Delegation of Learning from Multiple Sources of Information*

Angelos Diamantopoulos[†] Arina Nikandrova[‡]

March 20, 2024

Abstract

A principal delegates a decision to a biased expert. Before taking the decision, the expert may undertake incremental learning about the unknown binary state from two alternative information sources. There are no transfers but the principal retains the right to terminate the expert's learning to take the decision herself. The right to terminate learning benefits the principal when the preferences of the principal and the expert are sufficiently misaligned but may be detrimental when the preferences are sufficiently closely aligned.

Keywords: delegation of learning, Poisson process, learning in continuous time

JEL Codes: C73, D82, D83

1 Introduction

Good decision making requires good information, but it can often be the case that learning through information acquisition is inaccessible to the decision maker. Instead, learning may require engaging an expert who has access to some learning technology. The expert may be better equipped for undertaking learning due to specialist knowledge or know-how that he has accumulated over many years or because the expert has the

*We are grateful to Konrad Mierendorff for the encouragement at the early stages of the project, to Tatiana Mayskaya for invaluable advice and to the Editor, Martin Cripps, for clear editorial guidance. We also thank anonymous referees, Alkis Georgiadis-Harris, Francesco Nava, Romans Pancs, Marco Pelliccia, Emanuela Sciubba and Flavio Toxvaerd for comments and inspiring conversations at various stages of this project.

[†]Department of Economics, SOAS, University of London, UK, e-mail: ad67@soas.ac.uk.

[‡]Department of Economics, City, University of London, UK, e-mail: a.nikandrova@gmail.com.

exclusive access to key information sources. If the expert and the decision maker's incentives are perfectly aligned, then the latter can hire the former to learn about the state and also to make a decision if and when the expert finds fit. However, if the interests of the decision maker and the expert are not perfectly aligned, it is tempting to assume that the decision maker would always benefit from retaining at least some decision-making authority, particularly if the expert cannot misreport learning outcomes. In this paper, we show that this intuition is not entirely correct. When the decision maker cannot provide monetary rewards for learning, limiting the expert's decision-making authority may influence how the expert learns to the detriment of the decision maker.

This paper studies a dynamic model of learning by an expert (he) on behalf of a principal (she). The model builds on the sequential decision problem of Che and Mierendorff (2019) with two information sources generating signals about two mutually exclusive states of the world. We embed the special case of this decision problem into a game between the principal and the expert.

The players are interested in learning the prevailing binary state of the world as it determines the optimal binary decision. Conditional on the state, both the principal and the expert agree on the optimal decision, but in either state, the expert obtains a higher payoff than the principal from one of the decisions; that is, the expert is biased towards a particular decision.

The principal delegates learning and decision making to the expert. The principal cannot undertake learning herself as only the expert has access to two alternative information sources. Each information source generates a conclusive signal about one of the states of the world according to a Poisson process. The expert can utilise either information source for free[1] and must decide how to divide attention between the two sources. The principal observes the learning outcomes and also observes but cannot directly affect how the expert splits his attention between the information sources.

The principal and the expert discount future at a common rate. Hence, while learning is costless, it entails a costly delay to the decision. We assume that the discount rate is not too high, so that the expert, when sufficiently uncertain about the state, finds some learning optimal. At the same time, the discount rate is not too low, so that, left to his own devices, when uncertain, the expert would optimally allocate all attention to the information source that can conclusively reveal the currently less likely state — that is, while learning, the expert seeks a refutation of his current belief. In other words,

[1] It is convenient to assume that the expert can access information sources for free because it circumvents the problem of allocating the learning costs between the principal and the expert. In Section 5, we briefly discuss the implications of adding a flow cost of learning.

we restrict attention to the parameter values for which, in Che and Mierendorff (2019) terminology, "own-biased learning" heuristics is optimal.

The principal can choose between two simple contracting arrangements: complete delegation and flexible delegation. Under **complete delegation**, she commits to surrender completely the authority to make decisions; under **flexible delegation**, she retains the right to interrupt the expert's learning and make the decision herself whenever she pleases.[2]

We consider only two simple contracting arrangements because they naturally arise in real world settings. For example, in representative democracy, the government consists of experts who have time and resources to undertake due diligence on complex decisions on behalf of the electorate. While elected politicians are meant to represent the electorate, they may have their own vested interest in some decisions. Consequently, the interests of the government and electorate may not always be perfectly aligned. The electorate has very little (if any) power to fine-tune the set of actions available to the government or to dictate how to undertake due diligence on a particular matter. In fact, for most matters, the electorate is committed not to interfere in the government's learning and decision making until the next elections. Nevertheless, typically there are also statuary provisions for referenda — in many jurisdictions, a referendum on certain particularly important matters can be called if there is sufficient support for it among the electorate. A referendum takes the decision making out of the government's hands entirely and allows the electorate to act instead. Our model asks whether such provisions to deprive the expert from the decision-making authority are always beneficial for the principal.

Our main result shows that when preferences of the principal and the expert are sufficiently closely aligned, for a range of prior beliefs, the principal prefers complete delegation, thereby completely surrendering the authority to make the final decision, over flexible delegation, under which she retains the right to intervene. This, seemingly counter-intuitive, result is a consequence of the expert's strategic response when he expects the principal to intervene and terminate learning prematurely.

Under complete delegation the expert learns and acts according to his preferences. However, since preferences of the principal and the expert are misaligned, at some beliefs they disagree on the optimal learning policy. *Inter alia*, the misalignment of preferences means that sometimes the expert finds it optimal to continue learning, while the principal prefers to take the decision immediately; and sometimes he may prefer learning from an

[2] The paper compares the payoff that the principal derives from these two mechanisms for any prior. Therefore, we are not explicit whether the principal knows her prior at the time of choosing the contracting arrangement. We briefly discuss this further in Section 5.

information source that is sub-optimal from the principal's perspective.

The main benefit of flexible delegation is that it prevents the expert from prolonged learning, thus avoiding costly delay when the principal is already sufficiently confident of the right decision. However, flexible delegation cannot ensure that the expert undertakes learning from the information source preferred by the principal. In fact, the threat of early termination of learning under flexible delegation may make the expert's choice of information sources even more sub-optimal for the principal than it is under complete delegation. This adverse effect on the choice of the information source constitutes the cost of flexible delegation.

When the preferences of the principal and the expert are sufficiently closely aligned, the principal has to weigh the benefit of flexible delegation against its cost and at times may well find it optimal to completely surrender the decision-making authority to the expert. In contrast, when the preferences are sufficiently misaligned, the threat of premature learning termination does not change the expert's choice of information sources and so, flexible delegation has only upside and no downside for the principal.

Our paper is related to various strands of the literature on delegation and dynamic learning in environments without transfers.

The literature on optimal dynamic learning, pioneered by Wald (1947) has been rapidly growing in recent years (see, for example, Zhong (2022), Ke et al. (2016) and Ke and Villas-Boas (2019)). While our paper builds on Che and Mierendorff (2019), other closely related papers are Chaimanowong et al. (2023), Mayskaya (2024) and Nikandrova and Pancs (2018). In contrast to our paper, other papers in this genre study decision problems.

A central theme within the literature on delegation of learning is that of observability. Much of the literature focuses on unobservable learning by a biased agent who discloses his findings to the principal. In this context, the agent may be subject to a moral hazard (as in Guo (2016) and Garfagnini (2011)) or may not have incentives to disclose his findings truthfully (as in Argenziano et al. (2016), Deimen and Szalay (2019), Escobar and Zhang (2021), Herresthal (2022)). In our model, the expert's learning effort and outcomes are observable, but the principal has very limited tools for inducing the biased expert to pursue the learning policy the principal desires.

The rest of the paper is organized as follows. Section 2 describes the setup. Section 3 characterizes the expert's optimal learning policy under complete delegation and discusses the implications of this policy for the principal. Section 4 discusses the main result of the paper. Section 5 concludes by discussing some of the model assumptions.

2 Model

There are two players, a principal (she) and an expert (he). Time is continuous and indexed by $t \geq 0$. The time horizon is infinite. Both players discount future exponentially at rate $r > 0$.

Problem. The players are interested in taking the decision that is optimal for the unknown state of the world. There are two states of the world, $\omega \in \{A, B\}$ and two decisions $x \in \{a, b\}$.

Delegation. The expert, but not the principal, has access to an information technology that enables learning about ω. The expert can access this technology at no cost. At any time $t > 0$, the principal delegates learning as well as decision making to the expert thereby giving the expert authority to learn about ω and to take final decision $x \in \{a, b\}$. The principal observes but cannot directly affect, either through contingent payments or coercion, learning and the final decision taken by the expert.

We distinguish two types of delegation: *complete delegation* and *flexible delegation*. Under complete delegation, the principal surrenders completely the authority to make decisions to the expert. Under flexible delegation, at each time, t, while the expert has not taken action x yet, the principal retains the right to override the expert and take decision $y \in \{a, b\}$ herself.

