

# A Dynamic Game under Ambiguity: Repeated Bargaining with Interactive Learning<sup>1</sup>

David Besanko<sup>2</sup>

Northwestern University

Jian Tong<sup>3</sup>

University of Southampton

Jianjun Wu<sup>4</sup>

Princeton Economics Group

First version: December 30, 2011

This version: October 4, 2012

<sup>1</sup>The authors wish to thank seminar participants at the Southampton University, and the participants of the Econometric Society 2012 North American Summer Meeting at the Northwestern University for their comments and suggestions. We specially thank Charles Manski, Alessandro Pavan, Nabil Al-Najjar, and Jonathan Weinstein for stimulating discussions. The expertly advice on the history of probability theory and statistics from John Aldrich is gratefully acknowledged.

<sup>2</sup>Department of Management and Strategy, Kellogg School of Management, Northwestern University, Evanston, IL 60208. Email: d-besanko@kellogg.northwestern.edu

<sup>3</sup>Economics Division, School of Social Sciences, University of Southampton, Southampton, SO17 1BJ, UK. Email: j.tong@soton.ac.uk

<sup>4</sup>Princeton Economics Group, 707 State Rd, Suite 223, Princeton, NJ 08540. Email: j.wu@econgroup.com.

## Abstract

Conventional Bayesian games of incomplete information are limited in their ability to represent complete ignorance of an uninformed player about an opponent's private information. Using an illustrative example of repeated bargaining with interactive learning, we analyze a dynamic game of incomplete information that incorporates a multiple-prior belief system. We consider a game in which a principal sequentially compensates an agent for his effort on a novel experiment — a Poisson process with unknown hazard rate. The agent has knowledge to form a single prior over the hazard rate, but the principal has complete ignorance, represented by the set of all plausible prior distributions over the hazard rate. We propose a new equilibrium concept — Perfect Objectivist Equilibrium — in which the principal infers the agent's prior from the observed history of the game via maximum likelihood updating. The new equilibrium concept embodies a novel model of learning under ambiguity in the context of a dynamic game. The unique (Markov) equilibrium outcome determines a unique bargaining solution. The underlying Markov Perfect Objectivist Equilibria are all belief-free, in sharp contrast to Markov Perfect Bayesian Equilibria, which hinge on subjective pretense of knowledge and predict a continuum of equilibrium outcomes.

“Acknowledging what is known as known, what is not known as unknown, that is knowledge.”

— Analects of Confucius

“What has now appeared is that the mathematical concept of probability is inadequate to express our mental confidence or diffidence in making such inferences, and that the mathematical quantity which appears to be appropriate for measuring our order of preference among different possible populations does not in fact obey the laws of probability. To distinguish it from probability, I have used the term ‘likelihood’ to designate this quantity.”

— Sir R. A. Fisher (1925)

## 1 Introduction

The assumption that decision makers have unique subjective prior beliefs about unknown phenomena is commonplace in many economic models, particularly in Bayesian games in dynamic settings. Defenders of the Savage-Bayesian paradigm would usually invoke the argument that inferences must start from somewhere, so therefore a unique prior is a necessity.

At a fundamental level, this line of argument has been exposed to be questionable in light of the recent literature on multiple priors and ambiguity (Gilboa, Maccheroni, Marinacci and Schmeidler (GMMS), 2010; Manski, 2008), according to which, a multiple-prior belief system should be the rule and the unique-prior belief system is just a special case. Epstein and Schneider (2007) further illustrate that in a multiple priors setting, some form of likelihood inference can suitably replace the familiar Bayesian inference.

In this paper, we use the example of ‘repeated bargaining with interactive learning’ to illustrate that the *ad hoc* subjective element can be removed from the selection of prior distributions in a dynamic game of incomplete information. By allowing multiple prior distributions, arbitrary restriction on the set of priors, such as uniqueness, becomes unnecessary. As a consequence, the (iterative) selection of priors, in addition to updating beliefs based on them, becomes an ongoing process as the game is played. Learning (including the selection of priors and updating of beliefs) under ambiguity is based on observational facts and a likelihood function that is derived from the common knowledge of the game and suitable concept of equilibrium.<sup>1</sup>

Our illustrative model is based on the exponential bandit framework of Keller, Rady, and Cripps (2005). A principal uses sequential short-term contracts to compensate an expert agent to conduct a novel experiment. A breakthrough in the experiment brings benefits to both the principal and the agent, but it is *ex ante* uncertain whether the problem that must be “cracked” to achieve the breakthrough is actually solvable. The agent must continually

---

<sup>1</sup>If the observational data are not contaminated by random noises, as is the case in our model, the likelihood function degenerates and gives deterministic falsification of some subsets of hypotheses. The logical status of the inference degenerates to deductive inference.

decide how much to invest in solving the problem, and the investment of more funds can speed up the time of the breakthrough if the problem is indeed solvable.

As an expert, the agent (a ‘Bayesian’ player) has sufficient knowledge to form a unique prior probability about whether the problem is solvable, and he updates his posterior probability according to the investment history in the absence of a breakthrough. Bayes Rule dictates that the posterior probability decreases over time conditional on non-zero investment and absence of a breakthrough. The principal (a ‘non-Bayesian’ player) — who does not know the agent’s prior belief (his “type”) but can observe the investment history — relies on this history to infer the agent’s type. She must act on her ambiguous beliefs — represented by multiple priors and posteriors about the agent’s type — to determine how much to compensate the agent at each point in time. The principal’s deficiency of information about the true state of the world means that she cannot form complete pairwise rankings over alternative action plans (even without strategic uncertainty). In general, then, there may not exist any unanimous winners among all candidate action plans. How this potential indecision is resolved depends critically on the principal’s attitude toward ambiguity. Following GMMS (2010) we partially represent this factor of decision making by the max-min expected utility criterion.<sup>2</sup> More precisely, the preferences of the principal are lexicographical — they also require that no weakly dominated strategy should be chosen even if it (also) maximizes the worse-case utility function.<sup>3</sup>

Our paper makes five main contributions. First, we formulate and analyze a dynamic game of incomplete information with a multiple-prior belief system, as opposed to the conventional (single-prior) Bayesian game, *a la* Harsanyi (1967, 1968).<sup>4</sup> The game we analyze is rich in strategic interactions, full of potential for multiplicity of equilibria, and would be expected to generate equilibria that would be sensitive to the choice of a prior. Yet, we find that the equilibrium outcome can be implemented by a remarkably simple (pure strategy) cost-reimbursement contract with a time invariant reimbursement rate. The information transmission in equilibrium is minimal, but it is adequate to ensure that the investment intensity and the cumulative level of investment are Pareto efficient (allowing for compensation transfer).<sup>5</sup> Second, by developing a model that involves both learning (i.e., how a

---

<sup>2</sup>In the paper we also formally establish the equivalence between the max-min expected utility criterion and the max-min utility criterion for our setting, therefore we equivalently adopt the max-min criterion originated by Abraham Wald (1950) in his “Statistical Decision Functions”.

<sup>3</sup>The max-min expected utility criterion is only a necessary condition for a most preferred action plan. To ensure sufficiency, a most preferred action plan must not be weakly dominated by any feasible alternative. This additional necessary condition was proposed by Manski (2008), whose origin can be traced to Wald’s concept of admissibility. We propose that GMMS (2010), when suitably extended, could provide an axiomatic foundation for this condition.

<sup>4</sup>Inspired by the title of Harsanyi’s seminal paper – “Games with Incomplete Information Played by ‘Bayesian’ Players” – A fitting alternative title of the current paper could be “A Dynamic Game with Incomplete Information Involving A ‘Non-Bayesian’ Player”.

<sup>5</sup>The notion of socially efficient allocation adopted in this paper is based on Pareto efficiency and the compensation principle, which states that the winner(s) must compensate the loser(s). The compensation transfer is explicitly modelled. Whenever Pareto efficiency is subject to informational constraints, the latter are fully reflected in the representation of preferences.

‘non-Bayesian’ player learns) under ambiguity and strategic interactions between rational players, we add to and complement the emerging literature on learning under ambiguity (Epstein and Schneider, 2003, 2007). A desire to contribute to interactive epistemology (theory of knowledge) is a fundamental motivation of our research programme. Third, to analyze our dynamic game, we present a new equilibrium concept — the Perfect Objectivist Equilibrium (POE) — which extends and contrasts the familiar perfect Bayesian equilibrium (PBE) that is the standard in dynamic (single-prior) Bayesian games. The term ‘objectivist’ emphasizes the importance of objective inference rules in the learning process. The new equilibrium concept also allows us to deal with out-of-equilibrium updating of beliefs in a systematic way. Fourth, to solve for the simplest equilibrium outcome of the model, we define and characterize the Markov Perfect Objectivist Equilibria (MPOE), whose existence and uniqueness we establish. The MPOE in our dynamic bargaining game with incomplete information hinges on an (interactive learning-related) state variable whose transition exhibits the ratchet effect — it can only move in one direction. Fifth, we solve for Markov Perfect Bayesian Equilibria (MPBE) of the model (for a given specification of prior beliefs), and we also solve for the Belief-Free Equilibrium (BFE). A comparison of equilibria reveals that the MPOE of the game is also a BFE and thus is strongly robust with respect to the specification of subjective beliefs. By contrast, there is a multiplicity of MPBE outcomes which are sensitive to the specification of the subjective beliefs.

The efficiency, robustness and simplicity of the MPOE are especially noteworthy. Our analysis reveals that ambiguity and ambiguity aversion weaken the principal’s ability to minimize the agent’s information rent. As a result, the contract she offers does not sacrifice the efficiency of investment as part of a trade off to extract more expected surplus from trade. The MPOE is robust in the sense that it does not delicately depend on any *ad hoc* prior belief and out-of-equilibrium posterior belief. Due to Pareto efficiency it is also robust in that it is renegotiation-proof. The MPOE is also very simple both structurally and dynamically. The main reason for this is the principal’s lack of ability to fully commit to predetermined contractual terms after new information is revealed by the agent, which is a reasonable assumption that we make. This lack of commitment ability by the principal makes the agent wary of revealing sensitive information, which could be used by the principal against the agent’s interest in the future — the ratchet effect. With the information transmission from the agent to the principal being very limited (to avoid ratcheting), a pooling contract arises, and there is little incentive for the principal to alter the terms over time.

Our model of multiple-prior belief system is built upon the axiomatic foundation laid by GMMS (2010), which, in turn, is a synthesis of the pioneering works of Bewley (1986, 2002) and Gilboa and Schmeidler (1989). Crucially, GMMS (2010) interpret the former strand as representing objectively rational preferences, and the latter subjectively rational preferences. In comparison with the axiomatic foundation of the subjective probability theory (notably, the popular version by Anscombe and Aumann (1963)), GMMS (2010) propose less restrictive sets of axioms for a pair of rational preference relations (representing objective rationality and subjective rationality respectively).<sup>6</sup> Based on these more reasonable

---

<sup>6</sup>The subjective probability theory was pioneered by Ramsey (1926), de Finetti (1937) and Savage (1954).

sets of rationality axioms, GMMS (2010) prove that rational beliefs of a decision maker, as revealed by hypothetical betting behavior, are represented by a set of (multiple) prior distributions (hence ambiguity), as opposed to the Savage-Bayesian unique-prior belief system. Since the multiple prior belief system makes the subjective expected utility theory unworkable, GMMS (2010) also provide an axiomatic foundation for the minimum expected utility theory, which is a substitute for the subjective expected utility theory. Our model of decision under ambiguity is more explicit than GMMS (2010) in separating beliefs from tastes over ambiguity. In explicitly adopting lexicographical preferences, we require objective rationality to be primary and subjective rationality to be secondary. As a result, the beliefs system is entirely dealt with by objective rationality. Subjective rationality (which is a complete preference relation) is only involved in decision making in case objective rationality (which is an incomplete preference relation) is inadequate for the task.

Our model of learning under ambiguity is closely related to the seminal work of Epstein and Schneider (2007) that introduces the idea of using a likelihood ratio test as a procedure for reevaluating prior distributions. Our paper adopts the likelihood inference as a generic replacement of Bayesian inference to avoid the *ad hoc* choice of unique prior distribution. We show that for learning under ambiguity, the iterative selection of priors (through maximum likelihood test) duplicates Bayesian updating (through application of Bayes' Theorem). We also show, in our particular game-theoretic setting, the choice of critical value for the likelihood ratio test has no effect on inference. While Epstein and Schneider (2007) model learning about a memoryless (data generating) mechanism, we model a data generating mechanism that is history dependent. Therefore the two models are complementary. The multiple-prior likelihood inference we adopt can be seen as a novel synthesis of ideas from the three competing philosophies of statistics: Bayesian (e.g., Bayes' Theorem), frequentist (e.g., hypothesis testing theory) and Fisherian (e.g., likelihood and sufficiency).<sup>7</sup> A central theme of this new synthesis is the realization that a single prior (or posterior) probability distribution is not a sufficient statistic to summarize existent information<sup>8</sup>, while multiple prior distributions plus likelihoods more adequately represent knowledge (and ignorance).<sup>9</sup>

Our model also relates to the literature on belief-free games. In our characterization of the MPOE of the delegated experimentation model, we show that on any equilibrium path there is a state variable that serves as a sufficient statistic for the principal's beliefs, although off the equilibrium path this is not the case.<sup>10</sup> Because of the ratchet property of this state

---

<sup>7</sup> According to Efron (1998): “the development of modern statistical theory has been a three-sided tug of war between the Bayesian, frequentist and Fisherian viewpoints”. “In many ways the Bayesian and frequentist philosophies stand at opposite poles from each other, with Fisher's ideas being somewhat of a compromise”. “The world of applied statistics seems to need an effective compromise between Bayesian and frequentist ideas.” While the “Fisherian synthesis” is expected to continue to do very well in the 21th century, a “new synthesis” may also emerge.

<sup>8</sup>This statement is clearly Fisherian in spirit (see the quote from Fisher (1925).)

<sup>9</sup>To adequately represent knowledge, it is important to indicate what is known, but also what is not known (see the quote from Analects of Confucius). It is now well known (see Edwards 1992, pp. 57-61) that a single probability distribution cannot accurately represent complete ignorance.

<sup>10</sup>Sufficiency is defined in relation to the relevance to payoffs of the players.

variable, the equilibrium reimbursement rate offered by the principal in the MPOE reflects a bootstrap logic similar to equilibrium trigger strategies in an infinitely repeated game. The bootstrap property is so powerful that the optimality of the players' equilibrium strategies have little reliance on the players' beliefs about opponents' private information – they virtually become “belief-free” in the sense belief-free games (with incomplete information) and BFE (see Bergemann and Morris, 2007; Hörner and Lovo, 2009; Hörner, Lovo and Tomala, 2011). Our model of dynamic game with incomplete information and ambiguous beliefs and our solution concepts — POE and MPOE — share the goal of the literature on belief-free games: the pursuit of equilibrium outcome that is robust to specific restrictions on beliefs. Against this goal, in the setting of the delegated experimentation model, the MPOE performs exceedingly well – in our delegated experimentation game, the unique MPOE outcome is also the unique BFE. The procedure to solve for the MPOE also provides a practical procedure to identify the BFE. In contrast, there exist multiplicity of MPBE outcomes for our model, and although one of them coincides with the belief-free equilibrium outcome, there are a continuum of other equilibrium outcomes that do not. The concept of MPOE appears to be useful for filtering out the equilibria which can be rationalized by the concept of MPBE but which nevertheless rely on arbitrary restrictions on the specification of beliefs – subjective pretence of knowledge – and therefore may not be robust.

Our paper also contributes to the literature of sequential (noncooperative) bargaining (see Rubinstein 1982, 1987 for origin and survey), particularly, repeated bargaining with permanent private information (for surveys, see Fudenberg and Tirole 1991, Chapter 10; Kennan and Wilson 1993). Rubinstein (1982) formulated and characterized an alternating offers bilateral bargaining game with complete information. An important insight that emerged from the Rubinstein bargaining theory is that (taking bargaining procedure as given) the player's time preferences are a key determinant of bargaining power and outcome: it pays to be patient. Models of sequential bargaining under incomplete information allow the surplus from trade be private information of informed players, therefore are useful to describe many realistic bargaining situations. In a survey, Rubinstein (1987) commented on the “state of the art” of the literature:

In my opinion, we are far from having a definitive theory of bargaining with incomplete information for use in economic theory. The problems go deeper than bargaining theory and appear in the literature of refinement of S.E. [i.e., Sequential Equilibrium<sup>11</sup>], an issue explored thoroughly in the last few years. My intuition is that something is basically wrong in our approach to games with incomplete information and that the ‘state of the art’ of bargaining reflects our more general confusion.

As noted earlier, the issue of multiplicity of equilibria can be a severe problem for the conventional Bayesian approach, particularly when the type space of the informed player is

---

<sup>11</sup>In this paper, our discussion is focused on a closely related solution concept – Perfect Bayesian Equilibrium – to which Rubinstein's comments also apply.

very large. By contrasting a single prior belief system to a multiple priors system, our analysis illustrates that ambiguous beliefs combined with aversion to ambiguity may help overcome the problem of multiplicity of equilibria or indeterminacy of (noncooperative) bargaining solution. Another novel insight appears to be that superior information can be a source of bargaining power, endowed to the informed player when informational asymmetry is extreme, e.g., the uninformed player is completely ignorant about the opponent’s private information.

The paper is organized in seven sections. Section 2 describes the illustrative model of delegated experimentation with repeated bargaining that forms the basis of our analysis. Section 3 explains formally what we mean by ambiguity, and lays out the basic structure of the game we analyze. Section 4 defines the Perfect Objectivist Equilibrium and lays the foundations for characterizing it in our model. Section 5 proceeds through a set of steps to characterize Markov Perfect Objectivist Equilibrium outcome for our model. We first characterize and establish the existence of a Markov perfect objectivist equilibrium. We then show that this equilibrium is Pareto efficient (allowing for compensation transfer). Finally we establish the uniqueness of Markov equilibrium outcome in terms of bargaining solution. Section 6 broadens the perspectives to enable assessment of comparative performance of the current model relative to some existing models in a number of important literatures, including dynamic mechanism design with limited commitment and belief-free equilibrium. Section 7 summarizes and concludes. Proofs of all results are in the Appendix.

## 2 Model Formulation

We use the exponential bandit model of Keller, Rady, and Cripps (2005) to incorporate ambiguity into a dynamic game-theoretic setting with incomplete information. We assume that a principal hires an expert agent to solve a problem which may or may not be solvable. We refer to solving the problem as a “breakthrough.” If the agent achieves a breakthrough, the principal receives a “prize” equal to  $\Pi_P > 0$ . The agent also receives a prize equal to  $\Pi_A > 0$ .

Neither the principal nor the agent knows for certain that the problem is solvable. Specifically, let  $\omega$  be a binary variable, where  $\omega = 0$  means a breakthrough can never occur, and  $\omega = 1$  means a breakthrough is possible. The realization of  $\omega$  is unknown to both the principal and the agent unless a breakthrough occurs. However, the agent has expertise, grounded in past experience and knowledge of scientific facts, that enables him to form a prior probability  $p_0 \equiv \Pr(\omega = 1) \in [0, 1]$  that a solution exists. The experience and scientific facts that underpin  $p_0$  are unknown to the principal and are thus private information to the agent. In contrast to the conventional Bayesian approach in which the principal would have a given prior belief about  $p_0$ , we assume that the principal in general has multiple prior beliefs over possible values of  $p_0$ . This assumption, which has important implications for our analysis, captures the possibility that the principal does not know what to believe about  $p_0$ , and it incorporates, as a special case, the possibility of complete ignorance in which the principal believes that *any* prior belief about  $p_0$  is conceivable.

## 2.1 Investment and Belief Updating

Conditional on the problem having a solution, the time at which the breakthrough occurs is random, but it can be influenced by the intensity of the agent's investment in problem-solving activities. At each instant  $t$  in continuous time, the agent has at most one unit of a resource it can invest in problem solving. If the level of the agent's investment is  $k_t \in [0, 1]$ , then conditional on the problem being solvable, the hazard rate of a breakthrough is  $\lambda k_t dt$ . As the agent invests over time and a breakthrough does not occur, the agent updates his posterior belief about the likelihood that a solution to the problem exists. Expressed as a function of cumulative investment  $K_t \equiv \int_0^t k_\tau d\tau$  this posterior is given by:

$$p_t(K_t; p_0) = \Pr(\omega = 1 | \text{no breakthrough by time } t).$$

The law of motion for  $p_t(K_t; p_0)$  is determined by Bayes' Rule:

$$p_{t+dt}(K_{t+dt}; p_0) = \frac{p_t(K_t; p_0) e^{-\lambda k_t dt}}{(1 - p_t(K_t; p_0)) + p_t(K_t; p_0) e^{-\lambda k_t dt}},$$

where  $e^{-\lambda k_t dt} = \Pr(\text{no breakthrough in } (t, t+dt] | \text{no breakthrough by } t, \omega = 1)$ . The law of motion has a closed-form solution:

$$p_t(K_t; p_0) = \frac{1}{1 + \frac{1-p_0}{p_0} e^{\lambda K_t}}, \quad (1)$$

which, in the special case of constant investment  $k_\tau = 1$  for  $\tau \in [0, t]$  (so  $K_t = t$ ), reduces to  $p_t(t; p_0) = \frac{1}{1 + \frac{1-p_0}{p_0} e^{\lambda t}}$ .<sup>12</sup>

## 2.2 Compensation and Bargaining

The agent's investment is assumed to be observable and verifiable, so the agent's compensation can be conditioned on  $k_t$ . We assume that compensation is determined through short-term bargaining. At each instant of time, the principal makes a take-it-or-leave-it offer to reimburse a fraction of the agent's R&D cost  $C(k_t)$ , which is assumed be given by  $ak_t$ , where  $a > 0$ . More specifically, the principal's payment to the agent for the period  $[t, t+dt)$  is  $\phi_t a k_t$ , where  $\phi_t \in [0, 1]$  is the reimbursement rate. Because there is a time gap between the principal's offer and the agent's investment, these moves are sequential and therefore  $k_t$

---

<sup>12</sup> Interestingly, this Poisson process with uncertain binary hazard rates is equivalent to a Poisson process with a certain time-varying hazard rate given by  $\lambda k_t p_t(K_t; p_0)$ . To see this, let  $NB$  = "no breakthrough", then

$$\begin{aligned} \Pr(NB \text{ in } [t, t+dt] | NB \text{ by time } t) &= \frac{\Pr(NB \text{ by time } t+dt)}{\Pr(NB \text{ by time } t)} = \frac{p_0 e^{-\lambda K_t + dt} + (1-p_0)}{p_0 e^{-\lambda K_t} + (1-p_0)} \\ &= \frac{p_0 + (1-p_0) e^{\lambda K_t + dt}}{p_0 + (1-p_0) e^{\lambda K_t}} e^{-\lambda k_t dt} = \frac{p_t}{p_{t+dt}} e^{-\lambda k_t dt} = e^{-d \ln p_t} e^{-\lambda k_t dt} = e^{-\lambda p_t k_t dt}. \end{aligned}$$

The last equality uses  $d \ln p_t = -\lambda k_t (1 - p_t) dt$  which follows from the law of motion  $\frac{dp_t}{dt} = -\lambda k_t p_t (1 - p_t)$ .

can depend on  $\phi_t$ . Therefore, for any posterior probability  $p_t$  that the problem is solvable, the agent's (rate of flow of) instantaneous utility is

$$v_t = \lambda k_t p_t \Pi_A - a (1 - \phi_t) k_t, \quad (2)$$

while the principal's instantaneous utility is

$$u_t = \lambda k_t p_t \Pi_P - \phi_t a k_t. \quad (3)$$

Throughout the paper we maintain:

**Assumption 1**  $\lambda (\Pi_P + \Pi_A) \geq a$ .

Assumption 1 implies that the benefit-cost ratio for a problem known to be solvable is at least equal to 1.

### 2.3 Observable Actions

Because the principal moves first, we account for the beginning of the game by letting let  $0_{-1}$  denote almost the same calendar time as  $t = 0$  but with  $0_{-1} < 0$ . Denote by  $\mathcal{H}^t$  the set of all possible histories for the period  $\{0_{-1}\} \cup [0, t]$ , defined by

$$\mathcal{H}^t = \{h^t | h^t = (\phi^t, k^t), \phi^t = \{\phi_\tau | \tau \in \{0_{-1}\} \cup [0, t]\}, k^t \in \{k_{\tau'} | \tau' \in [0, t)\}\}$$

for  $t \in [0, \infty)$ .<sup>13</sup> For the special case  $t = 0_{-1}$ ,  $\mathcal{H}^{0_{-1}} = \{h^{0_{-1}} | h^{0_{-1}} = \phi_0 \in [0, 1]\}$ , where  $\phi_0$  is the principal's initial offer. The realized history of observable actions,  $h^t \in \mathcal{H}^t$  for  $t \in \{0_{-1}\} \cup [0, \infty)$  is public information.

