Title
Information and incentives in stochastic games, social learning and crowdsourcing

Author
Bohren, J. Aislinn

Publication Date
2012

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Information and Incentives in Stochastic Games, Social Learning and Crowdsourcing

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

J. Aislinn Bohren

Committee in charge:

Professor S. Nageeb Ali, Chair
Professor Vincent Crawford
Professor Dominique Lauga
Professor Craig McIntosh
Professor Joel Sobel
Professor Joel Watson

2012

The dissertation of J. Aislinn Bohren is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2012

DEDICATION

I dedicate this thesis to my family, who have fostered my intellectual curiosity and encouraged me to explore the unknown.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank the many professors, colleagues, friends and family who have provided helpful feedback and support during my graduate career, including:

The wonderful economics faculty at UCSD, who created a fun and exciting environment to nurture my development into a game theorist, including Vincent Crawford, Craig McIntosh, David Miller, Joel Sobel and Joel Watson. I would especially like to thank my advisor, Nageeb Ali, whose enduring energy, dedication and enthusiasm challenged me to pursue a demanding research agenda and strive to achieve a career in academia.

The graduate community in the economics department, who provided valuable comments on my research as well as a stimulating atmosphere in advising sessions, reading groups and seminars, including Kristy Buzard, Aeimit Lakdawala, Charles Lin and Leah Nelson. I would especially like to thank my co-author, Troy Kravitz, for being a terrific collaborator on Chapter 3: ”Incentives in Crowdsourcing”, as well as a close colleague throughout my time at UCSD.

The many other mentors and teacher who have shaped my academic journey, from my high school economics teacher, Carol Marin, who ignited my interest in the field, to my undergraduate professors at the University of Richmond, who inspired me to pursue a career in economics, especially Robert Dolan, Joseph Kent, William Owen, Robert Schmidt and Andrew Yates, to my colleagues at the Federal Deposit Insurance Corporation, especially Katherine Samolyk, who cemented my desire to pursue a research career.

My friend, Kevin Lingerfelt, whose anecdotes about the programming industry were the inspiration for the third chapter, ”Incentives in Crowdsourcing”.

And my parents, Valerie and Thomas Bohren, and sister, Meghan Bohren, who have been there to share in both my excitement and frustrations as I embarked on new research challenges.

I would also like to thank the many individuals and organizations who provided me with financial support, including Nageeb Ali, The Clive Granger Research Fellowship, The Institute for Humane Studies, Craig McIntosh, David Miller, Paul Niehaus and Yahoo! Research. I would especially like to thank the Center for Research on Educational Equity, Assessment, and Teaching Excellence (CREATE), including Larry McClure and Bud Mehan, whose generous research assistantship provided funding for the majority of my tenure at UCSD.

VITA

2004	B. S. in Mathematics and Economics *magna cum laude*, University of Richmond
2012	Ph. D. in Economics, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Information and Incentives in Stochastic Games, Social Learning and Crowdsourcing

by

J. Aislinn Bohren

Doctor of Philosophy in Economics

University of California, San Diego, 2012

Professor S. Nageeb Ali, Chair

My dissertation utilizes tools from game theory to derive novel economic insights in a variety of settings, including social learning with biased information processing beliefs, repeated games with persistent actions and contract design in crowdsourcing labor markets.

The first chapter explores how individuals learn from their predecessors when they are subject to biased beliefs about the information processing capabilities of oth-

ers. I consider a social learning environment in which individuals observe private signals, and learning is asymptotically efficient in the absence of information processing biases. Either underestimating or overestimating others' information processing capabilities can have important implications for the asymptotic efficiency of learning.

The second chapter studies a new class of stochastic games in which the actions of a long-run player have a persistent effect on payoffs. The setting is a continuous time game of imperfect monitoring between a long-run and a short-run player. The main result of this paper is to establish general conditions for the existence of Markovian equilibria and conditions for the uniqueness of a Markovian equilibrium in the class of all Perfect Public Equilibria. The existence proof is constructive and characterizes, for any discount rate, the explicit form of equilibrium payoffs, continuation values, and actions in Markovian equilibria. Action persistence creates a channel to provide intertemporal incentives, and offers a new and different framework for thinking about the reputations of firms, governments, and other long-run agents.

The final chapter examines information and incentive issues in a novel labor market setting: crowdsourcing. Our research focuses on how to optimally design labor contracts and the marketplace in a crowdsourcing setting to facilitate efficient information transmission between workers and firms. The structure of such contracts will depend on the institutional features of the specific job, including how costly it is for a worker to acquire information or how likely it is for the firm to observe when a worker incorrectly completes a task.

Chapter 1

Information-Processing Bias in Social Learning

1.1 Introduction

Observational learning plays an important role in the transmission of information, opinions and behavior. People may use a bestseller list to guide their purchase of a book or a car. Observing high participation rates amongst co-workers may increase the likelihood that a person contributes to their retirement plan. Social learning can also influence behavioral choices, such as whether to smoke or exercise regularly, or ideological decisions, such as which side of a moral or political issue to support. Given the wide range of situations influenced by observational learning, it is important to understand how biases in information processing affect learning. This paper explores how an information processing bias may interfere with the efficiency of social learning, and demonstrates that such biases can partially explain how inefficient choices can persist even when contradicted by public information.

This paper extends standard herding models in the tradition of Banerjee (1992) and Bikhchandani et al. (1992) to incorporate the idea of information-processing bias

(BIP). In the standard model with binary actions and signals, individuals have common-value preferences that depend on an unknown state of the world. Agents act sequentially, observing a private signal and the actions of previous agents before choosing an action. An information cascade occurs when it is optimal for an agent to ignore his private signal and act only on the basis of the information contained in the actions of previous agents. When this occurs, all subsequent agents follow suit and new information ceases to aggregate. With positive probability, agents herd on the suboptimal action and thus the equilibrium is inefficient.

A critical feature of this model is common knowledge of how individuals process information. Agents understand exactly how preceding agents incorporated the action history into their decision-making process, and are therefore aware of which actions contain no information. Since the herd is based on only a few initial signals, public beliefs about the state remain fragile and are easily reversed by the arrival of new information. Thus, if some agents don't observe prior actions and follow their private signal, or if public information is released periodically, social learning is asymptotically efficient.

However, what happens if agents are unsure about the information-processing capabilities of other agents? What if they believe the actions of previous agents reveal more information about private signals than is actually the case during a cascade, or what if they attribute too many actions to herding and are not sensitive enough to new information? This paper examines how a behavioral bias in information processing, which I refer to as information-processing bias (BIP), can interfere with optimal information aggregation even in settings where new information continues to arrive frequently during a cascade. Individuals subject to BIP are biased in their perception of the information-processing capabilities of others, and consequently fail to accurately disentangle repeated and new information.

In particular, suppose that a fraction of individuals do not observe preceding actions and select an action solely based on their private information. These uninformed agents simply do not have access to all available information, or are boundedly rational and unable to process multiple sources of information. Regardless of the justification for their presence, these uninformed agents always reveal their private signal. Individuals who incorporate the action history into their decision observe the full sequence of preceding actions but are uncertain about the information-processing capabilities of others. Consequently, these informed decision makers face an inferential challenge when extracting information from the actions of others, and their behavior will hinge on their beliefs about the population.

To fix ideas, suppose that each individual observes the history and is fully informed with probability p, and with probability $(1-p)$ is an uninformed type that only observes his private signal. Each informed individual believes that any other individual is informed with probability $\hat{p}$, where $\hat{p}$ need not coincide with p. The difference between p and $\hat{p}$ may arise because even very sophisticated individuals may underestimate or overestimate the information possessed by others, and so it is natural to allow for the distinction.

When $\hat{p} < p$ then an informed decision maker underestimates the fraction of preceding informed individuals. Accordingly, when this decision maker observes a series of identical actions, he incorrectly attributes too many of these actions to the private signals of uninformed individuals. This effect leads him to overweight information from the public history, and may allow public beliefs about the state to become entrenched. On the other hand, when $\hat{p} > p$, then an informed decision maker underweights the new information contained in correlated actions, rendering herds more fragile to contrary information.

To understand how BIP affects eventual efficiency and learning requires careful analysis of the rate of information accumulation. I characterize conditions that allow a herd to persist with positive probability, and conditions that ensure a herd breaks. Using these conditions and fixing the share of informed agents, I establish thresholds on beliefs about the share of informed agents, $\widehat{p}_1$ and $\widehat{p}_2$, such that when $\widehat{p} < \widehat{p}_1$

an incorrect herd can persist with positive probability and when $\widehat{p} > \widehat{p}_2$ a correct herd will always break. When beliefs fall between these two thresholds, $\widehat{p} \in (\widehat{p}_1, \widehat{p}_2)$, incorrect herds always break but correct herds can persist, so eventually a correct herd will persist. Herding will be efficient in that informed agents will choose the optimal action all but finitely often. Otherwise, there is positive probability that herding will be inefficient and informed agents will choose the suboptimal action infinitely often. Correct beliefs about agent types lead to efficient herding since $p \in (\widehat{p}_1, \widehat{p}_2)$.

During a herd, beliefs continue to strengthen. When a herd persists in the long run, public beliefs will converge to a point mass on the state matching the action agents are herding on. Thus, if a correct herd persists, then learning is complete, while if an incorrect herd persists, learning is fully incorrect. If no herd persists, then beliefs remain interior and fluctuate, and learning remains incomplete. Fully incorrect learning or the perpetual fluctuation of beliefs are possible because the conditional public likelihood ratio is no longer a martingale when BIP is present. In fact, when $\widehat{p} < p$, agents overweight herding actions and the conditional public likelihood ratio in an incorrect herd is a submartingale. This explains why fully incorrect learning is possible. On the other hand, when $\widehat{p} > p$, agents underweight herding actions. The conditional public likelihood ratio is a supermartingale in incorrect herds and a submartingale in correct herds. When beliefs are sufficiently incorrect, the submartingale diverges and both types of herds to eventually break.

BIP in the context of a herding model has important implications. A failure to recognize repeated information can confound learning by allowing an incorrect herd to persist even when new information counteracts the incorrect herd, whereas a failure to recognize new information can cause correct herds to continually break (as well as incorrect herds). Every time a correct herd is broken, there is a chance that an incorrect herd may form in its place. These results are robust to the inclusion of other sources of new information, such as public signals, gurus (perfectly informed agents) or continuous signals. Whenever BIP is severe enough that repeated information accumulates at a faster rate than new information, then incorrect herds persist with positive probability.

To illustrate the relevance of this result, consider a public health campaign to increase awareness about the risks of HIV. Agents need to decide if HIV is a threat to them, and whether to take appropriate precautions. They observe public signals from the government and other public health agencies, along with the actions of previous agents. If all agents are herding on the actions of a few initial agents who didn't believe that HIV was a significant threat, then the public health campaign should eventually overturn this herd. However, if agents are subject to BIP, then observing many preceding agents who didn't believe that HIV was a threat will lead to strong beliefs that this is the case, making it less likely that the public health campaign will successfully overturn the herd.[1] In this scenario, the best way to quash the false cascade is to release public signals immediately and frequently. This contrasts with the case of no BIP, in which the timing of public signal release is irrelevant.

Individuals may also use the history to learn about the information processing capabilities of other agents. In Section 1.4, I examine what happens when agents can also learn about p. Although fully incorrect learning is generally precluded in this setting, incorrect herds may still persist with positive probability. Herding is more likely to be efficient if several herds form and break before a herd persists or if a low share of agents observe the history.

BIP relates to the notion of persuasion bias first introduced by DeMarzo et al. (2001) in a model of opinion formation in networks. In their paper, decision makers embedded in a network graph treat correlated information from others as being independent, leading to informational inefficiencies. Although my paper studies a very different environment than theirs, BIP provides a natural analogue for considering persuasion bias in social learning.

Banerjee (1992) and Bikhchandani et al. (1992) first modeled social learning in

[1]This example abstracts from the payoff interdependencies of HIV transmission.

a sequential setting, as discussed above. Smith and Sorensen (2000) extend their models to include continuous signals. An unbounded signal space is sufficient to ensure complete learning, eliminating the possibility of inefficient cascades. Acemoglu et al. (2010) examines social learning in a general social network, which includes the sequential learning and uninformed agents networks as special cases. As such, BIP can also be viewed as a boundedly rational extension of the uninformed agents network topology in Acemoglu et al. (2010).

This paper is most closely related to concurrent work on social learning by Eyster and Rabin (2009). They extend a sequential learning model with continuous actions and signals to allow for “inferential naivety”: players realize that previous agents’ action choices reflect their signals, but fail to account for the fact that these actions are also based on the actions of agents preceding these players. While continuous actions lead to full revelation of players’ signals in the absence of inferential naivety, inferential naivety can confound learning by overweighing actions of the first few agents. Although similar in nature, inferential naivety and information-processing bias differ in generality and interpretation. Inferential naivety considers the case in which every repeated action is viewed as being independent with probability 1, whereas in the BIP model, most decision makers are sophisticated and recognize that some repeated actions may stem from herding behavior, but misperceive the exact proportion of repeated information. The analogue of inferential naivety in my environment corresponds to $\hat{p} = 0$ and $p = 1$. As such, both papers provide complementary explanations for the robustness of inefficient learning. Eyster and Rabin (2009) also embed inferential naive agents in a model with rational agents. When every nth player in the sequence is inferentially naive, rational agents achieve complete learning but inferentially naive agents do not. Augmenting the BIP and inferentially naive models with rational agents who do not know precisely which previous agents are also rational, naive or uninformed, and perhaps are even uncertain about the share of each type of agent is an interesting avenue left open for future research.

Guarino and Jehiel (2009) explore boundedly rational information processing in a sequential learning environment using the concept of analogy based expectation equilibrium (ABEE), in which agents best respond to the aggregate distribution of action choices. Learning is complete in a continuous action model - in an ABEE, the degree to which agents overweigh initial signals increases in a linear fashion, preventing these initial signals from permanently dominating subsequent new information. This contrasts with Eyster and Rabin (2009), the degree to which agents overweigh initial signals doubles each period, allowing a few early signals to overwhelm all future signals. As in the fully rational model, complete learning no longer obtains in an ABEE when actions are discrete.