Learning. At each time t, when he is given the authority to learn, the expert may allocate a unit of learning intensity between two information sources, A and B. Information source A reveals conclusively state A; that is, when the expert learns from A with intensity $\alpha \in [0, 1]$, source A generates a signal with a Poisson arrival rate α if $\omega = A$ and no signal if $\omega = B$. Symmetrically, information source B reveals conclusively state B; that is, when the expert learns from B with intensity $1 - \alpha$, source B generates a Poisson signal with arrival rate $(1 - \alpha)$ if $\omega = B$ and no signal otherwise. The learning process of the expert is denoted by $(\alpha_t)_{t \in \mathbb{R}_+}$. We refer to learning exclusively from source A — that is, choosing $\alpha = 1$ — as A-learning and to learning exclusively from source B — that is, choosing $\alpha = 0$ — as B-learning.

At $t = 0$, the principal and the expert share a common prior belief p_0 that the state is A. Subsequently, the principal observes $(\alpha_t)_{t \in \mathbb{R}_+}$ as well as learning outcomes. Hence, the belief that the state is A remains common throughout the game. Let p_t denote the belief at time t. By Bayes' rule, as long as the expert's learning generates no signal, p_t

evolves according to

$$\dot{p}_t = -(2\alpha_t - 1)\, p_t\,(1 - p_t)\,. \tag{1}$$

Payoffs. Let u_x^ω and v_x^ω denote the utility of the principal and the expert respectively, conditional on decision x and state ω. The states are labeled so that both the principal and the expert strictly prefer taking the decision that matches the state. Moreover, independently of the state, the expert prefers decision a relative to the principal and the principal prefers decision b relative to the expert. In particular, for $\Delta \geq 0$, the payoff profiles $\left(u_x^\omega, v_x^\omega\right)$ are

P, E	A	B
a	$1,\, 1+\Delta$	$-1,\, -1+\Delta$
b	$-1+\Delta,\, -1$	$1+\Delta,\, 1$

Parameter Δ plays a dual role in our model. First, it determines the importance of choosing the action that matches the state. As Δ increases, choosing the favoured action becomes more appealing and so the value of learning for the principal and the expert diminishes. When Δ exceeds 2, learning has no value at all because the favoured action delivers higher payoff irrespective of the state. Second, Δ is a measure of the disagreement between the principal and the expert. Throughout the paper, we assume that Δ is not too high because all interesting interactions between the principal and the expert take place when the disagreement between them is relatively small.

Assumption 1. $\Delta \leq 1$

Given belief p, let

$$U_x(p) \equiv p u_x^A + (1-p)\, u_x^B \text{ and } V_x(p) \equiv p v_x^A + (1-p)\, v_x^B \tag{2}$$

denote the expected payoff of the principal and the expert, respectively, from taking decision x immediately.

Aim. The aim is to compare the principal's payoffs under complete and flexible delegation.

Throughout the paper, quantities without any superscript correspond to flexible delegation, superscript E indicates quantities under complete delegation and superscript P indicates principal's individually optimal learning when the principal has access to the same learning technology as the expert.

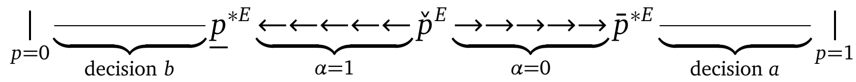


Figure 1: Expert's optimal learning strategy

3 Complete Delegation

Expert's optimal learning policy

Under complete delegation, the expert will pursue his individually optimal learning strategy. Indeed, since the principal cannot intervene, the expert has no reason to pursue a learning strategy different from the one he would pursue if he was facing an individual decision problem.

Che and Mierendorff (2019) fully characterise the solution to the expert's decision problem and show that the uniquely optimal learning policy depends on the cost of learning. In that paper, learning is associated with two types of costs. First, while learning, the decision maker incurs an additive flow cost. Second, learning involves a delay in making the decision and hence in receiving the payoff from this decision which is costly because the decision maker discounts future payoffs at an exponential rate. Our setup is a special case of Che and Mierendorff (2019) because there are no additive costs of learning; instead, the discount rate r constitutes the only learning cost.[3]

Furthermore, we restrict attention to the discount rate which is neither too high nor too low. Lemma 1 proves that for such intermediate values of discount rate, the expert optimally pursues *contradictory learning*, a learning policy that dictates devoting full attention to the source which generates signal against the state that the expert currently believes to be relatively more likely.[4] That is, the expert chooses $\alpha = 0$ thus learning from source B if state A is relatively more likely, i.e., $p > \check{p}^E$, and chooses $\alpha = 1$ thus learning from source A if state B is relatively likely, i.e., $p < \check{p}^E$, where $\check{p}^E$ is an optimally chosen reference belief.[5] In the absence of a signal, the expert's belief drifts in the direction which strengthens the original belief. If no signal arrives, the belief eventually reaches one of the optimally chosen boundary points $\underline{p}^{*E}$ or $\bar{p}^{*E}$, at which the expert is sufficiently convinced of the state to make an immediate decision without receiving a conclusive signal. Figure 1 depicts belief updating under contradictory learning that is

[3] On page 2998, Che and Mierendorff (2019) note that in their setting, either additive learning cost or discount rate can be zero, but not both.

[4] Lemma 1 corresponds to a special case of case (ii) in Theorem 1 in Che and Mierendorff (2019) with flow costs of learning switched off.

[5] Since $\Delta > 0$, $\check{p}^E$ is less than $1/2$.

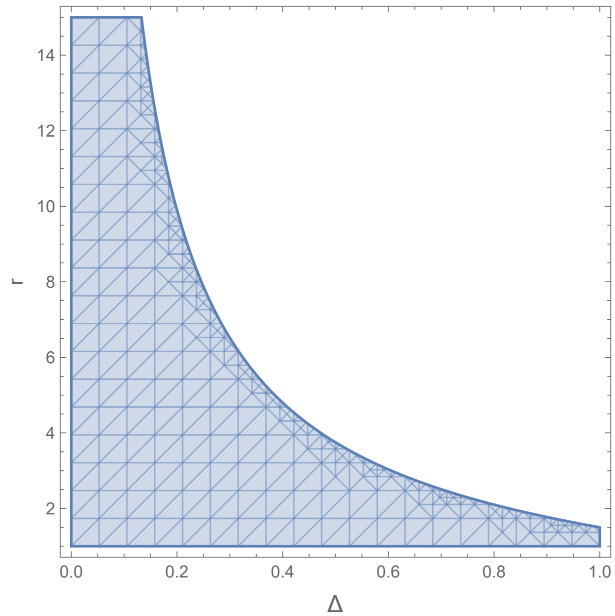


Figure 2: The parameter region in which Lemma 1 holds.

optimal for the expert.

Lemma 1. *For any $1 \geq \Delta \geq 0$, there exists a threshold $\bar{r}(\Delta) > 0$ such that for all $r \in [1, \bar{r}(\Delta)]$, the expert's optimal policy under complete delegation is contradictory learning with belief thresholds*

$$\underline{p}^{*E} = \frac{r}{2(1+r)+\Delta}, \tag{3}$$

$$\bar{p}^{*E} = 1 - \frac{r(1+\Delta)}{2(1+r)-\Delta}, \tag{4}$$

and $\check{p}^E$ implicitly defined as the unique solution of $g^E(p) = 0$, where

$$g^E(p) := p(1+\Delta)\left(1 - \left(\frac{p}{1-p}\right)^r \left(\frac{1-\bar{p}^{*E}}{\bar{p}^{*E}}\right)^r\right) + (1-p)\left(\left(\frac{1-p}{p}\right)^r \left(\frac{\underline{p}^{*E}}{1-\underline{p}^{*E}}\right)^r - 1\right). \tag{5}$$

Proof. See Appendix A.1. □

Figure 2 depicts the parameter region of interest to us in which the expert optimally pursues contradictory learning.[6] The figure shows that there is one-to-one relationship

[6]The statement of Lemma 1 can be weakened. Indeed, for any $1 \geq \Delta \geq 0$, there exist thresholds $1 \geq \underline{r}(\Delta) > 0$ and $\bar{r}(\Delta) > 0$ such that for all $r \in [\underline{r}(\Delta), \bar{r}(\Delta)]$, the expert's optimal policy under complete delegation is contradictory learning with belief thresholds (3) and (4). The lower bound $\underline{r}(\Delta)$ is equal to

between Δ and threshold $\bar{r}(\Delta)$. Hence, instead of fixing Δ and allowing r to vary, Lemma 1 can be equivalently restated fixing r and letting Δ to vary.

Corollary 1. *For any $r \geq 1$, there exists a threshold $1 \geq \bar{\Delta}(r) \geq 0$ such that for all $\Delta \in \left[0, \bar{\Delta}(r)\right]$, the expert's optimal policy under complete delegation is contradictory learning.*

Impact of complete delegation on the principal

Due to the symmetry in expert and principal's payoff structures, had the principal access to the expert's learning technology, at any $r \in [1, \bar{r}(\Delta)]$, just like the expert, she would also optimally choose to pursue contradictory learning policy — that is, the principal would choose to devote full attention to the source which generates signal against the state that *she* currently believes to be relatively more likely. [7] At a particular belief p, due to bias Δ, the principal and the expert may disagree on the source of information from which to learn or whether to learn at all. In particular, because irrespective of the state, the expert prefers decision a relative to the principal, the expert's learning region is shifted to the left relative to the principal's individually optimal learning region, as Lemma 2 proves and Figure 3a illustrates. Thus, for example, the autonomous expert optimally takes immediate decision a or b at lower p than the principal would have liked. Similarly, for the expert, the switch of learning from one source to another is shifted to the left relative to the principal.