The principal's action space  $\Phi_t$  at time  $t \geq 0$  is given by  $\Phi_t = \{\phi_{t+dt} | \phi_{t+dt} \in [0, 1]\}$ . The agent's action space  $\mathcal{K}_t$  at time  $t \geq 0$  is given by  $\mathcal{K}_t = \{k_t | k_t \in [0, 1]\}$ . Let  $\mathcal{K}_{0_{-1}} = \emptyset$ . Thus, prior to  $t = 0$ , only the principal makes a move, which is observable to the agent by time  $t = 0$ .

A pure strategy for the principal is denoted by  $s_P$ , while a pure strategy for the agent is denoted by  $s_A$ . A pure strategy profile  $s = (s_P, s_A)$  is a map

$$s : \{(h^t, p_0) | h^t \in \mathcal{H}^t, t \in \{0_{-1}\} \cup [0, \infty), p_0 \in [0, 1]\} \rightarrow \Phi_t \times \mathcal{K}_t,$$

such that  $s_P(h^t) \in \Phi_t$  and  $s_A(h^t, p_0) \in \mathcal{K}_t$  for all  $t \in [0, \infty)$ . Note, the principal's pure strategy  $s_P(h^t)$  does not depend on parameter  $p_0$ , which is not observable to her. In contrast, the agent's strategy is a function of  $p_0$ . Note that the history  $h^t \in \mathcal{H}^t$  at  $t \geq 0$  contains information on  $\phi_t$ , which is the offer for period  $[t, t + dt)$ , but the history  $h^t$  does not contain information on  $k_t$ , an asymmetry reflecting the sequential relation between  $\phi_t$  and  $k_t$ . The principal can commit to  $\phi_{t+dt}$  at time  $t$ , but the agent cannot commit to anything apart from

---

<sup>13</sup>Throughout the paper, we use subscript  $t$  to refer to actions or contract terms at a given point in time  $t$ , while we use superscript  $t$  to refer to histories of actions and contract terms up to time  $t$ .

playing the best responses. This asymmetry gives the principal a first-mover advantage. The agent's potential advantage is its hidden information about  $p_0$ .

Let  $\mathcal{S}_P$  and  $\mathcal{S}_A$  denote the set of all possible pure strategies of the principal and agent, respectively, and let  $\mathcal{S}$  denote the set of pure strategy profiles. (Thus,  $s_P \in \mathcal{S}_P$ ,  $s_A \in \mathcal{S}_A$ , and  $s \in \mathcal{S}$ .) In keeping with standard assumptions in the differential games literature, we impose the following regularity condition:<sup>14</sup>

**Assumption 2**  $\mathcal{S}_P$  and  $\mathcal{S}_A$  are such that  $s_P(h^t)$  and  $s_A(h^t, p_0)$  are continuous and differentiable with respect to  $t$  almost everywhere.

The learnable parameter space is  $\Theta = \{\theta = (\omega, p_0) \mid (\omega, p_0) \in \{0, 1\} \times [0, 1]\}$ . The principal's learning about  $\omega$  depends on observing a breakthrough, while her belief about  $p_0$  is inferred from observational data on history  $h^t$ . If a breakthrough occurs, the principal can infer that  $\omega = 1$ , but because the game ends, the ambiguity about  $p_0$  cannot be resolved. As the bargaining between the principal and agent occurs without a breakthrough, however, the principal can learn about  $p_0$  by observing the agent's action history from  $h^t$ .

The principal's inference rule involves a judgement about the equilibrium of the game. In principle, there could be multiple inference rules if there were multiple equilibrium outcomes or the data generating process involved off-equilibrium-path play. However, in this paper the data-generating mechanism is the play of the game along the equilibrium path of a pure strategy equilibrium. As a result, ambiguity arises solely because of multiple priors.

## 2.4 Limiting Case: Investment without Compensation

To shed light on the incentive issues in the model, we discuss the limiting case of investment when the principal does not compensate the agent for his effort. The principal would prefer that the agent invest no matter since solving the problem is a free option for the principal. By contrast, the agent's most preferred investment profile  $k^A(p_t)$  is given by the following proposition.

**Proposition 1** *In the absence of compensation from the principal, the agent's most preferred investment profile is given by*

$$k^A(p_t) = \begin{cases} 1 & \text{if } p_t \in (p^A, 1] \\ 0 & \text{if } p_t \in [0, p^A] \end{cases}, \quad (4)$$

where  $p^A$  is given by

$$p^A = \min \left\{ \frac{a}{\lambda \Pi_A}, 1 \right\} > 0.$$

This translates into a cumulative investment profile given by:

$$K^A(p_0) = \begin{cases} \frac{1}{\lambda} \ln \left[ \frac{\frac{(1-p^A)}{p^A}}{\frac{(1-p_0)}{p_0}} \right] & p_0 > p^A \\ 0 & p_0 \leq p^A \end{cases}, \quad (5)$$

---

<sup>14</sup>See Fudenberg and Tirole (1991), Chapter 13.

where the positive portion of  $K^A(p_0)$  is the solution for  $K$  to the equation  $p^A = \frac{1}{1 + \frac{1-p_0}{p_0} e^{\lambda K}}$ .

The agent's most preferred investment profile is maximum investment intensity as long as the posterior belief exceeds a belief cutoff, and no investment otherwise.

## 2.5 Short-term Bargaining with Full Information

Because the agent invests less than the principal prefers, the two parties may be able to achieve mutually beneficial gains if the principal compensates the agent for his effort. Here we sketch what the bargaining between the principal and agent would look like if the principal had perfect information about the prior  $p_0$ .<sup>15</sup>

To build intuition, consider first the case in which  $p_0 > p^A$ . The agent does not initially need to be compensated by the principal to be willing to invest. As long as a breakthrough is not achieved and the posterior  $p_t(K_t; p_0)$  exceeds  $p^A$ , then the principal does not offer reimbursement, i.e.,  $\phi_{t+dt} = 0$ . However, the moment that the posterior belief falls to  $p^A$ —or equivalently, the instant the agent's cumulative investment rises to  $K^A(p_0)$ —the agent must be reimbursed to be induced to invest. At that point, the principal will offer just enough reimbursement to make it worthwhile for the agent to invest. As we formalize below, the reimbursement rate that accomplishes this just satisfies the agent's stopping condition at  $t + dt$ :  $p_{t+dt}(K_{t+dt}; p_0) = \frac{a(1-\phi_{t+dt})}{\lambda \Pi_A}$ . If, despite the agent's investment, a breakthrough still does not occur,  $p_{t+dt}(K_{t+dt}; p_0)$  will keep falling and  $\phi_{t+dt}$  would need to rise to continue to motivate the agent to invest. However, this cannot continue indefinitely: the agent and the principal will reach the point at which the total surplus from continued trade falls to zero. At that point, additional reimbursement to the agent from the principal no longer can increase the principal's welfare. The belief cutoff corresponding to this point is

$$p^{**} = \frac{a}{\lambda(\Pi_A + \Pi_P)},$$

which, given Assumption 1 is less than or equal to 1. The corresponding reimbursement rate is

$$\phi^{**} = \frac{\Pi_P}{\Pi_A + \Pi_P}.$$

As we discuss in more detail in Section 5, this reimbursement rate has the property that it makes each party's share of cost equal to their share of benefits from a breakthrough.

If  $p^{**} < p_0 \leq p^A$ , the agent will immediately require reimbursement in order to invest. As described above, the reimbursement rate would need to be large enough to just satisfy the agent's stopping condition, but it would be bounded above by  $\phi^{**}$ . If  $p_0 \leq p^{**}$ , no trade would occur: at the highest reimbursement rate the principal would be willing to offer, the agent would be unwilling to invest.

---

<sup>15</sup>Because it is a special case of the more general analysis of bargaining under asymmetric information presented below, we defer the formal analysis of this bargaining until Section 5.

This discussion highlights the fundamental incentive problem when the agent knows  $p_0$  but the principal does not. If, for example, its actual  $p_0$  exceeds  $p^A$  the agent has the incentive to say otherwise. Such a claim would lead a “naive” principal to offer positive reimbursement even though the agent does not need it to be willing to invest. But, of course, this “story” is incomplete. In equilibrium, the agent’s investment behavior will reveal information about  $p_0$  to the principal, which the principal would presumably use to adjust its contract offers over time. How the principal learns from this observed behavior when it has multiple priors over  $p^0$  is a central theme of this paper. It is to this subject that we now turn.

### 3 Ambiguity in a Dynamic (Differential) Game of Incomplete Information

This section has two objectives. First, we develop the analytical approach for how a principal with multiple priors would learn over time. We do so by formulating the principal’s multiple priors over the agent’s private information  $p_0$  and then explaining the likelihood inference process that the principal uses to update her multiple prior beliefs. Second, we specify the principal’s objective function in light of the ambiguity she faces, and indicate what it means for the principal to follow an optimal strategy.

#### 3.1 Learning under Ambiguity

##### 3.1.1 The Principal’s Multiple Priors Over $p_0$

Each of the principal’s multiple priors over  $p_0$  is a function  $\mu_0 : \Sigma \rightarrow [0, 1]$ , where  $\Sigma$  is the Borel  $\sigma$ -algebra of subsets of  $[0, 1]$ . The set of all priors of the principal denoted by  $\mathcal{M}_0$ , is defined as  $\mathcal{M}_0 = \{\mu_0 | \mu_0 : \Sigma \rightarrow [0, 1]\}$ .

The principal’s knowledge about the process determining realized histories of the game is summarized by a set of likelihood functions

$$\mathcal{L} = \{l : \mathcal{H}^t \times \{t\} \times \mathcal{S} \times \{p_0 | p_0 \in [0, 1]\} \rightarrow [0, 1] | t \in \{0_{-1}\} \cup [0, \infty)\},$$

where  $l(\cdot; t, s, p_0)$  is a conditional probability measure<sup>16</sup> on  $\mathcal{H}^t$ . For example,  $l(h^t; t, s, p_0)$  is the probability of history  $h^t$  conditional on  $(t, s, p_0)$ . Each duple  $(\mu_0, l) \in \mathcal{M}_0 \times \mathcal{L}$  represents a theory, where  $l$  reflects the (structural) model specification, and  $\mu_0$  is a distribution on the value of the unknown parameter  $p_0$ . If  $\mu_0$  is a Dirac measure (i.e., an indicator function) given by

$$\delta_{p_0}(A) = \begin{cases} 1 & \text{if } p_0 \in A \\ 0 & \text{otherwise} \end{cases}, \forall A \subseteq [0, 1],$$

---

<sup>16</sup>The likelihood  $l(\cdot; t, s, p_0)$  is expressed as a probability over the observation space  $\mathcal{H}^t$ ; it is NOT a probability over the parameter space  $[0, 1]$ . Thus,  $l(h^t; t, s, p_0)$  does not obey the laws of probability when  $h^t$  is fixed and  $p_0$  is treated as a variable, e.g., the integration over  $p_0$  does not add up to one. That is why we use the term “likelihood” as opposed to “probability.”

then  $\mu_0 = \delta_{p_0}$  represents a unique specification of the parameter value.

One of the arguments of the likelihood is the strategy profile  $s$ . The likelihood is plausible only if  $s$  is plausible. In principle, the likelihood could reflect both “signal” (which is the equilibrium of the game) and “noise” (which could include measurement error or out-of-equilibrium play). We abstract from all noise components, so  $\mathcal{L}$  is a singleton and consists of a unique likelihood function  $l$  for any given  $(t, s, p_0)$ . Multiplicity of likelihood only arises from multiplicity of  $s \in \mathcal{S}$ . If the data generating mechanism is a pure strategy equilibrium, i.e.,  $s = (s_P, s_A)$ , then  $s$  determines an equilibrium path for each given  $p_0 \in [0, 1]$ , and thus,  $s$  gives rise to a unique conditional likelihood  $l(\cdot; \cdot, s, \cdot)$ . Only a unique conditional likelihood function  $l(\cdot; \cdot, s, \cdot)$  would then pass the likelihood ratio test (to be specified presently) at each point of time.

Denote by  $\mathcal{L}_0(s)$  the restriction of  $\mathcal{L}$  to  $s$ .  $\mathcal{L}_0(s)$  is a singleton of conditional likelihood, and  $l(\cdot; \cdot, s, \cdot) \in \mathcal{L}_0(s)$  represents the unique pure strategy equilibrium the principal and agent play.

### 3.1.2 Learning and Likelihood Inference — Iterative Selection of Priors and Updating

The learning process modeled here resembles the approach in Epstein and Schneider (2007). At each point in time, each theory duple  $(\mu_0, l) \in \mathcal{M}_0 \times \mathcal{L}_0(s)$  competes with other members to better explain the history  $h^t$ . In the absence of a breakthrough, the principal forms posteriors based on the realized history  $h^t$ . The principal (iteratively) selects the subset of prior distributions, each of which has to exceed a critical value of likelihood of generating the history  $h^t$ . Denote by  $\mu_t(\cdot | h^t; \mu_0, l)$  the conditional probability measure, which is calculated by Bayes’ Rule,

$$d\mu_t(\cdot | h^t; \mu_0, l) = \frac{l(h^t; t, s, \cdot) d\mu_0(\cdot)}{\int_0^1 l(h^t; t, s, \tilde{p}_0) d\mu_0(\tilde{p}_0)}, \quad (6)$$

where

$$L(h^t; t, s, \mu_0, l) = \int_0^1 l(h^t; t, s, \tilde{p}_0) d\mu_0(\tilde{p}_0) \quad (7)$$

is the (generalized) likelihood (over the distribution  $\mu_0$  as opposed to over  $p_0$ ). It follows that

$$\mu_t(A | h^t; \mu_0, l) = \frac{\int_{p_0 \in A} l(h^t; t, s, p_0) d\mu_0(p_0)}{L(h^t; t, s, \mu_0, l)} \quad (8)$$

for all  $A \in \Sigma$ . Note, the (variable) argument of the functional  $L(h^t; t, s, \mu_0, l)$  is  $\mu_0$ , while  $h^t$  is fixed. In general, a hypothesis is expressed as a distribution  $\mu_0$  instead of a single parameter value  $p_0$ . In the special case that  $\mu_0(\tilde{p}_0)$  degenerates to a Dirac measure, i.e.,  $\mu_0(\tilde{p}_0) = \delta_{p_0}(\tilde{p}_0)$ , then

$$L(h^t; t, s, \mu_0, l) = \int_0^1 l(h^t; t, s, \tilde{p}_0) d\delta_{p_0}(\tilde{p}_0) = l(h^t; t, s, p_0).$$

From (7) and (8) it is apparent that conditional probability measure  $\mu_t(\cdot|h^t; \mu_0, l)$  is well defined only if  $L(h^t; t, s, \mu_0, l) > 0$ ; i.e., the theory duple  $(\mu_0, l)$  is not contradicted by the observed history  $h^t$ . In principle, there can be a continuum of competing theories (hypotheses) that are candidate explanations for the observed history  $h^t$ . Likelihood inference can play a useful role to quantitatively assess the relative merits of these competing hypotheses.<sup>17</sup> Fixing the history  $h^t$  and underlying strategy profile  $s$ , the likelihood ratio between hypotheses  $(\mu_0, l)$  and  $(\tilde{\mu}_0, \tilde{l})$  is given by

$$\frac{L(h^t; t, s, \mu_0, l)}{L(h^t; t, s, \tilde{\mu}_0, \tilde{l})}.$$

We say that  $(\mu_0, l)$  has more support from the data than  $(\tilde{\mu}_0, \tilde{l})$  if this ratio exceeds 1.

Based on pairwise likelihood ratios, the principal can discriminate among all priors in  $\mathcal{M}_0$ . To formulate the likelihood inference procedure which is used by the principal, let  $\alpha \in [0, 1]$  be the critical value of the likelihood ratio test the principal uses. Let the set of all accepted posteriors against  $h^t$  be given by

$$\mathcal{M}_t^\alpha(h^t) = \left\{ \begin{array}{l} \mu_t(h^t; \mu_0, l) \mid \mu_0 \in \mathcal{M}_0, l \in \mathcal{L}_0(s), \\ L(h^t; t, s, \mu_0, l) \geq \alpha \max_{\substack{\tilde{\mu}_0 \in \mathcal{M}_0 \\ \tilde{l} \in \mathcal{L}_0(s)}} L(h^t; t, s, \tilde{\mu}_0, \tilde{l}) \end{array} \right\}. \quad (9)$$

Intuitively, the principal admits a posterior probability measure  $\mu_t(h^t; \mu_0, l)$  if and only if  $\mu_t(h^t; \mu_0, l)$  is a Bayesian update of  $\mu_0 \in \mathcal{M}_0$  such that there exists  $l \in \mathcal{L}_0(s)$  and the likelihood of  $(\mu_0, l)$  — given by  $L(h^t; t, s, \mu_0, l)$  — is at least a fraction  $\alpha$  of the possible maximum likelihood over all  $(\tilde{\mu}_0, \tilde{l}) \in \mathcal{M}_0 \times \mathcal{L}_0(s)$ . Note that the special case of  $\alpha = 0$  corresponds to Full Bayesian Updating (FBU):

$$\mathcal{M}_t^{FBU}(h^t) = \{\mu_t(h^t; \mu_0, l) \mid \mu_0 \in \mathcal{M}_0, l \in \mathcal{L}_0(s)\}. \quad (10)$$

By contrast, the special case of  $\alpha = 1$  corresponds to Maximum Likelihood Updating (MLU)<sup>18</sup>:

$$\mathcal{M}_t^{MLU}(h^t) = \left\{ \begin{array}{l} \mu_t(h^t; \mu_0, l) \mid \mu_0 \in \mathcal{M}_0, l \in \mathcal{L}_0(s), \\ L(h^t; t, s, \mu_0, l) = \max_{\substack{\tilde{\mu}_0 \in \mathcal{M}_0 \\ \tilde{l} \in \mathcal{L}_0(s)}} L(h^t; t, s, \tilde{\mu}_0, \tilde{l}) \end{array} \right\}. \quad (11)$$

In principle, the choice of critical value  $\alpha \in [0, 1]$  may potentially introduce an *ad hoc* subjective element into the inference rule. This turns out not to be the case in our game-theoretic context, since the inference is entirely deductive. As has been established previously, for  $\mu_t(h^t; \mu_0, l)$  to be well defined,  $L(h^t; t, s, \mu_0, l) > 0$  is necessary. Consequently, for

<sup>17</sup>For an authoritative account on likelihood inference, see Edwards (1992) and Hacking (1965).

<sup>18</sup>The terms FBU and MLU are borrowed from Gilboa and Marinacci (2011), which is the most recent comprehensive survey of the ambiguity literature.

$\mathcal{M}_t^\alpha(h^t)$  to be well defined,  $\max_{\substack{\tilde{\mu}_0 \in \mathcal{M}_0 \\ \tilde{l} \in \mathcal{L}_0(s)}} L\left(h^t; t, s, \tilde{\mu}_0, \tilde{l}\right) > 0$  is necessary; i.e., there must exist some  $(\tilde{\mu}_0, \tilde{l}) \in \mathcal{M}_0 \times \mathcal{L}_0(s)$  that is not contradicted by the history  $h^t$ .

We are interested in learning through inferences based on a pure strategy equilibrium of the game. For a given pure strategy equilibrium, a unique likelihood function can be established because a unique equilibrium path is determined for each realization of the state variable  $p_0$ . That is, there exists a (forecast) map  $f : \mathcal{S} \times \{p_0\} \rightarrow \mathcal{H}^\infty$  such that  $\forall (s, p_0) \in \mathcal{S} \times \{p_0\}$ ,  $f(s, p_0) = h^\infty \in \mathcal{H}^\infty$  predicts the entire equilibrium path of the game (conditional on the absence of a breakthrough). Let  $f^t(s, p_0)$  be the restriction of  $f(s, p_0)$  to the set  $\{0_{-1}\} \cup [0, t]$ . Therefore,  $f^t(s, p_0)$  predicts the on-equilibrium-path history for the period  $\{0_{-1}\} \cup [0, t]$ .

Define the (partial) identification correspondence, denoted by  $\mathcal{I}_d$ , such that<sup>19</sup>

$$\mathcal{I}_d(h^t) = \{p_0 \in [0, 1] \mid f^t(s, p_0) = h^t\}. \quad (12)$$

The correspondence  $\mathcal{I}_d : \mathcal{H}^t \rightarrow \{p_0\}$  maps any given on-equilibrium-path history to a subset of agent types. Since the agent type subset is rarely a singleton, the identification is usually partial. The next theorem states the updating of the principal's belief through pure strategy equilibrium (strategy) profile-based likelihood inference.

**Theorem 1** *Suppose  $s$  is a pure strategy equilibrium (strategy) profile. Then  $\mathcal{L}_0(s) = \{l\}$  is a singleton such that*

$$l(h^t; t, s, p_0) = \delta_{f^t(s, p_0)}(h^t) = \begin{cases} 1 & \text{if } h^t = f^t(s, p_0) \\ 0 & \text{otherwise} \end{cases}, \quad (13)$$

where  $\delta_{f^t(s, p_0)}(h^t)$  is a Dirac measure. For all on-equilibrium-path history  $h^t$ , for all  $\alpha \in [0, 1]$ .

$$\mathcal{M}_t^\alpha(h^t) = \mathcal{M}_t^{MLU}(h^t) = \{\mu_0 \in \mathcal{M}_0 \mid \mu_0(\mathcal{I}_d(h^t)) = 1\}. \quad (14)$$

It is interesting to note, by (9)  $\mathcal{M}_t^\alpha(h^t)$  is defined as a set of posterior distributions, which are the outcomes of applying Bayes' Theorem. According to (14),  $\mathcal{M}_t^\alpha(h^t)$  appears to be a set of prior distributions, which is the subset of  $\mathcal{M}_0$  that passes certain (iterative) selection. This comparison reveals the fact that the effect of Bayesian updating on beliefs is duplicated by the effect of selecting prior distributions by the maximum likelihood test. In this result, we see a seamless synthesis of the ideas from the Bayesian, frequentist and Fisherian schools of statistics, under the condition that the set of ambiguous priors  $\mathcal{M}_0$  is sufficiently large (inclusive).

The fact that the endogenous critical value of the likelihood ratio test can be set to  $\alpha = 1$  is because in a pure strategy equilibrium, in the absence of measurement error, the principal

---

<sup>19</sup>The term partial identification is borrowed from Manski (1995). In the current context, the pure strategy equilibrium (strategy) profile-based inferential problem is abstracted from statistical inference problem, therefore is purely a problem of identification *a la* Manski (1995).

only makes deductive inferences which rule out all possibilities that are contradicted by the observational evidence, but do not rule out any possibility that is consistent with such evidence.<sup>20</sup> In the language of classical hypothesis-testing theory, the probability of type I error is set to be zero. As a matter of fact, the updating from  $\mathcal{M}_0$  to  $\mathcal{M}_t^\alpha(h^t)$  is a truncation of the support of  $\mathcal{M}_0$  (which is  $[0, 1]$ ) to  $\mathcal{I}_d(h^t)$ .  $\mathcal{M}_t^\alpha(h^t)$  admits all probability measures whose supports lie within  $\mathcal{I}_d(h^t)$ . When  $\alpha = 1$  is chosen, the probability of type II error is uniformly minimized but still positive.<sup>21</sup> The fact (as revealed in the proof of Theorem 1) that  $\alpha \in [0, 1]$  is sufficient for representing deductive inference implies that inferential correctness is not sensitive to the choice of critical value  $\alpha$ . Without loss of generality, in the remainder of the paper, we confine our analysis to the case in which  $\alpha = 1$ . Hereafter, we omit reference to the value of  $\alpha$  and use the notation  $\mathcal{M}_t(h^t)$  to represent  $\mathcal{M}_t^1(h^t) = \mathcal{M}_t^{MLU}(h^t)$ .

So far we have only dealt with on-equilibrium-path histories. For any off-equilibrium path history  $h^t$ , by definition we must have  $\mathcal{I}_d(h^t) = \emptyset$ , which implies  $\mu_0(\mathcal{I}_d(h^t)) = 0$  for all  $\mu_0 \in \mathcal{M}_0$ . That is, the pure strategy equilibrium (strategy) profile-based identification process must fail if the history is off equilibrium path. To complete the updating of beliefs for off-equilibrium path histories, we propose the following definition.