Earlier work by Eyster and Rabin (2005) on cursed equilibrium also examines information processing errors. A cursed player doesn't understand the correlation between a player's type and his action choice, and therefore fails to realize a player's action choice reveals information about his type. A player with BIP understands the correlation between a player's type and their action choice, but incorrectly predicts the distribution of action choices in equilibrium.

BIP also relates to the recent literature on initial response models, including level-k analysis and cognitive hierarchy models.[2] The premise of these models is that agents best respond to their beliefs about how others act, but unlike equilibrium analysis, these beliefs are not required to be correct. Consider level-k analysis in the context of sequential learning. Anchoring level 0 types to randomize between the two possible actions, level 1 types best respond by following their private signal - this corresponds to uninformed types in the BIP model. Level 2 types believe all other agents follow their private signal, and thus act as BIP informed agents with beliefs $\widehat{p} = 0$. Consequently, the main difference between the two models stems from the beliefs informed agents have about other agents' types - BIP informed agents can place positive weight on other agents using a level 2 decision rule, whereas "informed agents" in a level k

[2] Costa-Gomes et al. (2009)Camerer et al. (2004)

analysis believe that all other agents use a level 1 decision rule. BIP itself stems from level 2 agents misperceiving the share of other agents who are level 2. There is no such misperception in level k models, as all level k agents place probability 1 on other agents being level k-1. The comparison to a cognitive hierarchy (CH) model is similar - level 1 agents correspond to BIP uninformed agents, while level 2 agents act like BIP informed agents with beliefs $\widehat{p} = 0$, who also believe some previous actions convey only noise (i.e. CH level 2 agents place positive probability on level 0 and level 1 types, but probability 0 on other level 2 types).

The organization of this paper proceeds as follows. Section 2.3 sets up the model and explores the conditions under which learning is confounded in the presence of BIP. Section 1.3 explores the robustness of the model to several extensions, including public signals, continuous signals and private values. Section 1.4 allows agents to also learn about p, while Section 1.5 discusses experimental evidence in support of BIP and concludes. All proofs are in the Appendix.

1.2 Model

The basic set-up of this model mirrors the standard sequential learning model with binary action and signal spaces. There are two payoff-relevant states of the world, $\omega \in \{L, R\}$ with common prior belief $P(\omega = L) = \pi^L \in (1/2, 1)$.[3] Nature selects one of these states at the beginning of the game. A countably infinite set of agents $T = \{1, 2, ...\}$ act sequentially and attempt to match the realized state of the world by making a single decision between two actions, $a_t \in \{L, R\}$. They receive a payoff of 1 if their action matches the realized state, and a payoff of 0 otherwise: $u(a_t, \omega) = 1_{a_t=\omega}$.

Before selecting an action, each agent privately observes a binary signal about the state of the world, $s_t \in \{l, r\}$, which is i.i.d. conditional on the state with precision

[3] An asymmetric prior obviates the need for breaking indifference.

$\pi^s \in (\pi^L, 1)$.[4] There are two types of agents. With probability $p > 0$, an agent is a socially informed type who observes the prior action choices of other agents. This agent uses her private signal and the action history to guide her action choice. The public history observed by informed agents is represented as $h_t = (a_1, ..., a_{t-1})$. With probability $1-p$, an agent is a socially uninformed type who only observes his private signal. An alternative interpretation for this uninformed type is a behavioral type who is not sophisticated enough to draw inference from the action history. This type's decision is solely guided by the information contained in his private signal.

Informed agents may misperceive the information-processing capabilities of others. Each informed individual believes that any other individual is informed with probability $\hat{p}$, where $\hat{p}$ need not coincide with p. This captures the fact that there is higher-order uncertainty over the level of information possessed by other agents, which we will refer to as information-processing bias (BIP). The difference between p and $\hat{p}$ may arise because even very sophisticated individuals may underestimate or overestimate the information possessed by others, and so it is natural to allow for the distinction. Incorrect beliefs about p can persist because no agent ever learns what the preceding agents actually observed or incorporated into their decision-making processes. Consequently, these informed decision makers face an inferential challenge when extracting information from the actions of others, and their behavior will hinge on their beliefs about the population. This bias interferes with optimal information aggregation if agents fail to accurately disentangle repeated and new information. An informed agent believes that other agents also hold the same beliefs about whether previous agents are informed or uninformed. Although requiring agents to hold identical misperceptions about others is admittedly restrictive, it is a good starting point to examine the possible implications of BIP. Extending the model to allow for heterogenous biases is left for future research.

Agents use Bayes rule to formulate beliefs about the state of the world. Denote

[4] π^s is defined such that $P(s_t = l|\omega = L) = P(s_t = r|\omega = R) = \pi^s$

public beliefs of informed agents at the beginning of period t by $\mu_t = P(\omega = L|h_t)$. Public beliefs depend on the history and beliefs about the share of informed agents. Denote private beliefs by μ_t^r if agent t observes a private r signal and μ_t^l if agent t observes a private l signal. Private beliefs for informed agents depend on public beliefs and their private signal realization, while private beliefs for agents who don't observe the history depend on only their private signal realization. Each agent maximizes expected payoffs with respect to their private beliefs about ω. For the uninformed type, this implies an agent chooses the action that corresponds to his private signal, while for an informed type, an agent chooses the action that corresponds to the state he believes is more likely, given his beliefs about p.

An *information cascade* occurs when it is optimal for an agent who observes the history to choose the same action regardless of his private signal. Throughout the paper, such an agent's action choice is described as *herding*. When herding arises, the agent's action reveals nothing about his private information, and social learning is impeded. An information cascade breaks when it becomes optimal for an informed agent to follow his private signal (i.e. in an L-herd, it is optimal for an informed agent to choose R if he receives an r signal). For a given sample path, we say the information cascade *persists in the limit* if it persists in every period for $t = 1, 2, ...$ and the information cascade *breaks* if $\exists \tau < \infty$ such that the information cascade breaks at period τ. The probability that such sample paths occur will determine the probability that a given information cascade persists or breaks.

This paper examines how the efficiency of information cascades depends on the relationship informed agents perceive between prior actions and signals - that is, their beliefs about the share of informed agents, $\hat{p}$, which determines the accuracy of the inference drawn from the history during a herd.

1.2.1 Cascades in the Benchmark Model

In the benchmark model with common knowledge of no informed agents, inefficient herding arises but the herds are not robust. To see this, consider the following. Define Δ_t as the difference between the number of L and R actions at the beginning of time t. The unique Nash Equilibrium when $p = \widehat{p} = 1$ is to herd on L whenever Δ_t reaches 1, and to herd on R whenever Δ_t reaches -2.[5] An information cascade begins with probability 1, and occurs on the suboptimal action with positive probability. However, these cascades are very fragile - because new information ceases to aggregate once a cascade begins, the cascade is easily reversed if additional information becomes available. For example, a bounded public signal could overturn the herd.

Augmenting the benchmark model with uninformed agents allows information from action choices to continue accumulating in a cascade.[6] We will see in section 1.2.2 that in the absence of BIP, the addition of uninformed agents ($p = \widehat{p} < 1$) leads to complete learning. However, what if players are uncertain about the share of agents who are informed? What if they believe the actions of previous agents reveal more information about private signals than is actually the case during a herd, or what if they attribute too many actions to herding and are not sensitive enough to the new information? The remainder of this section explores the conditions that allow an information cascade to persist and the conditions that ensure an information cascade breaks, and uses these conditions to examine the impact of BIP on learning.

[5] Note that the asymmetry in Δ_t required for the formation of an L-herd versus R-herd stems from the specification of the prior $\pi^L > 1/2$.

[6] The conditions for a herd to begin in the presence of uninformed agents are identical to the conditions in the benchmark case of $p = 1$, regardless of beliefs $\widehat{p}$. Before the formation of a herd, all agents are following their signal regardless of whether they observed the history, and the public likelihood ratio evolves in the same manner as in the benchmark case, leading to the same conditions for herd formation.

1.2.2 How Does BIP Affect Learning?

Before a herd forms, all agents follow their private signal. Since the decisions of informed and uninformed agents coincide, informed agents correctly infer that previous actions perfectly reveal private information and BIP doesn't affect behavior prior to the onset of a herd. However, BIP interferes with information aggregation during a herd.

Suppose a herd has begun on action L. Each subsequent L action during the herd is attributed to (i) an uninformed agent who followed his private signal with probability $(1-\widehat{p})$; and (ii) an agent who observed the history and followed the herd with probability $\widehat{p}$. The public likelihood ratio following an L action is updated as follows:

$$\frac{\mu_t}{1-\mu_t} = \left(\frac{\widehat{p}+(1-\widehat{p})\pi^s}{\widehat{p}+(1-\widehat{p})(1-\pi^s)}\right)\left(\frac{\mu_{t-1}}{1-\mu_{t-1}}\right)$$

Thus, an action that follows the herd still reveals some information. When $\widehat{p} < p$, informed agents overweight the informativeness of this action, leading to an upward bias in the likelihood ratio relative to correct beliefs. The opposite occurs when $\widehat{p} > p$: agents attribute too many L actions to herding rather than private signals, resulting in a downward bias in the likelihood ratio. Let $\phi^h = \left(\frac{\widehat{p}+(1-\widehat{p})\pi^s}{\widehat{p}+(1-\widehat{p})(1-\pi^s)}\right)$ represent the information accumulating from a supporting action, or an action that follows the herd.

When it is still optimal for informed agents to herd based on $\widehat{p}$ and the history, a decision-maker will attribute a contrary action to an uninformed agent. In an L-herd, each contrary action R is attributed to an agent who did not observe the history and received a private r signal. The public likelihood ratio following an R action is updated as follows:

$$\frac{\mu_t}{1-\mu_t} = \left(\frac{1-\pi^s}{\pi^s}\right)\left(\frac{\mu_{t-1}}{1-\mu_{t-1}}\right)$$

Let $\phi^c = \left(\frac{1-\pi^s}{\pi^s}\right)$ represent the information accumulating from a contrary action, or an action that doesn't follow the herd. Note that beliefs $\widehat{p}$ do not bias the informativeness of an R action.

The public likelihood ratio increases with each supporting action and decreases with each contrary action. Supporting actions are believe to reveal new information with probability $(1-\widehat{p})$, whereas contrary actions reveal new information with probability 1. Therefore, a contrary action is more informative than a supporting action.

Without loss of generality, normalize to zero the period in which the action that begins the herd is chosen, so the length of the herd at the beginning of period t is equal to t. Let $\Delta_t \in [0,1]$ be the fraction of contrary actions chosen after the onset of a herd.[7] In an L-herd, $\Delta_t = \frac{1}{t-1}\sum_{s=1}^{t-1} 1_{a_s=R}$ represents the share of R actions and $1-\Delta_t$ represents the share of L actions. In a herd, Δ_t is a sufficient statistic for the history when examining the evolution of the public likelihood ratio.

When Does a Herd Break?

A herd breaks when sufficient information accrues in favor of the alternative state such that an agent who observes the history finds it optimal to follow her signal.. When contrary actions are possible ($p < 1$), this happens with positive probability in any herd. A finite number of contrary actions can overturn a herd of any length. When $p = 1$, no contrary actions occur and the herd will never break. Theorem 1 demonstrates these results.

Theorem 1. *Suppose a herd is occurring in period t with contrary action share Δ_t. If $p < 1$ then there exists a set of sample paths that occur with positive probability along which the herd breaks within a finite number of periods after t. Otherwise, the herd will never break.*

This Theorem demonstrates that every herd breaks with positive probability when some agents don't observe the history, and the result holds for any belief $\widehat{p}$.

[7] If $p = 1$ then no contrary actions are observed and $\Delta_t^a = 0\ \forall t$.

When Does a Herd Persist?

Given that a herd breaks with positive probability when $p < 1$, I now examine when such a herd can also persist with positive probability, and when the herd breaks with probability 1. Define the *herd breaking threshold* $\Delta_t^*(\widehat{p})$ as the minimum share of contrary actions that will result in an agent following her private signal in period t. Whenever the actual share of contrary actions is greater than this threshold, the cascade will break. This threshold depends on informed agents' beliefs about uninformed agents.

The herd breaking threshold is calculated by finding the value of Δ_t such that the private likelihood ratio is equal to one when a contrary private signal is realized. Consider an L-herd. If an informed agent's private likelihood ratio remains greater than 1 when a private r signal is realized, then the agent will continue the herd. However, if the realization of a private r signal will flip her private likelihood ratio below 1, then the agent will choose action R when she receives an r signal and L when she receives an l signal. Thus, her action choice reveals her signal and the herd is broken. Lemma 1 formally characterizes the herd breaking threshold.

Lemma 1. *For each $t > 1$, the herd breaking threshold $\Delta_t^*(\widehat{p})$ can be represented as:*

$$\Delta_t^*(\widehat{p}) = \frac{\ln \Lambda_0 + (t-1)\ln \phi^h}{(t-1)\ln \phi^h - (t-1)\ln \phi^c}$$

where Λ_0 depends on beliefs at the beginning of the herd, μ_0. If Δ_t crosses above the threshold, the herd breaks, whereas if Δ_t remains below the threshold, the herd persists in period t.[8]

The herd breaking threshold depends on the relative rate of information accumulation from supporting and contrary actions. This threshold rises with an increase

[8] When $t = 1$, the herd persists by definition. Recall that the definition of a herd beginning in period 0 is that the subsequent agent (i.e. agent 1) chooses the same action regardless of his signal.

in the relative informativeness of supporting actions, and falls when the relative informativeness of contrary increases. If the sample path Δ_t lies above the herd breaking threshold, then the herd will break, whereas if Δ_t lies below, then the herd will persist.