Lemma 2. *For any $1 \geq \Delta > 0$ and all $r \in [1, \bar{r}(\Delta)]$, $\underline{p}^{*E} < \underline{p}^{*P}$ and $\bar{p}^{*E} < \bar{p}^{*P}$ and $\check{p}^{E} < \check{p}^{P}$.*

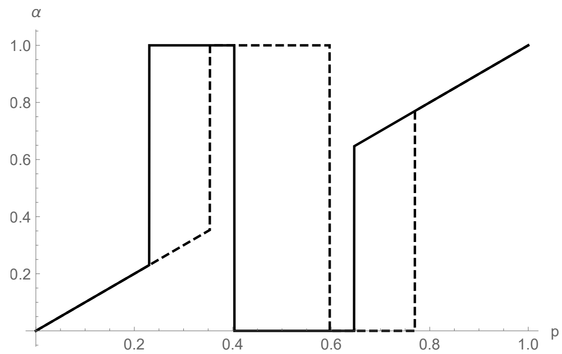
Proof. See Appendix A.2. □

Figure 3b compares the principal's value under complete delegation to principal's value in the first-best world where the principal has access to the same learning technology as the expert. As a result of disagreement on the optimal learning strategy at a given p, the principal's expected payoff when the expert undertakes learning on her behalf falls short of his first-best expected payoff.[8]

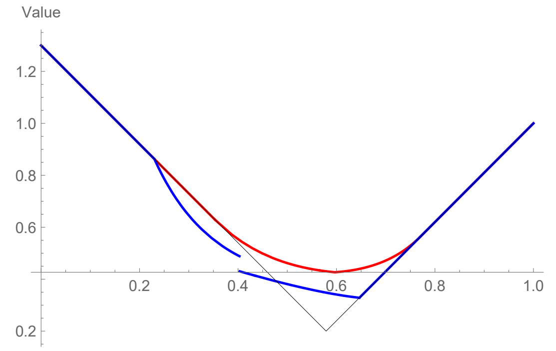
1 at $\Delta = 0$ and decreasing in Δ. Whenever the discount rate r is below $\underline{r}(\Delta)$, at some beliefs, the expert would optimally learn not to contradict but to confirm his current belief. We, however, restrict attention to $r \geq 1$ because our proof of Theorem 1 requires this assumption.

[7] Subject to re-labeling the states and actions, the payoffs of the principal are identical to those of the expert. However, the contradictory learning policy that is optimal for the expert is not optimal for the principal.

[8] The principal's value function under expert's optimal learning is discontinuous at a belief whenever the belief process is certain to move away from that belief. The beliefs with discontinuities correspond to thresholds where the expert switches to a different information source. In Figure 3b, the discontinuity occurs at $p = \check{p}^{E}$.



(a) Learning strategy. The optimal learning strategy from the perspective of the principal (dashed line) and the expert (solid line).When the player finds it optimal not to learn, α is set to be equal to belief p.



(b) The principal's value function. The red line is the value the principal would obtain if she had access to expert's learning technology; the blue line is the principal's value under complete delegation; the black thin line is the principal's value from immediate decision.

Figure 3: The optimal learning policy for each belief.

Figure 3b also demonstrates that whenever $r \in [1, \bar{r}(\Delta)]$, for some not-too-low p, delegation of learning allows the principal to obtain higher payoff than the payoff she obtains from immediate decision. However, for sufficiently low p, the principal prefers an immediate decision to any learning, while the expert finds it optimal to continue learning. Hence, for low p, the principal would prefer to take over the final decision making from the expert.

4 Main Result

Naturally, if the principal's and the expert's preferences are sufficiently misaligned, the principal is always better off under flexible delegation than under complete delegation. The main result of the paper is that, when the preferences are sufficiently closely aligned, one type of delegation does not unambiguously dominate the other. On the one hand, flexible delegation allows the principal to curtail expert's learning, thus eliminating the cost of delay in decision when she deems such delay excessive. On the other hand, the threat of the principal's intervention makes the expert favour the information source that avoids early termination of learning. Consequently, flexible delegation makes the expert switch to information source A at lower beliefs than under complete delegation, which makes the principal worse off.

Theorem 1 states the main result formally; below we provide intuition for the result. The statement of the theorem utilises Corollary 1 to fix the order of quantifiers.

Theorem 1. *For any $r \geq 1$, there exists a threshold $0 \leq \Delta^*(r) \leq \bar{\Delta}(r)$ such that*

- *for $\Delta < \Delta^*(r)$, there exist a subset of p for which the principal is strictly better off under complete delegation and a subset of p for which the principal is strictly better off under flexible delegation, the principal is indifferent between both types of delegation otherwise;*

- *for $\Delta \geq \Delta^*(r)$, the principal is strictly better off under flexible delegation for some p and is indifferent between both types of delegation otherwise.*

Proof. See Appendix A.3 □

Because the expert's learning regions are shifted to the left relative to the principal's individually optimal learning regions, under flexible delegation, any principal's intervention is necessarily one-sided. For low p, the principal stops the expert's learning from source A at higher p than the expert would have liked. For high p, the principal would like to prolong learning from source B relative to the expert. The principal, however, cannot do anything about this as she cannot coerce the expert to learn for longer than the expert finds optimal.

Since the principal's interventions are necessarily one-sided, flexible delegation does not affect the expert's B-learning continuation strategy starting from any belief. Under B-learning in the absence of a signal, belief p gradually increases. Hence, despite principal's intervention at some low p, at higher p the expert finds it optimal to pursue the same B-learning strategy under flexible and complete delegation — in either case, the expert optimally stops his B-learning when the posterior belief drifts to $\bar{p}^{*E}$. Thus, the merits of flexible delegation to the principal should be assessed on the basis of its impact on expert's A-learning.

Under flexible delegation, the ability to override the expert enables the principal to stop expert's A-learning whenever the posterior belief p is sufficiently low to make immediate decision b optimal. Eliminating the cost of the expert's prolonged learning is the main benefit of flexible delegation to the principal.

At the same time, due to the possibility of early termination, the value the expert derives from A-learning goes down and so the expert switches to learning from source A at a lower belief than he would have done absent the principal's intervention. The delay in the switch to A-learning hurts the principal as by Lemma 2, the principal wants the expert to switch to A-learning early on, that is, at any belief in $\left[\underline{p}^{*P}, \check{p}^{P}\right]$, with $\check{p}^{P} > \check{p}^{E}$.

Overall, whether at some beliefs, flexible delegation makes the expert's learning strategy even more sub-optimal for the principal than it is under complete delegation depends

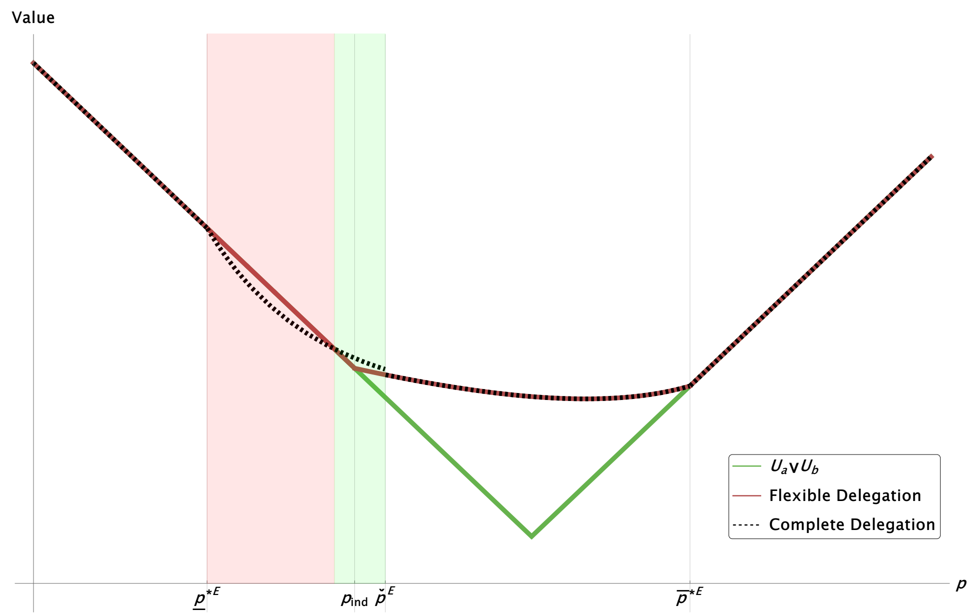


Figure 4: Principal's value function when Δ is low. The green line is the principal's value from immediate decision; the black dashed line is the principal's value under complete delegation; the red line is the principal's value under flexible delegation.

on whether at $\check{p}^E$, the belief at which the autonomous expert switches from A-learning to B-learning, the principal prefers the expert's B-learning to no learning. If so, then under flexible delegation, the principal intervenes in the belief region where the autonomous expert pursues A-learning, and such intervention makes the expert delay switching to source A to the detriment of the principal. In contrast, if at $\check{p}^E$ the principal prefers the immediate decision to expert's B-learning, then principal's intervention terminates expert's B-learning in the belief region where the autonomous expert pursues B-learning, and such intervention triggers no strategic response from the expert.