**Definition 1** *If  $l(h^t; t, s, p_0) = 0$  for all  $\mu_0 \in \mathcal{M}_0$  and  $l \in \mathcal{L}_0(s)$  (i.e.,  $\mathcal{I}_d(h^t) = \emptyset$ ), then  $\mathcal{M}_t(h^t)$  is defined by*

$$\mathcal{M}_t(h^t) = \mathcal{M}_0.$$

This definition captures the idea that the identification failure (i.e.,  $\mathcal{I}_d(h^t) = \emptyset$ ) indicates that the data  $h^t$  is generated by off-equilibrium path behavior, which implies that the pure strategy equilibrium (strategy) profile-based likelihood function is not valid for explaining the data. As a result, learning based on the invalid likelihood function should be undone, and the beliefs should reverse to the initial priors. In this case, all competing priors are contradicted by data  $h^t$ , and no prior prevails.

It follows from Theorem 1 and Definition 1 that the set of posterior beliefs is a subset of the set of all possible priors, i.e.,  $\mathcal{M}_t(h^t) \subseteq \mathcal{M}_0$ . Therefore the set of all posteriors is essentially a set of all priors that pass the maximum likelihood test. A prior  $\mu_0$  is rejected by  $h^t$  if and only if  $\mu_0 \in \mathcal{M}_0 \setminus \mathcal{M}_t(h^t)$ . If  $\mu_0$  is rejected, the posterior for  $\mu_0$  is not well-defined, therefore discarded.

As noted at the outset of this section, our model of learning about the ambiguous parameter is similar to Epstein and Schneider (2007). However, because learning in our model is embedded in a game theoretic interaction, our analysis differs from Epstein and Schneider (2007) in several important ways. For one thing, the likelihood function in our model is derived endogenously from the equilibrium of the game. In addition,  $\mathcal{L}(s)$  is a singleton for a given pure strategy equilibrium (strategy) profile in our model. Finally, while  $\alpha$  can

---

<sup>20</sup>Since the inferential problem pertinent to our model is confined to identification problem, it turns out that deductive inference is adequate for the task. Therefore our approach is in spirit close to Keynes's logical approach to probability (as a measure of rational beliefs), which is much concerned with the validity of inference (see Keynes 1921).

<sup>21</sup>Equivalently, the test is uniformly-most-powerful (U.M.P.).

be interpreted as a parameter reflecting the decision maker’s aversion to type II error in hypothesis testing (i.e.,  $\alpha = 1$  means maximum aversion, and  $\alpha = 0$  means maximum tolerance to type II errors), and thus ideally it should be explicitly modelled as attitude towards ambiguity, this issue does not arise in our analysis. This is because the value of  $\alpha$  does not affect the (deductive) inferential outcome in pure strategy equilibrium in our model, and we can simply assign  $\alpha = 1$  without loss of generality.

### 3.2 Principal’s Objective and Best Response

In Bayesian games in which a player has a prior belief about the environment, the specification of the player’s objective function is straightforward: conditional on the history of play, the player forms an expectation of the function it seeks to maximize based on the posterior beliefs implied by that history. However, if the player does not have a unique prior, the formulation of the player’s objective is more complex. Our specification of the principal’s objective function is inspired by Manski (2008) and GMMS (2010).

These papers deal with the situation faced by the principal in our model: how does one model decision making under uncertainty when the decision maker lacks the information to quantify uncertainty using a single probability measure? Manski (2008) formulates a two-step procedure in which the first step is to eliminate all weakly-dominated actions, and the second step is to maximize (over non-dominated actions) a not-uniquely specified utility function (which could be minimum, minimum-regret or expected utility function). GMMS (2010) axiomatize the problem of a decision maker who has a pair of preference relations: objectively rational preferences and subjectively rational preferences. If the decision maker chooses based on objectively rational preferences, he can defend his choice to others; if the decision maker chooses based on subjectively rational preferences, he cannot be convinced by others that his choice was wrong. Objectively rational preferences generate a unanimous but incomplete ordering of actions, while subjectively rational preferences generate a complete ordering of actions that can be represented by a minimum expected utility function (with respect to all priors in the set of the decision maker’s possible priors). GMMS (2010) demonstrate that given two plausible conditions (*consistency* and *caution*), there exists a common set of priors (that can be justified by a given set of inference rules) that enable the decision maker’s choices to be represented either by objectively rational preferences or by subjectively rational preferences. This provides a foundation for decision making based on the max-min rule. Basing decisions on a minimum expected utility function can be thought of as a way of completing an otherwise incomplete preference ordering based on objective rationality.

Our formulation applies GMMS (2010) to a dynamic game-theoretic setting. To develop this formulation, we define, derive, and characterize the principal’s minimum expected utility function—which we call the *worst-case value function*. To begin, note that the principal’s instantaneous conditional expected utility (conditional on  $(p_\tau(K_\tau; p_0), s, p_0)$ ) is given by

$$u_\tau(p_\tau(K_\tau; p_0) | s, p_0) = \lambda k_\tau(p_0) p_\tau(K_\tau; p_0) \Pi_P - \phi_\tau a k_\tau(p_0).$$

The principal's conditional value function can thus be expressed as

$$\begin{aligned} & W_t(p_t(K_t; p_0) | s, p_0) \\ &= \int_t^\infty u_\tau(p_\tau(K_\tau; p_0) | s, p_0) e^{-\lambda \int_t^\tau p_\psi k_\psi d\psi} e^{-r(\tau-t)} d\tau \\ &= u_t(p_t(K_t; p_0) | s, p_0) dt + e^{-(\lambda p_t k_t + r) dt} W_{t+dt}(p_{t+dt}(K_{t+dt}; p_0) | s, p_0), \end{aligned} \quad (15)$$

where  $e^{-\lambda \int_t^\tau k_\psi p_\psi d\psi}$  is the probability that no breakthrough occurs in the time interval  $[t, \tau]$  for  $\tau \geq t$  conditional on no breakthrough by time  $t$ . We can now define the principle's worst-case value function:

**Definition 2** *In the absence of a breakthrough prior to time  $t$ , the principal's worst-case value function  $U_t(h^t; s)$  is determined by the plausible posterior that, for a given pure strategy equilibrium  $s = (s_P, s_A)$ , minimizes the principal's conditional value; i.e.,*

$$\begin{aligned} & U_t(h^t; s) \\ &= \min_{\mu_0 \in cl(\mathcal{M}_t(h^t))} \int_0^1 W_t(p_t(K_t; p_0) | s, p_0) d\mu_0(p_0), \end{aligned} \quad (16)$$

where  $cl(X)$  denotes the closure of a set  $X$  (i.e., the smallest closed superset of  $X$ , or the set of all limit points of  $X$ ), and the definition of  $cl(\mathcal{M}_t(h^t))$  is based on convergence in distribution (CDF).

We can immediately establish the following equivalences:

**Theorem 2** *The principal's worst-case value function has the following equivalent expressions:*

- (i)  $U_t(h^t; s) = \min_{\mu_0 \in cl(\mathcal{M}_t(h^t))} \int_0^1 W_t(p_t(K_t; p_0) | s, p_0) d\mu_0(p_0);$
- (ii)  $U_t(h^t; s) = \min_{\delta_{p_0} \in cl(\mathcal{D}_t(h^t))} W_t(p_t(K_t; p_0) | s, p_0);$
- (iii)  $U_t(h^t; s) = \min_{p_0 \in cl(P(\mathcal{M}_t(h^t)))} W_t(p_t(K_t; p_0) | s, p_0),$

where  $\mathcal{D}_t$  is the largest Dirac subset of  $\mathcal{M}_t(h^t)$ ,  $P(\mathcal{M}_t(h^t))$  is the support of  $\mathcal{M}_t(h^t)$  and the definition of  $cl(\mathcal{D}_t(h^t))$  is based on CDF.

The significance of this result is that in describing the principal's worst-case value function we do not need worry about minimum *expected* utility. Instead, the worst case value function can be described by the minimum of the principal's *actual utility*  $W_t(p_t(K_t; p_0) | s, p_0)$  with respect to the values of  $p_0$  consistent with the principal's likelihood inferences given the history of play through time  $t$ .

To characterize the principal's best response conditional on her set of accepted posteriors, we begin by defining what it means for one pure strategy profile to weakly dominate another from the principal's perspective:

**Definition 3** *For a pair of pure strategy profiles  $(s_P, s_A), (\hat{s}_P, \hat{s}_A) \in \mathcal{S}_P \times \mathcal{S}_A$ , the profile  $(s_P, s_A)$  (weakly) dominates  $(\hat{s}_P, \hat{s}_A)$  conditional on  $\mathcal{M}_t(h^t)$ —denoted by  $(s_P, s_A) \succ_{\mathcal{M}_t(h^t)}^* (\hat{s}_P, \hat{s}_A)$ —iff*

$$\int_0^1 W_t(p_t(K_t; p_0) | s_P, s_A, p_0) d\mu_0(p_0) \geq \int_0^1 W_t(p_t(K_t; p_0) | \hat{s}_P, \hat{s}_A, p_0) d\mu_0(p_0),$$

for all  $\mu_0 \in cl(\mathcal{M}_t(h^t))$  and

$$\int_0^1 W_t(p_t(K_t; p_0) | s_P, s_A, p_0) d\mu_0(p_0) > \int_0^1 W_t(p_t(K_t; p_0) | \hat{s}_P, \hat{s}_A, p_0) d\mu_0(p_0),$$

for some  $\mu_0 \in cl(\mathcal{M}_t(h^t))$ .

We can represent conditional weak dominance in several equivalent ways:

**Proposition 2** For a given  $\mathcal{M}_t(h^t)$  and  $\forall (s_P, s_A), (\hat{s}_P, \hat{s}_A) \in \mathcal{S}_P \times \mathcal{S}_A$ , the following three statements are equivalent:

- (i)  $(s_P, s_A) \succ_{\mathcal{M}_t(h^t)}^* (\hat{s}_P, \hat{s}_A)$ .
- (ii)

$$W_t(p_t(K_t; p_0) | s_P, s_A, p_0) \geq W_t(p_t(K_t; p_0) | \hat{s}_P, \hat{s}_A, p_0), \forall \delta_{p_0} \in cl(\mathcal{D}_t(h^t)), \text{ and}$$

$$W_t(p_t(K_t; p_0) | s_P, s_A, p_0) > W_t(p_t(K_t; p_0) | \hat{s}_P, \hat{s}_A, p_0), \text{ for some } \delta_{p_0} \in cl(\mathcal{D}_t(h^t)).$$

(iii)

$$W_t(p_t(K_t; p_0) | s_P, s_A, p_0) \geq W_t(p_t(K_t; p_0) | \hat{s}_P, \hat{s}_A, p_0), \forall p_0 \in cl(P(\mathcal{M}_t(h^t))), \text{ and}$$

$$W_t(p_t(K_t; p_0) | s_P, s_A, p_0) > W_t(p_t(K_t; p_0) | \hat{s}_P, \hat{s}_A, p_0), \text{ for some } p_0 \in cl(P(\mathcal{M}_t(h^t))).$$

Proposition 2 tells us that the conditional weak dominance relation requires a consistent ordering of the principal's welfare for *all* possible beliefs that are consistent with her likelihood inferences. For example, Condition (iii) says that a pure strategy profile  $(s_P, s_A)$  weakly dominates another profile  $(\hat{s}_P, \hat{s}_A)$  if the principal's welfare under  $(s_P, s_A)$  is at least as large as it is under  $(\hat{s}_P, \hat{s}_A)$  for *all* values of  $p_0$  consistent with the principal's likelihood inferences and strictly greater for some values of  $p_0$ .

We can now define the principal's set of pure strategy best responses:

**Definition 4** For a given  $\mathcal{M}_t(h^t)$  and  $s_A \in \mathcal{S}_A$ , the principal's set of all best responses conditional on  $\mathcal{M}_t(h^t)$  is given by

$$\mathcal{S}_P^*(s_A; \mathcal{M}_t(h^t)) = \left\{ s_P \in \mathcal{S}_P \mid \begin{array}{l} s_P \in \arg \max_{s'_P \in \mathcal{S}_P} U_t(h^t; s'_P, s_A) \text{ and} \\ (\hat{s}_P, s_A) \not\succ_{\mathcal{M}_t(h^t)}^* (s_P, s_A), \forall \hat{s}_P \in \mathcal{S}_P \end{array} \right\}.$$

This definition entails that, informally speaking,  $s_P$  should not be conditionally weakly dominated by any pure strategy profile in  $\mathcal{S}_P$ . Given the agent's strategy  $s_A \in \mathcal{S}_A$  and the principal's belief represented by  $\mathcal{M}_t(h^t)$ , the principal's objective is to play her best response  $s_P \in \mathcal{S}_P^*(s_A; \mathcal{M}_t(h^t))$ . The principal's preferences over alternative strategies are determined by the worst-case-scenario performance of these strategies. A (continuation) strategy can be optimal only if it yields the best worst-case-scenario performance. If there are multiple

strategies that have the best worst-case-scenario performance, only the non-weakly dominated strategies can be optimal.<sup>22</sup> Overall, the principal's preferences are lexicographical.<sup>23</sup>

Before turning to an analysis of equilibrium, we specify the agent's value function. Using (2), and allowing for arbitrary (as opposed to optimal) investment  $k_\tau \in [0, 1]$ , the agent's value function is given by

$$\begin{aligned} V_t(p_t(K_t; p_0) | s, p_0) &= \int_t^\infty v_\tau(p_\tau(K_\tau; p_0) | s, p_0) e^{-\lambda \int_t^\tau k_\psi p_\psi d\psi} e^{-r(\tau-t)} d\tau \\ &= v_t(p_t(K_t; p_0) | s, p_0) dt + e^{-(\lambda k_t p_t + r) dt} V_{t+dt}(p_{t+dt}(K_{t+dt}; p_0) | s, p_0). \end{aligned} \quad (17)$$

## 4 Solution Concept

Our solution concept is an extension of the familiar perfect Bayesian equilibrium (PBE) to a setting in which the uninformed player has multiple priors. We call it the *perfect objectivist equilibrium* (POE). The term 'objectivist' is to emphasize the dominant role of objective inference, i.e., likelihood inference, in learning under ambiguity, as is modelled in the current paper, in contrast to Bayesian inference that underlines the PBE.

**Definition 5** A POE of the game is a tuple  $(s, \mathcal{M})$ , where  $\mathcal{M}$  represents the beliefs of the principal; i.e.,

$$\mathcal{M} = \{\mathcal{M}_t(h^t) | h^t \in \mathcal{H}^t, t \in \{0_{-1}\} \cup [0, \infty)\},$$

and the pure strategy profile  $s = (s_P, s_A)$  is a map

$$s : \{(h^t, p_0) | h^t \in \mathcal{H}^t, t \in \{0_{-1}\} \cup [0, \infty), p_0 \in [0, 1]\} \rightarrow \Phi_t \times \mathcal{K}_t,$$

such that,

(i) for any time  $t \in \{0_{-1}\} \cup [0, \infty)$  and realized history  $h^t$ , given the conditional posteriors  $\mathcal{M}_t(h^t)$ , the continuation strategies derived from  $s_A$  and  $s_P$  are mutual best responses. That is, for all history  $h^t \in \mathcal{H}^t$ , all  $t \in \{0_{-1}\} \cup [0, \infty)$ ,

$$V_t(p_t(K_t; p_0) | s_P, s_A, p_0) \geq V_t(p_t(K_t; p_0) | s_P, s'_A, p_0)$$

for all  $s'_A \in \mathcal{S}_A$  for all  $p_0 \in [0, 1]$ , and

$$s_P \in \mathcal{S}_P^*(s_A; \mathcal{M}_t(h^t)).$$

(ii) The initial set of posteriors is given by  $\mathcal{M}_{0_{-1}}(h^{0_{-1}}) = \mathcal{M}_0(h^0) = \mathcal{M}_0$ . The belief updating is through maximum likelihood inference as described in Section 3.1.2. Specifically, (a)  $\mathcal{L}_0(s)$ , the set of likelihood functions restricted to strategy profile  $s$ , is a singleton; (b) for all  $t \in \{0_{-1}\} \cup [0, \infty)$ , the set of all accepted posteriors  $\mathcal{M}_t(h^t)$  against any on-equilibrium-path history  $h^t$  is derived from maximum likelihood inference based on  $l \in \mathcal{L}_0(s)$  and  $\mathcal{M}_t(h^t) = \mathcal{M}_0$  off the equilibrium path.

<sup>22</sup>In the Appendix, we provide an example of a weakly dominated strategy. However, to understand the example, one must first understand the equilibrium analysis developed below. For this reason, the example is presented immediately after the proof of Theorem 3.

<sup>23</sup>See Section 6.4 for a further discussion on lexicographical (overall) preferences.

In Definition 5, condition (i) requires that given the posteriors for any given history (either on or off equilibrium path), all continuation strategies of the agent must be best responses to the principal's continuation strategies, and subject to this constraint, the continuation strategies of the principal must be best response too. The above condition for the principal must also hold for the special case: At  $t = 0_{-1}$ ,  $s_P(h^{0_{-1}}) = \phi_0 \in [0, 1]$ ,

$$s_P \in \mathcal{S}_P^*(s_A; \mathcal{M}_{0_{-1}}(h^{0_{-1}})).$$

Condition (ii) requires that belief updating follows the “likelihood inference” described in Section 3.1.2. In particular, it requires that for any history that is off the equilibrium path, every plausible equilibrium-based prediction of the play of the game is contradicted by the data, and the set of all posteriors reverts to equal the set of all priors.

Our definition of POE mirrors the definition of PBE, but it differs from the definition of the PBE in two ways. First, in the POE the principal's strategy cannot be conditionally weakly dominated (in the sense defined above), reflecting the principal's ambiguity aversion. By contrast, in a PBE, the principal's equilibrium strategy maximizes its expected utility function given the posterior beliefs that follow from its unique prior beliefs through Bayesian updating. Second, this Bayesian updating is replaced in the POE by likelihood inference on a set of multiple priors.

In the extreme case in which the initial set of priors is a singleton, the POE degenerates to a PBE. In this sense, the POE can nest the PBE as a limiting case. The POE can therefore be seen as an extension of the PBE to accommodate ambiguous beliefs. This extension is useful when an uninformed player lacks confidence in any single-prior belief. POE allows the solution concept not to hinge on the any subjective single-prior belief and therefore to be robust to subjective belief specification. The POE, however, differs from the belief-free equilibrium discussed in Section 6.2 in that it keeps track of beliefs and demands that equilibrium beliefs and strategies be mutually consistent after any history while the belief-free equilibrium does not.

In the POE, in principle the strategies of the players can be formulated as closed-loop strategies. In general a closed-loop strategy may depend on the full history  $h^t$  at time  $t$ . In a subset of POE of this model, for  $t \geq 0$  the players' strategies are Markovian. Furthermore, there exists an explicit set of (payoff-relevant) state variables that sufficiently summarize the payoff-relevant part of history  $h^t$ . They include  $K_t, \phi_t$ , which are known to both players. They also include a state variable  $q_t$ , which summarizes the principal's knowledge about the agent type and is common knowledge between the principal and agent. This state variable (which will be defined in section 5.2), is also known to both players. In addition, the agent knows  $p_0$ , which is his private information. Therefore, the players' strategies can be explicitly expressed as Markov strategies. We call each POE in this subset a Markov perfect objectivist equilibrium (MPOE), i.e., a POE in which all players (and player types) use Markov strategies.

We conclude this section by defining a Pareto dominant equilibrium outcome:

**Definition 6** *A Pareto dominant equilibrium outcome is an equilibrium outcome which Pareto dominates all other (different) equilibrium outcomes, that is, in comparison with any*

other (different) equilibrium outcome, it does not make the principal or any type of the agent worse off; and it makes either the principal or at least one type of the agent strictly better off. “To make the principal worse off” means, from the principal’s perspective, the outcome is either weakly conditionally dominated by, or has lower worst-case value than, the outcome in comparison. “To make the principal strictly better off” means, from the principal’s perspective, the outcome weakly conditionally dominates the outcome in comparison.

## 5 Characterization of the MPOE

In this section, we focus on MPOE where the payoff relevant history can be summarized by an array of state variables  $(K_t, \phi_t, q_t)$ , where  $q_t$  represents the commonly known estimate of the agent’s private type,  $p_0$ . Before we formally define  $q_t$ , let us first consider a benchmark case in which the principal knows the agent’s type, in other words,  $q_t = p_0$  for all  $t$ .

### 5.1 The Benchmark: Principal Knows $p_0$

In this case, the principal’s strategy is represented by  $\phi_{t+dt}(K_t; p_0)$  and the agent’s,  $k_t(K_t, \phi_t; p_0)$ , and there is no ambiguity over the agent’s type  $p_0$ . From the first-order approximation of (17), we can write the agent’s Bellman equation as

$$rV = \frac{\partial V}{\partial \phi_t} \frac{d\phi_t}{dt} + \max_{k_t \in [0, 1]} \left\{ \left[ -a(1 - \phi_t) + \lambda p_t \Pi_A - \lambda p_t V + \frac{\partial V}{\partial K_t} \right] k_t \right\}, \quad (18)$$

where  $p_t = p_t(K_t; p_0)$  is the agent’s posterior belief.

Since the expression in the curly bracket in (18) is linear in  $k_t$ , the optimal  $k_t$  must satisfy the following condition:

$$\tilde{k}_t^*(K_t, \phi_t; p_0) = \begin{cases} 1 & \text{if } p_t(K_t; p_0) \geq \frac{a(1 - \phi_t) - \frac{\partial V}{\partial K_t}}{\lambda(\Pi_A - V)} \\ 0 & \text{otherwise,} \end{cases} \quad (19)$$

i.e., the agent’s optimal investment intensity  $k_t$  only takes values 1 or 0. Given the fact that the principal uses a Markovian strategy which does not directly depend on time, there is no incentive for the agent to delay investment in order to benefit from an improved reimbursement rate because delay can only alter time but not the state variables. As a result,  $k_t = 0$  means termination of investment altogether, which occurs if and only if  $V_t = 0$  and  $\frac{\partial V}{\partial K_t} = 0$ . Using this insight, along with the expression for  $p_t(K_t; p_0)$  in (1), we can reformulate the agent’s optimal  $k_t$  as

$$\tilde{k}_t^*(K_t, \phi_t; p_0) = \begin{cases} 1 & \text{for } p_0 \geq \bar{p}(K_t, \phi_t), \\ 0 & \text{otherwise,} \end{cases} \quad (20)$$

where

$$\bar{p}(K_t, \phi_t) \equiv \frac{1}{1 + \left[ \frac{\lambda \Pi_A}{a(1 - \phi_t)} - 1 \right] e^{-\lambda K_t}}. \quad (21)$$

In other words, if  $K_t$  has already been invested, then as long as the agent's type  $p_0$  is above the threshold  $\bar{p}(K_t, \phi_t)$ , he will continue to invest  $k_t = 1$ . Note that solving equation  $p_0 = \bar{p}(K_t, \phi_t)$  for  $K_t$  gives

$$\bar{K}(\phi_t, p_0) \equiv \frac{1}{\lambda} \ln \left[ \left( \frac{\lambda \Pi_A}{a(1 - \phi_t)} - 1 \right) \frac{p_0}{1 - p_0} \right],$$

which then allows us to represent the agent's optimal investment  $k_t$  as

$$\tilde{k}_t^*(K_t, \phi_t; p_0) = \begin{cases} 1 & \text{for } K_t \leq \bar{K}(\phi_t, p_0), \\ 0 & \text{otherwise.} \end{cases} \quad (22)$$

Equation (22) represents the agent's investment threshold in terms of cumulative investment, which will facilitate the derivation of the principal's optimal strategy. To see this, recall from the discussion in Section 2.4 that for  $K_t \in [0, K^A(p_0)]$ , the principal, who knows  $p_0$ , can set  $\phi_t = 0$ , because the agent is willing to invest without compensation. But when  $K_t > K^A(p_0)$ , the principal must provide some compensation, or the agent will not invest. Given  $K_t$ , the principal needs to pick a reimbursement rate  $\phi_t$  such that  $p_0 \geq \bar{p}(K_t, \phi_t)$  holds in order to induce the agent to invest  $k_t = 1$ , which implies

$$\phi_t \geq 1 - \frac{\lambda \Pi_A}{a} p_t(K_t; p_0). \quad (23)$$

In addition,  $\phi_t$  must be nonnegative, so the reimbursement rate must satisfy the following incentive compatibility constraint:

$$\phi_t \geq \tilde{\phi}_t(K_t; p_0) = \max \left\{ 1 - \frac{\lambda \Pi_A}{a} p_t(K_t; p_0), 0 \right\}, \quad (24)$$

which will bind because the principal can make a take-it-or-leave it offer to the agent.