To examine how the behavior of the herd evolves across periods, we need to also characterize the limit behavior of the herd breaking threshold. Lemma 2 shows that the herd breaking threshold monotonically converges to a finite limit.

Lemma 2. *The herd breaking threshold $\Delta_t^*(\widehat{p})$ monotonically converges to a finite limit, which can be represented as:*

$$\Delta^*(\widehat{p}) = \frac{\ln \phi^h}{\ln \phi^h - \ln \phi^c}$$

The actual share of contrary actions in a herd, Δ_t, converges a.s. to its expected value, which is finite and depends on the state, by the strong Law of Large Numbers. Let $\Delta_\infty = \lim_{t\to\infty} \Delta_t$ represent this limit. Comparing the expected share of contrary actions to the limit of the herd breaking threshold allows us to determine whether the herd breaking threshold is crossed with probability one.

Recall that during a herd, the share of contrary actions lies below the herd breaking threshold. If the limit share of contrary actions lies above the limit of the herd breaking threshold, and thus lies in the region where a herd is broken, then almost surely every sample path $\{\Delta_t\}_{t=0}^{\infty}$ crosses the herd breaking threshold at some point as it converges to its limit. When this is the case, a herd breaks with probability 1.

On the other hand, if the expected share of contrary actions lies below the limit of the herd breaking threshold, then the information accumulating *on average* from supporting actions outweighs the information accumulating *on average* from contrary actions and a herd persists in the limit with positive probability. This result is due to the Law of the Iterated Logarithm Sheu (1974), which bounds the rate at which the sequence $\{\Delta_t\}_{t=0}^{\infty}$ converges to its expected value. The probability that $\{\Delta_t\}_{t=0}^{\infty}$ crosses outside this bound infinitely often is zero. This is used to show that there exists a set of sample paths of positive measure such that the actual share of contrary actions never crosses the herd breaking threshold as it converges to its limit, and on such sample paths the herd will never break.

Theorem 2 outlines the conditions under which a herd can persist with positive probability in the long run, and the conditions under which a herd almost always breaks.

Theorem 2. *Given state ω, suppose a herd has formed on action a. Let Δ_∞ represent the limit of the share of contrary actions.*

(i) If beliefs $\widehat{p}$ are such that Δ_∞ lies below the limit of the herd breaking threshold,

$$\Delta_\infty < \Delta^*(\widehat{p})$$

where $\Delta^*(\widehat{p})$ is as defined above, then there exists a set of sample paths $\widetilde{\mathbf{\Delta}}$ that occur with positive probability such that the herd breaking threshold is never crossed: for $\{\Delta_t\}_{t=0}^\infty \in \widetilde{\mathbf{\Delta}}$, $\Delta_t < \Delta_t^*(\widehat{p})\ \forall t$. On such sample paths, the herd is never broken.

(ii) If beliefs $\widehat{p}$ are such that Δ_∞ lies above the limit of the herd breaking threshold,

$$\Delta_\infty > \Delta^*(\widehat{p})$$

then for almost every sample path $\{\Delta_t\}_{t=0}^\infty$, there exists a period τ such that the herd breaking threshold is crossed at τ and the herd breaks: $\Delta_\tau > \Delta_\tau^*(\widehat{p})$. Thus, the herd is broken with probability 1.

This result is illustrated in Figure 1 for an R-herd, using arbitrary parameter values. The conditions outlining when a herd can persist and when a herd breaks with probability one will be used in the next section to determine how BIP affects the efficiency of herding.

Efficiency of Herding

The optimal action choice is the action that matches the state. Herding is *efficient* when informed agents choose the optimal action for all but finitely many periods. Thus far, the analysis has proceeded without specifying whether the herd is on the optimal action. In order to determine whether herding is efficient, it is now necessary to consider incorrect and correct herds separately.

Suppose agents are herding on action a. If the herd is correct, then the limit of the share of contrary actions is $\Delta_\infty^{\omega=a} = (1-p)(1-\pi^s)$ and if the herd is incorrect, the limit of the share of contrary actions is $\Delta_\infty^{\omega\neq a} = (1-p)\pi^s$. These limits combined with Theorem 2 can be used to determine which herds persist. The expected share of contrary actions is higher when the herd is incorrect, so if an incorrect herd persists with positive probability then so does a correct herd, and if a correct herd almost always breaks then so does an incorrect herd.[9] It is precisely when incorrect herds break but correct herds can persist that herding will be efficient.

The position of the herd breaking threshold depends on beliefs about the share of informed agents, $\widehat{p}$. An increase in $\widehat{p}$ means less information is accumulating from supporting actions. Beliefs that the herd is correct do not strengthen as quickly and a lower share of contrary actions is required to overturn the herd. Therefore, the herd breaking threshold shifts down as $\widehat{p}$ increases.

Define $\widehat{p}_1$ as the cutoff point such that when $\widehat{p} > \widehat{p}_1$, the expected share of contrary actions in an incorrect herd lies in the herd breaking region, causing incorrect herds to break with probability 1, and when $\widehat{p} < \widehat{p}_1$, the expected share of contrary actions in an incorrect herd lies below the herd breaking region, allowing incorrect herds persist with positive probability. At $\widehat{p}_1$, $\Delta_\infty^{\omega\neq a}$ lies on the limit of the herd breaking threshold, so $\widehat{p}_1$ solves:

[9] Suppose $\Delta_\infty^{\omega=a} < \Delta_\infty^{\omega\neq a}$. Then $\Delta_\infty^{\omega\neq a} < \Delta_\infty^* \Rightarrow \Delta_\infty^{\omega=a} < \Delta_\infty^*$ and if $\Delta_\infty^{\omega=a} > \Delta_\infty^* \Rightarrow \Delta_\infty^{\omega\neq a} > \Delta_\infty^*$

$$(1-p)\pi^s = \Delta^*(\widehat{p}_1)$$

Likewise, let $\widehat{p}_2$ be the cutoff point such that when $\widehat{p} > \widehat{p}_2$, the expected share of contrary actions in a correct herd list in the herd breaking region, breaking a correct herd with probability 1, and when $\widehat{p} < \widehat{p}_2$, the expected share of contrary actions lies below the herd breaking region, allowing correct herds persist with positive probability. Similarly, $\Delta_\infty^{\omega=a}$ lies on the limit of the herd breaking threshold at $\widehat{p}_2$, so $\widehat{p}_2$ solves:

$$(1-p)(1-\pi^s) = \Delta^*(\widehat{p}_2)$$

Since $(1-p)(1-\pi^s) < (1-p)\pi^s$ for $p \in [0,1)$ and the limit of the herd breaking threshold decreases with $\widehat{p}$, the cutoff point for correct herds to break is higher than the cutoff point for incorrect herds to break $(\widehat{p}_2 > \widehat{p}_1)$. Theorem 3 uses these cutoff points to characterize the efficiency of herding for any beliefs $\widehat{p}$.

Theorem 3. *Let $\widehat{p}_1$ and $\widehat{p}_2$ represent the cutoff points such that incorrect herds break with probability 1 when $\widehat{p} > \widehat{p}_1$ and correct herds break with probability 1 when $\widehat{p} > \widehat{p}_2$. Consider beliefs $\widehat{p} \in [0,1)$[10]:*

(i) If $\widehat{p} < \widehat{p}_1$ then incorrect and correct herds both persist with positive probability. There is positive probability that informed agents' action choices converge on the suboptimal action and herding is inefficient.

(ii) If $\widehat{p} \in (\widehat{p}_1, \widehat{p}_2)$ then incorrect herds break but correct herds persist with positive probability. A correct herd forms and persists in the limit with probability 1, so

[10]In the case where $\widehat{p} = 1$ and $p < 1$, observing contrary actions would be inconsistent with informed agents' beliefs. Therefore, interpreting this case necessitates assumptions on how informed agents interpret contrary actions. For example, if agents attributed contrary actions to a crazy type, then they would ignore these actions. A herd would always persist since no new information accumulates, but learning would be incomplete.

informed agents' action choices converge to the optimal action with probability 1 *and herding is efficient.*

(iii) If $\widehat{p} > \widehat{p}_2$ *then incorrect herds and correct herds both break with probability* 1*. No herd persists in the limit, so informed agents choose the suboptimal action infinitely often and action choices are inefficient.*

(iv) Suppose $p < 1$*. Then* $p \in (\widehat{p}_1, \widehat{p}_2)$ *and herding is efficient when beliefs are correct (*$\widehat{p} = p$*).*

Part (iv) of Theorem 3 demonstrates that when beliefs about the share of informed agents are correct ($\widehat{p} = p$), then herding is efficient. Informed agents attribute the correct share of supporting actions to herding, and the correct share to new information from uninformed agents. In an incorrect herd, agents don't overestimate the informativeness of supporting actions, and the public likelihood ratio doesn't become too extreme. This allows the incorrect herd to break. During a correct herd, agents attribute enough supporting actions to private signals and the correct herd is not oversensitive to contrary actions, allowing a correct herd to form and persist in the long run. This result is a special case of Theorem 4 of Acemoglu et al. (2010), which establishes complete learning for the network topology in which all agents are rational, and some agents only observe their own signal.

In fact, as long as $\widehat{p}$ is approximately correct, herding will be efficient. Although agents slightly overestimate or underestimate the informativeness of herding actions, when $\widehat{p}$ lies in the window of efficient herding then this bias is not significant enough to outweigh the accurate information accumulating from uninformed agents. An incorrect herd may persist for longer or a correct herd may break more often than would have been the case if beliefs were correct, but ultimately a correct herd will form and persist. The herding model is robust to perturbations of beliefs about the share of informed agents, and efficient information aggregation is still achieved in the long run.

When $\widehat{p}$ is too extreme in either direction, then efficient herding will no longer obtain. If agents significantly underestimate the share of informed agents (part (i) of Theorem 3), they overestimate the informativeness of supporting actions. The repeated information from actions of agents who herded accumulates at a faster rate than the new information from uninformed agents. If an incorrect herd forms, then it may persist in the long run and agents will perpetually choose the suboptimal action.

On the other hand, if agents significantly overestimate the share of agents who are informed (part (iii) of Theorem 3), they attribute new information from uninformed agents to herding. Too little information accumulates from supporting actions, preventing the herd from persisting. Both incorrect and correct herds form with positive probability, so when a herd breaks, the next herd to form may be correct or incorrect. Beliefs about the true state remain fragile, and agents oscillate between correct and incorrect herds. Both types of cascades occur infinitely often, which results in agents choosing both the optimal and suboptimal actions infinitely often. Figure 1.7 illustrates the areas corresponding to parts (i) - (iii) of Theorem 3 for an R-herd.

The robustness of efficient herding depends on the actual share of informed agents. When p is low, many agents reveal their private signal so accurate information accumulates at a faster rate. Misinterpreting the informativeness of supporting actions has a small impact on the public likelihood ratio, and the interval of efficient herding is large. On the other hand, when the actual share of informed agents is large, most agents are herding so new information accumulates slowly. The public likelihood ratio is very sensitive to any information, and inaccurate information can have a significant impact. Thus, efficient herding is robust to larger perturbations over beliefs when p is small. Figure 1.7 shows how the regions of beliefs over agent types depend on the share of informed agents.

Whenever a herd persists in the long run, public beliefs about ω will converge to a point mass on the state matching the cascade action. Learning is *complete* if public beliefs converge to a point mass on the true state, whereas learning is *fully incorrect* if public beliefs converge to a point mass on the incorrect state. If public beliefs about the state remain interior ($\mu_t \in (0,1)$), then learning is *incomplete*. Corollary 1 demonstrates that learning is complete when a correct herd persists in the limit, but learning is fully incorrect when an incorrect herd persists in the limit. If no herd persists, then the public likelihood ratio does not converge and learning is incomplete.

Corollary 1. *Learning is as follows:*

(i) If a correct herd persists, then learning is complete.

(ii) If an incorrect herd persists, then learning is fully incorrect.

(iii) If no herd persists, then the public likelihood ratio perpetually oscillates, and learning is incomplete.

Usually, the conditional public likelihood ratio is a martingale and therefore converges, preventing fully incorrect learning or perpetually fluctuating beliefs. However, this result hinges on correct beliefs about the share of informed agents. When $\widehat{p} \neq p$, the conditional public likelihood ratio is no longer a martingale. Suppose $\omega = R$ and $\widehat{p} < p$. In an incorrect L-herd, Λ_t^R is a submartingale because agents overweight L actions, and L actions increase the likelihood ratio. Therefore, when the L-herd persists, Λ_t^R diverges to infinity and learning is fully incorrect. In contrast, in a correct R-herd, Λ_t^R is a supermartingale because agents overweight R actions, which decrease the likelihood ratio. Non-negative supermartingales converge, so Λ_t^R does converge in an R-herd. In fact, when an R-herd persists, Λ_t^R converges to 0 and learning is complete.

Now suppose that $\widehat{p} > p$. In an incorrect L-herd, Λ_t^R is now a supermartingale because agents underweight L actions relative to R actions, so Λ_t^R decreases in expectation. Because $\Lambda_t^R > 1$ in an L-herd, eventually Λ_t^R crosses below 1 and the herd breaks. The opposite happens in a correct R-herd. Agents underweight R actions, so Λ_t^R is a submartingale and increases in expectation. Provided beliefs are far enough away from the truth, the Λ_t^R eventually crosses above 1, breaking the herd. Thus, both types of herds break before the conditional public likelihood ratio can converge or diverge. Once a herd breaks, Λ_t^R oscillates until another herd forms, at which point the process repeats.

1.2.3 Numerical Example

The following example illustrates the potential for BIP to confound learning. Consider a model where nature selects state L with probability $\pi^L = 0.51$. With probability $p = 0.7$, agents are "socially informed", and with probability $1 - p = 0.3$, agents are "socially uninformed" and do not observe the history. Both types of agents observe a private binary signal which reveals the true state with probability $\pi^s = 2/3$. "Socially informed" agents hold a common belief $\widehat{p}$ that previous agents are "socially informed".