Figure 4 depicts the principal's payoffs when Δ is sufficiently low so that at $\check{p}^E$, the principal's payoff under expert's B-learning is higher than the payoff from the immediate decision. In this case, the principal prefers complete delegation for initial beliefs in the green region and flexible delegation for beliefs in the red region. In other words, when the interests of the principal and the expert are sufficiently closely aligned, the principal's choice between flexible delegation and complete delegation involves a trade off between improving payoff for sufficiently low p, at which flexible delegation allows terminating expert's prolonged learning, and decreasing payoffs for some higher beliefs, at which the principal would be better off if the expert pursued his first-best A-learning policy.

In contrast, Figure 5 depicts the principal's payoffs when Δ is sufficiently high so that at $\check{p}^E$, the principal's payoff under the expert's B-learning is lower than her payoff from the

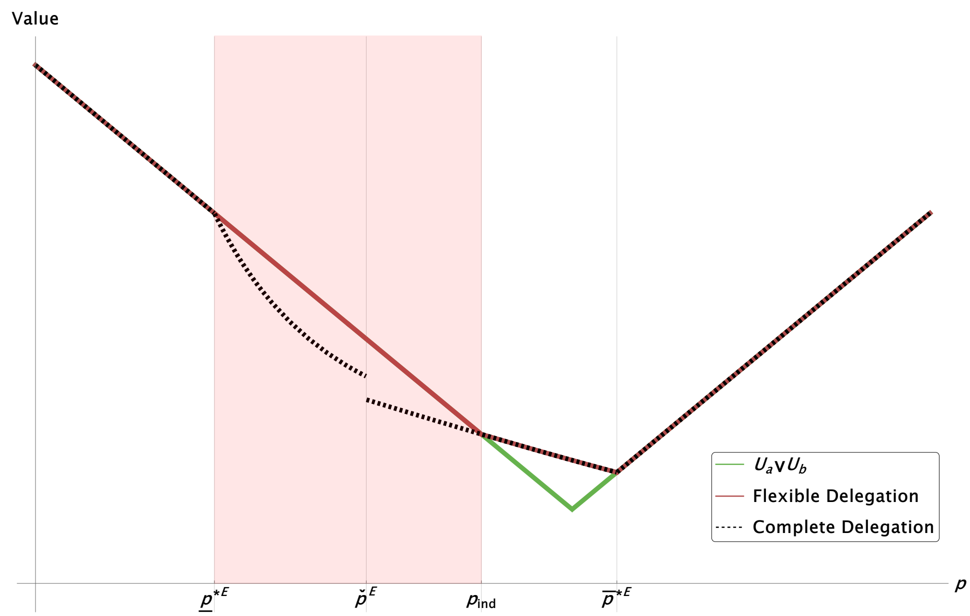


Figure 5: Principal's value function when Δ is high. The green line is the principal's value from immediate decision; the black dashed line is the principal's value under complete delegation; the red line is the principal's value under flexible delegation.

immediate decision. In this case, under flexible delegation the principal intervenes when the expert pursues B-learning. Such intervention does not trigger a strategic response from the expert. Since flexible delegation does not affect the expert's learning on the region where the principal never intervenes, flexible delegation does not have a cost, but it still has the benefit of improving the principal's payoff at low p; that is, in the figure, there is no green region, but there is a red region.

Intuitively, as r increases, any delay in decision becomes costlier and so the principal and the expert derive less value from learning. As a result, the principal and expert's learning regions shrink, thereby magnifying their disagreement about the information source from which to learn. Consequently, it can be expected that as r increases, the possibility of principal's intervention makes expert's learning strategy more sub-optimal for a smaller set of Δ. Indeed, Figure 6 illustrates that threshold $\Delta^*(r)$ decreases with r.

5 Discussion and Concluding Remarks

In our model, the expert accesses information sources for free and so discount rate r, which is common to the principal and the expert, is the only cost of learning. It is natural to consider situations in which the expert additionally has to pay some flow cost, $c \neq 0$, for utilising an information source. In a decision problem, the impact of flow cost c

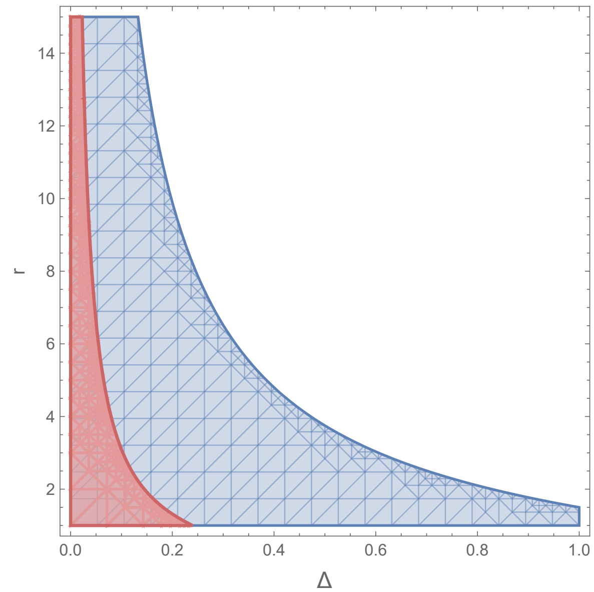


Figure 6: The union of the blue and the red regions corresponds to parameters Δ and r that are admissible in Theorem 1. In the red region, for some p the principal is better off under complete delegation than under flexible delegation.

and discount rate r on the optimal learning policy is qualitatively the same — both costs reduce the value of learning, thus compressing from both sides the interval of beliefs in which some learning is optimal. In our model, the flow cost that is borne by the expert alone reduces the scope for the principal's intervention under the flexible delegation. The main benefit of the flexible delegation is that it allows the principal to curtail expert's A-learning when the principal is already sufficiently confident that b is the right decision to make. With flow cost c, the autonomous expert, who bears this cost, is naturally averse to prolonged learning — that is, the expert's individually optimal stopping threshold $\underline{p}^{*E}$ increases with c — and so, the range of beliefs at which the principal benefits from the ability to terminate the expert's learning shrinks. For sufficiently high c, the expert may even want to terminate A-learning at higher beliefs than the principal, who does not internalise c. Clearly, in this case, flexible and complete delegation are payoff equivalent for the principal.[9] We conjecture, however, that at lower c, the results from Theorem 1 remain qualitatively unchanged and when Δ is small, the strategic response of the expert may still sometimes make the principal worse off under flexible delegation than under complete delegation.

We did not explicitly specify whether the principal knows her prior belief when she makes the choice between complete and flexible delegation. Our analysis provides in-

[9]The cost c above which $\underline{p}^{*E} \geq \underline{p}^{*P}$ can be easily derived and is given by $\bar{c} = \frac{r(\Delta+2r+4)\Delta}{2r+2-\Delta}$.

sights that are relevant regardless of this assumption. In some situations, the principal may need to determine a contractual arrangement for learning about a specific issue and so will be aware of her prior assessment at the time of contracting. For example, a client may ask a financial consultant for an advice on whether to sell a particular stock, knowing the current market sentiment or the company's latest financial statement.[10] In such situations, to avoid unnecessary delay, the principal will not engage the expert when she is already sufficiently confident about the right decision; she will only engage the expert if she derives some value from the expert's learning, in which case complete delegation is superior to flexible delegation (see green region in Figure 4). In other words, in cases like this, the principal's payoff from delegation of learning will be the upper envelope of payoffs under flexible and complete delegation. In other situations, the principal may be required to decide on a contractual arrangement before her prior belief is revealed. For example, the government undertakes learning and has the decision-making authority on all issues, and the electorate has little scope to fine tune the contractual arrangements to the prior belief on each specific issue. According to Theorem 1, in situations like this, while flexible delegation unambiguously dominates complete delegation when the disagreement between the principal and the expert is large, when the disagreement is small, the choice between the contractual arrangements is less clear cut and will generally depend on the distribution of priors and the relative importance of various issues.

[10]This example fits well our model. For the most part, it is reasonable to expect that the consultant wants to make the right decision given the current sentiment of the stock market. Nevertheless, he may have a bias towards one of the two actions, stemming perhaps from the risk exposure of his overall portfolio. Furthermore, such financial advice is often provided for free to retail bank clients who do not necessarily have to heed this advice.

Appendix A Technical Appendix

A.1 Proof of Lemma 1

Under complete delegation, the expert solves an optimal control free-boundary problem. Let V denote the expert's value function. By the dynamic programming principle, wherever it is differentiable, V can be characterized through the Hamilton-Jacobi-Bellman (HJB) equation:

$$rV(p) \geq \max_{\alpha}\Big\{p\alpha\left(v_a^A - V(p)\right) + (1-p)(1-\alpha)\left(v_b^B - V(p)\right) \\ -(2\alpha-1)p(1-p)V'(p)\Big\},\ p\in(0,1) \quad \text{(A.1)}$$

subject to the boundary condition[11]

$$V(p) = V_a(p) \vee V_b(p). \quad \text{(A.2)}$$

Since the HJB equation is linear in α, the optimal policy is a bang-bang solution $\alpha(p)\in\{0,1\}$, except for posteriors where the derivative of the objective vanishes.