To complete the derivation of the Markov perfect equilibrium under full information, we must find the principal's preferred termination threshold for  $K_t$  given  $\tilde{\phi}_t(K_t; p_0)$ . From equation (15), we can formulate the principal's Bellman equation as

$$rW = \frac{\partial W}{\partial \tilde{\phi}_t} \frac{d\tilde{\phi}_t}{dt} + \max_{\tilde{k}_t \in \{0, 1\}} \left\{ \left[ -\tilde{\phi}_t a + \lambda p_t \Pi_P - \lambda p_t W + \frac{\partial W}{\partial K_t} \right] \tilde{k}_t \right\},$$

where  $\tilde{k}_t$  represents the principal's preferred investment by the agent. The maximization problem has the following bang-bang solution:

$$\tilde{k}_t(K_t; p_0) = \begin{cases} 1 & \text{if } p_t(K_t; p_0) > \frac{a\tilde{\phi}_t - \frac{\partial W}{\partial K_t}}{\lambda(\Pi_P - W)}, \\ 0 & \text{otherwise.} \end{cases} \quad (25)$$

Note that (25) is similar to the agent's optimal investment strategy (19) except for the cut-off threshold for the posterior belief  $p_t(K_t; p_0)$ . The following proposition shows that the principal can choose the reimbursement rate  $\phi_t$  appropriately such that the agent's investment strategy coincides with the principal's preferred investment path.

**Proposition 3** Suppose  $p_0$  is known to the principal. The Markov perfect equilibrium strategy of the principal—denoted by  $\tilde{\phi}_{t+dt}^*(K_t; p_0)$ —is given by:

$$\tilde{\phi}_{t+dt}^*(K_t; p_0) = \begin{cases} 0 & \text{if } K_t \in [0, K^A(p_0)], \\ 1 - \frac{\lambda \Pi_A}{a} p_{t+dt}(K_t + dt; p_0) & \text{if } K_t \in [K^A(p_0), K^{**}(p_0)], \\ \leq \phi^{**} & \text{if } K_t \geq K^{**}(p_0), \end{cases} \quad (26)$$

where  $\phi^{**} \equiv \frac{\Pi_P}{\Pi_P + \Pi_A}$  and  $K^{**}(p_0) \equiv \bar{K}(\phi^{**}, p_0)$ . The Markov perfect equilibrium strategy for the agent is  $\tilde{k}_t^*(K_t, \tilde{\phi}_t^*(K_t; p_0); p_0)$  as described in (22). On the equilibrium path, the investment outcome is

$$\tilde{k}_t^*(K_t, \tilde{\phi}_{t+dt}^*(K_t; p_0); p_0) = \begin{cases} 1 & \text{if } K_t \leq K^{**}(p_0), \\ 0 & \text{otherwise,} \end{cases} \quad (27)$$

which is also Pareto efficient.

Summing up, when the principal knows  $p_0$ , the bargaining game unfolds as follows. The principal begins by offering no compensation. If  $p_0 < p^A$  the agent would still invest. As time passes with no breakthrough, the principal would continue to offer no compensation as long as cumulative investment is less than  $K^A(p_0)$ . But as more time passes without a breakthrough, and cumulative investment increases above  $K^A(p_0)$ , the principal begins to offer compensation, but only enough so the agent is just indifferent between investing and not investing. If a breakthrough still is not realized, the posterior probability that the problem is solvable goes down, necessitating a higher reimbursement to keep the agent willing to invest. But when the reimbursement rate rises to  $\phi^{**}$ , the principal no longer increases it, and investment then terminates. This point of termination is socially efficient. The termination reimbursement rate  $\phi^{**}$  can be interpreted as the principal's reservation price in the following sense: if the reimbursement rate was above it, then the agent would be induced to overinvest, exceeding the socially efficient level  $K^{**}(p_0)$ . This would be suboptimal for the principal who would be the sole bearer of net loss. To prevent this, the reimbursement rate must be capped at  $\phi^{**}$ .

A key advantage that the principal has under full information about  $p_0$  is that she can hold off offering compensation until it is absolutely needed (because she knows  $K^A(p_0)$ ). If the principal does not know  $p_0$ , then an agent has good reason not to disclose it truthfully. For example, if an agent with  $p_0 > p^A$  disclosed that fact, he would initially receive no compensation from the principal, and when he eventually did receive compensation offers, they would be ones that eliminated the agent's entire surplus from trade.

## 5.2 MPOE: Existence and Uniqueness

The benchmark equilibrium derived in the previous subsection is *full-information Pareto efficient* (allowing for compensation transfer), taking into account the agent's investment incentives. We now study the MPOE for our general model in which the principal does

not know the agent's type  $p_0$ . With ambiguity over the agent's type, the principal cannot offer a reimbursement rate  $\phi_t$  conditional on  $p_0$ . However, in this section we will show that there exists a unique MPOE outcome which is surprisingly simple — the principal offers a constant reimbursement rate  $\phi^{BO}$  from the start to the end of the game and the agent adopts a bang-bang investment policy  $k_t^{BO}$ , where  $k_t^{BO}$  and  $k_t^{BO}$  are defined as follows:

**Definition 7** Let the bargaining outcome  $(\phi^{BO}, k_t^{BO})$  be defined as

$$\phi^{BO} \equiv \phi^{**} = \frac{\Pi_P}{\Pi_P + \Pi_A}, \quad (28)$$

$$k_t^{BO} = \begin{cases} 1 & \text{if } K_t \leq K^{**}(p_0), \\ 0 & \text{otherwise.} \end{cases} \quad (29)$$

Unlike the full-information Pareto efficient equilibrium (Proposition 3), the principal offers her reservation reimbursement rate  $\phi^{**}$  throughout the game. The left hand side of (28) is the reimbursement rate  $\phi^{**}$ , representing the principal's share of investment cost, while the right hand side of (28),  $\frac{\Pi_P}{\Pi_P + \Pi_A}$ , represents the principal's percentage of total benefit from breakthrough. In other words, the equilibrium reimbursement contract has the property:

$$\text{Share of cost} = \text{Share of benefit.}$$

Intuitively, this result reflects a balance between two sources of bargaining power: the principal has the power to make take-it-or-leave-it offers while the agent has the power to conceal private information that causes ambiguity for the principal. This results in an equilibrium that divides the cost by benefit shares.

We will establish that this is the unique MPOE outcome in three steps. First, we show that the principal's knowledge of the agent's type can be represented by a scalar state variable. Second, we derive the unique MPOE for the class of equilibrium when such representation is possible. Finally, we show that the equilibrium we derived in the first step applies to all other classes of MPOE.

### 5.2.1 Principal's Knowledge about the Agent's Type

The principal has multiple priors over the agent's type — the agent's prior belief  $p_0$ . However, the agent's investment actions reveal (partial) information about his type, which is used by the principal to form a (biased) point estimate of the agent's type. The information revealed through the agent's action can be summarized by a scalar variable  $q_t$ , which will be formally defined presently. For now, we note that the Markov strategy of the agent  $k_t(K_t, \phi_t, q_t; p_0)$  is a function of state variables:  $K_t$ ,  $\phi_t$ , and  $q_t$ . Similar to the full information case, we can write the Bellman equation for the agent's optimal investment strategy as

$$rV = \frac{\partial V}{\partial \phi_t} \frac{d\phi_t}{dt} + \max_{k_t \in [0,1]} \left\{ \left[ -a(1 - \phi_t) + \lambda p_t \Pi_A - \lambda p_t V + \frac{\partial V}{\partial K_t} \right] k_t + \frac{\partial V}{\partial q_t} \frac{dq_t}{dt} \right\}. \quad (30)$$

Unlike the full information case, this Bellman equation has an extra term  $\frac{\partial V}{\partial q_t} \frac{dq_t}{dt}$ . If this term depends on action  $k_t$  (negatively), then there exists an informational strategic effect that affects the agent's investment behavior. In particular, the agent may underinvest in order to mimic a lower type. But first, consider a naive agent who ignores this strategic effect. The agent would not invest at date  $t$  if

$$-a(1 - \phi_t) + \lambda p_t \Pi_A - \lambda p_t V + \frac{\partial V}{\partial K_t} < 0.$$

For  $V \geq 0$  and  $\frac{\partial V}{\partial K_t} \leq 0$ , a sufficient condition for the above inequality is  $p_t \leq \frac{a(1 - \phi_t)}{\lambda \Pi_A}$ , which is equivalent to  $p_0 \leq \bar{p}(K_t, \phi_t)$ , where  $\bar{p}(K_t, \phi_t)$  is defined in (21). In other words,  $\bar{p}(K_t, \phi_t)$  is a naive agent's investment threshold at date  $t$ . If the principal observes positive  $k_t$ , she can infer that the agent's type  $p_0$  is at least  $\bar{p}(K_t, \phi_t)$ . In other words,  $\bar{p}(K_t, \phi_t)$  is the lower bound of the agent's type. Because this lower bound can be inferred only if the agent makes positive investment, the state variable  $q_t$  can be formally defined as

$$\begin{aligned} q_{t+dt} &= \max_{\tau \in [0, t]} \{ \bar{p}(K_\tau, \phi_\tau) 1_{k_\tau > 0} \}, \\ q_0 &= 0, \end{aligned} \tag{31}$$

where  $1_{k_\tau > 0}$  is an indicator function. Note,  $q_t$  has two important properties: (i) the Markov property

$$q_{t+dt} = \begin{cases} \bar{p}(K_t, \phi_t) 1_{k_t > 0} & \text{if } \bar{p}(K_t, \phi_t) 1_{k_t > 0} > q_t, \\ q_t & \text{otherwise;} \end{cases} \tag{32}$$

and (ii) the ratchet property

$$\frac{dq_t}{dt} \geq 0, \tag{33}$$

i.e.,  $q_t$  can only move in one direction. As the lower bound of the agent's type,  $q_t$  is a (biased) point estimate. However, the principal when formulating her strategy would treat the agent as if his type is  $q_t$ .<sup>24</sup> Because  $\bar{p}(K_\tau, \phi_\tau)$  is the marginal naive type that is just indifferent between accepting and not accepting  $\phi_\tau$ , we can interpret  $\bar{p}(K_\tau, \phi_\tau)$  as a measure of the concession (controlling for  $K_\tau$ ) the agent would be prepared to make in bargaining with the principal over the reimbursement rate  $\phi_\tau$ . To explain, note that a strategically sophisticated agent mimics some target marginal naive type in order to conceal his true type. Controlling for  $K_\tau$ , each marginal naive type maps one-to-one into a reservation price of the agent. Formally,  $\bar{p}(K_\tau, \phi_\tau^R) = p_0^T$  is equivalent to  $\phi_\tau^R = 1 - \frac{\lambda \Pi_A}{a} p_\tau(K_\tau; p_0^T)$ , which gives the expression of the reservation price of the agent who intends to mimic the target type  $p_0^T$ .

<sup>24</sup>This state variable does not have to reflect the correct beliefs of the principal, which must be the product of equilibrium-based valid inference. This is because, off the equilibrium path, the set of posterior beliefs equals the initial set of beliefs  $\mathcal{M}_0$  and becomes stationary. In contrast, the state variable  $q_t$  may still evolve over time and affect the principal's action. The fact  $q_t$  is a biased estimate of the agent's type and that it is used to guide the players' strategies does not imply that the players are not (sequentially) rational, or use invalid inference to form beliefs in any way. The state variable  $q_t$  and the common knowledge about it are simply part of a coordination device used by the players.

Therefore (controlling for  $K_\tau$ ) there exists a one-to-one mapping between a concession (in reservation price of bargaining) and a ratcheting-up of the target type (for misrepresentation). Now, from (32), it follows that  $q_t$  represents the highest marginal naive type based on the history  $h^t \in \mathcal{H}^t$  for  $t > 0$ . Thus,  $q_t$  is the maximum concession the agent had ever been prepared to make in his history of bargaining with the principal over the reimbursement rate  $\phi_\tau$ . By treating  $q_t$  as a (biased) point estimate of the agent's type, the principal treats the agent as if he would be willing to accept the same concession at any future time  $t' > t$ . If she wants to induce the agent to invest at time  $t'$ , she would be unwilling to demand a smaller concession. This ratchet mechanism effectively consolidates any temporary concession by the agent into a permanent concession (controlling for  $K_{t'}$ ). The implication is that if the agent deviates from an equilibrium level of concession by conceding more (e.g., behaving naively), then this deviation will lead to reduction in his information rent as well as the value of his project.

We have assumed there is no strategic effect ( $\frac{\partial V}{\partial q_t} \frac{dq_t}{dt} = 0$ ). We now formally establish that that along the equilibrium path, the principal indeed eschews the strategic effect and makes it safe for the agent to play naively.

**Proposition 4** *If there exists an MPOE such that the principal's Markov strategy takes the form  $\phi_{t+dt} = \phi(K_t, q_t)$ , then  $\phi_t = \phi^{**}$  for all states  $(K_t, q_t)$  that are on an equilibrium path. Consequently,*

$$\frac{\partial V}{\partial q_t} = 0, \quad (34)$$

and

$$\frac{dq_t}{dt} > 0 \text{ if and only if } k_t > 0. \quad (35)$$

This result implies that along the equilibrium path of an MPOE of the form  $\phi_{t+dt} = \phi(K_t, q_t)$  the principal always makes an offer equal to her reservation reimbursement rate  $\phi^{**}$ .<sup>25</sup> The basic intuition is that as long as  $\phi_t < \phi^{**}$  termination of the trading relationship (which occurs at the first moment at which when  $k_t = 0$ ) is suboptimal for the principal. To ensure that the principal's offer  $\phi_{t+dt} = \phi(K_t, q_t)$  is truly a take-it-or-leave-it offer, the principal's offer price must reach her reservation reimbursement rate  $\phi^{**}$ .

### 5.2.2 Existence of MPOE

We now prove the existence of an MPOE and its properties. First, we define a candidate equilibrium:

---

<sup>25</sup>Recall that  $\phi^{**}$  is the reservation price in that it is the highest reimbursement rate the principal is willing to offer under full information. This upper bound helps prevent overinvestment under full information. This is also true under ambiguity, where the ambiguity-averse principal is extremely averse to the possibility of inducing a low-type agent to overinvest with  $\phi_t > \phi^{**}$ .

**Definition 8** Let the tuple  $(s^{**}, \mathcal{M}^{**})$  be such that

(a) at  $t = 0_{-1}$ ,

$$s_P^{**}(h^{0_{-1}}) = \phi^{**};$$

(b) for all  $t \geq 0$ ,

$$\begin{aligned} s_P^{**}(h^t) &= \phi_{t+dt}^{**}(K_t, q_t) \\ &= \begin{cases} 0 & \text{if } K_t \in [0, K^A(q_t)], \\ 1 - \frac{\lambda_{\Pi_A}}{a} p_{t+dt}(K_t + dt; q_t) & \text{if } K_t \in [K^A(q_t), K^{**}(q_t)], \\ \phi^{**} & \text{if } K_t \geq K^{**}(q_t), \end{cases} \end{aligned} \quad (36)$$

where  $K^{**}(q_t) = \bar{K}(\phi^{**}, q_t)$ ;

$$\begin{aligned} s_A^{**}(h^t, p_0) &= k_t^{**}(K_t, \phi_t, q_t; p_0) \\ &= \begin{cases} 1 & \text{if } K_t \leq \bar{K}(\phi_t, p_0) \text{ and either } q_t \geq \bar{p}(K_t, \phi_t) \text{ or } \phi_t \geq \phi^{**}, \\ 0 & \text{otherwise; } \end{cases} \end{aligned} \quad (37)$$

$\mathcal{L}_0(s^{**})$  is derived from  $s^{**}$ ; the derivation of  $\mathcal{M}_t(h^t) \in \mathcal{M}^{**}$  is as described in Section 3.1.2,  $\mathcal{M}_{0_{-1}}(h^{0_{-1}}) = \mathcal{M}_0(h^0) = \mathcal{M}_0$ .

It is easily verifiable that along the predicted path of  $(s^{**}, \mathcal{M}^{**})$ , the reimbursement is at the principal's reservation rate, i.e.,  $\phi_t = \phi^{**}$ . Consequently the marginal naive type (of agent) is  $\bar{p}(K_t, \phi^{**})$ . The agent's strategy  $s_A^{**}(h^t, p_0)$  given by (37) is characterized by the feature that for  $t > 0$  all higher types  $p_0 \in (\bar{p}(K_t, \phi^{**}), 1]$  mimic the behavior of the marginal type  $\bar{p}(K_t, \phi^{**})$ . If the marginal type's participation constraint is violated (i.e.,  $\phi_t < \phi^{**}$ ) then all higher types stop investing. So  $s_A^{**}(h^t, p_0)$  is a partial pooling strategy. This strategy essentially sets the agent's reservation rate for  $\phi_t$  (at  $\phi^{**}$ ), below which the agent would reject the offer through suspending investment. By such strategy, the agent also withholds sensitive information about his true type.

The candidate equilibrium under ambiguity  $(s^{**}, \mathcal{M}^{**})$  bears some remarkable resemblance with the Markov perfect equilibrium under full information, as described by Proposition 3. For the Markov strategy of the principal, the only essential difference is to replace the true type  $p_0$  with the point estimate  $q_t$ . That is, the principal acts as if she knew the agent's true type to be  $q_t$ . For the strategy of the agent, under full information MPE the agent always plays naively, while under  $(s^{**}, \mathcal{M}^{**})$ , the agent plays the naive strategy only if the principal eschews the informational strategic effect voluntarily, which technically entails either  $q_t \geq \bar{p}(K_t, \phi_t)$  or  $\phi_t \geq \phi^{**}$ . Along the predicted path this condition is always satisfied, that is, the principal indeed voluntarily eschews the strategic effect. In a counterfactual scenario off the predicted path, the principal may deviate from this voluntary restriction, only to find that the agent will decline the low offer  $\phi_t$  and stop investing.

We now present the main existence result of the paper. We establish that the candidate equilibrium in Definition 8 is indeed an MPOE, and we show that the equilibrium is Pareto efficient and thus Pareto dominates other POE outcomes.

**Theorem 3** (a) The tuple  $(s^{**}, \mathcal{M}^{**})$  given by Definition 8 is a pure strategy MPOE

(b) The equilibrium  $(s^{**}, \mathcal{M}^{**})$  generates a Pareto efficient outcome, where the principal and agent's actions are given by Definition 7.

(c) The equilibrium given by Definition 8 generates the Pareto dominant POE outcome if there exist multiple POE outcomes.

The equilibrium in Definition 8 is significant because each possible type of the agent receives his largest possible payment, while the principal is indifferent between the payoff she receives under this equilibrium and the outcomes under other Markov equilibria if they exist. To see why, let  $V^{**}(K_t, \phi^{**}, p_0)$  denote the value function of the agent along an equilibrium path in the equilibrium in Definition 8. Given that investment intensity in this equilibrium is socially efficient at each point of time and the surplus given to the agent (through the expected benefit of breakthrough and compensation transfer) is maximized,  $V^{**}(K_t, \phi^{**}, p_0)$  must set an upper bound to the value function of the agent for all MPOE. Furthermore, part (c) of Theorem 3 establishes that this equilibrium generates a Pareto dominant equilibrium outcome in case there exist multiple equilibrium outcomes.

### 5.2.3 Ambiguity Resolution through Likelihood Inference on Equilibrium Paths

We have established the existence and the uniqueness of the MPOE outcome. We now describe how the principal's objective learning process works along the equilibrium path, i.e., how the principal resolves her ambiguity through likelihood inference on the equilibrium path.

Along the equilibrium path, the principal offers a constant reimbursement rate  $\phi_t = \phi^{**}$  for all  $t \geq 0$ . The agent's best response is to invest  $k_t = 1$  as long as  $p_0 \geq \bar{p}(K_t, \phi^{**})$ .<sup>26</sup>

At  $t = dt$ , which is immediately after the game begins, if  $k_0 = 1$  is observed, then the principal can infer  $p_0 \geq p^{**} = \frac{a(1-\phi^{**})}{\lambda\Pi_A}$ , because the likelihood of an agent with type  $p_0 < p^{**}$  making positive investment is zero. Therefore,  $p^{**}$  is the (inclusive) maximum lower bound of  $p_0$ , i.e.,  $p^{**} = \inf P_t = \min P_t$  and  $p^{**} \notin P_{dt}$ . In addition, the principal understands that an agent with type  $p_0 \geq p^{**}$  would have invested  $k_0 = 1$  in the project. As a result, the principal can infer that  $p_0 \in [p^{**}, 1]$ . In this case, 1 can be inferred as the (inclusive) least upper bound of  $p_0$ , i.e.,  $p_0 \leq 1 = \sup P_{dt} = \max P_{dt}$  and  $1 \in P_{dt}$ , where  $P_{dt}$  is the support of  $\mathcal{M}_t(h_t)$  at time  $dt$ . On the other hand, if  $k_0 = 0$  is observed immediately after the game begins, the principal can infer that  $p_0 < p^{**}$ , that is,  $p_0 \in [0, p^{**})$ , because an agent with type  $p_0 \geq p^{**}$  would have made positive investment. Furthermore, because the agent expects the principal to offer  $\phi_t = \phi^{**}$  throughout the game, he would not delay his investment if his type  $p_0 \geq p^{**}$ . Consequently, if no investment is observed immediately after the game begins, no investment will be observed at any time later and the game is effectively ended.

---

<sup>26</sup>It can be shown that (37) is equivalent to

$$k_t^{**}(K_t, \phi_t, q_t; p_0) = \begin{cases} 1 & \text{if } \phi_t \geq \phi_t^{**}(K_{t-dt}, q_{t-dt}) \text{ and } p_0 \geq \bar{p}(K_t, \phi^{**}), \\ 0 & \text{otherwise.} \end{cases}$$

where  $\bar{p}(K_t, \phi^{**})$  is the cutoff type defined by (21).

The same likelihood inference is repeated at any time  $t > 0$ . In particular, at time  $t + dt$ , if  $k_t = 1$  is observed, from the agent's investment best response (37), the principal infers  $p_0 \geq \bar{p}(K_t, \phi^{**})$ . This implies the support  $P_{t+dt} = [\bar{p}(K_t, \phi^{**}), 1]$ , where  $\inf P_{t+dt} = \bar{p}(K_t, \phi^{**})$ . In fact, as the agent continues to invest in the absence of a breakthrough, the lower bound of the support  $P_t$  gradually increases from  $p^{**} = \inf P_{dt}$  to the upper bound  $\sup P_t = 1$  in a Logistic (diffusion) process. To see this, note that

$$\frac{d \inf P_t}{dt} \frac{1}{\inf P_t} = \frac{\lambda}{1 - p^{**}} (\sup P_t - \inf P_t),$$

which implies the ambiguity of the agent's type is reduced smoothly along the equilibrium path. On the other hand, if investment is not observed at time  $t + dt$ , then the principal can infer  $p_0 \leq \bar{p}(K_t, \phi^{**})$ . Previously at time  $t$ , the principal inferred that  $p_0 \geq \bar{p}(K_{t-dt}, \phi^{**})$ . Let  $dt \rightarrow 0$ , the principal can infer that  $\inf P_{t+dt} = \sup P_{t+dt} = \bar{p}(K_t, \phi^{**})$ . Consequently, only  $p_0 = \bar{p}(K_t, \phi^{**})$  is plausible. In addition, the agent will not make any positive investment at any time  $t' > t$ . Hence, ambiguity is resolved abruptly,

$$\inf P_{t'} = \sup P_{t'} = \bar{p}(K_t, \phi^{**}) \quad (38)$$

for all  $t' > t$ . That is, the true value of  $p_0$  becomes known to the principal after time  $t$ . However, knowing the exact type of the agent will not changed the principal's strategy at this point, because the principal has already induced the maximum investment from the agent and it is also in the principal's interest to terminate the project.

The following proposition established the link between the above heuristic discussion of the ambiguity resolution with the state variable  $q_t$  defined in MPOE strategies.