Suppose that an L herd has formed. In the case of an incorrect herd, the probability of a contrary R action is equal to the probability that the agent is uninformed times the probability that this agent observes a correct signal: $\Delta_\infty^{\omega=R} = 0.2$. If the herd is correct, the probability of an R action is equal to the probability that the agent is uninformed times the probability that this agent observes an incorrect signal: $\Delta_\infty^{\omega=L} = 0.2$

We can use these parameters and Theorem 3 to characterize the efficiency of herding. If $\widehat{p} < 0.59$ then incorrect and correct herds both persist with positive probability. Learning may or may not be complete, depending on whether an incorrect or correct herd persists. If $\widehat{p} \in (0.59, 0.79)$ then incorrect herds break, but correct herds persist with positive probability. Eventually a correct herd will form and persist, leading to complete learning. If $\widehat{p} > 0.79$ then beliefs about the true state are too fragile - both incorrect and correct herds break, so learning is incomplete. Note that correct beliefs about the share of informed agents ($\widehat{p} = 0.7$) fall in the interval that leads to complete learning, as established in Theorem 3.

1.2.4 Comparative Statics

The precision of the private signal and the probability that agents are informed affects the relative positions of the herd breaking threshold and the expected share of contrary actions. Therefore, a change in any of these parameters will affect whether an incorrect herd breaks or a correct herd persists.

An increase in the probability that agents are informed, p, reduces the frequency of contrary actions since more agents observe the history and follow the herd. This can move the limit of the sample path for a correct or incorrect herd into the herd persisting region. In the former case, the increase is beneficial because it allows correct herds to persist, while in the latter case, the increase introduces inefficiency by allowing incorrect herds to persist.

An increase in the precision of the private signal, π^s, has an ambiguous effect. This change affects information accumulation through two channels: the information accumulating from each individual action, and the frequency of each type of action. More informative contrary actions decreases the herd breaking threshold, which makes it more likely that both types of herds break. The frequency of contrary actions decreases in a correct herd, and increases in an incorrect herd. Incorrect herds are less likely to persist as there are more informative and more frequent contrary actions. The overall impact on correct herds is ambiguous, as there are fewer contrary actions but each of these actions have a larger impact on beliefs.

This comparative static presents an interesting insight: more precise information may not always improve welfare. If more precise information increases the likelihood that a correct herd breaks, then herding is more likely to be inefficient. However, more precise information may also increase the probability that a correct herd forms in the first place, which would increase the efficiency of herding. The tradeoff between the efficiency gains and losses from more precise information leaves open an interesting question for future research.

1.3 Extensions

Thus far, we have established how the efficiency of information cascades depends on the relationship informed agents perceive between prior actions and signals. These results are robust to different modifications of the benchmark model, several of which are discussed informally below.

1.3.1 Public Signals

Suppose that in addition to learning from their own private information and the actions of others, informed agents also observe a sequence of public signals. We allow a public signal $\sigma_t \in \{l, r\}$ of precision $\pi^\sigma \in (1/2, 1)$ to be released with probability $\varepsilon > 0$ each period, and examine whether inefficient herding can still persist in the presence of this infinite sequence of new information.

Now the public likelihood ratio evolves to incorporate new information from action choices and public signal realizations. An r public signal multiplies the public likelihood ratio by $\phi^\sigma = \left(\frac{1-\pi^\sigma}{\pi^\sigma}\right)$, while an l public signal multiplies the likelihood ratio by $1/\phi^\sigma$. Define the contrary public signal lead $\Delta_t^\sigma \in [-1, 1]$ as the difference between the share of contrary and supporting public signals. In an L-herd, $\Delta_t^\sigma = \frac{1}{t}\sum_{s=1}^{t}\left(1_{\sigma s=r} - 1_{\sigma_s=l}\right)$ represents the difference between the share of r and l public signals.[11] In a herd, Δ_t and Δ_t^σ are sufficient statistics for the history when examining the evolution of the public likelihood ratio.

Theorem 1 is still valid with the addition of public signals, so all cascades break with positive probability.[12] In a similar fashion to Section 2.3, we characterize

[11] The additive structure of information accumulating from public signals is due to equal precision of signals across states. An l signal exactly cancels an r signal, so the difference between the number of r and l signals is a sufficient statistic for the public signal history. In the case of unequal signal precision across states, two variables would be necessary to keep track of the public signal history.

[12] With public signals, it is now possible for a herd to break even when $p = 1$, provided that the public signal is more informative than a supporting action $\left(\pi^\sigma > \frac{\widehat{p}+(1-\widehat{p})\pi^s}{1+\widehat{p}}\right)$. When agents correctly believe all previous agents observed the history ($\widehat{p} = 1$), this condition corresponds to $\pi^\sigma > 1/2$ and a herd can

when cascades can also persist with positive probability. When information accumulates from two sources, the herd breaking threshold at time t is represented as a line in (Δ, Δ^σ) space, such that an agent will follow her private signal when $(\Delta_t, \Delta_t^\sigma)$ lies in the half-plane above this threshold. The slope of the herd breaking threshold is negative and independent of t, capturing the tradeoff between contrary public signals and contrary actions: as the contrary public signal lead increases, fewer contrary actions are necessary to reach the herd breaking threshold.

The actual contrary public signal lead converges to its finite expected value, conditional on the state. Comparing $(\Delta_\infty, \Delta_\infty^\sigma,)$ to the limit of the herd breaking threshold allows us to determine whether the herd breaking threshold is crossed with probability one. If $(\Delta_\infty, \Delta_\infty^\sigma)$ lies in the half-plane above the limit of the herd breaking threshold, and thus lies in the region where a herd is broken, then almost surely every sample path $\{(\Delta_t, \Delta_t^\sigma)\}_{t=0}^\infty$ crosses the herd breaking threshold at some point as it converges to its limit and a herd breaks with probability 1. On the other hand, if the limit $(\Delta_\infty, \Delta_\infty^\sigma)$ lies in the half-plane below the limit of the herd breaking threshold, then a herd persists in the limit with positive probability. This result is due to the Law of the Iterated Logarithm for two-dimensional processes, which bounds the sequence $\{(\Delta_t, \Delta_t^\sigma)\}_{t=0}^\infty$ by a sequence of disks of decreasing radius, centered around the limit $(\Delta_\infty, \Delta_\infty^\sigma)$.[13] Figure 1.7 illustrates when a herd can persist.

As in the previous section, these conditions can be used to characterize the efficiency of herding by characterizing cut-off points. The results of Theorem 3 extend directly to the case of public signals. Thus, although the addition of public signals may reduce the scope for inefficient herding, it is not eliminated entirely. A formal characterization of the results from this section is available in a Supplementary Appendix.

These results demonstrate that BIP in the context of a sequential learning model

break as long as the public signal isn't pure noise. At the other extreme, if agents believe no preceding agents observe the history ($\widehat{p} = 0$), then the public signal needs to be more informative than the private signal, $\pi^\sigma > \pi^s$, for a herd to break with positive probability.

[13](Sheu, 1974)

with public signals has important implications. If agents overestimate the amount of new information contained in the history, BIP can confound learning by allowing an incorrect herd to persist even when public signals are released to counteract the incorrect herd. In this scenario, the best way to quash a false herd is to release public signals immediately and frequently, and a rumor may be near impossible to break once it becomes entrenched. This contrasts with the case of no BIP, in which the timing of public signal release is irrelevant - public information that breaks a cascade at time t will also break the herd at time $t+\tau$ for any τ.

1.3.2 Continuous Signals

In another variation on the model of Section 2.3, suppose that rather than receiving a binary private signal, agents receive a signal drawn from a continuous support. Smith and Sorensen (2000) show that allowing for unbounded continuous signals eliminates the possibility of incomplete learning. I examine whether this result remains true in the presence of BIP.

Let $s_t = P(\omega = L|\sigma_t)$ represent an agent's private belief that $\omega = L$ after receiving signal σ_t, computed using Bayes rule. Conditional on the state, s_t is i.i.d. with conditional distribution $F^{\omega}(s)$.and support $(0,1)$, so private signals are unbounded but no signal perfectly reveals the state. There exists a cutoff $s^*(\mu) = 1 - \mu$ such that an informed agent chooses L for $s \geq s^*(\mu)$ and R for $s < s^*(\mu)$. Note $s^*(\mu)$ is decreasing in μ - an agent chooses action L for a broader range of private signals when public beliefs are more in favor of state L. An uninformed agent, who only observes her private signal, uses the cutoff $s^* = 1/2$ to determine his action choice, independent of current public beliefs.

An information cascade forms when informed agents choose the same action for every signal in the support of F. With unbounded signals, there is always a signal that will overturn an interior public belief and cascades only occur in the limit.

However, with continuous signals, informed and uninformed agents act differently even when no herd is occurring. Repeated actions become less and less informative as beliefs strengthen, because informed agents choose this action for a wider range of private signals.

In order to establish how BIP influences the limiting properties of the public likelihood ratio, I first consider how beliefs about the share of informed agents affects the rate at which information accumulates from actions. When state L is more likely, attributing more actions to uninformed agents dampers the impact of R actions and raises the impact of L actions. An R action from an uninformed type indicates a private signal $s_t < 1/2$ while an R action from an informed type indicates a stronger private signal $s_t < 1 - \mu_t < 1/2$. On the other hand, an L action from an uninformed type is more informative than an L action from an informed type as the former indicates a private signal stronger that falls in the interval $[1/2, 1)$, whereas the latter indicates a private signal in the interval $[1 - \mu_t, 1)$, which is wider interval. The opposite is true when state R is more likely.

For the remainder of the analysis, consider the case where $\omega = L$ and define the likelihood ratio as $\Lambda_t = \left(\frac{1-\mu_t}{\mu_t}\right)$. Given that the signal space is unbounded, the only stationary limit beliefs about the state are placing probability 1 on either state L or state R. When $\widehat{p} \neq p$, Λ_t is no longer a martingale and convergence may not obtain. The following theorem characterizes which stationary limit points of Λ_t are reached with positive probability, as a function of $\widehat{p}$.

Theorem 4. *Let $\overline{\mathcal{L}}$ represent the set of stationary limit points that Λ_t converges to with positive probability. There exists cutoff points $\widehat{p}_1$ and $\widehat{p}_2$ such that*

1. *If $\widehat{p} < \widehat{p}_1$, then $\overline{\mathcal{L}} = \{0, \infty\}$*

2. *If $\widehat{p} \in (\widehat{p}_1, \widehat{p}_2)$, then $\overline{\mathcal{L}} = \{0\}$*

3. *If $\widehat{p} > \widehat{p}_2$ then $\overline{\mathcal{L}} = \{\emptyset\}$*

where $\widehat{p}_2 \in (0,1)$ *for all* p*, and* $\widehat{p}_1 \in (0,1)$ *for* $p > \frac{1-2F^L(1/2)}{2(1-F^L(1/2))}$

First consider the scenario when agents believe actions reveal more private information than is actually the case. When beliefs favor L over R ($\Lambda < 1$), the conditional likelihood ratio decreases in expectation and converges to 0 with positive probability 0, so complete learning is possible. In the case where beliefs favor R over L ($\Lambda > 1$), the conditional likelihood ratio increases in expectation. When $\widehat{p}$ is far enough away from the truth, the conditional likelihood ratio also converges to infinity with positive probability, and fully incorrect learning is also possible.

When agents believe actions contain less private information than is actually the case, the results are flipped. If beliefs favor L over R, then the conditional likelihood ratio increases in expectation, and if beliefs favor R over L, the conditional likelihood ratio decreases in expectation. When $\widehat{p}$ is far enough away from the truth, the conditional likelihood ratio converges to infinity when it is less than 1, and converges to 0 when it is greater than 1. Therefore, neither fixed point is stable and the likelihood ratio perpetually fluctuates. Provided agents' beliefs about the information content of actions are approximately correct, 0 is the only stable fixed point of the likelihood ratio and learning is correct. These results are presented formally in the Supplemental Appendix.

1.3.3 Private Values Types

Suppose there are two private value types, θ_L and θ_R, who choose actions L and R, respectively, regardless of the history, and let both types occur with positive probability. The result: less information accumulates from both supporting and contrary actions, but the conclusions of Section 2.3 are still valid. In fact, allowing for uncertainty over the share of private value types may lead to similar conclusions as information-processing bias: if agents underestimate the share of private value types, they will overestimate the informational content of actions; and if agents overestimate the share of private value types, they will underestimate the informational content of actions.

1.4 Learning About BIP

In the previous section, agents begin with exogenous and possibly incorrect beliefs about the information processing capabilities of other agents. Informed agents use these beliefs and the history to update their beliefs about the state of the world. This section will use a simple example to examine what happens when agents can also learn about the information processing capabilities of other agents.

Suppose p is a random variable distributed according to a common prior. Upon observing the history, informed agents use Bayes rule to update their beliefs about p. Let $g(p, \omega)$ represent the common prior beliefs held by informed agents (after observing their own type). For this example, suppose there are two possible shares of informed agents, $p \in \{0.4, 0.8\}$, and (p, ω) are independent of each other with marginal distributions $P(\omega = L) = \pi^L$ and $P(p = 0.4) = \pi^p$. Let the precision of the private signal be $\pi^s = 0.75$.

I will examine whether incomplete learning is possible in the case where $\omega = L$ and $p = 0.8$. Now informed agents will use the history to learn about both p and ω. Let $g(p, \omega|h_t)$ represent an agent's joint probability distribution over p and ω after observing history h_t. Then the conditional likelihood ratio between $(0.4, L)$ and any other pair (ω, p) is a martingale, represented as:

$$\Lambda_t(p, \omega) = \frac{g(p, \omega|\Delta_t^R, \Delta_t^L)}{g(.8, L|\Delta_t^R, \Delta_t^L)} = \Lambda_{t-1}(p, R)\left(\frac{P(a_t|p, \omega)}{P(a_t|.4, L)}\right)$$

where Δ_t^R is the share of contrary R actions observed in L-herds and Δ_t^L is the share of contrary L actions observed in R-herds as of time t. The probability of a given action depends on which type of herd is occurring.