Theorem 1 in Che and Mierendorff (2019) shows that the expert's value function is the upper envelope of values from no learning, A-learning, B-learning as well as a learning policy that we call *confirmatory learning*.[12] In what follows, we provide functional forms for the quantities characterising the optimal learning policy which we require for the subsequent analysis.

Contradictory A-learning

Setting $\alpha = 1$ corresponds to A-learning. Value derived from A-learning satisfies the ordinary differential equation

$$rV_1(p) = p\left(v_a^A - V_1(p)\right) - p(1-p)V_1'(p) \quad \text{(A.3)}$$

The solution to this ODE with boundary condition $V_1(q) = v$ is well-defined for $q\in(0,1)$ and is given by

$$V_1(p;q,v) = \frac{1}{1+r}\left(pv_a^A + \frac{1-p}{1-q}\left(\frac{q(1-p)}{p(1-q)}\right)^r\left((1+r)v - qv_a^A\right)\right). \quad \text{(A.4)}$$

Let $\underline{p}^{*E}$ denote the belief at which the expert optimally chooses to terminate A-learning. At $\underline{p}^{*E}$, the expert is sufficiently pessimistic about state A that he makes the immediate decision b.

[11] Operator $\vee$ is the binary max operator.
[12] Che and Mierendorff (2019) refer to this learning policy as opposite-biased learning.

Value matching and smooth pasting at $\underline{p}^{*E}$ pins down the boundary belief:

$$\underline{p}^{*E} = \frac{r v_b^B}{v_a^A - v_b^A + r\left(v_b^B - v_b^A\right)}, \tag{A.5}$$

which, upon substituting the expert's payoffs, is equivalent to (3). Hence, the value from the optimally terminated A-learning is

$$V_1^*(p) = \frac{1}{1+r}\left(p v_a^A + (1-p) v_b^B \left(\frac{1-p}{p}\right)^r \left(\frac{\underline{p}^{*E}}{1-\underline{p}^{*E}}\right)^r\right). \tag{A.6}$$

By direct computations, function $V_1^*(p)$ is strictly convex on $\left[\underline{p}^{*E}, 1\right]$.

Contradictory B-learning

By setting $\alpha = 0$ and going through the similar steps, we can derive the value from the optimally terminated B-learning. In particular, value derived from B-learning satisfies the ordinary differential equation

$$r V_0(p) = (1-p)\left(v_b^B - V_0(p)\right) + p(1-p) V_0'(p). \tag{A.7}$$

The solution to this ODE with boundary condition $V_0(q) = v$ is well-defined for $q \in (0,1)$ and is given by

$$V_0(p; q, v) = \frac{1}{1+r}\left((1-p) v_b^B + \frac{p}{q}\left(\frac{(1-q)p}{(1-p)q}\right)^r ((1+r)v - (1-q) v_b^B)\right). \tag{A.8}$$

The expert optimally terminates B-learning at belief $\bar{p}^{*E}$, at which he is sufficiently optimistic about state A to make decision a. By smooth pasting and value matching,

$$\bar{p}^{*E} = 1 - \frac{r v_a^A}{r(v_a^A - v_a^B) + v_b^B - v_a^B}, \tag{A.9}$$

which, upon substituting the expert's payoffs, is equivalent to (4). Hence, the value from the optimally terminated B-learning is

$$V_0^*(p) = \frac{1}{1+r}\left((1-p) v_b^B + p v_a^A \left(\frac{p}{1-p}\right)^r \left(\frac{1-\bar{p}^{*E}}{\bar{p}^{*E}}\right)^r\right). \tag{A.10}$$

By direct computations, function $V_0^*(p)$ is strictly convex on $\left[0, \bar{p}^{*E}\right]$.

Confirmatory learning

Under confirmatory learning policy the expert focuses all his attention on the source capable of generating a conclusive signal about state he thinks to be more likely. If no signal arrives, beliefs converge to the absorbing belief, p^{*E}, where $\alpha = 1/2$ and the expert does not take an action unless a signal arrives. We denote the value of the learning policy with $\alpha = 1/2$, by $V^*(p)$. It is immediate that

$$V^*(p) = p\mathbb{E}\left[e^{-r\tau}v_a^A \mid \omega = A\right] + (1-p)\mathbb{E}\left[e^{-r\tau}v_b^B \mid \omega = B\right] \tag{A.11}$$

$$= \frac{pv_a^A + (1-p)v_b^B}{1+2r} \tag{A.12}$$

Further, by smooth pasting and value matching of the above equation and either (A.3) or (A.7) we can derive the absorbing belief p^{*E}:

$$p^{*E} = \frac{v_b^B}{v_b^B + v_a^A} \tag{A.13}$$

Taking $V^*(p^{*E})$ as a boundary condition, the value from confirmatory learning for the expert is:

$$V_{cf}^E(p) = \begin{cases} \frac{1}{1+r}\left((1-p)v_b^B + \frac{1}{1+2r}(1-p)^{-r}p^{1+r}v_a^A\left(\frac{v_a^A}{v_b^B}\right)^r\right) & p < p^{*P} \\ \frac{2v_a^A v_b^B}{(v_a^A+v_b^B)(1+2r)} & p = p^{*P} \\ \frac{1}{1+r}\left(pv_a^A + \frac{1}{1+2r}\left((1-p)^{1+r}p^{-r}v_b^B\left(\frac{v_b^B}{v_a^A}\right)^r\right)\right) & p > p^{*P} \end{cases} \tag{A.14}$$

Optimal learning policy

Let $\bar{r}^E(\Delta)$ denote the smallest discount rate r such that at any prior belief, the expert prefers making decision immediately to some learning. The expected payoff from the immediate decision is the lowest when the expert is indifferent between decisions a and b — that is, at a belief $\hat{p}$ such that

$$V_a(\hat{p}) = V_b(\hat{p}) \quad \Longrightarrow \quad \hat{p} = \frac{v_b^B - v_a^B}{v_a^A - v_b^A + v_b^B - v_a^B} = \frac{2-\Delta}{4} \tag{A.15}$$

At this belief, equating the expected value of immediate decision to the value from learning for $\delta \to 0$ units of time and only then making the decision yields

$$\bar{r}^E(\Delta) = \frac{\left(v_b^B - v_a^B\right)\left(v_a^A - v_b^A\right)}{v_b^B v_a^A - v_a^B v_b^A} = \frac{4-\Delta^2}{2\Delta}. \tag{A.16}$$

Note that

$$\lim_{\Delta \to 0^+} \bar{r}^E(\Delta) = \infty.$$

That is, for very small bias Δ the expert prefers some learning for any discount rate r.

Let $\underline{r}^E(\Delta)$ denote the highest discount rate at which the optimal learning policy features confirmatory learning. According to Che and Mierendorff (2019), confirmatory learning is part of the optimal policy if and only if at p^{*E} it is optimal to pursue the confirmatory strategy.

In order to characterize $\underline{r}^E(\Delta)$, we make the following observation.

Lemma A.1. *$V_0^*(p^{*E}) > V_1^*(p^{*E})$ for $\Delta > 0$.*

Proof. By direct computations,

$$V_0^*(p^{*E}) - V_1^*(p^{*E}) = \frac{1+\Delta}{(1+r)(2+\Delta)}\left(\left(\frac{r}{2-\Delta+r(1-\Delta)}\right)^r - \left(\frac{r(1+\Delta)}{2+r+\Delta}\right)^r\right). \tag{A.17}$$

For $1 > \Delta > 0$ and $r > 0$, the expression above is positive because

$$\frac{r}{2-\Delta+r(1-\Delta)} > \frac{r(1+\Delta)}{2+r+\Delta}.$$

□

Hence, the confirmatory learning policy will not be part of the individually optimal strategy for the expert if and only if $V_0^*(p^{*E}) \geq V^*(p^{*E})$ or, equivalently,

$$\left(\frac{r v_a^A}{v_a^A - (1+r)v_b^A}\right)^r - \frac{1}{1+2r} \geq 0. \tag{A.18}$$

The value of r that makes the above condition hold with equality is the threshold value $\underline{r}^E(\Delta)$. We require that $\underline{r}^E(\Delta) \leq 1$ for all Δ. After substituting the expert's payoffs, the left-hand side of (A.18) becomes

$$f(r) := \left(\frac{r}{2-\Delta+r(1-\Delta)}\right)^r - \frac{1}{1+2r}. \tag{A.19}$$

At $\Delta = 0$,

$$f(r) = \left(\frac{r}{2+r}\right)^r - \frac{1}{1+2r}, \tag{A.20}$$

which is equal to zero if $r = 1$ and is strictly positive if $r > 1$. Thus, for $\Delta = 0$, inequality (A.18) is satisfied for all $r \geq 1$. Furthermore, by inspection, $f(r)$ is increasing in Δ and so inequality (A.18) is also satisfied for all $\Delta > 0$ and all $r \geq 1$.

Upper envelope of contradictory A- and B-learning

Since $V_1^*(\underline{p}^{*E}) > V_0^*(\underline{p}^{*E})$, $V_0^*(\bar{p}^{*E}) > V_1^*(\bar{p}^{*E})$, V_1^* intersects V_0^* at least once. Lemma A.2 shows that the intersection is unique.