**Proposition 5** *For all MPOE such that the principal's Markov strategy takes the form  $\phi_{t+dt} = \phi(K_t, q_t)$ , there is a close relationship between the state variable  $q_t$  and the principal's beliefs about the agent's type such that for all  $t \geq 0$  with  $K_t > 0$*

$$\inf P_t = \begin{cases} q_t & \text{on an equilibrium path,} \\ 0 & \text{off equilibrium path.} \end{cases}$$

*It follows that for any state along an equilibrium path such that  $t \geq 0$  with  $K_t > 0$ ,*

$$\begin{aligned} \inf P_t &= q_t, \\ \inf P_{t+dt} &= \begin{cases} \bar{p}(K_t, \phi^{**}) & \text{if } k_t > 0, \\ \inf P_t & \text{if } k_t = 0, \end{cases} \end{aligned} \quad (39)$$

*and*

$$\frac{d \inf P_t}{dt} = \begin{cases} \lim_{dt \rightarrow 0} \frac{\bar{p}(K_t, \phi^{**}) - \inf P_t}{dt} & \text{if } k_t > 0, \\ 0 & \text{if } k_t = 0. \end{cases} \quad (40)$$

As has been demonstrated, the state variable  $q_t$  compresses the information from observed history and forms a meaningful (lower bound) estimate about the agent's private type  $p_0$ . Along any equilibrium path, the meaning of  $q_t$  is exact. Thereby, it is a minimum sufficient (payoff-relevant) statistic of the principal's belief (about the agent's type) - the maximum lower bound. Off the equilibrium path,  $q_t$  also has a meaningful interpretation: it represents the maximum bargaining concession the agent has ever made by revealing sensitive information about its type to the principal. The fact that the agent only suffers the consequences of such revelation off the equilibrium path implies that along an equilibrium path, the agent is strategically sophisticated enough not to reveal any sensitive information.

Does such restriction on information transmission impede the allocational outcome? Does it affect the distributional outcome? Even though the principal does not know  $p_0$ , the cumulative investment level delivered by a type- $p_0$  agent in the POE described in Theorem 3 is the full-information Pareto efficient cumulative investment. Thus, even though the information transmission in equilibrium is minimal, the allocational properties of the equilibrium are as if the principal actually knew the agent's prior belief. The extreme initial and persistent informational asymmetry, however, does have distributional consequences through its effect on the allocation of bargaining power as will be elaborated in the next section.

#### 5.2.4 Uniqueness of MPOE Outcome

So far we have focused on a specific class of MPOE which involves the state variable  $q_t$ . The Markov strategy status thereof is justified by the fact that  $q_t$  is a minimum sufficient (payoff-relevant) statistic of the principal's belief about the agent's type (along all equilibrium paths). To derive general results about all MPOE, we either need to establish that there is no other class of MPOE or show that properties of the equilibrium in Definition 8 apply to all other classes of MPOE if they exist. In this section we do the latter. The next lemma extends the result in Proposition 4— $\phi_t = \phi^{**}$  along the equilibrium path—to any MPOE:

**Lemma 1** *For any equilibrium path of an MPOE, the principal's reimbursement rate is  $\phi_t = \phi^{**}$ , and the agent's investment is  $k_t^{**}$  for all  $t \geq 0$  as given by Definition 7.*

This lemma hinges on the definition of Markov perfect equilibrium for incomplete information, which emphasizes that (roughly speaking) in a Markov equilibrium the players' strategies must only depend on the coarsest sufficient statistics of payoff-relevant history. The key idea of this lemma's proof is that a point estimate of the agent type, such as the state variable  $q_t$  is indispensable in a Markov perfect equilibrium. Therefore if any other MPOE exists, it must be similar to the MPOE involving  $q_t$ , and it must entail an on-equilibrium-path property similar to Proposition 4, notably,  $\phi_t = \phi^{**}$  for all  $t \geq 0$ .

Since the equilibrium outcome only depends on the on-equilibrium-path play of the game, even if there exist multiple MPOE, the equilibrium outcome in terms of (non-cooperative) bargaining solution must (still) be unique:

**Theorem 4** *There exists a unique MPOE outcome in terms of a (non-cooperative) bargaining solution, which is given by  $(\phi^{BO}, k_t^{BO})$ .*

As already established by Theorem 3, this (non-cooperative) bargaining solution is also socially efficient (i.e., Pareto efficient, allowing compensation transfer).

Theorem 4 implies that, in any MPOE the agent's information rent reaches its upper bound, just as in the MPOE established in Theorem 3. Intuitively, the reason why the agent's information rent is maximized is because he never lets the principal know unambiguously that the reimbursement rate  $\phi^{**}$  is unnecessarily "generous." Letting  $P_t \equiv P(\mathcal{M}_t(h_t))$ , i.e., the support of  $\mathcal{M}_t(h_t)$ , notice that  $p_t(K_t; \inf P_t) = p^{**}$  (and equivalently  $K_t = K^{**}(\inf P_t)$ ) holds for all  $t \geq 0$ . The ambiguity-averse principal therefore can never justify a reduction of  $\phi_t$  to below  $\phi^{**}$ . If the principal wanted to control the agent's information rent, she would have to be able to induce the agent to reveal sensitive information such as  $p_t(K_t; \inf P_t) > p^{**}$  (when that is the case). In all MPOE such that  $\phi_{t+dt} = \phi(q_t, K_t)$ , the agent would reject such an attempt by the principal and could do so credibly.

Intuitively the principal faces a trade off between inducing the low marginal type to invest in a socially efficient way and reducing the information rents of the higher types. The initial and persistent extreme asymmetry of information between the principal and agent, combined with ambiguity aversion of the principal, tilts the balance toward trading with the marginal low type efficiently and tolerating the information rents earned by the higher types.<sup>27</sup> This is because the principal is especially concerned with the worst-case estimation of social surplus from trade (conditional on trade occurring), which is zero. In this case, an attempt to extract positive surplus from trade can only mean forcing the trading partner to take a loss, which is ruled out by the partner's participation constraint. When the social surplus from trade is zero, the only bargaining solution that respects both parties' participation constraints is  $\phi_t = \phi^{**}$  (i.e., share of cost = share of benefit).

The above analysis delivers two important novel findings. First, the possession of exclusive information under conditions of ambiguity is extremely asymmetric can be a source of significant bargaining power (endowed to the informed party). Second, prudence against ambiguity (i.e., ambiguity averse preferences) can be an important determinant of the (non-cooperative) bargaining solution. In general, prudence entails a conservative estimate of social surplus from trade. On the one hand, this reduces the uninformed party's motivation to aggressively extract surplus from trade. On the other hand, this also prevents the uninformed party from paying the informed party too high a price. In the current example of repeated bargaining with interactive learning, the balance of these two effects pins down a unique bargaining solution. This result will appear much more striking when it is compared with the bargaining solutions predicted by Markov perfect Bayesian equilibria (MPBE), as characterized in Section 6.1.

---

<sup>27</sup> Although the principal has a first mover advantage, it is entirely offset by the disadvantage of extreme informational asymmetry.

## 6 Broader Perspectives

In this section, we broaden the perspectives in order to compare the insights of the current model to existing models in several important literatures, including dynamic mechanism design with limited commitment and belief-free equilibrium. We also leverage the insights of our analysis to offer thoughts on ambiguity aversion and learning under ambiguity.

### 6.1 Conventional Bayesian Dynamic Mechanism Design with Limited Commitment

To conserve space, we do not undertake a full review of the literature on conventional Bayesian dynamic mechanism design. This literature is vast (see Bolton and Dewatripont (2005) for a survey). Here we focus on the problem of lack of commitment (in bilateral contracting under dynamic adverse selection), because this is the reason why an optimal long term contract cannot simply be the repetition of an optimal static contract (when the agent's hidden type is time-invariant). Suppose that each type of the agent truthfully reveals his true type in the first period of a multiple-period model. Without commitment, there will be re-contracting between the principal and the agent in each period (due to unilateral violation of the extended optimal static contract). This then implies that the agent will not be able to enjoy any information rent after the initial period. Taking this prospect into account, the agent will not be induced by the optimal static contract to fully truthfully reveal his type. In general, the optimal dynamic contract will be affected by what is well known in the literature as the "ratchet effect". Particularly, some degree of "pooling" may arise from equilibrium dynamic contracting.

Should the ratchet effect not arise (hypothetically), the optimal dynamic contract could have been the repetition of an optimal static contract leading to a separating equilibrium. In the presence of a ratchet effect, it becomes harder for the principal to profitably induce each type of the agent to fully truthfully reveal his true type. Depending on the strength of the ratchet effect, it may bias the equilibrium of the principal's optimal mechanism towards more or less pooling. If one could rank the possible types of equilibrium of the model by the initial degree of separation at time  $dt \rightarrow 0$ , say, measured the lowest agent type from whom the principal is unable to separate the type  $p_0 = 1$ , then one could measure in clearly-defined terms the strength of the ratchet effect. Hence, the higher the degree of separation, the weaker the ratchet effect. Given the structure of a dynamic mechanism design (with fixed hidden agent's type) problem, it is interesting to know how the prior belief of the principal affects the strength of the ratchet effect. Particularly, do less informative prior beliefs lead to stronger ratchet effect? As far as we know, this question has not been explicitly and systematically addressed in the literature. In what follows we approach this question in a particular way that will provide a partial answer. Given that our model assumes the least informative prior beliefs and induces a unique MPOE, we ask whether the degree of separation increases in a pure strategy Markov perfect Bayesian equilibrium (MPBE) if we replace the belief of the principal by a particular single-prior belief system (holding all else

fixed). The answer is affirmative and stated in the following proposition.

**Proposition 6** *Consider the delegated experimentation model in which the principal has a prior belief about  $p_0$  that is a uniform probability density  $\frac{1}{p^{**}+1-\vec{p}}$  over closed intervals  $[0, p^{**}]$  and  $[\vec{p}, 1]$ , and zero probability (a hole) over open interval  $(p^{**}, \vec{p})$ . Then there exists a pure strategy MPBE for each of a continuum of games indexed by the parameter  $\vec{p} \in [p^{**}, 1]$  where the equilibrium  $(\check{s}, \check{\mathcal{M}})$  is such that.<sup>28</sup>*

$$\check{\mathcal{M}}_0 = \left\{ \mu | d\mu(p_0) = \begin{cases} \frac{1}{p^{**}+1-\vec{p}} dp_0 & \text{for } p_0 \in [0, p^{**}] \\ 0 & \text{for } p_0 \in (p^{**}, \vec{p}) \\ \frac{1}{p^{**}+1-\vec{p}} dp_0 & \text{for } p_0 \in [\vec{p}, 1] \end{cases} \right\}.$$

$$\check{s}_P(h^{0-1}) = \phi_0 = \max \left( 1 - \frac{\lambda \Pi_A}{a} \vec{p}, 0 \right).$$

And  $\forall t \geq 0$  and  $\forall h^t \in \mathcal{H}^t$ ,

$$\check{s}_P(h^t) = \check{\phi}_{t+dt}(K_t, q_t) =$$

$$\begin{cases} 0 & \text{for } K_t \in [0, K^A(\max(\vec{p}, q_t))] \\ 1 - \frac{\lambda \Pi_A}{a} p_{t+dt}(K_t + dt; \max(\vec{p}, q_t)) & \text{for } K_t \in (K^A(\max(\vec{p}, q_t)), K^{**}(\max(\vec{p}, q_t))) \\ \phi^{**} & \text{for } K_t \geq K^{**}(\max(\vec{p}, q_t)). \end{cases}$$

(where  $\mathcal{M}_t(h^t)$  is a singleton) and

$$\check{s}_A(h^t; p_0) = \check{k}_t(K_t, \phi_t, q_t; p_0) = \begin{cases} 1 & \text{if } \phi_t \geq \check{\phi}_t \text{ and } K_t \leq \bar{K}(\phi_t, p_0) \\ 0 & \text{otherwise.} \end{cases}$$

All types in  $[\vec{p}, 1]$  pool together initially at time  $dt \rightarrow 0$  and it becomes suboptimal for the principal to violate the participation constraint of the (initial) marginal type, i.e.,  $p_0 = \vec{p}$ . In general all higher types understand that by mimicking the behavior of the (current) marginal type (including the initial one and each new one) they can maximize their information rents.

Note, the unique prior distribution is subjective. The principal does not have to justify it on any objective basis (which does not have to exist). It is possible that the principal's subjective prior belief is wrong. Such an error, however, can never be proven by the observational evidence generated by the equilibrium described above. A subjective Bayesian model of dynamic game with incomplete information allows this. Features of such non-falsifiable subjective prior belief, e.g., parameter  $\vec{p}$ , can play a crucial role in affecting the strength of the ratchet effect, as is illustrated by the following proposition.

Let the parameter  $\vec{p} \in [p^{**}, 1]$  represent the informativeness of the principal's single-prior belief system. It also represents the initial degree of separation in the MPBE. The variable  $(1 - \vec{p}) \in [0, 1 - p^{**}]$  represents the strength of the ratchet effect.

<sup>28</sup>Note that  $\check{\mathcal{M}}$ , whose members are all singletons, represents the single-distribution beliefs of the principal corresponding to this specific equilibrium. Coupling this belief system with POE allows POE to nest perfect Bayesian equilibrium (PBE).

**Proposition 7** *The initial degree of separation in the MPBE is strictly increasing in the informativeness of the principal's single-prior belief system, represented by  $\vec{p}$ , and the strength of the ratchet effect is strictly decreasing in  $\vec{p}$ . In the limiting case,  $\vec{p} = p^{**}$ , the least informative single-prior belief system results in the largest ratchet effect, and the same as the result of the least informative multiple-prior belief system. Ceteris Paribus,  $\dot{\phi}_t$  is non-increasing in  $\vec{p}$ .*

The above result shows that a player's subjective pretense of knowledge (i.e.,  $\vec{p} > p^{**}$ ), if publicly known, can possibly influence the strategy of an opponent and hence the equilibrium outcome. Such subjective pretense of information can be false but not observationally falsifiable in an equilibrium. While such subjective belief may benefit the believer in a bargaining situation (if it happens to be compatible with the true state) by providing a means of commitment, if it is false (i.e.,  $p_0 \in (p^{**}, \vec{p})$ ), such commitment becomes a source of Pareto inefficient outcome and harms the believer as well.<sup>29</sup>

There is another problem with the subjective probabilistic beliefs described above—they make the bargaining problem nontrivial and lead to multiplicity of MPBE. In the unique MPOE outcome we have established, the ambiguity-averse principal's decisions are influenced by the pessimistic expectation of zero value of surplus, making the bargaining problem trivial. As a result, a unique MPOE outcome arises. By contrast, subjective probabilistic beliefs allow the principal to entertain arbitrary expected values of surplus from trade. A Bayesian principal may expect a strictly positive value of surplus to be divided between herself and agent. As a result, the bargaining problem is not trivial and multiplicity of MPBE arrises. The following proposition gives some concrete examples.

**Proposition 8** *In the delegated experimentation model with a Bayesian principal, there exists a continuum of MPBE indexed by parameter  $\dot{\phi} \in (\phi^{**}, 1]$ , where the equilibrium  $(\dot{s}, \dot{\mathcal{M}})$  is such that*

$$\dot{\mathcal{M}}_0 = \left\{ \mu \mid \frac{d\mu(p_0)}{dp_0} = 1 \text{ for } p_0 \in [0, 1] \right\}$$

i.e., the set of prior beliefs is a singleton uniform distribution,

$$\dot{s}_P(h^{0-1}) = \phi_0 = \dot{\phi};$$

and for  $\forall t \geq 0$  and  $\forall h^t \in \mathcal{H}^t$ ,

$$\dot{s}_P(h^t) = \dot{\phi}_{t+dt}(c_t) = c_t \dot{\phi} + (1 - c_t) \phi^{**}$$

where  $\dot{\phi}$  is such that

$$u_0 \left( \frac{\bar{p}(0, \dot{\phi}) + 1}{2} \mid \dot{\phi}, 1, p_0 \right) \geq 0, \quad (41)$$

---

<sup>29</sup>See Crawford (1982) for some insightful observations on commitment and disagreement.

$$\begin{aligned} c_{t+dt} &= \left(1 - \frac{1}{2} \mathbf{1}_{(\dot{\phi}_t - \phi_t) k_t > 0}\right) c_t, \\ c_0 &= 1 \end{aligned} \quad (42)$$

and

$$\dot{s}_A(h^t; p_0) = \dot{k}_t(K_t, \phi_t, c_t; p_0) = \begin{cases} 1 & \text{if } K_t \leq \bar{K}(\phi_t, p_0) \text{ and } \phi_t \geq \dot{\phi}_t \\ 0 & \text{otherwise.} \end{cases}$$

The inequality (41) is the principal's participation constraint at  $t = 0$ , which ensures that (conditional on  $k_0 = 1$ ) the principal's expected surplus from trade is non-negative. Note, the principal's participation constraint (conditional on  $k_t = 1$ ) at  $t \geq 0$  is given by

$$\begin{aligned} E_t \left[ u_t | \bar{p}(K_t, \dot{\phi}) \right] &= \int_{\bar{p}(K_t, \dot{\phi})}^1 \frac{u_t(p_t(K_t; p_0) | \dot{\phi}, 1, p_0)}{1 - \bar{p}(K_t, \dot{\phi})} dp_0, \\ &= \frac{u_t(E_t[p_t(K_t; p_0) | \bar{p}(K_t, \dot{\phi})] | \dot{\phi}, 1, p_0)}{1 - \bar{p}(K_t, \dot{\phi})} = \frac{u_t\left(\frac{\bar{p}(0, \dot{\phi}) + 1}{2} | \dot{\phi}, 1, p_0\right)}{1 - \bar{p}(K_t, \dot{\phi})} \\ &\geq 0 \text{ if and only if } u_0\left(\frac{\bar{p}(0, \dot{\phi}) + 1}{2} | \dot{\phi}, 1, p_0\right) \geq 0. \end{aligned} \quad (43)$$

For  $t = 0$ , if  $\dot{\phi} \rightarrow \phi^{**}$  then  $\bar{p}(K_t, \dot{\phi}) \rightarrow p^{**}$  the left hand side of (41) must be strictly positive. Then there must exist  $\dot{\phi} \in (\phi^{**}, 1]$  such that for all  $\dot{\phi} \in [\phi^{**}, \dot{\phi}]$  inequality (41) is satisfied. The state variable  $c_t$  is the "reputation" or "credibility" index.  $c_t = 1$  means the reputation for being tough on one's bargaining position has never been damaged by any precedence of concession from the reservation price. The expression  $(-\log_2 c_t)$  represents the total number of incidences of previous concessions, each of which halves the credibility index. The properties of  $c_t$  include the Markov property (as shown in the law of motion) and the ratchet property:

$$dc_t = \begin{cases} -\frac{c_t}{2} < 0 & \text{if } (\dot{\phi}_t - \phi_t) k_t > 0, \\ 0 & \text{otherwise,} \end{cases}$$

i.e., the movement of  $c_t$  is unidirectional (non-increasing). Thus, in this MPBE there is a ratchet effect with respect to credibility. The credibility (state variable) is not so closely linked to the type of the agent. It is essentially related to the agent's determination/resolution not to concede from his intended bargaining position. If the alleged resolution has been compromised in history, then it loses its power to influence the opponent's incentive. Everytime the agent damages his credibility, the principal will ratchet up her own expected surplus from trade, and hence reduce the agent's surplus by a finite positive value. For the agent, it is always worth to refrain from current investment in order to protect his credibility. The strategies of both the principal and agent are "trigger" strategies: every instance (of accepting an lower offer) that damaged the agent's credibility would trigger the principal to become tougher and the agent softer afterwards. This strategies have the bootstrap (self sufficiency) property that makes them mutual best responses.

**Proposition 9** For  $\vec{p} = p^{**}$ , we have

$$\check{\phi}_{t+dt}(c_t) \geq \phi^{**} = \check{\phi}_{t+dt}(K_t, q_t)$$

for all  $t \geq 0$ ; and the inequality is strict for some  $t \geq 0$ , particularly,

$$\check{\phi}_0 > \phi^{**} = \check{\phi}_0.$$

The analysis in this section clearly demonstrates two weaknesses of PBE as a solution concept. First, it is prone to rationalizing epistemologically absurd outcomes, with beliefs that are wrong but never falsifiable in the equilibrium, as is shown by Proposition 6. Second, it is prone to multiplicity of equilibrium, as is illustrated by Proposition 8.

The comparison between the MPBE and MPOE is very revealing. While the MPOE outcome is unique, there exist a continuum of MPBE for the Bayesian version of the model with uniform priors. In practice, the uniform distribution is often treated as uninformative, at least approximately (though, in general, this is recognized to be incorrect; see, for example, Edwards 1992). The analysis above demonstrates that the set of MPBE based on (unambiguous) uniformly distributed prior belief is very different from the MPOE based on truly uninformative ambiguous beliefs. In this instance, the approximate representation of ignorance by the uniform distribution turns out to represent an “informative” (or rather a “misinformative”) pretence of knowledge that may bias the prediction of the outcome of the game. This example vindicates our claim that Bayesian model is inadequate to represent complete ignorance by some player about an opponent’s private information, and hence understates informational asymmetry.

## 6.2 Belief-Free Equilibrium

There is a recent literature on belief-free equilibrium (BFE), which aims at characterizing equilibria of games with incomplete information that are robust to specification of beliefs (see Hörner and Lovo (2009), Hörner, Lovo and Tomala (2011)). Hörner, Lovo and Tomala (2011) define a BFE as a strategy profile such that, “after every history, every player’s continuation strategy is optimal, given her information, and independently of the information held by the other players. That is, it must be a subgame-perfect equilibrium for every game of complete information that is consistent with the player’s information.” Our discussion below is based on this definition. BFE also relates to the concept of ex post (Nash) equilibrium (see Crémer and McLean (1985), Kalai (2004)). Bergemann and Morris (2007) introduce the notion of (static) belief-free incomplete information game, compare and look into the epistemic foundations of three belief-free solution concepts (including belief-free equilibrium). In general, the solution concept of BFE applies to belief-free games with incomplete information, for which no priors are specified and keeping track of beliefs is not required.

We can view the goal of non-cooperative game theory as to uncover stable structures of human interactions. Stable structures need to have the ability to survive potential deviations driven by self interests, as well as critical examinations entailed by rational beliefs and

learning which are based on evidence and reason. Subjective beliefs or pretence of knowledge cannot provide a reliable epistemic foundation for a stable structure. Viewed in this light, the BFE condition rids the structure of human interaction unreliable subjective pretence of knowledge, thus makes it stable. In the same vein, perfect objectivist equilibrium (POE), which is based on objectively established rational beliefs and learning, is also free from unreliable subjective pretence of knowledge. It is thus natural to further compare POE with belief-free equilibrium. As it transpires that (confining to pure strategy profile in game) POE is a weaker solution concept than BFE. POE represents weaker ambiguity aversion by the uninformed player than in a BFE. The former is based on max-min plus non-weak-dominance criterion while the latter is based on the criterion of unanimous preferability (with respect to all plausible beliefs).

As a result, if there exists a pure strategy BFE for a belief-free game with incomplete information that corresponds to a dynamic game under ambiguity, then a BFE pure strategy profile must coincide with strategy profile of a pure strategy POE. Therefore, if a pure strategy BFE exists, than looking for the corresponding pure strategy POE can be a useful way to approach it. As an example, the strategy profile of the pure strategy MPOE in our model is a pure strategy BFE of the corresponding belief-free game with incomplete information.

**Proposition 10** *The pure strategy profile  $s^{**}$ , as defined in Definition 8, constitutes a BFE of the corresponding belief-free game with incomplete information.*

The converse does not hold in general, i.e., the existence of a pure strategy POE may not ensure the existence of a pure strategy BFE.

Applying the notion of BFE to our model suggests that the BFE may be driven by the bootstrap property: the principal offers her reservation rate  $\phi^{**}$  because every (sufficiently high) type of the agent always demands this reservation rate along an equilibrium path. If a (sufficiently high type) agent did not insist on this demand, he would be punished by the principal (and his future self), thereby reducing the reimbursement rate and his information rent.

In general, if there exists a BFE, then the BFE profile coupled with full-support single prior belief system constitutes a PBE. As an example, we have:

**Proposition 11** *The BFE (of the corresponding belief-free game with incomplete information) – the pure strategy profile  $s^{**}$ , as defined in Definition 8 – constitutes a perfect Bayesian equilibrium (PBE) in the corresponding Bayesian game if the principal's belief is represented by any single prior distribution with full support.<sup>30</sup>*

---

<sup>30</sup>The condition of full support may be necessary. For example, consider the principal's prior distribution is uniform apart from a hole in the interval  $(p^{**}, \bar{p})$ , as discussed in Section 6.1 with  $\bar{p} > p^{**}$ . The coupling of the pure strategy profile  $s^{**}$  with such a prior belief of the principal cannot constitute a PBE, because, otherwise, an zero-probability event could be observed on an equilibrium path and lead to contradiction with the prior itself, which is absurd.