By the Martingale Convergence Theorem, this likelihood ratio converges. Whether complete learning obtains depends on several factors, including whether there are pairs (p, ω) that are indistinguishable from each other, the relative weight that the prior places on these indistinguishable pairs, and the duration of previous herds on the opposite action.

If a herd persists, then there are two realizations of (p, ω) that may be indistinguishable. Consider an R-herd: if it persists, the share of L actions converges to its expected value, $\Delta_{\infty}^{L}(0.8, L) = 0.15$ (recall $\Delta_{\infty}^{L} = (1-p)\pi^{s}$ in an incorrect herd). There are two possible pairs (p, ω) that would give rise to this share of L actions, as $\Delta_{\infty}^{L}(0.4, R)$ is also equal to 0.15 (recall $\Delta_{\infty}^{L} = (1-p)\left(1-\pi^{s}\right)$ if the state is R). It is impossible to distinguish between $(0.8, L)$ and $(0.4, R)$ in an R-herd since both of these pairs result in the same expected share of contrary actions. Therefore $\frac{P(a_t|0.4,R)}{P(a_t|.8,L)} = 1$ in an R-herd and no information is gained about the relative likelihood of these two events. The probability of all other pairs (p, ω) converges to 0 when the R-herd persists, since $\left(\frac{P(a_t|p,\omega)}{P(a_t|.4,L)}\right) \neq 1$ and zero is the only finite stationary point of $\Lambda_t(p, \omega)$ for such pairs.

Next consider information from previous L-herds. The share of contrary R actions in previous L-herds also yields information about p, and these contrary R actions help distinguish between $(0.8, L)$ and $(0.4, R)$. The only pair that is indistinguishable from $(0.8, L)$ in an L-herd is $(0.93, R)$, which differs from the pair that is indistinguishable in an R-herd (and in this example, is not in the support of (p, ω)). Thus, the true value of p would be identified if both expected shares are observed. Even a finite number of observations from a previous L-herd helps distinguish between $(0.8, L)$ and $(0.4, R)$. Suppose previous L-herds occurred for τ_L periods and yielded a share Δ^{R} contrary actions. Then the information gleaned from these L-herds multiplies the relative likelihood of $(0.8, L)$ and $(0.4, R)$ by

$$\phi^{\tau_L} = \left[\left(\frac{P(a = L|0.4, R)}{P(a = L|0.8, L)} \right)^{(1-\Delta^R)} \left(\frac{P(a = R|0.4, R)}{P(a = R|0.8, L)} \right)^{\Delta^R} \right]^{\tau_L}$$

When Δ^R is close to its expected value, $\Delta^R_\infty = .05$, this expression is less than 1, and therefore increases the relative likelihood of $(0.8, L)$ compared to $(0.4, R)$

We can now characterize the limit of $\Lambda_t(0.4, R)$ in an R-herd:

$$\Lambda_\infty(0.4, R) = 3 \left(\frac{\pi^p}{1 - \pi^p} \right) \left(\frac{1 - \pi^L}{\pi^L} \right) \phi^{\tau_L}$$

If $\Lambda_\infty(0.4, R) > 3$ then an R-herd can persist with positive probability. In the limit, agents believe that state R is more likely even when they receive a private l signal. Because Λ_t is a martingale, fully incorrect learning is not possible as in the previous section.[14] However, learning is incomplete and beliefs remain interior at $\Lambda_\infty(0.4, R)$. If $\Lambda_\infty(0.4, R) < 3$ then the R-herd breaks with probability 1 as agents will eventually come to believe state L more likely and end the herd. Note that the duration of previous L-herds affects whether a given herd can persist, as longer previous L-herds decrease $\Lambda_\infty(0.4, R)$.

In this simple example, there are no indistinguishable pairs in L-herds. Thus, an L-herd always persists with positive probability, and learning will be complete when this occurs. However, a more general prior over p allows the possibility of indistinguishable pairs in both correct and incorrect herds, precluding complete learning for both cases. This section illustrates that, when agents can learn about others information processing capabilities, the scope for inefficient herding is reduced and the possibility of fully incorrect learning is generally eliminated.

[14] Fully incorrect learning about ω is not generally possible when agents can learn about p. This will only occur if agents have a common posterior that puts no weight on the correct value of p.

1.5 Discussion and Conclusion

This paper demonstrates that a bias about how others' process information can significantly affect the efficiency of learning. Particularly, it is possible for agents to continue to choose the suboptimal action despite the release of new information contradicting the herd. In the benchmark model, this would be impossible. Inefficient herding occurs because information ceases to aggregate; when even the smallest amount of information continues to accumulate, inefficient herding no longer occurs. Experimental evidence from Goeree et al. (2007) suggests that new information does indeed continue to accumulate in a herd: regardless of how many previous agents chose the same action, some agents still follow their private signal. In the benchmark model, this off-the-equilibrium-path action would be ignored since it is not rational. However, it seems plausible that subsequent agents would recognize these off-the-equilibrium-path actions are likely to reveal an agent's private signal, and therefore contain information. Thus, BIP allows new information to continue to enter the model, and provides an explanation for inefficient herding when this is the case. Inefficient herding occurs because the rate of information accumulation from repeated information outweighs the rate of information accumulation from new information, and these herds can persist even when contradicted by public information. Additionally, it explains how convergence may fail to obtain even when public information is released that supports the correct herd.

Experimental evidence from Koessler et al. (2008) supports the possibility of BIP in an observational learning model. They examine the fragility of cascades in a model where one agent receives a more precise signal than others. The unique Nash equilibrium of such a model is for the high informed agent to follow her signal. Thus, receipt of a contrary signal overturns a cascade. Koessler et al. (2008) find that highly informed agents rarely overturn a cascade when equilibrium prescribes that they do so. As the length of the cascade increases, highly informed agents become even less likely to follow their signal: highly informed participants break 65% of cascades when there are two identical actions, but only 15% of cascades when there are 5 or more identical actions. This phenomenon is likely explained by the evolution of participants' beliefs. The evolution of elicited beliefs is similar to the belief process that would arise if all agents followed their signal, and thus conveyed their private information. In addition, Koessler et al. (2008) find that off-the-equilibrium-path play frequently occurs, and these non-equilibrium actions are informative, providing support for the actual presence of some uninformed agents, in addition to a strong belief about their presence.

Kubler and Weizsacker (2004) also find evidence consistent with BIP. They conclude that subjects do learn from their predecessors, but are uncertain about the share of previous agents who also learned from their predecessors. Particularly, agents underestimate the share of previous agents who herded, and therefore overestimate the amount of new information contained in previous actions.

Another interesting consequence of BIP is that agents may actually be worse off if more information accumulates than was expected. As p decreases, more private information accumulates since fewer agents observe the history. However, if p is far enough below $\widehat{p}$, correct herds will become too fragile and herding will be inefficient.

Conformity preferences is another bias that could make agents more likely to herd as the length of the herd increases, but equilibrium play in such a model differs significantly from a model with BIP and public signals. With BIP, if the contrary public signal lead is high enough to break a herd of a given length, then subsequent agents do not continue to herd. This model has a unique equilibrium where agents choose whichever state that they believe is more likely based on the history and the degree of BIP. However, with conformity preferences, if the preference to conform is large enough, then it is irrelevant whether or not the contrary public signal lead is high enough to make agents believe the alternative state more likely. Agents simply want to choose the action that the majority of other agents will choose. So there are multiple equilibria where agents choose the state that they believe the majority of other agents will choose, independent of the likelihood that this state is the true state.

This model leaves open interesting questions for future research on information processing capabilities. Individuals may differ in their depth of reasoning and their ability to combine different information sources. Such biases may have important implications for the way information is aggregated. Examining the implications of BIP in a more general model may yield interesting insights into this issue. While the assumption of common beliefs over the informational content of the history is a good starting point, a valid criticism is that this model requires implausible levels of belief coordination. Thus, examining how the model fares with heterogenous beliefs about the information processing capabilities of other agents is another avenue for future research. Allowing partial observability of histories would be natural extensions to generalize the model, while introducing payoff interdependencies would make the model applicable to election and financial market settings.

I thank Vince Crawford, Ernesto Dal Bo, Frederic Koessler, Craig McKenzie, Matthew Rabin, Joel Sobel, and especially Nageeb Ali for useful comments. I also thank participants of the UCSD theory lunch for helpful feedback.

1.6 Appendix: Proofs

Proof of Theorem 1 on pp. 13

Let $\left(\frac{\mu_0}{1-\mu_0}\right)$ represent public beliefs before the action that begins the herd.

Suppose $p < 1$. Let there be an L-herd in period t with $y = (t-1)(1-2\Delta_t)$ net L actions (i.e. the number of L actions minus the number of R actions). Each R action decreases the likelihood ratio by a factor ϕ^c and each L action increases the likelihood ratio by a factor ϕ^h. Observe $\phi^h * \phi^c \leq 1$, so the net effect of an R action and an L action decreases the likelihood ratio. Then $y > 0$ R actions will outweigh the y net L actions. Let κ be the greatest k such that $\left(\frac{\mu_0}{1-\mu_0}\right)\left(\frac{\pi^s}{1-\pi^s}\right)(\phi^c)^k < 1$. Then

κ R actions outweigh initial public beliefs. Note κ is finite since $\left(\frac{\mu_0}{1-\mu_0}\right) < \infty$. Then $y + \kappa$ R actions will break this herd. The probability that the next $y + \kappa$ actions are R is:

$$
\begin{aligned}
[(1-p)(1-\pi^s)]^{y+\kappa} &> 0 \text{ if the herd is correct} \\
[(1-p)\pi^s]^{y+\kappa} &> 0 \text{ if the herd is incorrect}
\end{aligned}
$$

This is a lower bound on the probability that the herd breaks. Similar analysis yields the same results for R-herds. Q.E.D.

Proof of Lemma 1 on pp. 14

Let μ_0 represent public beliefs before the action that begins the herd and suppose agents are herding on action a^h. Let $\Lambda_t = \frac{P(\omega = a^h | h_t)}{P(\omega \neq a^h | h_t)}$ be the public likelihood ratio of the probability that the herd is correct to the probability that the herd is incorrect, let $\Lambda_t^{s^c}$ be the updated private likelihood ratio after a contrary private signal, and let $\Lambda_t^{s^h}$ be the updated private likelihood ratio after a supporting private signal. In an L-herd, $\Lambda_0 = \left(\frac{\mu_0}{1-\mu_0}\right)$ while in an R-herd, $\Lambda_0 = \left(\frac{1-\mu_0}{\mu_0}\right)$. Note $\Lambda_0 > 1$ by definition. Provided it is still optimal for agents who observe the history to herd, the likelihood ratio in a herd evolves as follows:

$$
\begin{aligned}
\Lambda_t &= \Lambda_0 \left(\frac{\pi^s}{1-\pi^s}\right) \left(\phi^h\right)^{(t-1)(1-\Delta_t)} \left(\phi^c\right)^{(t-1)\Delta_t} \\
&= \Lambda_0 \left(\phi^h\right)^{(t-1)(1-\Delta_t)} \left(\phi^c\right)^{(t-1)\Delta_t - 1} \\
s_t = s^c \Rightarrow \Lambda_t^{s^c} &= \Lambda_0 \left(\phi^h\right)^{(t-1)(1-\Delta_t)} \left(\phi^c\right)^{(t-1)\Delta_t} \\
s_t = s^h \Rightarrow \Lambda_t^{s^h} &= \Lambda_0 \left(\phi^h\right)^{(t-1)(1-\Delta_t)} \left(\phi^c\right)^{(t-1)\Delta_t - 2}
\end{aligned}
$$

The herd breaking threshold Δ_t^* is calculated by finding the value of Δ_t that satisfies $\Lambda_t^{s^c} = 1$, so the private likelihood is equal to one when a contrary private signal is realized:

$$\begin{aligned}
& \Lambda_t^{s^c} = 1 \\
\Rightarrow \quad & \Lambda_0 \left(\phi^h\right)^{(t-1)(1-\Delta_t^*)} \left(\phi^c\right)^{(t-1)\Delta_t^*} = 1 \\
\Rightarrow \quad & \ln \Lambda_0 + (t-1)(1-\Delta_t^*) \ln \phi^h + (t-1)\Delta_t^* \ln \phi^c = 0 \\
\Rightarrow \quad & \Delta_t^* = \frac{\ln \Lambda_0 + (t-1) \ln \phi^h}{(t-1) \ln \phi^h - (t-1) \ln \phi^c}
\end{aligned}$$

Q.E.D.

Proof of Lemma 2 on pp. 15

$$\Delta_\infty^* = \lim_{t \to \infty} \Delta_t^* = \lim_{t \to \infty} \frac{\ln \Lambda_0 + (t-1) \ln \phi^h}{(t-1) \ln \phi^h - (t-1) \ln \phi^c} = \frac{\ln \phi^h}{\ln \phi^h - \ln \phi^c}$$

Also note that Δ_t^* monotonically decreases to its limit, a fact which will be used in the proof of Theorem 2.

$$\begin{aligned}
\frac{d}{dt}\Delta_t^* &= \frac{d}{dt}\left[\left(\frac{1}{t-1}\right) * \frac{\ln \Lambda_0 + (t-1) \ln \phi^h}{\ln \phi^h - \ln \phi^c}\right] \\
&= \left(\frac{1}{t-1}\right)^2 * \frac{-\ln \Lambda_0}{\ln \phi^h - \ln \phi^c} < 0
\end{aligned}$$

Q.E.D.

Proof of Theorem 2 on pp. 16

(a) Let $\Delta_\infty < \Delta^*$.