Lemma A.2. *For $r \geq 1$, V_1^* intersects V_0^* once and from above.*

Proof. The argument follows closely the proof of Lemma 2 in Che and Mierendorff (2019). If at

p, $V_0^*(p) = V_1^*(p) = V(p)$, then summing (A.3) and (A.7) yields:

$$V_0'(p) - V_1'(p) = \frac{1}{p(1-p)}\left((1+2r)V(p) - \left(pv_a^A + (1-p)v_b^B\right)\right) \tag{A.21}$$
$$= \frac{1+2r}{p(1-p)}\left(V(p) - V^*(p)\right), \tag{A.22}$$

where V^* is the value of the learning policy with $\alpha = 1/2$ defined in (A.12).

Since whenever $r \geq 1$, V lies above V^*, whenever V_1^* and V_0^* intersect, $V_0' > V_1'$. Consequently, V_1^* intersects V_0^* once and from above. □

Function V_1^* intersects V_0^* at belief $\check{p}^E$ implicitly defined as the unique solution in $\left[\underline{p}^{*E}, \bar{p}^{*E}\right]$ to $g^E(p) = 0$, where function

$$g^E(p) := V_1^*(p) - V_0^*(p), \tag{A.23}$$

which, upon substituting the expert's payoffs, is equivalent to (5).

A.2 Proof of Lemma 2

Following similar steps as in the proof of Lemma 1, we can derive the belief $\underline{p}^{*p}$ at which the principal would terminate A-learning if she had access to the same information technology as the expert:

$$\underline{p}^{*p} = \frac{ru_b^B}{u_a^A - u_b^A + r(u_b^B - u_b^A)} = \frac{r(1+\Delta)}{2(1+r) - \Delta} \tag{A.24}$$

Similarly, the principal would terminate her individually optimal B-learning at belief

$$\bar{p}^{*p} = 1 - \frac{ru_a^A}{ru_a^A - u_a^B - ru_a^B + u_b^B} = 1 - \frac{r}{2(1+r) + \Delta} \tag{A.25}$$

Given the analytical expressions for the thresholds, by direct computation:

$$\underline{p}^{*P} - \underline{p}^{*E} = \frac{\Delta r(\Delta + 4 + 2r)}{4(1+r)^2 - \Delta^2} > 0 \tag{A.26}$$

since by assumption, $1 > \Delta > 0$ and $r > 0$.

Similarly, by direct computation:

$$\overline{p}^{*P} - \overline{p}^{*E} = \frac{\Delta r(\Delta + 4 + 2r)}{4(1+r)^2 - \Delta^2} > 0 \tag{A.27}$$

as above.

At $\Delta = 0$, the payoffs of the principal and the expert coincide and so $\check{p}^E = \check{p}^P$. By Lemma A.3, which is stated below, $\check{p}^E$ is decreasing in Δ. Since subject to re-labeling the states and actions, the payoffs of the principal are identical to those of the expert, an equivalent argument can be used to show that $\check{p}^P$ is increasing in Δ. Hence, for any $\Delta > 0$, $\check{p}^E < \check{p}^P$.

Lemma A.3. *Let $\check{p}^E \in \left[\underline{p}^{*E}, \bar{p}^{*E}\right]$ be implicitly defined as the unique solution of $g^E(p) = 0$, where $g^E(p)$ is given in (5). Then, $\check{p}^E$ is decreasing in Δ.*

Proof. By the implicit function theorem,

$$\frac{\partial}{\partial \Delta}\check{p}^E = -\left.\frac{\partial g^E/\partial \Delta}{\partial g^E/\partial p}\right|_{p=\check{p}^E}. \tag{A.28}$$

By Lemma A.2, at $p = \check{p}^E$, $\partial g^E/\partial p < 0$ and so the sign of the derivative $\partial \check{p}^E/\partial \Delta$ is the same as the sign of the derivative $\partial g^E/\partial \Delta$ at $p = \check{p}^E$, where $g^E(p) = 0$.

It will be convenient to change variables to

$$\rho := \frac{p}{1-p}, \quad \underline{\rho} := \frac{\underline{p}^{*E}}{1-\underline{p}^{*E}}, \quad \text{and} \quad \bar{\rho} := \frac{\bar{p}^{*E}}{1-\bar{p}^{*E}}. \tag{A.29}$$

After the change of variables, function $g^E(p)$ is equal to $(1-p)h(\rho)$, where

$$h(\rho) := \rho(1+\Delta)\left(1-\left(\frac{\rho}{\bar{\rho}}\right)^r\right)+\left(\frac{\underline{\rho}}{\rho}\right)^r - 1. \tag{A.30}$$

Consequently, the requirement that $g^E(p) = 0$ is equivalent to $h(\rho) = 0$. Since $\check{p}^E \in \left[\underline{p}^{*E}, \bar{p}^{*E}\right]$, the solution of $h(\rho) = 0$ must belong to the interval $\left[\underline{\rho}, \bar{\rho}\right]$. Furthermore, the derivative $\partial g^E/\partial \Delta$ has the same sign as the derivative

$$\frac{\partial h}{\partial \Delta} = \rho\left(1-\left(\frac{\underline{\rho}}{\rho}\right)^{1+r} - \left(1+\frac{3+2r}{1+\Delta}\frac{1}{\bar{\rho}}\right)\left(\frac{\rho}{\bar{\rho}}\right)^r\right). \tag{A.31}$$

When $\Delta = 0$, after substituting in $\underline{\rho}$ and $\bar{\rho}$,

$$h(\rho) = \rho\left(1-\left(\frac{r\rho}{2+r}\right)^r\right)+\left(\frac{r}{(2+r)\rho}\right)^r - 1, \tag{A.32}$$

and so the unique solution of $h(\rho) = 0$ is $\rho = 1$. Then, the derivative

$$\left.\frac{\partial h}{\partial \Delta}\right|_{\substack{\Delta=0,\\ h(\rho)=0}} = 1-(1+2r)\left(\frac{r}{2+r}\right)^r < 0 \tag{A.33}$$

for all $r \geq 1$.

When $\Delta = 1$, after substituting in $\underline{\rho}$ and $\bar{\rho}$,

$$\left.\frac{\partial h}{\partial \Delta}\right|_{\Delta=1} = \rho\left(1-\left(\frac{r}{(3+r)\rho}\right)^{1+r} - 2^r(1+r)(1+2r)(r\rho)^r\right). \tag{A.34}$$

Evaluating this derivative at $\rho = \underline{\rho}$,

$$\left.\frac{\partial h}{\partial \Delta}\right|_{\substack{\Delta=1,\\ \rho=\underline{\rho}}} = -\frac{2^r}{r}(1+r)(1+2r)\left(\frac{r^2}{3+r}\right)^{1+r} < 0 \tag{A.35}$$

for all $r \geq 1$. Furthermore, the derivative of $\left.\frac{\partial h}{\partial \Delta}\right|_{\Delta=1}$ with respect to ρ is negative at $\rho = \underline{\rho}$ and decreasing for all $r \geq 1$:

$$\left.\frac{\partial}{\partial \rho}\left(\left.\frac{\partial h}{\partial \Delta}\right|_{\Delta=1}\right)\right|_{\rho=\underline{\rho}} = 1 + r - 2^r(1+2r)(1+r)^2\left(\frac{r^2}{3+r}\right)^r < 0 \tag{A.36}$$

and

$$\frac{\partial^2}{\partial \rho^2}\left(\left.\frac{\partial h}{\partial \Delta}\right|_{\Delta=1}\right) = -\frac{r(1+r)}{(3+r)\rho^2}\left(r\left(\frac{r}{(3+r)\rho}\right)^r + 2^r(1+r)(3+r)(1+2r)\rho(r\rho)^r\right) < 0. \tag{A.37}$$

Hence, $\left.\frac{\partial h}{\partial \Delta}\right|_{\Delta=1} < 0$ at $h(\rho) = 0$ for all $r \geq 1$.

Given that when $\Delta = 0$ and $\Delta = 1$, $\frac{\partial h}{\partial \Delta}$ evaluated at $h(\rho) = 0$ is negative for all $r \geq 1$, if this derivative never changes its sign, it is negative everywhere. By contradiction, suppose that there exit ρ and Δ that satisfy $h(\rho) = 0$ and

$$\frac{\partial^2 h}{\partial \Delta^2} = r(1+r)\frac{\rho^2}{r^2}\left(\left(\frac{\underline{\rho}}{\rho}\right)^{r+2} - \frac{(3+2r)^2}{(1+\Delta)^3\rho^3}\left(\frac{\rho}{\bar{\rho}}\right)^{r+2}\right) = 0. \tag{A.38}$$

After some algebraic manipulations, requiring that $h(\rho) = 0$ and $\frac{\partial^2 h}{\partial \Delta^2} = 0$ hold simultaneously is equivalent to solving the system of non-linear equations:

$$\rho(1+\Delta) = \frac{1-\left(\underline{\rho}/\rho\right)^r}{1-(\rho/\bar{\rho})^r} \tag{A.39}$$

$$\rho(1+\Delta) = (3+2r)^{2/3}\left(\frac{\rho/\bar{\rho}}{\underline{\rho}/\rho}\right)^{\frac{r+2}{3}} \tag{A.40}$$