Given this result, one naturally wonders if imposing the restriction that “probabilistic beliefs should have full support” is sufficient to ensure all PBE deliver this BFE outcome. The answer is negative – it can be shown (as in Proposition 8) that even with full support, there exist other MPBE outcomes that are different from the BFE outcome. What these other multiple MPBE predict are unlikely to be genuinely stable structures of human interactions.

### 6.3 Markov Perfect Equilibrium for Games with Incomplete Information

According to Maskin and Tirole (2001), one of the practical reasons for the prominence of Markov strategy and equilibrium in Markov strategies, e.g., Markov Perfect Equilibrium (MPE), in applied game theory is that “MPE is often quite successful in eliminating or reducing a large multiplicity of equilibria in dynamic games, and thus in enhancing the predictive power of the model.” Behind such practical virtual, there is also the philosophical appeal that “Markov strategies prescribe the *simplest form of behavior that is consistent with rationality*”. For dynamic games with observable actions (i.e., perfect or almost perfect information), Maskin and Tirole (2001) demonstrate that this philosophical principle can be formulated precisely and consistently as to give exact formal definitions of Markov strategy and MPE. “Informally, a Markov strategy depends only on payoff-relevant past events. More precisely, it is measurable with respect to the coarsest partition of histories for which, if all other players use measurable strategies, each player’s decision-problem is also measurable.”

For dynamic games with incomplete information, there is, however, no established formal definition of equilibrium in Markov strategies, e.g., Markov Perfect Bayesian Equilibrium (MPBE). For games with incomplete information, the evaluation of payoff depends on (posterior) belief, the inference about which depends on equilibrium. Therefore the judgement about whether a past event is payoff-relevant can also depend on the equilibrium. Without specification of an equilibrium it is not clear what is meant by the “coarsest” partition of histories. Therefore the definition provided by Maskin and Tirole (2001) does not apply to games with incomplete information in general, and dynamic games (with incomplete information) under ambiguity in particular.

Our application of the concepts Markov strategy, Markov perfect objective equilibrium (MPOE) and Markov perfect Bayesian equilibrium (MPBE) proceed on a case-by-case basis, without the guidance of a generic definition. Our usage of the concept of Markov strategy requires that it must “prescribe the *simplest form of behavior that is consistent with rationality*”, which is interpreted in the specific contexts of the models. For the example MPOE analyzed in this paper, it is impossible to reduce the list of state variables any further. The only potential candidate for reduction is  $q_t$ . If  $q_t$  were removed, then the agent could safely behave naively without fearing revealing sensitive information because the principal’s Markov strategy did not depend on such information. Given the agent’s naive investment strategy, it would be optimal for the principal to experiment with reduced reimbursement rates and learn about the agent’s true type and benefit from this sensitive information. That means an MPOE must have a state variable like  $q_t$ , which summarizes the payoff-relevant knowledge of

the uninformed principal about the private information of her opponent. This state variable can capture the informational strategic effect (e.g., ratchet effect) on the players' behavior. One potential benefit of focusing the analysis on Markov equilibrium is to avoid multiplicity of equilibrium outcomes. This is achieved by the MPOE of the delegated experimentation model (see Theorem 4). In contrast, the same cannot be said about the MPBE, as Proposition 8 demonstrates, there exist a continuum of trigger mechanisms represented by state variable  $c_t$ . The two state variables  $q_t$  and  $c_t$ , representing the ratchet mechanisms, are both of the simplest form of behavior that is consistent with rationality.

One of the original motivation for favoring Markov strategy was the view that it would necessarily help avoid variables which are not (directly) payoff-relevant, but facilitate the “bootstrapping” property. According to Maskin and Tirole (2001), “[Markov] strategies depend on as few variables as possible; they involve no complex ‘bootstrapping’ in which each player conditions on a particular variable only because others do the same.” Although this view may be valid for dynamic games with observable actions (i.e., perfect or almost perfect information), it is not generally valid for dynamic games with incomplete information, as has been illustrated by the counter example used in the paragraph above. Taking another (counter-)example – the model described by Proposition 8, if  $c_t$  is eliminated from the set of state variables, then there exists no MPBE which only depends on the remaining state variables. More elaborately, if both players did not condition on the variable  $c_t$ , then the solution to the agent's dynamic programming problem (i.e., his Bellman equation) would be the naive investment strategy like (22), in which case, the principal would have incentive to experiment with reduced reimbursement rates and learn about the agent's true type and benefit from this sensitive information. That means the principal would condition on (at least) a new state variable that could keep track of his knowledge about the agent's type. As a result, no MPBE exists that only depends on the reduced set of state variables (excluding  $c_t$ ). This result is in contrast with the benchmark model analyzed in Section 5.1, where the principal knows the agent's type and the game is with observable actions (including nature's action). In the complete information benchmark model, the Markov perfect equilibrium strategies do not depend on variables like  $q_t$  and  $c_t$ , and indeed do not involve “bootstrapping”.

In conclusion, for dynamic games with incomplete information, Markov strategy and the “bootstrapping” property are not (necessarily) mutually exclusive. Consequently, focusing on Markov strategies *per se* may not guarantee success in “eliminating or reducing a large multiplicity of equilibria in dynamic games, and thus in enhancing the predictive power of the model”, as is demonstrated by Proposition 8.

## 6.4 Ambiguity Aversion: Tastes Separated from Beliefs

The new equilibrium concept proposed in this paper – perfect objectivist equilibrium (POE) – emphasizes the objectivity of players' beliefs and the validity of their inferences whenever learning is involved. In the current formulation, our definition of POE may seem to be tied up with the max-min decision criterion, which undoubtedly involves personal tastes, which

are subjective. In this section we discuss the issue of separation of beliefs from tastes, and explain how it is achieved by our theory.

In a critical assessment of the ambiguity aversion literature, Al-Najjar and Weinstein (2009) suggest that it is essential for any satisfactory generalization of the Bayesian methodology to preserve a *substantive* separation of tastes from beliefs, in the sense best embodied by the following quote from Aumann (1987): “[U]tilities directly express tastes, which are inherently personal. It would be silly to talk about ‘impersonal tastes,’ tastes that are ‘objective’ or ‘unbiased’. But it is not at all silly to talk about unbiased probability estimates, and even to strive to achieve them. On the contrary, people are often criticized for wishful thinking – for letting their preferences color their judgement. One cannot sensibly ask for expert advice on what one’s tastes should be; but one may well ask for expert advice on probabilities.”

As has been established in Section 3.2, the players’ preferences underpinning our definition of POE satisfy the criterion that an optimal solution for a player must not be weakly dominated (within updated set of beliefs). This when coupled with the max-min criterion is stronger than the max-min criterion. This additional criterion does have force: In Section A.8 of the Appendix, we give an example of a weakly dominated strategy that satisfies the max-min criterion. This was proposed by Manski (2008) as an intuitive criterion. Since the dual preference relations proposed by GMMS (2010) form the basis for a lexicographical (overall) preference relation, we propose that suitable extension of GMMS (2010) could easily provide an axiomatic foundation for this (lexicographical) criterion. The key idea is that whenever the objectively and subjectively rational preference relations give different ordering between a pair of action plans, i.e., the former gives a strict preference while the latter gives an indifference, the former should dominate the latter in the overall ordering. As has been alluded in the introduction, objective rationality is the primary preference relation and subjective rationality is secondary. In contrast, although GMMS (2010) require that subjectively rational preferences be consistent with the objectively rational preferences (in a weak sense), this requirement is not strong enough to rule out all inconsistency between them. In GMMS (2010), subjective rationality is solely responsible for decision making, therefore the preference relation is not lexicographical. The benefit of the GMMS (overall) preference relation is technical – it can be fully represented by a utility function. The cost, though, is epistemic – it has less respect to objectivity. We believe it is compelling to argue that strict objectively rational preference relations should matter (for pairwise choices) whenever they are well-defined. Therefore, even when we adopt the max-min criterion, we treat it only as a necessary condition for rational decision making.

We therefore agree with Al-Najjar and Weinstein (2009) on the essentiality of this separation criterion. We further suggest that for a theory of decision making under (non-degenerate) ambiguity to meet this criterion, it is essential to have a lexicographical overall preference relation, part of which is the objectively rational preference relation formulated by GMMS (2010), and originally proposed by Bewley (1986, 2002). On the one hand, the ambiguous beliefs of a decision maker should be derived from this (incomplete) preference

relation alone.<sup>31</sup> On the other hand, the tastes of the decision maker over ambiguity should be derived only from the subjectively rational preference relation, which can be as formulated either by GMMS (2010), or a well-founded alternative. Given the clear division of labor between the objectively and subjectively rational relations in the lexicographical (overall) preference relation, the *substantive* separation of tastes from beliefs is achieved.

In the limiting case of degenerate ambiguous beliefs, i.e., single distribution belief, both objective and subjective rationality degenerate into the identical Bayesian rationality, where a *substantive* separation of tastes from beliefs is achieved if the single prior is objectively established according to Aumann (1987).

It could be argued that the max-min criterion for decision under ambiguity may be too cautious or pessimistic.<sup>32</sup> Actually, Ellsberg (1961) – the seminal article that introduced the notion of “ambiguity” – already considered the first model that was capable of representing a continuum of varying degrees of ambiguity aversion.<sup>33</sup> Now there is a rich variety of models of decision making under ambiguity that consider weaker representations of aversion to ambiguity. (See Gilboa and Marinacci (2011) for a survey of axiomatic decision theoretic models.) There is yet no known compelling argument against extending the definition of POE to allow it to be compatible to some other well-founded criterion for decision under ambiguity. For games with more than one players who have ambiguous beliefs, it is desirable for a general definition of POE to be able to accommodate different attitudes toward ambiguity across players.

The admission of a subjective element in acting upon ambiguous beliefs is not inevitable, if and only if a belief-free equilibrium exists and is played. In that case, objectively rational preferences happen to be adequate for decision making, and therefore subjectively rational preferences are not involved. In general, the admission of subjective elements opens a challenging new question – How free should the preferences over ambiguity be allowed? This is a generic open question. In what follows we narrow it to more specific settings. Obviously, all rational preferences over ambiguity must satisfy certain consistency requirements. Some set of such consistency requirements can be mapped into (or represented as) a commitment to certain methodology of handling ambiguous beliefs. As long as the methodology has a definitive characterization (i.e., is well-defined), it can be critically assessed. Then a key specific question one faces is: Should objectivity be a necessary admission criterion for the methodology?<sup>34</sup> Take the extreme example of complete ignorance (as in the current model).

---

<sup>31</sup> It is noteworthy that the objectively rational preference relation satisfies the independence axiom (or the Savage “sure-thing” principle), the lack of which is the target of many criticism levelled against the ambiguity-aversion literature (see for example, Al-Narjjar and Weinstein, 2009).

<sup>32</sup> This view can be traced back to a remark by Savage (1951, p. 63), which was part of his review of Wald (1950): “Application of minimax [corresponding to max-min here] rule … is indeed ultra-pessimistic; no serious justification for it has ever been suggested, and it can lead to the absurd conclusion in some cases that no amount of relevant experimentation should deter the actor from behaving as though he were in complete ignorance.”

<sup>33</sup> Ellsberg (1961) used a utility function which is a convex combination of the minimum expected utility function and the expected utility function based on the best-estimated probability distribution.

<sup>34</sup> The term ‘objective’ (as opposed to ‘personal’) means ‘inter-personal’, or ‘agreeable to any (reasonable)

The max-min plus not-weakly-dominated criterion for decision making ( $\text{max-min}^+$ ) does not entail any subjective methodology of handling beliefs. In contrast, subjective expected utility maximization does entail a methodology which cannot meet the criterion of objectivity. Should methodological rationality exclude the latter while admit the former (in the case of complete ignorance)?

The  $\text{max-min}^+$  criterion is the simplest model of ambiguity aversion, or prudence. Whether this model is too cautious or pessimistic, and prone to “lead to the absurd conclusion”, really depends on the context where the model is applied. In the current model of repeated bargaining with interactive learning, it does not seem to be the case.<sup>35</sup> Here it makes more interesting comparison, between the two extreme cases: max-min versus subjective expected utility maximization. Note, the conventional Bayesian variant of the model can also be seen as replacing  $\text{max-min}^+$  criterion with less ambiguity-averse tastes over ambiguity (while keeping the initial underlying ambiguity the same). While the principal’s rational beliefs are represented by the set of multiple probability measures, she deliberately compresses the set of beliefs into a singleton without (valid) epistemic justification but in order to appear to be ambiguity-neutral. In this light, the unique subjective probability measure no longer represents her (objectively established) rational beliefs, but a product of personal tastes over ambiguity. In meeting the criterion proposed by Al-Najjar and Weinstein (2009) on separation of tastes from beliefs, the conventional Bayesian model appears to be less satisfactory after all. As has been shown in Section 6.1, it also suffers from the problem of extreme multiplicity of equilibrium. From a normative point of view, this less ambiguity-averse model is probably much worse than the  $\text{max-min}^+$  model. From a descriptive (positive) point of view, the less ambiguity-averse model certainly has extremely weaker predictive power or testability.

## 6.5 Learning under Ambiguity and Dynamic Consistency

In a recent authoritative survey of the ambiguity literature, Gilboa and Marinacci (2011) suggest that Bayes’s rule can be extended to non-Bayesian beliefs in more than one ways. Two typical examples are *full Bayesian updating (FBU)* and *maximum likelihood updating (MLU)*. If beliefs are given by a set of priors  $C$ , and event  $B$  is known to have occurred,

---

person’. In this sense, the classical (Fisherian and frequentist) statistics can be seen as a more objective methodology for statistic inference than the Bayesian methodology (excluding the “robust Bayesian”). (For philosophical discussion about objective knowledge, see Popper (1972).)

<sup>35</sup>In the current model, the experience (observation) data is not noisy and thus allows learning to abstract from the statistical inference problem and focus on the identification problem. In this case, talking about logical refutation (in stead of statistical rejection) of a hypothesis makes perfect sense. (For philosophical discussions of the meaning of ‘refutation’, see Popper 1959 and 1963.)

Ironically indeed, in the unique MPOE outcome, the principal does behave as if she were in complete ignorance about the agent’s type. Furthermore, this kind of result is precisely the pursuit of the literature on belief-free equilibrium. There is, however, nothing absurd about this.

In contrast, if in a learning process the experience data is intrinsically noisy (such as learning about population property from small sample), then the notion of refutation of a hypothesis may not apply, and therefore the max-min criterion may appear absurdly ultra-pessimistic.

then the FBU set of posteriors (on  $B$ ) is given by<sup>36</sup>

$$C_B = \{p(\cdot|B) | p \in C\}.$$

The MLU set of posteriors is given by

$$C_B^M = \left\{ p(\cdot|B) | p \in \arg \max_{q \in C} q(B) \right\}.$$

By definition,  $C_B^M \subseteq C_B$  for all  $B$ . A sufficient condition for  $C_B^M = C_B$  for all  $B$  is that the set of priors  $C$  is large enough. More precisely, we have the following definition and proposition.

**Definition 9** *The ambiguous beliefs represented by the set of priors  $C$  is sufficiently modest if for any  $B$  (which is possible to be proven to have occurred), the following conditions are satisfied: (i)  $\max_{p \in C} p(B) = 1$ ; (ii)  $C_B \subseteq C$ .*

Condition (i) says that  $C$  is sufficiently large such that for any  $B$  (which is possible to be proven to have occurred) there must exist  $p \in C$  such that  $p(B) = 1$ . Condition (ii) says that  $C$  is sufficiently large such that the Bayesian update of any prior on any conditioning event must coincide with a prior (this can be itself or another one) belonging to  $C$ .

**Theorem 5** *If  $C$  represents sufficiently modest beliefs, then  $C_B = C_B^M$  for all  $B$ .*

The idea for proof is the following: If condition (i) is satisfied then for any  $B$ , we have

$$C_B^M = \{p(\cdot|B) | p(B) = 1, p \in C\} = \{q \in C | q(B) = 1\}.$$

If condition (ii) is also satisfied then for any  $B$  and any  $p(\cdot|B) \in C_B$ , we have  $p(\cdot|B) \in C$ ,  $p(B|B) = 1$ , and hence  $p(\cdot|B) \in \{q \in C | q(B) = 1\}$ . It follows that  $C_B \subseteq \{q \in C | q(B) = 1\} = C_B^M$ . Since  $C_B \supseteq C_B^M$  for all  $B$ , we must have

$$C_B = C_B^M \text{ for all } B.$$

In the modelling of learning under ambiguity in the current paper, we assume the set of the priors is  $\mathcal{M}_0$ , which is already the largest feasible set of priors. Therefore the conditions for sufficiently modest beliefs are satisfied. Furthermore, and we establish that FBU and MLU are equivalent for our model. There is another general criterion for saying  $C$  is large enough, the concept of rectangularity proposed by Epstein and Schneider (2003). The set of

---

<sup>36</sup>Notations used in this section are adopted from Gilboa and Marinacci (2011), independently of the rest of our paper. Here the definition of the observable conditioning event  $B$  is based on the state space, i.e.,  $B$  is subset of the state space. Therefore, each (non-degenerating) conditioning event leads to some (partially) identification of the true state. The implicitly assumption here is that the observation has identification power, or equivalently falsification (or refutation) power; and the inference does not involve the statistical inference problem.

priors  $C$  is rectangular if “it can be decomposed into a set of current-period beliefs, coupled with next-period conditional beliefs, in such a way that any combination of the former and the latter is in  $C$ ” (Gilboa and Marinacci, 2011).<sup>37</sup> It is an open question as to the relationship between the sufficient modesty condition stated above and the rectangularity condition, particularly, whether inclusion relation applies one way or the other, or both.

The concept of rectangularity has another important application; i.e., it is closely related to the axiom of dynamic consistency (Epstein and Schneider, 2003). It has been recognized in the literature of dynamic models of ambiguous beliefs that the axiom of dynamic consistency is necessary for Bayes’s rule to be respected in updating of beliefs. In our model, rectangularity is satisfied and Bayes’s rule is respected in the sense that it is never violated, but also becomes redundant (or trivial) when we use maximum likelihood updating for likelihood inference. The Bayesian updates of the priors that generate the maximum likelihood are simply themselves.

## 7 Summary and Conclusions

In this paper, we study a dynamic game of incomplete information in a model of delegated experimentation that features repeated bargaining. The principal seeks a solution to a problem and contracts with an agent to find it, but a solution to the problem may not exist. Neither the principal nor the agent knows for certain whether the problem can be solved, but the agent’s expertise allows him to formulate a prior probability  $p_0$  that the problem is solvable. This prior belief is private information to the agent. Unlike standard models of contracting with asymmetric information, we do not assume that the principal has a unique prior distribution over the agent’s prior. Instead, we allow the principal to hold multiple prior distributions over  $p_0$ . This gives rise to a process by which the principal iteratively selects beliefs over  $p_0$  based on observational facts and a likelihood function that is derived from the common knowledge about the game and the concept of equilibrium. Following the work of Manski (2008) and GMMS (2010), the principal is assumed to maximize a worst-case value function that reflects an aversion to ambiguity. As a solution concept for the game, we propose an extension of perfect Bayesian equilibrium to a setting in which the uninformed principal has multiple priors. We call this extension the perfect objectivist equilibrium, or POE. We fully characterize the Markov perfect objectivist equilibrium or MPOE for the game of delegated experimentation, and show that there is a unique MPOE outcome, in which the principal offers the agent a simple linear contract that involves a time-invariant reimbursement of the agent’s costs of investment in solving the problem. This induces the agent to choose the investment profile and cumulative investment that is Pareto efficient (allowing for compensation). The reimbursement rate simply equates the share of cost to the share of benefit.

Our analysis of the simple game of repeated bargaining with interactive learning (about

---

<sup>37</sup>We have not yet been able to locate any formal proof of the claim suggested by Gilboa and Marinacci (2011) that  $C_B^M = C_B$  if  $C$  is rectangular.

permanent private information) delivers some powerful messages. From a descriptive point of view, the message is that the fact some economic players hold ambiguous beliefs and are prudent against ambiguity could be an important determinant of bargaining solutions. From a normative point of view, the moral lesson is that objectively and truthfully acknowledging what is known as well as what is not known, and being prudent about the resulting ambiguity could be a constructive way toward resolving conflict of interests in bargaining situations.

The novel findings of the current paper will, inevitably, arouse questions about the generality of the main results. For example, in the (generic) interest in game theory, two obvious questions can be raised: How general are the results on existence of (pure strategy) POE, and the uniqueness of MPOE outcome? What general conditions do these results entail? Given the early stages of this research programme, these important questions are open; and need to be explored in the next step. Other questions concerning generality can also be directed toward the results of Pareto efficiency of the bargaining outcome and the existence of belief-free equilibrium. For these, it is rather intuitive to speculate that the results are unlikely to be general. For example, if there is ambiguity about multiple parameters of the model (say,  $p_0$  as well as  $\Pi_A$ ), then the conservative reservation price set by the prudent principal will likely to be below the socially efficient level.<sup>38</sup> Although the agent knows that the bargaining outcome is not Pareto efficient, he has no credible means to transmit this information because of conflict of interests. Counterfactually, if the principal knew the true type of the agent and realized that the price were too low, she would increase the offer price. This implies that the MPOE strategy profile can no longer be a belief-free equilibrium. The (potential) lack of generality for the efficiency result and the existence of belief-free equilibrium should neither be seen as a surprise, nor viewed as a disappointment of this new research programme. Eventually, this research programme should be assessed by its overall success (or lack of it) in *appropriately* formalizing the role of incomplete knowledge (including complete ignorance), asymmetric information and interactive learning in the emergence of stable structures of human interactions.

The idea that the principal may not be able to form a single prior over private information has potentially broad applicability. In principle, it allows better representation of ignorance in a player's belief and hence more adequate account of the extent and scope of informational asymmetry between players. This is a breakthrough relative to the conventional Bayesian mechanism design literature. For example, in mechanism design models of optimal regulation (Baron and Myerson, 1982; Laffont and Tirole, 1986) or of non-linear pricing (Maskin and Riley, 1984), the principal's (single) prior belief plays a critical role in shaping the nature of the mechanism. We intend to explore how the possibility multiple priors held by an ambiguity-averse principal changes the implications of these important benchmark models. In preliminary work (Besanko, Tong, Wu, 2011), we find that the Baron and Myerson mechanism does not emerge as a POE when the principal can hold multiple

---

<sup>38</sup>Suppose the principal only knows the range of  $\Pi_A \in [\underline{\Pi}_A, \bar{\Pi}_A]$  with  $\underline{\Pi}_A > 0$ , then the principal's reservation price will be  $\phi = \frac{\underline{\Pi}_A}{\underline{\Pi}_A + \Pi_P} \leq \frac{\bar{\Pi}_A}{\underline{\Pi}_A + \Pi_P} \equiv \phi^{**}$ . If priced above  $\phi$ , conditional on the agent accepting the offer, the principal would fear that the agent were a low  $p_0$  but high  $\Pi_A$  type who would cause her a net loss.

priors.<sup>39</sup>

More generally, we hope that the concept of multiple priors and the POE can provide a valuable framework for analyzing delegated experimentation with even richer structures than the setting explored here. Among the interesting extensions in this direction would be exploration of general contract spaces, unobservable action profiles that give rise to noisy signals, noise due to measurement errors, and the possibility that a given agent may be replaced by other agents if performance in achieving a breakthrough seems unsatisfactory.

## A Appendix

### A.1 Proof of Proposition 1

The proof is essentially identical to that of Proposition 3.1 in Keller, Rady and Cripps (2005) and is thus omitted.