Define a numeric representation of the action choice space as an infinite sequence of spaces $\{Y_i\}, i = 1, 2, ...$ over which a probability measure is defined, where $Y_i = \{0, 1\}\ \forall i$. Let $Y = Y_1 \times Y_2 \times ... \times Y_i \times ...$ be the product of such spaces. Let $\{y_i\}$ be an infinite sequence of mutually independent and identically distributed random variables drawn from this space, with distribution function $g(y)$ defined as follows, depending on whether the herd is correct or incorrect:

Correct Herd	Incorrect Herd
$g(0) = p + (1-p)\pi^s$	$g(0) = p + (1-p)(1-\pi^s)$
$g(1) = (1-p)(1-\pi^s)$	$g(1) = (1-p)\pi^s$

A supporting action corresponds to $y_i = 0$, and a contrary action corresponds to $y_i = 1$. Let μ_y represent the expected value of y_i and σ_y^2 represent the variance of y_i.The expected value and variance of y_i are as follows:

Correct Herd	Incorrect Herd
$\mu_y = (1-p)(1-\pi^s)$	$\mu_y = (1-p)\pi^s$
$\sigma_y^2 = (1-p)(1-\pi^s) - (1-p)^2(1-\pi^s)^2$	$\sigma_y^2 = (1-p)\pi^s - (1-p)^2(\pi^s)^2$

Note the equivalence of $\sum_{i=1}^t y_i$ and $t\Delta_{t+1}$, and therefore the equivalence of μ_y and $\Delta_\infty = E[\Delta_t]$. Transform y_i as follows:

$$\overline{y}_i = y_i - \mu_y$$

and calculate $E[\overline{y}_i] = 0$ and $Var(\overline{y}_i) = \sigma_y^2$ and $\sum_{i=1}^t Var(\overline{y}_i) = t\sigma_y^2$. We have $|\overline{y}_i|$

bounded above by 1, which is independent of t. So trivially,

$$\sup_{i \le t} \text{l.u.b. } |\overline{y}_i| = o\left(\frac{t\sigma_y^2}{\log\log t\sigma_y^2}\right)^{\frac{1}{2}}$$

for each t as $t \to \infty$. The necessary assumptions for the law of the iterated logarithm (LIL) applied to a one-dimensional independent random variable are satisfied.[15] The LIL can be used to bound $\sum_{i=1}^{t} \overline{y}_i$:

$$\lim_{t\to\infty} \sup \frac{\sum_{i=1}^{t} \overline{y}_i}{\sqrt{2t\sigma_y^2 \log\log t\sigma_y^2}} = 1 \text{ a.s.}$$

Thus, for $\delta > 0$.

$$P\left[\sum_{i=1}^{t} \overline{y}_i \ge (1+\delta)\sqrt{2t\sigma_y^2 \log\log t\sigma_y^2} \text{ i.o.}\right] = 0$$

Define

$$B_t = (1+\delta)\sqrt{\frac{2\sigma_y^2 \log\log t\sigma_y^2}{t}}$$

This means that for almost all realizations of $\{y_i\}$, there exist only *finitely many t* such that $\frac{1}{t}\sum_{i=1}^{t} y_i$ lies outside $\Theta_t = [\mu_y + B_t, \mu_y - B_t]$ Define

$$\zeta = \left\{\{\widehat{y}_i\} \mid \frac{1}{t}\sum_{i=1}^{t} \widehat{y}_i > \mu_y + B_t \text{ for some } t\right\}$$

[15] The necessary assumptions are:

(i) $\overline{y}_1, \overline{y}_2, ...$ is an infinite sequence of real-valued independent random variables of class L^2

(ii) $E[\overline{y}_i] = 0$

(iii) $\sum_{i=1}^{n} Var(\overline{y}_i) \to \infty$ as $n \to \infty$

(v) $\sup_{i \le t}(\text{l.u.b. } |\overline{y}_i||) = o\left(\frac{\sum_{i=1}^{n} Var(\overline{y}_i)}{\log\log \sum_{i=1}^{n} Var(\overline{y}_i)}\right)^{\frac{1}{2}}$ for each n as $n \to \infty$

as the set of realizations of $\{y_i\}$ such that $\frac{1}{t}\sum_{i=1}^{t} y_i$ crosses its upper bound at least once. To show that the measure of ζ is strictly less than 1, consider the following. For each $\{\widehat{y}_i\} \in \zeta$, form a corresponding sample path $\{y'_i\}$ by changing $\widehat{y}_\tau$ to $y'_\tau = 0$ for each τ such that $\frac{1}{\tau}\sum_{i=1}^{\tau} \widehat{y}_\tau > \mu_y + B_\tau$ (for any $\{\widehat{y}_i\} \in \zeta$ there are only finitely many such τ). Then each element in ζ has a unique corresponding element in $Z \setminus \zeta$. So the measure of the set $Z \setminus \zeta$ is at least as large as the measure of ζ and this implies that the measure of ζ is strictly less than 1. Therefore, the measure of $Z \setminus \zeta$ is strictly positive. So there exists a set of realizations of $\{y_i\}$ that occur with positive probability such that $\frac{1}{t}\sum_{i=1}^{t} y_i$ never crosses outside $\mu_y + B_\tau$.

Given that $\{r_t\}$ and $\{\Delta_t^*\}$ are monotonic with respect to t,

$$\begin{aligned} \{B_t\} &\rightarrow 0 \\ \{\Delta_t^*\} &\rightarrow \Delta^* \\ \Delta_\infty &= \mu_y < \Delta^* \end{aligned}$$

there are at most a finite number of periods k such that Δ_t^* lies inside

$$\Theta_t = [\mu_y + B_t, \mu_y - B_t]$$

The probability that Δ_t doesn't cross Δ_t^* during these k periods is bounded below by $g(0)^k > 0$ (the probability of k supporting actions, which will never break a herd), and thus is strictly positive. Once Δ_t^* lies above Θ_t, all realizations in the set $Z \setminus \zeta$ never cross outside Θ_t, and therefore never cross Δ_t^*. So there is a set of sample path realizations that occur with positive probability such that the actual share of contrary actions never crosses the threshold required to break the herd, allowing the herd to persist with positive probability in the limit. QED. Hartman and Wintner (1941)

(b): Suppose $\Delta_\infty > \Delta^*$. Then Δ_∞ lies in the region that breaks a herd. By the law of large numbers, almost all sample paths of $\{\Delta_t\}_{t=0}^{\infty}$ converge to Δ_∞, so the threshold required to break the herd is crossed with probability 1. Thus the herd is broken with probability 1. Q.E.D.

Proof of Theorem 3 on pp. 18

Recall $\phi^h = \left(\frac{\widehat{p}+(1-\widehat{p})\pi^s}{\widehat{p}+(1-\widehat{p})(1-\pi^s)}\right)$. Note $\frac{d\phi^h}{d\widehat{p}} = \frac{1-2\pi^s}{(1-\pi^s+\widehat{p}\pi^s)^2} < 0$ and $\ln(\phi^c) = \ln\left(\frac{1-\pi^s}{\pi^s}\right) < 0$

$$\frac{d\Delta^*(\widehat{p})}{d\widehat{p}} = \frac{-\ln(\phi^c)\frac{d\phi^h}{d\widehat{p}}}{\phi^h\left(\ln(\phi^h) - \ln(\phi^c)\right)^2} < 0$$

(i) By definition, $\Delta_\infty^{\omega\neq a} = \Delta^*(\widehat{p}_1)$ so the limit of the sample path for an incorrect herd lies on the herd breaking threshold at $\widehat{p}_1$. Since $\frac{d\Delta^*(\widehat{p}_1)}{d\widehat{p}} < 0$, and the limit of the sample path doesn't depend on $\widehat{p}$, for $\widehat{p} < \widehat{p}_1$, $\Delta_\infty^{\omega\neq a} = \Delta^*(\widehat{p}_1) < \Delta^*(\widehat{p})$ so $\Delta_\infty^{\omega\neq a}$ lies below the herd breaking threshold and incorrect herds persist with positive probability, by theorem 2. Thus, there is positive probability that an incorrect herd persists, and agents choose only the suboptimal action infinitely often. Correct herds also persist with positive probability since $\Delta_\infty^{\omega=a} < \Delta_\infty^{\omega\neq a} < \Delta^*(\widehat{p}) \Rightarrow \Delta_\infty^{\omega=a}$ also lies below the herd breaking threshold. Thus, agents' action choices also converge on the optimal action with positive probability.

(iii) By definition, $\Delta_\infty^{\omega\neq a} = \Delta^*(\widehat{p}_2)$ so the limit of the sample path for a correct herd lies on the herd breaking threshold at $\widehat{p}_2$. Since $\frac{dk(\widehat{p})}{d\widehat{p}} < 0$, and the limit of the sample path doesn't depend on $\widehat{p}$, for $\widehat{p} > \widehat{p}_2$, $\Delta_\infty^{\omega=a} = \Delta^*(\widehat{p}_2) > \Delta^*(\widehat{p})$ so $\Delta_\infty^{\omega=a}$ lies above the herd breaking threshold and correct herds break with probability 1, by theorem 2. Since $\Delta_\infty^{\omega\neq a} > \Delta_\infty^{\omega=a} > \Delta^*(\widehat{p})$, incorrect herds also break with probability 1. Thus, no herd persists in the limit. Each time a herd breaks, correct and incorrect herds both form with positive probability. A new herd will form if the same action is played in the two periods subsequent to the herd breaking. The probability of two correct actions is $(\pi^s)^2 > 0$ and the probability of two incorrect actions is $(1-\pi^s)^2 > 0$. Since neither type of herd persists in the limit, both correct and incorrect herds form infinitely often. Therefore both the optimal and suboptimal action are chosen infinitely often, and herding is inefficient.

(ii) If $\widehat{p} \in (\widehat{p}_1, \widehat{p}_2)$ then $\Delta_\infty^{\omega \neq a} = \Delta^*(\widehat{p}_1) > \Delta^*(\widehat{p})$ and $\Delta_\infty^{\omega = a} = \Delta^*(\widehat{p}_2) < \Delta^*(\widehat{p})$ so incorrect herds break with probability 1 but correct herds persist with positive probability. Let A represent the event where no herd is occurring. Let $q_C(\mu)$ represent the probability that a correct herd forms and $q_I(\mu)$ represent the probability that an incorrect herd forms, given that no herd is occurring. Let $r_C(\mu)$ represent the probability that a correct herd persists in the limit. These probabilities depend on current public beliefs, but all are positive. Suppose event A is occurring in period t; that is, in period t, no herd is occurring. Then the probability that event A occurs again at some future period is $1 - q_C(\mu_t) r_C(\mu_t) < 1$. Event A occurs again if (a) no herd forms in period t (then A occurs in $t+1$) (b) an incorrect herd forms in period t (since this herd breaks with probability 1) (c) a correct herd forms in period t and breaks. The periods that A occur in form an increasing sequence $\tau_1 < \tau_2 < \ldots$ and for each τ_k, the probability that A occurs again is $1 - q_C(\mu_{\tau_k}) r_C(\mu_{\tau_k}) < 1$. Thus, the probability that event A occurs infinitely often is $\lim_{n \to \infty} \prod_{k=1}^{n} 1 - q_C(\mu_{\tau_k}) r_C(\mu_{\tau_k}) = 0$. So with probability 1, A occurs only a finite number of times. Thus, a correct herd forms and persists with probability 1, and agents will choose only the optimal action infinitely often.

(iv) (a) First show that $p > \widehat{p}_1$ by showing that the expected share of contrary actions in an incorrect herd lies above the herd breaking threshold when $\widehat{p} = p$ i.e. show $\Delta_\infty^{\omega \neq a} > \Delta^*(p)$

Consider the following equations:

$$\begin{aligned}
f(p) &= \frac{(1-p)\pi^s}{(1-(1-p)\pi^s)} \ln\left(\frac{\pi^s}{1-\pi^s}\right) \\
g(p) &= \ln(\phi^h) \\
&= \ln\left(\frac{p+(1-p)\pi^s}{p+(1-p)(1-\pi^s)}\right) \\
&= \ln\left(\frac{(1-\pi^s)p+\pi^s}{1-\pi^s+\pi^s p}\right)
\end{aligned}$$

At $p = 1$

$$f(1) = g(1) = 0$$

At $p = 0$

$$f(0) = \frac{\pi^s}{1-\pi^s} \ln\left(\frac{\pi^s}{1-\pi^s}\right) > g(0) = \ln\left(\frac{\pi^s}{1-\pi^s}\right)$$

since $\frac{\pi^s}{1-\pi^s} > 1$. Take the derivative of each expression with respect to p:

$$\begin{aligned}
\frac{d}{dp} f(p) &= -\frac{\pi^s}{(1-\pi^s+\pi^s p)^2} \ln\left(\frac{\pi^s}{1-\pi^s}\right) < 0 \\
\frac{d}{dp} g(p) &= \frac{-(2\pi^s - 1)}{((1-\pi^s)p + \pi^s)(1-\pi^s+\pi^s p)} < 0
\end{aligned}$$

Take the second derivative of each expressions with respect to p:

$$\begin{aligned}
\frac{d^2}{dp^2} f(p) &= \frac{2(\pi^s)^2}{(1-\pi^s+\pi^s p)^3} \ln\left(\frac{\pi^s}{1-\pi^s}\right) > 0 \\
\frac{d^2}{dp^2} g(p) &= \frac{((1-\pi^s)^2 + (\pi^s)^2 + 2p\pi^s(1-\pi^s))(2\pi^s - 1)}{[((1-\pi^s)p + \pi^s)(1-\pi^s+\pi^s p)]^2} > 0
\end{aligned}$$

Given $f(0) > g(0)$, $f(1) = g(1)$ and both functions monotonically decrease at a decreasing rate, we can conclude that $f(p) > g(p)$ over the interval $[0, 1)$.

$$
\begin{aligned}
& f(p) > g(p) \\
\Rightarrow \quad & \frac{-(1-p)\pi^s}{(1-(1-p)\pi^s)} \ln\left(\phi^c\right) > \ln\left(\phi^h\right) \\
\Rightarrow \quad & (1-p)\pi^s > \frac{\ln\left(\phi^h\right)}{\ln\left(\phi^h\right) - \ln\left(\phi^c\right)} \\
\Rightarrow \quad & \Delta_\infty^{\omega \neq a} > \Delta^*(p) \\
\Rightarrow \quad & p > \widehat{p}_1
\end{aligned}
$$

Thus, $p > \widehat{p}_1$ for all $p \in [0, 1)$.