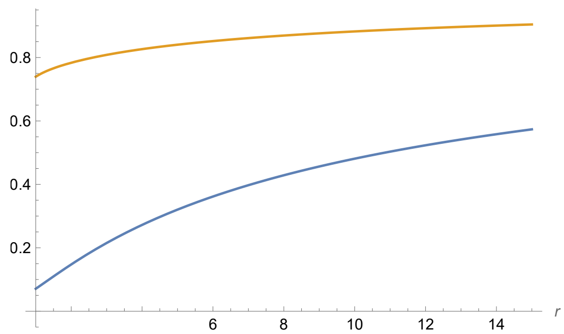
Define new variables

$$\underline{x} = \frac{\underline{\rho}}{\rho} \quad \text{and} \quad \bar{x} = \frac{\rho}{\bar{\rho}}, \tag{A.41}$$

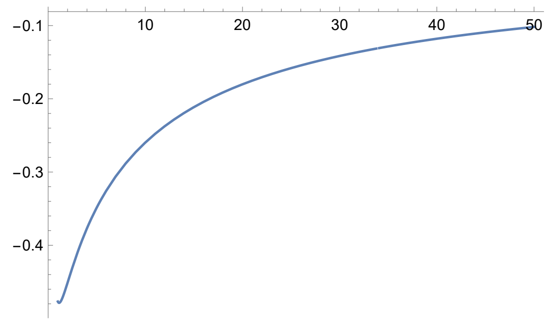
where $0 < \underline{x} \leq \bar{x} < 1$. Expressing ρ and Δ in the new variables yields[13]

$$\Delta = \frac{-r(r+(2+r)\bar{x}\underline{x}) + \sqrt{r^4 + 2r^2(r^2-2)\bar{x}\underline{x} + (2+r)^4\bar{x}^2\underline{x}^2}}{2(1+r)\bar{x}\underline{x}} \tag{A.42}$$

[13] Δ is a solution to a quadratic equation. One root of that equation is always negative and so can be discarded.



(a) Solution $(\underline{x}, \bar{x})$ is admissible as $0 < \underline{x} < 1$ (blue line), $0 < \bar{x} < 1$ (orange line) and $\underline{x} < \bar{x}$.



(b) Parameter Δ implied by the solution $(\underline{x}, \bar{x})$ is not admissible as it is less than 0.

Figure 7: Numerical solution to system (A.44)-(A.45).

and

$$\rho = \frac{r}{(2 + r + \Delta)\underline{x}}. \tag{A.43}$$

In new variables, system (A.39)-(A.40) is

$$\rho(1 + \Delta) = \frac{1 - \underline{x}^r}{1 - \bar{x}^r} \tag{A.44}$$

$$\rho(1 + \Delta) = (3 + 2r)^{2/3} \left(\frac{\bar{x}}{\underline{x}} \right)^{\frac{r+2}{3}}. \tag{A.45}$$

Numerical methods yield a unique solution $(\underline{x}, \bar{x})$ satisfying the requirement that $0 < \underline{x} \leq \bar{x} < 1$ (see Figure 7a). However, at this solution, Δ is negative for all $r \geq 1$ (see Figure 7b). Since there is no admissible solution to system (A.44)-(A.45), derivative $\frac{\partial h}{\partial \Delta}$ evaluated at $h(\rho) = 0$ never changes its sign and is negative for all admissible parameter values.

□

A.3 Proof of Theorem 1

Preliminary observation

Lemma A.4 proves that for $r \in [1, \overline{r}(\Delta)]$, under flexible delegation, the expert would never pursue the confirmatory learning policy. This observation reduces the types of effects flexible delegation can have on the expert's learning strategy.

Lemma A.4. *If $1 \leq r(\Delta) \leq \overline{r}(\Delta)$, then $V_0^*(p) \geqslant V_{cf}^E(p)$ for all p.*

Proof. Since, $1 \leq r(\Delta) \leq \overline{r}(\Delta)$, under complete delegation, the expert uses only contradictory learning policy for any p. Hence, $V_0^*\left(\check{p}^E\right) = V_1^*\left(\check{p}^E\right) \geqslant V_{cf}^E\left(\check{p}^E\right)$. In addition, $V_0^*(p) \geqslant V_{cf}^E(p)$ for any belief $p \geqslant \check{p}^E$, since contradictory B-learning is the expert's optimal learning strategy under complete delegation for these beliefs.

From Lemma A.1, we know that $\check{p}^E \leqslant p^{*E}$, which implies that $V_0^*\left(p^{*E}\right) > V_{cf}^E\left(p^{*E}\right)$. For any belief $p < p^{*E}$ both $V_{cf}^E(p)$ and $V_0^*(p)$ have identical learning policies with $\alpha = 0$. Since the confirmatory strategy at p^{*E} has a continuation utility that is dominated by the contradictory B-learning, $V_0^*(p) \geqslant V_{cf}^E(p)$ for $p < p^{*E}$, which concludes the proof. □

Lemma A.5 implies that if, under flexible delegation, at some p, the principal finds allowing the expert to pursue B-learning optimal, she also finds it optimal to allow B-learning at all higher beliefs at which the autonomous expert optimally learns.

To state the result, recall that $U_b(p)$ is the value the principal derives from immediate decision b and is given in (2). The value that the principal derives from B-learning is the solution to the principal's equivalent of the HJB equation (A.7). When boundary condition is $U_0(q) = v$, the solution is well-defined for $q \in (0,1)$ and is given by the principal's equivalent of (A.8):

$$U_0(p;q,v) = \frac{1}{1+r}\left((1-p)u_b^B + \frac{p}{q}\left(\frac{(1-q)p}{(1-p)q}\right)^r ((1+r)v - (1-q)u_b^B)\right). \tag{A.46}$$

Lemma A.5. *There exits a threshold $\tilde{\Delta} < 1$ such that for all $\Delta \leq \tilde{\Delta}$, equation*

$$U_b(p) = U_0(p;\bar{p}^{*E}, U_a(\bar{p}^{*E})) \tag{A.47}$$

has a unique solution $p_{ind} \in \left[0, \bar{p}^{*E}\right]$ and for all $\Delta > \tilde{\Delta}$, equation (A.47) has no solution in $\left[0, \bar{p}^{*E}\right]$. When it exists, the solution p_{ind} is increasing in Δ.

Proof. The second derivative of $U_0(p;q,v)$ with respect to p is

$$\frac{\partial^2 U_0}{\partial p^2} = \frac{r}{(1-p)^2 pq}\left(\frac{(1-q)p}{(1-p)q}\right)^r \left((1+r)v - (1-q)u_b^B\right). \tag{A.48}$$

The sign of this derivative depends on the sign of the bracket $\left((1+r)v - (1-q)u_b^B\right)$, which does not depend on p. Hence the function is either globally convex or globally concave in p. In either case, the linear function $U_b(p)$ can cross $U_0(p, \bar{p}^{*E}, U_a(\bar{p}^{*E}))$ at most twice.

At $p = 0$,

$$U_b(0) = 1 + \Delta > \frac{1+\Delta}{1+r} = U_0(0;\bar{p}^{*E}, U_a(\bar{p}^{*E})) \tag{A.49}$$

At $p = \bar{p}^{*E}$,

$$U_b(\bar{p}^{*E}) = \bar{p}^{*E}u_b^A + (1-\bar{p}^{*E})u_b^B = \frac{(3+4r-\Delta)\Delta - 2}{2+2r-\Delta} \tag{A.50}$$

and

$$U_0(\bar{p}^{*E};\bar{p}^{*E}, U_a(\bar{p}^{*E})) = U_a(\bar{p}^{*E}) = \bar{p}^{*E}u_a^A + (1-\bar{p}^{*E})u_a^B = \frac{2-\Delta-2r\Delta}{2+2r-\Delta}. \tag{A.51}$$

Hence, from (A.50) and (A.51) it follows that $U_b(\bar{p}^{*E}) \leq U_0(\bar{p}^{*E};\bar{p}^{*E}, U_a(\bar{p}^{*E}))$ iff

$$\Delta^2 - (4+6r)\Delta + 4 \geq 0. \tag{A.52}$$

Equation $\Delta^2-(4+6r)\Delta+4=0$ has two solutions for Δ but only one in $[0,1]$:

$$\tilde{\Delta}=2+3r-\sqrt{3r(4+3r)}. \tag{A.53}$$

Consequently, on $\left[0,\bar{p}^{*E}\right]$, $U_b(p)$ crosses $U_0(p;\bar{p}^{*E},U_a(\bar{p}^{*E}))$ once and from above when $\Delta\leq\tilde{\Delta}$ and $U_b(p)$ either crosses $U_0(p;\bar{p}^{*E},U_a(\bar{p}^{*E}))$ twice or not at all when $\Delta>\tilde{\Delta}$.