### A.2 Proof of Theorem 1

Given the pure strategy equilibrium with strategy profile  $s$ , for each realization of  $p_0$  the equilibrium path is given by  $f^t(s, p_0)$  and the likelihood function is uniquely given by  $l(h^t; t, s, p_0) = \delta_{f^t(s, p_0)}(h^t)$ , which implies (13). We are left to prove (14) holds. First of all, given the realized on-equilibrium-path history  $h^t$  it can be inferred that all conditional plausible states belong to  $\mathcal{I}_d(h^t)$ . That is,  $\mathcal{I}_d(h^t)$  is the conditioning event for Bayesian updating. As a result, for all  $\mu_0 \in \mathcal{M}_0$  such that  $\mu_0(\mathcal{I}_d(h^t)) > 0$ , the posterior  $\mu_t(h^t; \mu_0, l)$  is well-defined and for all  $A \subseteq \mathcal{I}_d(h^t)$  we have

$$\mu_t(A|h^t; \mu_0, l) = \frac{\int_{p_0 \in A} \delta_{f^t(s, p_0)}(h^t) d\mu_0(p_0)}{\int_0^1 \delta_{f^t(s, p_0)}(h^t) d\mu_0(\tilde{p}_0)} = \frac{\int_{p_0 \in A} \delta_{f^t(s, p_0)}(h^t) d\mu_0(p_0)}{\mu_0(\mathcal{I}_d(h^t))},$$

which implies

$$\mu_t(\mathcal{I}_d(h^t) | h^t; \mu_0, l) = \frac{\int_{p_0 \in \mathcal{I}_d(h^t)} \delta_{f^t(s, p_0)}(h^t) d\mu_0(p_0)}{\mu_0(\mathcal{I}_d(h^t))} = 1.$$

Since  $\mathcal{M}_0$  includes all plausible probability distribution over  $\Sigma$ , i.e.,  $\mathcal{M}_0$  is sufficiently large, we must have  $\mu_t(h^t; \mu_0, l) \in \mathcal{M}_0$ . This implies for all  $\alpha \in [0, 1]$ ,

$$\mathcal{M}_t^\alpha(h^t) \subseteq \{\mu_0 \in \mathcal{M}_0 | \mu_0(\mathcal{I}_d(h^t)) = 1\}.$$

Second, for any on-equilibrium-path history  $h^t$ , we have

$$\begin{aligned} \max_{\tilde{\mu}_0 \in \mathcal{M}_0} L\left(h^t; t, s, \tilde{\mu}_0, \tilde{l}\right) &= \max_{\tilde{\mu}_0 \in \mathcal{M}_0} \int_0^1 \delta_{f^t(s, p_0)}(h^t) d\tilde{\mu}_0(p'_0) \\ &= \int_0^1 \delta_{f^t(s, p_0)}(h^t) d\delta_{p_0}(p'_0) \Big|_{p_0 \in \mathcal{I}_d(h^t)} = \delta_{f^t(s, p_0)}(h^t) \Big|_{p_0 \in \mathcal{I}_d(h^t)} = \delta_{h^t}(h^t) = 1. \end{aligned}$$

---

<sup>39</sup>More precisely we refer to the counterpart of “perfect objectivist equilibrium” in a static game, which is called “objectivist Nash equilibrium” (ONE), as an extension to conventional “Bayesian Nash equilibrium”.

As a result, for all  $\mu_0 \in \mathcal{M}_0$  such that  $\mu_0(\mathcal{I}_d(h^t)) = 1$ , we have

$$\mu_t(h^t; \mu_0, l) = \mu_0,$$

for  $l \in \mathcal{L}_0(s)$ , and

$$L(h^t; t, s, \mu_0, l) = 1 \geq \alpha \max_{\substack{\tilde{\mu}_0 \in \mathcal{M}_0 \\ \tilde{l} \in \mathcal{L}_0(s)}} L\left(h^t; t, s, \tilde{\mu}_0, \tilde{l}\right) = \alpha.$$

It follows that

$$\mathcal{M}_t^\alpha(h^t) \supseteq \{\mu_0 \in \mathcal{M}_0 | \mu_0(\mathcal{I}_d(h^t)) = 1\}.$$

Overall, we have for all  $\alpha \in [0, 1]$

$$\mathcal{M}_t^\alpha(h^t) = \{\mu_0 \in \mathcal{M}_0 | \mu_0(\mathcal{I}_d(h^t)) = 1\}.$$

### A.3 Proof of Theorem 2

Before we prove the main claims, we start with some preliminary results.

Let  $\mathcal{A}$  be a set of probability measures. By definition, the support of  $\mathcal{A}$  is represented by

$$P(\mathcal{A}) = \bigcup_{\mu \in \mathcal{A}} \{x \in [0, 1] | \mu(x) > 0\}$$

and  $cl(\mathcal{A})$  includes all  $\mu \in \mathcal{A}$  and all limit points  $\mu'$  of  $\mathcal{A}$ . That is  $\mu' = \lim_{n \rightarrow \infty} \mu_n$  where  $\mu_n \in \mathcal{A}$ . Hence

$$P(\mu') = \left\{x \in [0, 1] \mid \lim_{n \rightarrow \infty} \mu_n(x) > 0\right\}.$$

As a preliminary step, we establish that if  $\mathcal{A}$  includes the set of all Dirac measures over support  $P(\mathcal{A})$ , then

$$P(cl(\mathcal{A})) = cl(P(\mathcal{A})).$$

First, we show  $cl(P(\mathcal{A})) \subseteq P(cl(\mathcal{A}))$ .

Let  $x \in cl(P(\mathcal{A}))$ . If  $x \in P(\mathcal{A})$ , then there exists  $\mu \in \mathcal{A}$  such that  $\mu(x) > 0$ . Because  $\mu \in cl(\mathcal{A})$ , we must have  $x \in P(cl(\mathcal{A}))$ .

Suppose  $x \in cl(P(\mathcal{A})) \setminus P(\mathcal{A})$ , i.e.,  $x$  is a boundary point of  $P(\mathcal{A})$ , then there exists  $x_N \in P(\mathcal{A})$  and  $x_N \rightarrow x$ . Define  $\mu_N = \delta_{x_N} \in \mathcal{A}$ , we must have  $\mu_N \rightarrow \mu = \delta_x$  (convergence in distribution), which implies  $\mu \in cl(\mathcal{A})$  and thus  $x \in P(cl(\mathcal{A}))$ .

Second, we show  $P(cl(\mathcal{A})) \subseteq cl(P(\mathcal{A}))$ .

Let  $x \in P(cl(\mathcal{A}))$ , then there exists  $\mu \in cl(\mathcal{A})$  such that  $\mu(x) > 0$ . If  $\mu \in \mathcal{A}$ , then we must have  $x \in P(\mathcal{A}) \subseteq cl(P(\mathcal{A}))$ . Now suppose  $\mu \notin \mathcal{A}$ , then  $\mu$  is a boundary point of  $\mathcal{A}$  and there exists  $\mu_N \in \mathcal{A}$  such that  $\mu_N \rightarrow \mu$  (convergence in distribution). If  $\mu_N(x) > 0$ , then  $x \in P(\mathcal{A}) \subseteq cl(P(\mathcal{A}))$ . If  $\mu_N(x) = 0$  for all  $N$ , then we must have  $\lim_{N \rightarrow \infty} \mu_N(x) = 0$ . Then  $\mu_N \rightarrow \mu$  implies  $x$  is a mass point for  $\mu$ .

Suppose  $x \notin cl(P(\mathcal{A}))$ , that is,  $x$  is an exterior point of  $P(\mathcal{A})$ . Then there must exist  $\varepsilon > 0$  such that  $(x - \varepsilon, x + \varepsilon) \not\subseteq P(\mathcal{A})$  and  $F_\mu$  is continuous at  $(x - \varepsilon, x)$  or  $(x, x + \varepsilon)$ , where the CDF  $F_\mu$  is defined by

$$F_\mu(x') \equiv \int_0^{x'} d\mu(z)$$

for all  $x' \in [0, 1]$ .

We distinguish three cases: (a)  $x = 0$ ; (b)  $x \in (0, 1)$ ; (c)  $x = 1$ .

(a)  $x = 0$ : There must exist  $\varepsilon > 0$  such that  $(0, \varepsilon) \not\subseteq P(\mathcal{A})$  and  $F_\mu(x')$  is continuous at  $x'$  for all  $x' \in (0, \varepsilon)$ . On the one hand,

$$\lim_{N \rightarrow \infty} F_{\mu_N}(x') = F_\mu(x') > 0.$$

On the other hand, since  $F_{\mu_N}(x') = F_{\mu_N}(0) = 0$  for all  $N \in \mathbb{N}$ , we must have

$$\lim_{N \rightarrow \infty} F_{\mu_N}(x') = 0.$$

The above is a contradiction, which implies case (a) is empty.

(b)  $x \in (0, 1)$ : There must exist  $\varepsilon > 0$  such that  $(x - \varepsilon, x + \varepsilon) \not\subseteq P(\mathcal{A})$ ,  $F_\mu(x')$  is continuous at  $x'$  for all  $x' \in (x - \varepsilon, x)$  and  $F_\mu(x'')$  is continuous at  $x''$  for all  $x'' \in (x, x + \varepsilon)$ . On the one hand,

$$\lim_{N \rightarrow \infty} F_{\mu_N}(x') = F_\mu(x') < F_\mu(x'') = \lim_{N \rightarrow \infty} F_{\mu_N}(x'').$$

On the other hand, since  $F_{\mu_N}(x) = F_{\mu_N}(x') = F_{\mu_N}(x'')$ , we must have

$$\lim_{N \rightarrow \infty} F_{\mu_N}(x) = \lim_{N \rightarrow \infty} F_{\mu_N}(x') = \lim_{N \rightarrow \infty} F_{\mu_N}(x'').$$

We have therefore derived a contradiction which implies that case (b) is also empty.

(c)  $x = 1$ : There must exist  $\varepsilon > 0$  such that  $(1 - \varepsilon, 1) \not\subseteq P(\mathcal{A})$  and  $F_\mu(x')$  is continuous at  $x'$  for all  $x' \in (1 - \varepsilon, 1)$ . On the one hand,

$$\lim_{N \rightarrow \infty} F_{\mu_N}(x') = F_\mu(x') < 1.$$

On the other hand, since  $F_{\mu_N}(x') = F_{\mu_N}(1) = 1$ , we must have

$$\lim_{N \rightarrow \infty} F_{\mu_N}(x') = \lim_{N \rightarrow \infty} F_{\mu_N}(1) = 1.$$

We have therefore derived a contradiction which implies that case (c) is empty too.

We therefore have established that  $x \in cl(P(\mathcal{A}))$ .

Note, By Theorem 1,  $\mathcal{M}_t(h^t)$  is the set of all probability measures over the set  $\mathcal{I}_d(h^t)$ , where  $\mathcal{I}_d(h^t)$  is the set of  $p_0$  such that  $h^t$  is predicted by  $f^t(s, p_0)$ . This implies

$$P(\mathcal{M}_t(h^t)) = \mathcal{I}_d(h^t),$$

$$P(\mathcal{D}_t(h^t)) = \mathcal{I}_d(h^t),$$

and that  $\mathcal{D}_t(h^t)$  is the set of all Dirac measures over the set  $\mathcal{I}_d(h^t)$ .

Let  $\mathcal{A} = \mathcal{M}_t(h^t)$  and  $\mathcal{A} = \mathcal{D}_t(h^t)$  respectively, we therefore establish respectively that

$$P(cl(\mathcal{M}_t(h^t))) = cl(P(\mathcal{M}_t(h^t))),$$

$$P(cl(\mathcal{D}_t(h^t))) = cl(P(\mathcal{D}_t(h^t))).$$

Next we note that  $cl(\mathcal{D}_t(h^t))$  is the set of all Dirac measures over the set  $cl(P(\mathcal{D}_t(h^t)))$ . This is obvious because by definition  $cl(\mathcal{D}_t(h^t))$  is the set of all Dirac measures over the set  $P(cl(\mathcal{D}_t(h^t)))$ .

Now we prove the main claims. First, (i)  $\Leftrightarrow$  (ii) is due to the fact that  $cl(\mathcal{M}_t(h^t))$  includes  $cl(\mathcal{D}_t(h^t))$ , which contains all Dirac measures over  $cl(P(\mathcal{M}_t(h^t)))$ . All members of  $cl(\mathcal{M}_t(h^t))$ , including all solutions of the minimization problem in expression (i), are a convex combination of some Dirac measures over  $cl(P(\mathcal{M}_t(h^t)))$ . There must exist a corner solution to the minimization problem in expression (i), that is a Dirac measure that belongs to  $cl(\mathcal{D}_t(h^t))$  and also minimizes  $W_t(p_t(K_t; p_0) | s, p_0)$ . To see this, let  $\tilde{\mu}_0 \in cl(\mathcal{M}_t(h^t))$  be a worst-case distribution that solves the minimization problem in expression (i). Let  $P(\tilde{\mu}_0)$  be the support of  $\tilde{\mu}_0$ . Then there must exist  $\tilde{p}_0 \in P(\tilde{\mu}_0)$  such that  $\delta_{\tilde{p}_0}$  is also a worst-case distribution that solves the minimization problem in expression (i). Suppose the opposite, that is,

$$W_t(p_t(K_t; \tilde{p}_0) | s, \tilde{p}_0) > \int_0^1 W_t(p_t(K_t; p_0) | s, p_0) d\tilde{\mu}_0(p_0)$$

for all  $\tilde{p}_0 \in P(\tilde{\mu}_0)$ . Integrating both sides of the above inequality w.r.t.  $\tilde{\mu}_0(\tilde{p}_0)$  over  $P(\tilde{\mu}_0)$ , we have

$$\begin{aligned} & \int_{\tilde{p}_0 \in P(\tilde{\mu}_0)} W_t(p_t(K_t; \tilde{p}_0) | s, \tilde{p}_0) d\tilde{\mu}_0(\tilde{p}_0) > \int_0^1 W_t(p_t(K_t; p_0) | s, p_0) d\tilde{\mu}_0(p_0) \\ & \Leftrightarrow \int_0^1 W_t(p_t(K_t; p_0) | s, p_0) d\tilde{\mu}_0(p_0) > \int_0^1 W_t(p_t(K_t; p_0) | s, p_0) d\tilde{\mu}_0(p_0) \end{aligned}$$

which is a contradiction. Since  $\delta_{\tilde{p}_0} \in cl(\mathcal{D}_t(h^t)) \subseteq cl(\mathcal{M}_t(h^t))$ ,  $\delta_{\tilde{p}_0}$  must also solve the minimization problem in expression (ii). As a result, (i)  $\Leftrightarrow$  (ii).

Finally, (ii)  $\Leftrightarrow$  (iii) is obvious because  $cl(\mathcal{D}_t(h^t))$  is the set of all Dirac measures over the set  $cl(P(\mathcal{M}_t(h^t)))$ .

## A.4 Proof of Proposition 2

The proof of Theorem 1 has already established the following::

$$\begin{aligned} P(cl(\mathcal{M}_t(h^t))) &= cl(P(\mathcal{M}_t(h^t))), \\ P(\mathcal{M}_t(h^t)) &= P(\mathcal{D}_t(h^t)), \\ P(cl(\mathcal{D}_t(h^t))) &= cl(P(\mathcal{D}_t(h^t))). \end{aligned}$$

The rest of the result follows immediately from Definition 3.

## A.5 Proof of Proposition 3

We must show that the principal's Markov strategy is optimal if and only if the principal chooses the reimbursement rate  $\tilde{\phi}_{t+dt}^*(K_t; p_0)$ . To show the sufficiency, we note that given the agent's strategy  $\tilde{k}_t^*(K_t, \phi_t; p_0)$ , the principal chooses  $\tilde{\phi}_t^*(K_t; p_0)$  and induces the agent to follow the socially efficient investment strategy specified by (25). Hence  $\tilde{\phi}_{t+dt}^*(K_t; p_0)$  must be optimal because it allows the principal to extract the maximum surplus.

To show the necessity, we note that for  $K_t \in [0, K^A(p_0)]$ , the principal must set  $\tilde{\phi}_t^*(K_t; p_0) = 0$  otherwise she can increase her payoff by lowering  $\tilde{\phi}_t^*(K_t; p_0)$  without changing the agent's investment level. Further more, we know

$$\lim_{K_t \rightarrow K^{**}(p_0)} \tilde{\phi}_t^*(K_t; p_0) = \phi^{**}$$

where  $K^{**}(p_0)$  is the optimal termination threshold value of  $K_t$ . This implies for  $K_t \in (K^A(p_0), K^{**}(p_0))$ , we also have  $\tilde{\phi}_t^*(K_t; p_0) = \tilde{\phi}_t(K_t; p_0)$ . Finally, for  $K_t \geq K^{**}(p_0)$ , we need to have  $\tilde{\phi}_t^*(K_t; p_0) \leq \phi^{**}$  otherwise the agent will overinvest. As a result<sup>40</sup>,

$$\begin{aligned} & \tilde{\phi}_{t+dt}^*(K_t; p_0) \\ &= \begin{cases} 0 & \text{if } K_t \in [0, K^A(p_0)] \\ 1 - \frac{\lambda \Pi_A}{a} p_{t+dt}(K_t + dt; p_0) & \text{if } K_t \in (K^A(p_0), K^{**}(p_0)) \\ \leq \phi^{**} & \text{if } K_t \geq K^{**}(p_0), \end{cases} \end{aligned}$$

is a necessary condition for  $\tilde{\phi}_{t+dt}^*(K_t; p_0)$  to be optimal.

From (22), the principal's compensation strategy  $\tilde{\phi}_{t+dt}^*(K_t; p_0)$  induces the investment policy given by (27). This keeps the agent's value the same as if there is no compensation transfer, and it strictly improves and maximize the principal's value for all  $K_t \geq 0$  in the absence of a breakthrough. The outcome is thus Pareto efficient for all  $K_t \geq 0$  in the absence of a breakthrough.

The proof of the rest of the claim is trivial.

## A.6 Proof of Proposition 4

We establish the first sentence of the Proposition by proving two preliminary claims.

**Claim (i):** *If there exists an MPOE such that the principal's Markov strategy takes the form  $\phi_{t+dt} = \phi(K_t, q_t)$ , then  $\phi_0 = \phi(0, q_0) = \phi^{**}$  for all state  $(0, q_0)$  that is on an equilibrium path.*

Suppose  $\phi_{t+dt} = \phi(K_t, q_t)$  and  $k_t = k(K_t, \phi_t, q_t; p_0)$  are the equilibrium Markov strategies of the principal and agent respectively. If the state (event)  $(K_{t-dt}, q_{t-dt})$  is on an equilibrium

<sup>40</sup>At a first glance, this formulation of the principal's Markov strategy may seem capable of contradicting the important claim that the Markov strategy does not directly depend on time. Notice that if  $K_{t+dt} = K_t$ , then there is no change to the state variables; nevertheless,  $\tilde{\phi}_{t+2dt}^*(K_{t+dt}; p_0)$  can differ from  $\tilde{\phi}_{t+dt}^*(K_t; p_0)$ , and this change is purely due to a difference in time. This seeming contradiction, however, disappears in the limit as  $dt \rightarrow 0$ . Thus the time-invariance property of the Markov strategy holds asymptotically.

path, then the offer  $\phi_t = \phi(K_{t-dt}, q_{t-dt})$  must be a credible take-it-or-leave-it offer, i.e., it is genuinely not renegotiable. Formally,  $k_t = 0$  implies  $\frac{dq_t}{dt} = 0$  and  $\frac{dK_t}{dt} = 0$ , and it follows  $\frac{d\phi_t}{dt} = 0$ . As a result,  $(K_{t'}, \phi_{t'}, q_{t'})$  becomes stationary for  $t' \geq t$ . In this case, the agent would reject the offer only if the net flow benefit is negative. Therefore, if  $k_t = 0$  then it must be valid to infer that

$$p_0 < \bar{p}(K_t, \phi_t)$$

and hence

$$\sup P_{t+dt} = \bar{p}(K_t, \phi_t).$$

Suppose  $\phi_0 < \phi^{**}$  in an MPOE, then  $k_0 = 0$  implies  $\frac{dq_t}{dt} = 0$  and  $\frac{dK_t}{dt} = 0$ , and it follows  $\frac{d\phi_t}{dt} = 0$ .  $(K_t, \phi_t, q_t)$  becomes stationary for  $t > 0$ . This strategy is not sequentially rational given that  $p_0 < \bar{p}(0, \phi_0)$  because  $\phi_{2dt} = \phi^{**}$  weakly dominates  $\phi_{2dt} = \phi_0$  for  $p_0 \in [0, \bar{p}(\phi_0, 0))$ . Note the former strictly dominates the latter for  $p_0 \in [p^{**}, \bar{p}(\phi_0, 0))$ , while they are equivalent for  $p_0 \in [0, p^{**}]$ . Thus the Markov strategy  $\phi(0, q_0) = \phi_0$  cannot be an equilibrium strategy.

Now suppose  $\phi_0 > \phi^{**}$ , then the type  $p_0 = p^{**} - \varepsilon$  would have incentive to invest, but the social surplus is negative and the type  $p_0$  agent at least breaks even, the expected loss must be borne by the principal. The worst case must be even worse, and the worst case value function for the principal must be negative. For all types  $p_0 \geq p^{**}$ ,  $\phi_0 = \phi^{**}$  is sufficient to induce  $k_0 = 1$  there  $\phi_0 > \phi^{**}$  is unnecessarily too high. In conclusion,  $\phi_0 > \phi^{**}$  must be suboptimal.

**Claim (ii):** *If there exists an MPOE such that the principal's Markov strategy takes the form  $\phi_{t+dt} = \phi(K_t, q_t)$ , then  $\phi_{t+dt} = \phi(K_t, q_t) = \phi^{**}$  for all state  $(K_t, q_t)$  that is on an equilibrium path.*

Suppose  $\phi_t < \phi^{**}$  for some  $t > 0$ , and  $\phi_\tau = \phi^{**}$  and  $k_\tau > 0$  for all  $\tau < t$ . Then  $k_t = 0$  implies  $\frac{dq_t}{dt} = 0$  and  $\frac{dK_t}{dt} = 0$ , and it follows  $\frac{d\phi_t}{dt} = 0$ . As a result,  $(K_{t'}, \phi_{t'}, q_{t'})$  becomes stationary for  $t' \geq t$ . This strategy is not sequentially rational given that  $p_0 < \bar{p}(K_t, \phi_t)$  because  $\phi_{t+2dt} = \phi^{**}$  weakly dominates  $\phi_{t+2dt}$  for  $p_0 < \bar{p}(K_t, \phi_t)$ . Thus the Markov strategy  $\phi(K_{t-dt}, q_{t-dt}) = \phi_t < \phi^{**}$  cannot be an equilibrium strategy.

Now suppose  $\phi_t > \phi^{**}$ , then the worst case value function for the principal is negative, which is suboptimal. Also, for all types such that  $p_t(K_t; p_0) \geq p^{**}$ ,  $\phi_t = \phi^{**}$  is sufficient to induce  $k_t = 1$  there  $\phi_t > \phi^{**}$  is suboptimally too high.

To show (34), we note that  $\phi_t = \phi^{**}$  means  $\phi_t$  is time-invariant and not dependent on  $q_t$ . Consequently,  $\frac{\partial V}{\partial q_t} = 0$ . Finally, (35) is due to the fact that if and only if  $k_t > 0$  we have

$$\bar{p}(K_t, \phi_t) = \bar{p}(K_{t-dt} + k_t dt, \phi^{**}) > \bar{p}(K_{t-dt}, \phi^{**}) = q_t.$$

## A.7 Proof of Theorem 3

### A.7.1 Preliminary Lemmas

**Lemma 2** *For  $t \geq 0$ , if for  $\forall \tau \geq t - dt$  the principal always plays the Markov strategy  $\phi_{\tau+dt}^{**}(K_\tau, q_\tau)$ , as defined in (36), then the following conditions hold:  $\phi_{t+dt}^{**}(K_t, q_t) \leq \phi^{**}$  and  $\frac{dq_t}{dt} > 0$  only if  $\phi_{t+dt}^{**}(K_t, q_t) = \phi^{**}$ .*

**Proof.** It is straightforward to verify that  $\phi_{t+dt}^{**}(K_t, q_t) \leq \phi^{**}$ . The rest of the proof takes three steps:

First, we note that for  $K_t \in [K^A(q_t), K^{**}(q_t)]$  and  $k_t = 1$

$$q_t = \bar{p}(K_t + dt, \phi_{t+dt}^{**}) \geq \bar{p}(K_{t+dt}, \phi_{t+dt}^{**}).$$

The equality follows from the definition of strategy  $\phi_{t+dt}^{**}$ . The inequality follows from the fact  $K_{t+dt} = K_t + k_t dt \leq K_t + dt$ . In the limit we have

$$q_t \geq \bar{p}(K_t, \phi_t^{**}).$$

Equation (32) implies  $\frac{dq_t}{dt} = 0$  (even if  $k_t = 1$ ).

Second, for  $K_t \in [0, K^A(q_t)]$ , we have

$$K_t < K^A(q_t)$$

which implies

$$\bar{p}(K_t, 0) < \bar{p}(K^A(q_t), 0) = q_t,$$

Again, equation (32) implies  $\frac{dq_t}{dt} = 0$ .

Finally, for  $K_t \geq K^{**}(q_t)$ ,  $\phi_{t+dt}^{**}(K_t, q_t) = \phi^{**}$ .