(b) Next show that $p < \widehat{p}_2$ by showing that the limit of the realized sample path for a correct herd lies below the herd breaking threshold limit when $\widehat{p} = p$ i.e. show $\Delta_\infty^{\omega = a} < \Delta^*(p)$. Consider the following equations:

$$
\begin{aligned}
f(p) &= \frac{(1-p)(1-\pi^s)}{(1-(1-p)(1-\pi^s))} \ln\left(\frac{\pi^s}{1-\pi^s}\right) \\
&= \frac{1-p-\pi^s+\pi^s p}{(p+\pi^s-\pi^s p)} \ln\left(\frac{\pi^s}{1-\pi^s}\right) \\
g(p) &= \ln(\phi^h) \\
&= \ln\left(\frac{(1-\pi^s)p+\pi^s}{1-\pi^s+\pi^s p}\right)
\end{aligned}
$$

At $p = 1$

$$
f(1) = g(1) = 0
$$

At $p = 0$

$$f(0) = \frac{1-\pi^s}{\pi^s} \ln\left(\frac{\pi^s}{1-\pi^s}\right) < g(0) = \ln\left(\frac{\pi^s}{1-\pi^s}\right)$$

since $\frac{1-\pi^s}{\pi^s} < 1$. Take the derivative of each expression with respect to p:

$$\begin{aligned} \frac{d}{dp} f(p) &= \frac{-(1-\pi^s)}{(p+\pi^s-\pi^s p)^2} \ln\left(\frac{\pi^s}{1-\pi^s}\right) < 0 \\ \frac{d}{dp} g(p) &= \frac{-(2\pi^s-1)}{((1-\pi^s)p+\pi^s)(1-\pi^s+\pi^s p)} < 0 \end{aligned}$$

Take the second derivative of each expressions with respect to p:

$$\begin{aligned} \frac{d^2}{dp^2} f(p) &= \frac{2(1-\pi^s)^2}{(p+\pi^s-\pi^s p)^3} \ln\left(\frac{\pi^s}{1-\pi^s}\right) > 0 \\ \frac{d^2}{dp^2} g(p) &= \frac{((1-\pi^s)^2+(\pi^s)^2+2p\pi^s(1-\pi^s))(2\pi^s-1)}{[((1-\pi^s)p+\pi^s)(1-\pi^s+\pi^s p)]^2} > 0 \end{aligned}$$

Given $f(0) < g(0)$, $f(1) = g(1)$ and both functions monotonically decrease at a decreasing rate, we can conclude that $f(p) < g(p)$ over the interval $[0, 1)$.

$$\begin{aligned} & f(p) < g(p) \\ \Rightarrow\ & \frac{(1-p)(1-\pi^s)}{(1-(1-p)(1-\pi^s))} \ln\left(\frac{\pi^s}{1-\pi^s}\right) < \ln\left(\phi^h\right) \\ \Rightarrow\ & (1-p)(1-\pi^s) < \frac{\ln\left(\phi^h\right)}{\ln\left(\phi^h\right) - \ln(\phi^c)} \\ \Rightarrow\ & \Delta_\infty^{\omega=a} < \Delta^*(p) \\ \Rightarrow\ & p < \widehat{p}_2 \end{aligned}$$

Thus, $p < \widehat{p}_2$ for all $p \in [0, 1)$. This illustrates that correct herds persist and incorrect herds break when $\widehat{p} = p$. Q.E.D.

Proof of Corollary 1 on pp. 21

Suppose an L-herd forms and persists. In order for a herd to persist,

$$\Delta_t < \Delta_t^*(\widehat{p}) \; \forall t$$

Let $\Delta_t = \Delta_t^* - c_t$ for some $c_t > 0$. Note $\Lambda_0 \left(\phi^h\right)^{(t-1)(1-\Delta_t^*)} \left(\phi^c\right)^{(t-1)\Delta_t^*} = 1 \; \forall t$ since this represents beliefs at the herd breaking threshold. We can rewrite the public likelihood ratio in an L-herd as:

$$\begin{aligned}
\frac{\mu_t}{1-\mu_t} &= \Lambda_0 \left(\frac{\pi^s}{1-\pi^s}\right) \left(\phi^h\right)^{(t-1)(1-\Delta_t)} \left(\phi^c\right)^{(t-1)\Delta_t} \\
&= \Lambda_0 \left(\frac{\pi^s}{1-\pi^s}\right) \left(\phi^h\right)^{(t-1)(1-\Delta_t^*+c_t)} \left(\phi^c\right)^{(t-1)(\Delta_t^*-c_t)} \\
&= \Lambda_0 \left(\frac{\pi^s}{1-\pi^s}\right) \left(\phi^h\right)^{(t-1)(1-\Delta_t^*)} \left(\phi^c\right)^{(t-1)(\Delta_t^*)} \left(\phi^h\right)^{(t-1)c_t} \left(\phi^c\right)^{-(t-1)c_t} \\
&= \left(\frac{\pi^s}{1-\pi^s}\right) \left(\frac{\phi^h}{\phi^c}\right)^{(t-1)c_t}
\end{aligned}$$

Note $\left(\frac{\phi^h}{\phi^c}\right) > 1$.

(i) If the L-herd is correct, then learning is complete when the inverse public likelihood ratio converges to zero. The limit of the public likelihood ratio is as follows:

$$\lim_{t\to\infty} \frac{1-\mu_t}{\mu_t} = \lim_{t\to\infty} \left(\frac{1-\pi^s}{\pi^s}\right) \left(\frac{\phi^c}{\phi^h}\right)^{(t-1)c_t} = 0$$

Similar analysis shows $\lim_{t\to\infty} \frac{\mu_t}{1-\mu_t} = 0$ in a correct R-herd. Thus, when a correct herd persists, learning is complete.

(ii) If the L-herd is incorrect, then learning is fully incorrect when the public likelihood ratio converges to infinity:

$$\lim_{t\to\infty}\frac{\mu_t}{1-\mu_t}=\lim_{t\to\infty}\left(\frac{\pi^s}{1-\pi^s}\right)\left(\frac{\phi^h}{\phi^c}\right)^{(t-1)c_t}=\infty$$

Similar analysis shows $\lim_{t\to\infty}\frac{1-\mu_t}{\mu_t}=\infty$ in an incorrect R-herd. Thus, learning is fully incorrect when an incorrect herd persists.

(iii) Suppose incorrect and correct L-herds both break. Then $\exists\tau$ s.t. $\Delta_\tau>\Delta^*_\tau(\widehat{p})$ and the public likelihood ratio falls below 1. If an L-herd forms again, this will repeat whereas if an R-herd forms, then the public likelihood ratio eventually rises above 1, at which point the R-herd breaks. Thus, the public likelihood ratio never converges in either an L-herd or an R-herd and public beliefs remain interior, resulting in incomplete learning. Q.E.D.

1.7 Figures

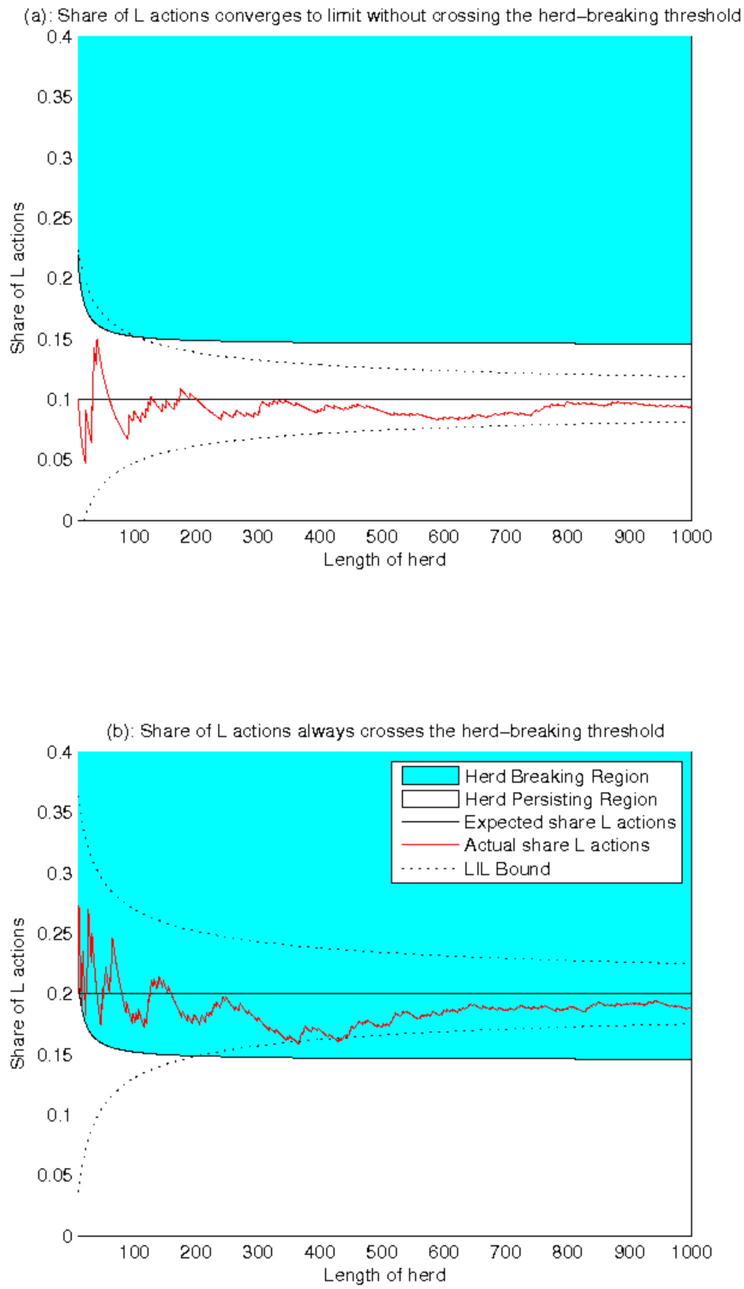

Figure 1.1: The Herd Breaking Threshold

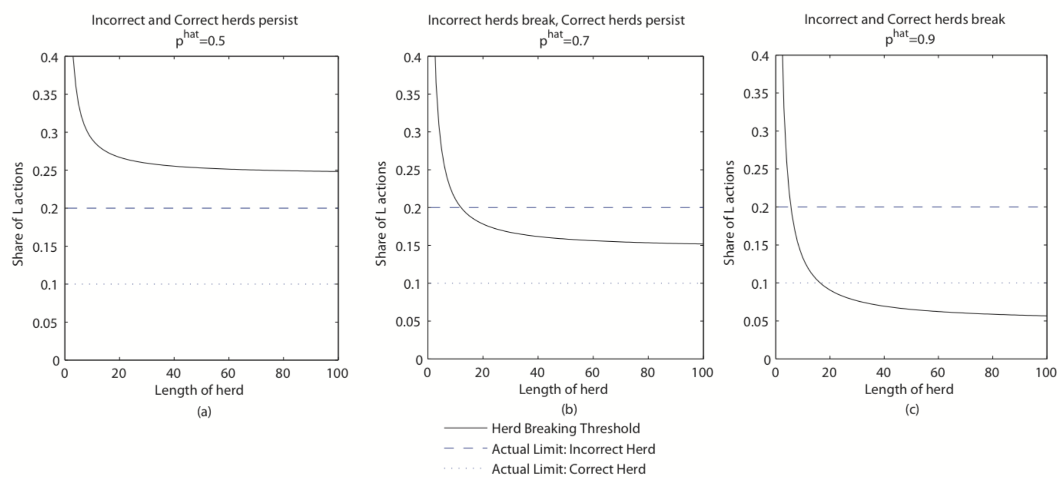

Figure 1.2: When Can a Herd Persist?

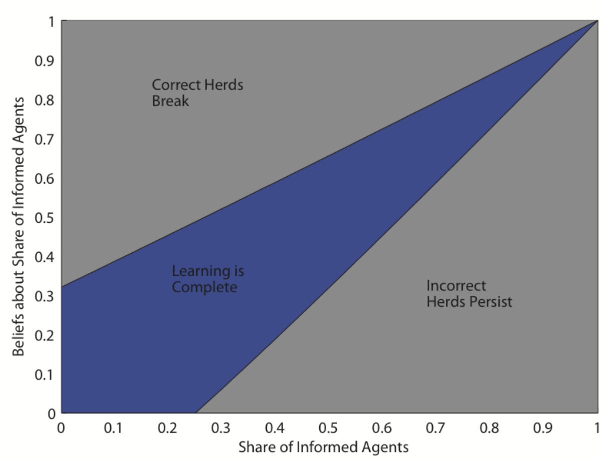

Figure 1.3: Learning is complete when beliefs are approximately correct

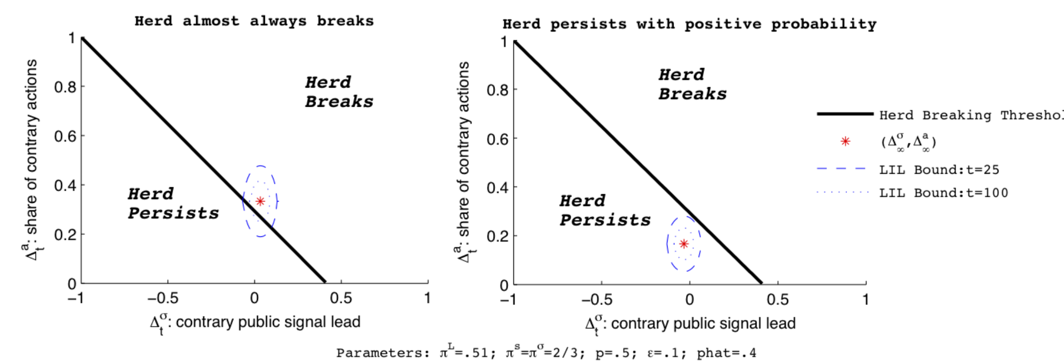

Figure 1.4: When Does a Herd Persist?