By the implicit function theorem,

$$\frac{\partial}{\partial\Delta}p_{ind}=-\left.\frac{\partial U_b/\partial\Delta-\partial U_0/\partial\Delta}{\partial U_b/\partial p-\partial U_0/\partial p}\right|_{p=p_{ind}} \tag{A.54}$$

Consider $p=p_{ind}$ at which $U_b(p)$ crosses $U_0(p;\bar{p}^{*E},U_a(\bar{p}^{*E}))$ from above. Then,

$$\left.\frac{\partial U_b}{\partial p}-\frac{\partial U_0}{\partial p}\right|_{p=p_{ind}}=-\frac{r(1-2p_{ind}+\Delta)}{p_{ind}(1-p_{ind})}<0, \tag{A.55}$$

and the sign of the derivative (A.54) is the same as the sign of

$$\left.\frac{\partial U_b}{\partial\Delta}-\frac{\partial U_0}{\partial\Delta}\right|_{p=p_{ind}}=\frac{1}{1+r}\left(r+p+\frac{p}{\bar{p}^{*E}}\left(\frac{(1-\bar{p}^{*E})p}{(1-p)\bar{p}^{*E}}\right)^r w(r,\Delta)\right), \tag{A.56}$$

where

$$w(r,\Delta):=\left(\frac{(1+r)(2+r)^2}{2+r(1-\Delta)-\Delta}-\frac{4+r(5+2r)}{2+2r-\Delta}-\frac{r(1+r)}{1+\Delta}-r\right). \tag{A.57}$$

The derivative (A.56) is positive if $w(r,\Delta)$ is positive. The derivative

$$\frac{\partial w}{\partial\Delta}=\frac{r(1+r)}{(1+\Delta)^2}-\frac{4+r(5+2r)}{(2+2r-\Delta)^2}+\frac{(1+r)^2(2+r)^2}{(2+r(1-\Delta)-\Delta)^2} \tag{A.58}$$

is equal to 0 at $r=0$ and is increasing in r. Indeed, the first term $r(1+r)/(1+\Delta)^2$ is increasing in r by inspection. The second term is also increasing in r:

$$\frac{\partial}{\partial r}\left(-\frac{4+r(5+2r)}{(2+2r-\Delta)^2}\right)=\frac{6+5\Delta+r(2+4\Delta)}{(2+2r-\Delta)^3}>0. \tag{A.59}$$

The derivative of the third term

$$\frac{\partial}{\partial r}\left(\frac{(1+r)^2(2+r)^2}{(2+r(1-\Delta)-\Delta)^2}\right)=\frac{2(1+r)(2+r)^3-2(1+r)^3(2+r)\Delta}{(2+r(1-\Delta)-\Delta)^3} \tag{A.60}$$

is also positive because it decreases with Δ and is $2(1+r)(2+r)(3+2r)>0$ at $\Delta=1$. Hence, $w(r,\Delta)$ is increasing in Δ for all $r>0$. At $\Delta=0$, $w(r,\Delta)=r/(2+2r)>0$ and so $w(r,\Delta)>0$ for all $r>1$ and $0\leq\Delta\leq1$. Hence, $\partial p_{ind}/\partial\Delta>0$ and so, when $\Delta>\tilde{\Delta}$, $U_b(p)$ does not cross $U_0(p;\bar{p}^{*E},U_a(\bar{p}^{*E}))$ on $\left[0,\bar{p}^{*E}\right]$. □

Lemma A.5 is the key step in the proof of Theorem 1. Lemma A.5 implies that, when Δ is sufficiently small, there exists a unique threshold p_{ind} such that for $p < p_{ind}$, the principal prefers the immediate decision b to letting the expert learn from source B until stopping at $\bar{p}^{*E}$ and taking action a upon stopping. Furthermore, when Δ is sufficiently large, the principal prefers the immediate action b to expert's optimal B-learning at any $p \in \left[0, \bar{p}^{*E}\right]$.

Proof of the theorem

The proof proceeds in three steps. First, it is argued that any principal's intervention at $p \geq \check{p}^E$ does not change the experts learning strategy compared to the optimal strategy of an autonomous expert. Second, it is argued that any intervention at $p < \check{p}^E$ makes the principal worse off. In the final step, it is argued that there exists a unique threshold Δ^* such that under flexible delegation, the principal intervenes at $p < \check{p}^E$ if and only if $\Delta < \Delta^*$.

STEP 1: By Lemma A.5, if $\min\left\{p_{ind}, \bar{p}^{*E}\right\} \geq \check{p}^E$, at $\check{p}^E$ the principal prefers action b to expert's optimal B-learning. Consequently, under flexible delegation, the principal optimally intervenes to terminate expert's B-learning at beliefs $p > \check{p}^E$. Because under B-learning in the absence of signal, belief moves towards higher p, the principal's intervention at some low $p \in \left(\check{p}^E, \min\left\{p_{ind}, \bar{p}^{*E}\right\}\right)$ does not change expert's B-learning continuation strategy at any higher belief compared to his strategy under complete delegation. Hence, the intervention does not change the principal's continuation payoff relative to her continuation payoff under complete delegation.

STEP 2: If $p_{ind} < \check{p}^E$, the principal intervenes at beliefs $q \in (\underline{p}^{*E}, \check{p}^E)$ at which the autonomous expert would have pursued A-learning. When the principal terminates A-learning at some belief $q \in (\underline{p}^{*E}, \check{p}^E)$ and takes action b, the expert's value from A-learning is $V_1(p; q, V_b(q))$, where V_1 is defined in (A.4) and V_b is defined in (2).

For $q(\underline{p}^{*E}, \check{p}^E)$, $V_1(p; q, V_b(q))$ is bounded above by $V_1(p; \underline{p}^{*E}, V_b(\underline{p}^{*E})) = V_1^*(p)$, where $V_1^*(p)$ is defined in (A.6), because, by construction, $V_1^*(p)$ is the value the expert derives from the optimally terminated A-learning. Since by Lemma A.2, $V_1^*(p)$ crosses $V_0^*(p)$ only once and from above, $V_1(p; q, V_b(q))$ cannot cross $V_0^*(p)$ at $p > \check{p}^E$; it either crosses $V_0^*(p)$ at some $\widehat{p} \in [q, \check{p}^E)$ or not at all. Consequently, the expert switches to information source A at a lower belief $\widehat{p}$ than under complete delegation, if at all. Switching to A-learning at a lower belief $\widehat{p} < \check{p}^E$ makes the principal worse off because by Lemma 2, at $\check{p}^E$ the principal prefers A-learning.

STEP 3: There exists a unique threshold Δ^* such that inequality $p_{ind} < \check{p}^E$ holds for all $0 \leq \Delta < \Delta^*$ and does not hold otherwise. Indeed, p_{ind} is implicitly defined as a solution to $U_b(p) = U_0(p; \bar{p}^{*E}, U_a(\bar{p}^{*E}))$, and by Lemma A.5, $\partial p_{ind}/\partial \Delta > 0$; threshold $\check{p}^E$ is implicitly defined as a solution to $V_1^*(p) = V_0^*(p)$ and by Lemma 2, $\partial \check{p}^E/\partial \Delta < 0$. When $\Delta = 0$, the principal and the expert derive the same value from learning and so $p_{ind} < \check{p}^E$ by construction of the optimal learning policy. When $\Delta = \bar{\Delta}$, $p_{ind} > \check{p}^E$ because $\underline{p}^{*E} = \bar{p}^{*E} < \underline{p}^{*P} = \bar{p}^{*P}$ and the principal derives no value from expert's learning. Then, monotonicity of the derivatives $\partial p_{ind}/\partial \Delta > 0$ and $\partial \check{p}^E/\partial \Delta < 0$ implies that there exists a unique $0 \leq \Delta^* \leq \bar{\Delta}$ that solves $p_{ind} = \check{p}^E$.

References

Argenziano, R., S. Severinov, and F. Squintani (2016). Strategic information acquisition and transmission. *American Economic Journal: Microeconomics 8*(3), 119–155. 4

Chaimanowong, W., T. Ke, and J. M. Villas-Boas (2023). Searching for breakthroughs. *Available at SSRN 4425117*. 4

Che, Y.-K. and K. Mierendorff (2019). Optimal dynamic allocation of attention. *American Economic Review 109*(8), 2993–3029. 2, 3, 4, 7, 16, 19

Deimen, I. and D. Szalay (2019). Delegated expertise, authority, and communication. *American Economic Review 109*(4), 1349–1374. 4

Escobar, J. F. and Q. Zhang (2021). Delegating learning. *Theoretical Economics 16*(2), 571–603. 4

Garfagnini, U. (2011). Delegated Experimentation. *Unpublished Manuscript*, 1–46. 4

Guo, Y. (2016). Dynamic delegation of experimentation. *American Economic Review 106*(8), 1969–2008. 4

Herresthal, C. (2022). Hidden testing and selective disclosure of evidence. *Journal of Economic Theory*, 105402. 4

Ke, T. T., Z.-J. M. Shen, and J. M. Villas-Boas (2016). Search for information on multiple products. *Management Science 62*(12), 3576–3603. 4

Ke, T. T. and J. M. Villas-Boas (2019). Optimal learning before choice. *Journal of Economic Theory 180*, 383–437. 4

Mayskaya, T. (2024). Following beliefs or excluding the worst? the role of unfindable state in learning. *European Economic Review 162*, 104653. 4

Nikandrova, A. and R. Pancs (2018). Dynamic project selection. *Theoretical Economics 13*(1), 115–143. 4

Wald, A. (1947). Foundations of a general theory of sequential decision functions. *Econometrica*, 279–313. 4

Zhong, W. (2022). Optimal dynamic information acquisition. *Econometrica 90*(4), 1537–1582. 4