In conclusion, we have either  $\frac{dq_t}{dt} = 0$  or  $\phi_{t+dt}^{**}(K_t, q_t) = \phi^{**}$ . As a result,  $\frac{dq_t}{dt} > 0$  only if  $\phi_{t+dt}^{**}(K_t, q_t) = \phi^{**}$ . ■

**Lemma 3** *The principal's Markov strategy  $\phi_{t+dt}^{**}(K_t, q_t)$  and the agent's Markov strategy  $k_t^{**}(K_t, \phi_t, q_t; p_0)$  are mutual best responses following all histories for all  $t \geq 0$ .*

**Proof.** The proof of Lemma 2 shows that when Lemma 2 applies, we have either  $\frac{dq_t}{dt} = 0$  or  $\phi_{t+dt}^{**}(K_t, q_t) = \phi^{**}$ . The latter condition applies only for  $K_t \geq K^{**}(q_t)$ . In this subcase,  $\phi_{t+dt}^{**}(K_t, q_t) = \phi^{**}$  is satisfied, and implies  $q_\tau = \bar{p}(K_{\tau-dt}, \phi^{**})$  (equivalently,  $K_{\tau-dt} = \bar{K}(\phi^{**}, q_\tau) = K^{**}(q_\tau)$ ) for all  $\tau > t$ . Consequently,  $K_\tau \geq K^{**}(q_\tau)$  and  $\phi_{\tau+dt}^{**}(K_\tau, q_\tau) = \phi^{**}$  hold for all  $\tau > t$ . That is,  $\phi_t$  is constant throughout in this subcase. Since the state variable  $q_t$  affects the agent's value function only through affecting future values of  $\phi_\tau$ , which is constant in this subcase, we must thereby have  $\frac{\partial V}{\partial q_t} = 0$ . Overall, we have  $\frac{\partial V}{\partial q_t} \frac{dq_t}{dt} = 0$ . That is, there is no informational strategic effect if the principal always plays  $\phi_{t+dt}^{**}(K_t, q_t)$ .

We first show that  $k_t^{**}(K_t, \phi_t, q_t; p_0)$  is the agent's best response to  $\phi_{t+dt}^{**}(K_t, q_t)$ . There are three cases to be considered separately.

**Case 1.** Given  $t \geq 0$ , suppose for  $\forall \tau \geq t - dt$  the principal always plays the Markov strategy  $\phi_{\tau+dt}^{**}(K_\tau, q_\tau)$ . That is, the offer  $\phi_t$  (made at time  $t - dt$ ) is not a point deviation from  $\phi_t^{**}(K_{t-dt}, q_{t-dt})$ , and it is assumed that there will be no point deviation ever after either. In this case, Lemma 2 applies, and we must also have  $\frac{\partial V}{\partial q_t} \frac{dq_t}{dt} = 0$ . The best response for the agent solves the Bellman equation (30) (ignoring the informational strategic effect). The solution is the bang-bang investment strategy:

$$k_t(K_t, \phi_t, q_t; p_0) = \begin{cases} 1 & \text{if } K_t \leq \bar{K}(\phi_t, p_0), \\ 0 & \text{otherwise.} \end{cases} \quad (44)$$

Given  $t \geq 0$ , suppose the offer  $\phi_t$  (made at time  $t - dt$ ) is a point deviation from  $\phi_t^{**}(K_{t-dt}, q_{t-dt})$  such that  $\phi_t > \phi_t^{**}(K_{t-dt}, q_{t-dt})$  or  $\phi_t < \phi_t^{**}(K_{t-dt}, q_{t-dt})$ . We require the point deviation being uniform, that is, the absolute difference  $|\phi_t - \phi_t^{**}(K_{t-dt}, q_{t-dt})|$  cannot be contingent on the magnitude of  $dt$ , which is an undefined infinitesimal quantity. As a result, in the limit  $dt \rightarrow 0$  we must have

$$\frac{|\phi_t - \phi_t^{**}(K_{t-dt}, q_{t-dt})|}{dt} \rightarrow \infty. \quad (45)$$

It is verifiable (from (36), (1) and (21)) that whenever a (uniform) point deviation  $\phi_t \neq \phi_t^{**}(K_{t-dt}, q_{t-dt})$  is feasible, we have the following relations:

$$\text{If } \phi_t = \phi_t^{**}(K_{t-dt}, q_{t-dt}), \text{ then } q_t \rightarrow \bar{p}(K_t, \phi_t) \text{ as } dt \rightarrow 0, \text{ or } q_t > \bar{p}(K_t, \phi_t). \quad (46)$$

$$\text{If } \phi_t \gtrless \phi_t^{**}(K_{t-dt}, q_{t-dt}), \text{ then } q_t \gtrless \bar{p}(K_t, \phi_t). \quad (47)$$

**Case 2.** Given  $t \geq 0$ , suppose the offer  $\phi_t$  (made at time  $t - dt$ ) is a point deviation from  $\phi_t^{**}(K_{t-dt}, q_{t-dt})$  such that  $\phi_t > \phi_t^{**}(K_{t-dt}, q_{t-dt})$ . And it is assumed that there will be no more point deviations ever after. In this case, we have  $q_t > \bar{p}(K_t, \phi_t)$ , therefore  $\frac{dq_t}{dt} = 0$ . The strategy (44) is still a best response.

**Case 3.** Given  $t \geq 0$ , suppose the offer  $\phi_t$  (made at time  $t - dt$ ) is a point deviation from  $\phi_t^{**}(K_{t-dt}, q_{t-dt})$  such that  $\phi_t < \phi_t^{**}(K_{t-dt}, q_{t-dt})$ . And it is assumed that there will be no more point deviations ever after. In this case we have  $q_t < \bar{p}(K_t, \phi_t)$ , therefore  $\frac{dq_t}{dt} > 0$ . It then follows from (45) and (47) that

$$\frac{\bar{p}(K_t, \phi_t) - q_t}{dt} \rightarrow \infty, \quad (48)$$

which (using (32)) implies  $\frac{dq_t}{dt} \rightarrow \infty$  if  $k_t > 0$ . That is, if  $k_t > 0$  the state variable  $q_t$  would increase by a finite positive real number. Given that  $\phi_{\tau+dt}^{**}(K_\tau, q_\tau)$  is played for all  $\tau \geq t$ , this would mean a reduction of the agent's information rent by a finite positive real number. Formally, we have  $\frac{\partial V}{\partial q_t} \frac{dq_t}{dt} \rightarrow -\infty$ . The alternative choice by the agent is  $k_t = 0$ , which would keep  $(K_\tau, q_\tau)$  unchanged between time  $t$  and  $t + dt$ , consequently no reduction of information rent. The only loss of the alternative choice is due to discounting of the project value by the factor  $e^{-rdt}$ . The loss is approximately by a fraction of  $rdt \rightarrow 0$ , clearly outweighed by the finite reduction of project value following the choice  $k_t > 0$ . The best response, which is the solution to the Bellman equation (30) (taking the informational strategic effect into account), then must be  $k_t = 0$ .

To synthesize all the best responses under cases 1, 2 and 3 into a single formulation, the agent's optimal strategy must be given by  $k_t^{**}(K_t, \phi_t, q_t; p_0)$ . To verify this, note strategy  $k_t^{**}(K_t, \phi_t, q_t; p_0)$  degenerates to the naive strategy (44) for cases 1 and 2; and it degenerates to  $k_t = 0$  for case 3.

To verify that  $\phi_{t+dt}^{**}(K_t, q_t)$  is a best response to  $k_t^{**}(K_t, \phi_t, q_t; p_0)$ , note that  $\phi_{t+dt}^{**}(K_t, q_t)$  minimizes  $\phi_{t+dt}$  subject to the agent's 'participation' constraint:  $K_t \leq \bar{K}(\phi_t, p_0)$  and either  $q_t \geq \bar{p}(K_t, \phi_t)$  or  $\phi_t \geq \phi^{**}$  (because  $\phi_{t+dt}^{**}$  is the minimum reimbursement rate that the

agent is willing to continue ‘participation’.). Any deviation from  $\phi_{t+dt}^{**}(K_t, q_t)$  must either increase or decrease  $\phi_{t+dt}$ . The former cannot be profitable for the principal because it either unnecessarily transfers some expected surplus (from trade) from her to the agent, or may induce socially inefficient investment (if  $\phi_{t+dt} > \phi^{**}$ ). The latter cannot be profitable either because, given the agent’s strategy  $k_t^{**}(K_t, \phi_t, q_t; p_0)$  the (deviation) offer will be rejected by all types of the agent, while  $\phi_{t+dt}^{**}$  can guarantee non-negative expected value for the principal (note,  $\phi_{t+dt}^{**}$  is always below or equal to the principal’s reservation price  $\phi^{**}$ ). The deviation cannot help the principal to improve the deal relative to  $\phi_{t+dt}^{**}$ . Overall,  $\phi_{t+dt}^{**}(K_t, q_t)$  is sequentially optimal for any realization of the agent’s type. ■

### A.7.2 Proof of Theorem

Part (a): To establish  $(s^{**}, \mathcal{M}^{**})$  constitutes a POE, we need to establish part (i) and (ii) of Definition 5. Part (ii) has been established by discussion in Section 3.1.2 and Definition 8. Hence we are left to prove part (i), which requires us to establish  $s_P^{**}$  and  $s_A^{**}$  are mutual best responses for all  $t \in \{0_{-1}\}$  and  $[0, \infty)$ . We will proceed in two steps.

Claim (i.1):  $s_A^{**}$  and  $s_P^{**}$  are mutual best responses for all  $t \geq 0$  and all  $h^t \in \mathcal{H}^t$ .

This claim is an immediate implication of Lemma 3.

Claim (i.2): For  $t = 0_{-1}$ ,  $s_P^{**}$  is optimal for the principal. The logical of proof is the same as for claim (i.1). Given  $s_A^{**}$ , the strategy  $s_P^{**}(h^{0_{-1}}) = \phi^{**}$  is unanimously optimal.

Part (b): For  $\phi = \phi^{**}$ , the termination investment level is given by  $\bar{K}(\phi, p_0) = K^{**}(p_0)$ . The equilibrium given by Definition 8 therefore generates the same investment policy (i.e., given by (27)) that is induced by compensation strategy  $\tilde{\phi}_{t+dt}^{*}(K_t, p_0)$  under full information. Since the investment policy is a Pareto efficient allocation of resource, the difference between the two outcomes in question is purely distributional and due to the difference between the two compensation transfers. Both outcomes are Pareto efficient because in each case it is impossible to make one player better off without making the other worse off.

Part (c): For all equilibrium outcomes (if there exist multiple of them),  $U_{0_{-1}} = 0$  and none of them is weakly dominated by any other equilibrium outcome. At  $t = 0_{-1}$ , the principal is “indifferent” between them. Among all the equilibrium outcomes, the agent gets the highest reimbursement rate and hence the largest amount of payment from the principal from the equilibrium given by Definition 8 or its outcome-equivalent equilibria. This outcome makes each type of the agent best off. In any other equilibrium outcome (if it exists), some type of the agent is strictly worse off than in the equilibrium given by Definition 8. Therefore the equilibrium given by Definition 8 generates the Pareto dominant equilibrium outcome.

## A.8 An Example of Weakly Dominated Strategy Which Satisfies the $\max - \min$ Criterion

Consider the (artificial) set of agent types:  $[0, p^A]$ . We argue that the following strategy of the principal, which is to offer no compensation to the agent, is a weakly dominated strategy which satisfies the max-min criterion. To see this, note that the aforementioned the strategy

is weakly dominated by the time-invariant reimbursement strategy  $\phi_t = \phi^{**}$ , conditional on the agent's investment strategy being given by (37). For all  $p_t \in [0, p^{**}]$  and  $t \geq 0$  both strategies give the same value of  $W_t$ , which is zero and also the minimum value. For all  $p_t \in (p^{**}, p^A]$  and all  $t \geq 0$ , strategy  $\phi_t = \phi^{**}$  gives strictly larger value of  $W_t$  than the strategy of no compensation.

## A.9 Proof of Proposition 5

By Proposition 4  $\frac{\partial V}{\partial q_t} = 0$  on an equilibrium path, then Bellman equation (30) implies that for all  $p_0 < \bar{p}(K_t, \phi^{**})$  choosing  $k_t = 0$  is optimal for the agent. If  $k_t > 0$  occurs, it can be validly inferred that the true type  $p_0 \geq \bar{p}(K_t, \phi^{**})$  and hence  $\inf P_{t+dt} = \bar{p}(K_t, \phi^{**})$ . We also know that  $q_{t+dt} = \bar{p}(K_t, \phi^{**})$  if  $k_t > 0$ .

If  $k_t = 0$ , but  $K_t > 0$ , then we must have  $k_0 > 0$ ; otherwise,  $k_0 = 0$  would imply an stationary state such that  $(q_{t'}, K_{t'}, \phi^{**}) = (0, 0, \phi_0)$  for all  $t' \geq 0$ , which is a contradiction to the assumption  $K_t > 0$ . It follows from  $k_0 > 0$  that  $\inf P_{dt} = q_{dt} = \bar{p}(0, \phi^{**}) = p^{**}$  and  $\inf P_t = q_t \geq p^{**}$  for  $t > 0$  with  $K_t > 0$ . For any time  $\tau > 0$ , if  $k_\tau > 0$  both  $q_\tau$  and  $\inf P_\tau$  will ratchet up in an identical way; if  $k_\tau = 0$  both  $q_\tau$  and  $\inf P_\tau$  will remain unchanged and stay at the identical level for ever. This implies for all  $t \geq 0$  with  $K_t > 0$

$$\inf P_{t+dt} = q_{t+dt} \text{ on an equilibrium path.}$$

Since  $\inf P_0 = q_0 = 0$ , we must also have

$$\inf P_t = q_t \text{ on an equilibrium path.}$$

The rest of the proof is trivial.

## A.10 Proof of Lemma 1

Proposition 4 has already established that on any equilibrium path of an MPOE such that the state variable  $q_t$  is an argument of Markov strategies,  $\phi_t = \phi^{**}$ . What remains to be shown is that if there exists any other MPOE, the result  $\phi_t = \phi^{**}$  should still hold on any equilibrium path. Without loss of generality, suppose there exists another MPOE with state variable  $\tilde{q}_t$  as argument of Markov strategies, where  $\tilde{q}_t$  is one dimensional or a scalar variable. Should  $\tilde{q}_t$  be multidimensional, then the resulting state-contingent strategy would not be the “*simplest form of behavior that is consistent with rationality*”<sup>41</sup> since it would be less simple than that associated with  $q_t$ .

As has been shown in the proof of Proposition 4, the MPOE associated with the state variable  $q_t$  has the property that the principal's offer  $\phi_t$  along any equilibrium path is a credible take-it-or-leave-it offer, if it is rejected along the equilibrium path, then the trading

<sup>41</sup>The quotation is from Maskin and Tirole (2001), which was cited as a main philosophical consideration that underpins the notion of Markov strategies. See Section 6.3 for further discussion of the definition of Markov strategy and Markov equilibrium.

relation is terminated permanently and the play of the game reaches a steady state – all state variables become stationary (over time). This property entails that  $q_{t+dt} = q_t$  if  $k_t = 0$ . For the MPOE associated with the state variable  $\tilde{q}_t$  to prescribe the simplest forms of behavior, it is necessary that thereby the principal’s offer  $\phi_t$  along any equilibrium path is also a credible take-it-or-leave-it offer. (Otherwise, the behavior would not be the simplest). Technically, the requirement entails  $\tilde{q}_{t+dt} = \tilde{q}_t$  if  $k_t = 0$ .

Given the similarity between  $\tilde{q}_t$  and  $q_t$ , the argument in proof of Proposition 4 applies to establishing  $\phi_t = \phi^{**}$  for the case of  $\tilde{q}_t$  as well.

## A.11 Proof of Theorem 4

Trivial.

## A.12 Proof of Proposition 6

The proof is similar to the proof of Theorem 3 (apart from that given the agent’s strategy the principal’s strategy is always unanimously optimal), therefore is omitted.

## A.13 Proof of Proposition 7

It is an immediate implication of Proposition 6.

## A.14 Proof of Proposition 8

The proof of existence of  $\phi$  has been given in the main text following the proposition, which also establishes the “trigger” strategy component and the bootstrap property. The rest of the proof is similar to the proof of Theorem 3 (apart from that given the agent’s strategy the principal’s strategy is always unanimously optimal), therefore is omitted.

## A.15 Proof of Proposition 9

Trivial.

## A.16 Proof of Proposition 10

Given the agent’s (partial pooling) strategy  $s_A^{**}$ , as defined in Definition 8, the principal’s strategy  $s_P^{**}$  is unanimously optimal, therefore, a best response. Similarly, given the principal’s strategy  $s_P^{**}$ , as defined in Definition 8, the agent’s strategy  $s_A^{**}$  is also optimal, hence, a best response.

## A.17 Proof of Proposition 11

Trivial (similar to the proof of Proposition 10).

## A.18 Proof of Theorem 5

The idea for proof has been given in the main text.

## References

- [1] Al-Najjar, N. I. and J. Weinstein (2009). “The Ambiguity Aversion Literature: A Critical Assessment”, *Economics and Philosophy*, 25, pp. 249–284.
- [2] Anscombe, F. J. and R. J. Aumann (1963), “A Definition of Subjective Probability”, *Annals of Mathematical Statistics*, 34 (1), pp. 199-205.
- [3] Aumann, R. J. (1987), “Correlated equilibrium as an expression of Bayesian rationality”. *Econometrica*, 55(1), pp. 1–18.
- [4] Baron, David P. and Roger Myerson, (1982) “Regulating a Monopolist with Unknown Costs,” *Econometrica*, 50 (4), pp. 911-930.
- [5] Bergemann, Dirk and Stephen Morris (2007), “Belief Free Incomplete Information Games”, Cowles Foundation Discussion Paper No. 1629.
- [6] Besanko, David, Jian Tong, and Jianjun Wu, “A Static Mechanism Design Model with Hidden Information and Ambiguity,” (working paper).
- [7] Bewley, Truman F. (2002), “Knightian Decision Theory. Part I”, *Decisions in Economics and Finance*, 25, pp. 79-110. (The working paper version appeared in 1986.)
- [8] Bolton, Patrick and Mathias Dewatripont (2005), *Contract Theory*, MIT Press, Cambridge, Massachusetts.
- [9] Crawford, Vincent P. (1982), “A Theory of Disagreement in Bargaining”, *Econometrica*, 50 (3), pp. 607-637.
- [10] Crémer, J. and R. McLean (1985), “Optimal Selling Strategies Under Uncertainty for a Discriminating Monopolist When Demands Are Interdependent,” *Econometrica*, 53, pp. 345–361.
- [11] de Finetti, Bruno (1937), “La Prévision: Ses Lois Logiques, Ses Sources Subjectives”, *Annales de l'Institute Henri Poincaré*, 7, pp.1-68.
- [12] Edwards, A.W.F. (1992), *Likelihood* (Expanded Edition), Johns Hopkins University Press, London.
- [13] Efron, Bradley (1998), “R. A. Fisher in the 21th Century”, *Statistical Science*, 13 (2), pp. 95-114.

- [14] Ellsberg, Daniel (1961), “Risk, Ambiguity and the Savage Axioms”, *Quarterly Journal of Economics*, 75 (4), pp. 643-669.
- [15] Epstein, Larry G. and Martin Schneider (2003), “Recursive Multiple-Priors”, *Journal of Economic Theory*, 113, pp. 1-31.
- [16] Epstein, Larry G. and Martin Schneider (2007), “Learning under Ambiguity”, *Review of Economic Studies*, 74, pp. 1275-1303.
- [17] Fisher, R. A. (1925), *Statistical Methods for Research Worker*, Oliver and Boyd, Edinburgh.
- [18] Fudenberg, Drew and Jean Tirole (1991), *Game Theory*, MIT Press, Cambridge, Massachusetts.
- [19] Ghirardato, Paolo and Fabio Maccheroni and Massimo Marinacci (2004), “Differentiating Ambiguity and Ambiguity Attitude”, *Journal of Economic Theory*, 118, pp. 133-173.
- [20] Gilboa, Itzhak and David Schmeidler (1989), “Maxmin Expected Utility with Non-unique Prior”, *Journal of Mathematical Economics*, 18, pp. 141-153.
- [21] Gilboa, Itzhak and Fabio Maccheroni and Massimo Marinacci and David Schmeidler (2010), “Objective and Subjective Rationality in a Multiple Prior Model”, *Econometrica*, 78 (2), pp. 755-770.
- [22] Gilboa, Itzhak and Massimo Marinacci (2011), “Ambiguity and the Bayesian Paradigm”, Working Paper.
- [23] Hacking, Ian (1965) *Logic of Statistical Inference*, Cambridge University Press, Cambridge, England.
- [24] Hacking, Ian (2001), *An Introduction to Probability and Inductive Logic*, Cambridge University Press, Cambridge, England.
- [25] Harsanyi, John C. (1967), “Games with Incomplete Information Played by ‘Bayesian’ Players, I-III. Part I. The Basic Model”, *Management Science*, 14 (3), Theory Series, pp. 159-182.
- [26] Harsanyi, John C. (1968), “Games with Incomplete Information Played by ‘Bayesian’ Players, I-III. Part II. Bayesian Equilibrium Points”, *Management Science*, 14 (5), Theory Series, pp. 320-334.
- [27] Harsanyi, John C. (1968), “Games with Incomplete Information Played by ‘Bayesian’ Players, I-III. Part I. I-III. Part III. The Basic Probability Distribution of the Game”, *Management Science*, 14 (7), Theory Series, pp. 486-502.

- [28] Hörner, Johannes and Stefano Lovo (2009), “Belief-free Equilibria in Games with Incomplete Information”, *Econometrica*, 77 (2), pp. 453-487.
- [29] Hörner, Johannes and Stefano Lovo and Tristan Tomala (2011), “Belief-free Equilibria in Games with Incomplete Information: Characterization and Existence”, *Journal of Economic Theory*, 146, pp. 1770-1795.
- [30] Kalai, Ehud (2004), “Large Robust Games,” *Econometrica*, 72, pp. 1631–1665.
- [31] Keller, Godfrey, Sven Rady, and Martin Cripps (2005), “Strategic Experimentation with Exponential Bandits,” *Econometrica*, Vol. 73 (1), pp. 39-68.
- [32] Kennan, John and Robert Wilson (1993), “Bargaining with Private Information,” *Journal of Economic Literature*, 31, pp. 45-104.
- [33] Keynes, John Maynard (1921), *A Treatise on Probability*, MacMillan and Co., London.
- [34] Laffont, Jean-Jacques and Jean Tirole (1986), “Using Cost Observation to Regulate Firms,” *Journal of Political Economy*, 94 (3), pp. 614-641.
- [35] Manski, Charles F. (1995), *Identification Problems in Social Sciences*, Harvard University Press, Cambridge, Massachusetts.
- [36] Manski, Charles F. (2008), “Partial Prescriptions for Decisions with Partial Knowledge”, NBER Working Paper 14396.
- [37] Manski, Charles F. (2011), “Choosing Treatment Policies under Ambiguity”, *Annual Review of Economics*, 3.
- [38] Maskin, Eric and John Riley (1984), “Monopoly with Incomplete Information,” *RAND Journal of Economics*, 15(2), pp. 171-196.
- [39] Maskin, Eric and Jean Tirole (2001), “Markov Perfect Equilibrium I. Observable Actions”, *Journal of Economic Theory*, 100, pp. 191-219.
- [40] Popper, Karl (1959), *The Logic of Scientific Discovery*, Hutchinson & Co., London.
- [41] Popper, Karl (1963), *Conjectures and Refutations: The Growth of Scientific Knowledge*, Routledge, London.
- [42] Popper, Karl (1972), *Objective Knowledge: An Evolutionary Approach*, Oxford University Press, Oxford.
- [43] Ramsey, Frank P. (1926), “Truth and Probability”, in *The Foundations of Mathematics and Other Logical Essays*, London, Kegan Paul, and New York, Harcourt, Brace and Co., 1931.

- [44] Rubinstein, Ariel (1982), “Perfect Equilibrium in a Bargaining Model”, *Econometrica*, 50 (1), pp. 97-109.
- [45] Rubinstein, Ariel (1987), “A Sequential Strategic Theory of Bargaining”, *Advances in Economic Theory: Fifth World Congress*, 5, edited by Truman Fassett Bewley, Chapter 5, Cambridge University Press, Cambridge, England.
- [46] Savage L. (1951), “The Theory of Statistical Decision”, *Journal of the American Statistical Association*, 46 (253), pp. 55-67.
- [47] Savage L. (1954), *The Foundations of Statistics*, New York: Wiley.
- [48] Wald, Abraham (1950), *Statistical Decision Functions*, John Wiley & Sons, New York.