &= \begin{cases} \mathrm{pu}\left(\gamma - x\left(p(N-1)+1\right)\right) + (1-p)u\left(\gamma - x\left((1-p)(N-1)+1\right)\right),\ if\ i=f \\ (1-p)u\left(\gamma + x\left(p(N-1)\right)\right) + pu\left(\gamma + x\left((1-p)(N-1)\right)\right),\ if\ i=a \end{cases} \end{aligned}$$

Observe that in such separating equilibrium with $\overline{\lambda}_F = p$, $\overline{\lambda}_A = 1 - p$:

$$x(a) = \beta E\left[t|a\right] = \beta\overline{\gamma}\frac{p^a(1-p)^{N-a} - (1-p)^a p^{N-a}}{p^a(1-p)^{N-a} + (1-p)^a p^{N-a}} = \beta\overline{\gamma}\frac{1 - \left(\frac{p}{1-p}\right)^{N-2a}}{1 + \left(\frac{p}{1-p}\right)^{N-2a}} = \beta\overline{\gamma} - \frac{2\beta\overline{\gamma}}{\left(\frac{1-p}{p}\right)^{N-2a} + 1}$$

As $N \to \infty$, with $p > \frac{1}{2}$:

$$\lim_{N\to\infty} x\left(p(N-1)\right) = \lim_{N\to\infty} x\left(p(N-1)+1\right) = \beta\overline{\gamma}$$

$$\lim_{N\to\infty} x\left((1-p)(N-1)\right) = \lim_{N\to\infty} x\left((1-p)(N-1)+1\right) = -\beta\overline{\gamma}$$

Thus, simplify $E\left[u|\theta, t_i\right]$ to:

$$E\left[u|t_f\right] = E\left[u|t_a\right] \to pu\left(\gamma - \beta\overline{\gamma}\right) + (1-p)u\left(\gamma + \beta\overline{\gamma}\right).$$

Any excluded policy is such that $x \in [-\beta\overline{\gamma},\ \beta\overline{\gamma}]$. Similar to the above derivations, obtain that for any given x, the ex ante expected utility for each $U(x)$ is equal to $U(x) = pu(\gamma - x) + (1-p)u(\gamma + x)$, with $x \in [0, \beta\overline{\gamma}]$ covering all possible policies.

Finally, study which one maximizes $U(x)$:

$$U^{'}(x) = -pu^{'}(\gamma - x) + (1-p)u^{'}(\gamma + x)$$

$$U^{''}(x) = pu^{''}(\gamma - x) + (1-p)u^{''}(\gamma + x)$$

With $u^{''}(\bullet) < 0$ by construction, $U^{''}(x) < 0 \implies U^{'}(x^*) = 0$ solves for a maximum $x^* : \frac{u^{'}(\gamma - x^*)}{u^{'}(\gamma + x^*)} = \frac{1-p}{p}$. Note that x^* is well defined since $u^{'}(\bullet) < 0$ by construction and $\left|u^{'}(\gamma - x^*)\right| < \left|u^{'}(\gamma + x^*)\right|$.

As $x = 0$, $U^{'}(0) = -pu^{'}(\gamma) + (1-p)u^{'}(\gamma) = (1-2p)u^{'}(\gamma) > 0$. Hence, marginally increasing x from 0 always increases $U(x)$. Thus, $x = 0$ never maximizes the ex ante utility, which proves part 1.

If $x^* \geq \beta\overline{\gamma}$, by monotonicity $U(x)$ increases in $x\ \forall\ x \in [0, \beta\overline{\gamma}]$. Thus, $x = \beta\overline{\gamma}$ maximizes utility. If $x^* < \beta\overline{\gamma}$, then $x = \beta\overline{\gamma}$ is suboptimal. ■