Chapter 2

Stochastic Games in Continuous Time: Persistent Actions in Long-Run Relationships

2.1 Introduction

A rich and growing literature on repeated games and reputation has studied how the shadow of the future affects the present. Yet, in many instances, a long-run player is influenced not only by considerations of its future but also by decisions it has made in the past. For example, a firm's ability to make high quality products is a function of not only its effort today but also its past investments in developing technology and training its workforce. A government's ability to offer efficient and effective public services to its citizens depends on its past investments in improving its public services. A university cannot educate students through instantaneous effort alone, but needs to have made past costly investments in hiring faculty and building research infrastructure. In all of these settings, and many others, a long-run player is directly influenced by choices it has made in the past: past actions influence key characteristics of the long-run player's environment, such as the quality of a firm's product or the level of a policy instrument; in turn, these characteristics play a central role in determining current and future profitability. This paper studies a new class of stochastic games in which the actions of a long-run player have a persistent effect on payoffs, and studies how its incentives are shaped by its past and future.

In analyzing this class of stochastic games, the paper develops a new understanding of reputational dynamics. Since Kreps et al. (1982), the canonical framework has modeled a long-run player's reputation as the belief that others have that the firm is a behavioral type that takes a fixed action in each period. This framework has been very influential and led to a number of insights across the gamut of economics. Nevertheless, it is unclear across many settings that reputational incentives are driven exclusively by the possibility that players may be non-strategic and are absent when there is common knowledge that the long-run player is rationally motivated by standard incentives. In contrast, my environment returns to a different notion of reputation as an asset (Klein and Leffler, 1981) in which a firm's reputation is shaped by its present and past actions. Persistent actions not only capture an intuitive notion of reputation as a type of capital, but also connect reputation to the aspects of a firm's choices that are empirically identifiable. This environment provides insights on important questions about the dynamics of reputation formation, including: when does a firm build its reputation and when does it allow it to decay; when do reputation effects persist in the long-run, and when are they temporary; how does behavior relate to underlying parameters such as the cost and depreciation rate of investment or the volatility of quality.

I study a continuous-time model with persistent actions and imperfect monitoring between a single long-run player and a continuum of small anonymous players. At each instant, each player chooses an action, which is not observed by others; instead, the long-run player's action generates a noisy *public signal*, and the long-run player observes the aggregate behavior of the short-run players. The stage game varies across time through its dependence on a state variable, whose evolution depends on the long run player's action through the public signal. This state variable determines the payoff structure of the associated stage game, and captures the current value of past investments by the long-run player.

When the long-run player chooses an action, it considers both the impact that this action has on its current payoff and its continuation value via the evolution of the state variable. For example, a firm may bear the cost of investment today, while reaping the rewards through higher sales and prices tomorrow. This link between current play and future outcomes creates intertemporal incentives for the long-run player. From Faingold and Sannikov (2011) and Fudenberg and Levine (2007), we know that in the absence of this state variable, intertemporal incentives fail: the long-run player cannot attain payoffs beyond those of its best stage game equilibrium. I am interested in determining when action persistence leads the firm to choose an action apart from that which maximizes its instantaneous payoff, and thus investigate whether persistent actions can be used to provide non-trivial intertemporal incentives in settings where those from standard repeated games fail.

The key contributions of this paper are along three dimensions. The theoretical contribution is to establish general conditions for the existence of Markovian equilibria, and conditions for the uniqueness of a Markovian equilibrium in the class of all Perfect Public Equilibria. An explicit characterization of the form of equilibrium payoffs, continuation values and actions, for any discount rate, yields insights into the relationship between the structure of persistence and the decisions of the long-run player. An application of the existence and uniqueness results to a stochastic game without action persistence shows that the long-run player acts myopically in the unique Perfect Public Equilibria of this setting. Lastly, I use these results to describe several interesting properties relating equilibrium payoffs of the stochastic game to the structure of the underlying stage game.

The conceptual contribution of this paper is to illustrate that action persistence creates a channel for effective intertemporal incentive provision in a setting where this is not possible in the absence of such persistence. A stochastic game has two potential channels through which intertemporal incentives can be used to guide behavior. First, the stage game varies across time in response to players' actions, and thus a long run player's actions directly impact payoffs in future periods as well as the current period. Second, as in repeated games, players can be rewarded or punished based on the public signal: actions today can affect, in equilibrium, how others behave in the future. In a Markovian equilibrium, intertemporal incentives can only flow through this first channel, as the public signal is ignored. When the unique Perfect Public Equilibria is Markovian, it precludes the existence of any equilibria that use the public signal to generate intertemporal incentives via punishments and rewards. As such, the ability to generate effective intertemporal incentives in a stochastic game of imperfect monitoring stems entirely from the impact that action persistence has on future payoffs.

Lastly, the results of this paper have practical implications for equilibrium analysis in a wide range of applied settings known to exhibit persistence and rigidities, ranging from industrial organization to political economy to macroeconomics. Markovian equilibria are a popular concept in applied work. Advantages of Markovian equilibria include their simplicity and their dependence on payoff relevant variables to specify incentives. Establishing that non-Markovian equilibria do not exist offers a strong justification for focusing on this more tractable class of equilibria.

Additionally, this paper derives a tractable expression to construct Markovian equilibria, which can be used to formulate empirically testable predictions about equilibrium behavior. Equilibrium continuation values are specified by the solution to a *nonstochastic* differential equation defined over the state space, while the long-run player's action is determined by the sensitivity of its future payoffs to changes in the state variable (the first derivative of this solution). This result provides a tool that can be utilized for equilibrium analysis in applications. Once functional forms are specified for the underlying game, it is straightforward to derive the relevant differential equation, calibrate it with realistic parameters, and use numerical methods to estimate its solution. This solution is used to explicitly calculate equilibrium payoffs and actions, as a function of the state variable. Note these numerical methods are used for estimation in an equilibrium that has been characterized analytically, and not to simulate an approximate equilibrium; an important distinction.

It may be helpful to describe the contributions of this paper and its relationship to the existing literature using one application that can be studied with the tools developed in this paper: the canonical product choice game. Consider a long-run firm interacting with a sequence of short-run consumers. The firm has a dominant strategy to invest low effort, but would have greater payoffs if it could somehow commit to high quality (its "Stackelberg payoff"). Repeated interaction in discrete time with imperfect monitoring generates a folk theorem Fudenberg and Levine (1994), but the striking implication from Faingold and Sannikov (2011) and Fudenberg and Levine (2007) is that such intertemporal incentives are absent in continuous-time games. Since Fudenberg and Levine (1992), we know that if the firm could build a reputation for being a *commitment type* that produces only high quality products, a patient normal firm can approach these payoffs in every equilibrium. Faingold and Sannikov (2011) shows that this logic remains in continuous-time games, but that as in discrete-time, these reputation effects are temporary: eventually, consumers learn the firm's type, and reputation effects disappear in the long-run (Cripps et al., 2007).[1]. Departing from standard repeated and reputational games, I consider a simple and realistic modification in which the firm's current product quality is a noisy function of past investment. Recent investment has a larger impact on current quality than investment further in the past, which is captured by a parameter θ that can be viewed as the rate at which past investment decays. Product quality, X_t, is modeled as a stochastic process:

[1] Mailath and Samuelson (2001) show that reputational incentives can also come from a firm's desire to *separate* itself from an incompetent type. Yet, these reputation effects are also temporary unless the type of the firm is replaced over time.

$$X_t = X_0 e^{-\theta t} + \theta \int_0^t e^{-\theta(t-s)} \left(a_s ds + dZ_s\right)$$

given some initial value X_0, investment path $\left(a_s\right)_{s\leq t}$ and Brownian motion $\left(Z_s\right)_{s\leq t}$. The evolution of product quality can be derived using this stochastic process, and takes the form:

$$dX_t = \theta \left(a_t - X_t\right) dt + \sigma dZ_t$$

When investment exceeds the current product quality, the firm is in a reputation building phase, and the product quality drifts upward (perturbed by a Brownian motion). When product quality is high and the firm chooses a lower investment level, it enters a period of reputational decay characterized by declining product quality. In this model, I show that there is a unique Perfect Public Equilibrium, which is Markovian in product quality. The firm's reputation for quality will follow a cyclical pattern, characterized by phases of reputation building and decay. Importantly, this cyclical pattern does not dissipate with time and reputation effects are permanent; this contrasts with the temporary reputation effects observed in behavioral types models. The product choice game is one of the many settings that can utilize the tools and techniques of this paper to shed light on the relationship between persistent actions and equilibrium behavior.

This paper makes an important modeling choice by employing a continuous-time framework. Recent work has shown that the continuous time framework often allows for an explicit characterization of equilibrium payoffs, for any discount rate (Sannikov, 2007); this contrasts with the folk-theorem results that typify the discrete time repeated games literature, and characterize the equilibrium payoff set as agents become arbitrarily patient Fudenberg and Levine (1994). Additionally, continuous time allows for an explicit characterization of equilibrium behavior; an important feature if one wishes to use the model to generate empirical predictions and relate the model to observable behavior. In discrete time, results in similar settings have generally been limited to identifying equilibrium payoffs.

Related Literature: This paper uses tools developed by Faingold and Sannikov (2011), and so I comment on how I generalize their insights. Their setting can be modeled as a stochastic game in which the state variable is the belief that the firm is a commitment type, and the transition function follows Bayes rule. Faingold and Sannikov (2011) characterize the unique Markov equilibrium of this incomplete information game using an ordinary differential equation. This paper extends these tools to characterize conditions for the existence and uniqueness of Markov equilibria in a setting with an arbitrary transition function between states, which can have a stochastic component independent of the public signal, where the long run player's payoffs may also depend on the state variable and the state space may be unbounded or have endpoints that are not absorbing states.

My work is also conceptually related to Board and Meyer-ter-Vehn (2011). They model a setting in which product quality takes on high or low values and is replaced via a Poisson arrival process; when a replacement occurs, the firm's current effort determines the new quality value. Consumers learn about product quality through noisy signals, and reputation is defined as the consumers' belief that the current product quality is high. Realized product quality in their setting is therefore discontinuous (jumping between low and high), and this discontinuity plays a key role in determining intertemporal incentives. In the product choice application of my setting, the quality of a firm's product is a smooth function of past investments and its investment today, and thus, the analysis is very different.

The role of persistence in intertemporal incentives can also be contrasted with our understanding of other continuous-time repeated games. Sannikov and Skrzypacz (2010) show that burning value through punishments that affect *all players* is not effective for incentives in settings with imperfect monitoring and Brownian signals, and that in these cases, it is more effective to punish by transferring value from some players to others. But in many settings, including those between long-run and myopic players, it would be impossible to avoid burning value and so intertemporal incentives collapse.[2]. Fudenberg and Levine (2007) examine a product choice game between a long-run and short-run player and demonstrate that it is not possible to earn equilibrium payoffs above the payoffs corresponding to repeated play of the static Nash equilibrium when the volatility of the Brownian component is independent of the long-run player's action. Thus, the intertemporal incentives that persistent actions induce could not emerge with standard continuous-time repeated games.

The organization of this paper proceeds as follows. Section 2.2 explores two simple examples to illustrate the main results of the model. Section 2.3 sets up the model. Section 2.4.4 analyzes equilibrium behavior and payoffs, while the final section concludes. All proofs are in the Appendix.

2.2 Examples

2.2.1 Persistent Investment as a Source of Reputation

Suppose a single long-run firm seeks to provide a continuum of small, anonymous consumers with a service. At each instant t, the firm chooses an unobservable investment level $a_t \in [0, \overline{a}]$. Consumers observe a noisy public signal of the firm's investment each instant, which can be represented as a stochastic process with a drift term that depends on the firm's action and a volatility term that depends on Brownian noise

$$dY_t = \theta a_t dt + \sigma dZ_t.$$

Investment is costly for the firm, but increases the likelihood of producing a high quality product. The stock quality of a product at time t, represented as X_t, captures the link between past investment levels and current product quality. This stock evolves

[2] Sannikov and Skrzypacz (2007) show how this issue also arises in games between multiple long-run players in which deviations between individual players are indistinguishable.

according to a mean-reverting stochastic process where the change in stock quality at time t is

$$\begin{aligned} dX_t &= \theta dY_t - \theta X_t dt \\ &= \theta \left(a_t - X_t \right) dt + \sigma dZ_t. \end{aligned}$$

Stock quality is publicly observed. The expected change in quality is increasing when investment exceeds the current quality level, and decreasing when investment is below the current quality level. The parameter θ captures the persistence of investment: recent investment has a larger impact on current quality than investment further in the past. Thus, θ embodies the rate at which past investment decays: as it increases, more recent investments play a larger role in determining current product quality relative to investments further in the past. This stochastic process, known as the Ornstein-Uhlenbeck process, has a closed form that gives an insightful illustration of how past investments of the firm determine the current product quality. Given a history of investment choices $(a_s)_{s \leq t}$, the current value of product quality is

$$X_t = X_0 e^{-\theta t} + \theta \int_0^t e^{-\theta(t-s)} a_s ds + \sigma \int_0^t e^{-\theta(t-s)} dZ_s,$$

given some initial value of product quality X_0. As shown in this expression, the impact of past investments decays at a rate proportional to the persistence parameter θ and the time that has elapsed since the investment was made.

Consumers simultaneously choose a purchase level $b_t^i \in [0, 10]$. The aggregate action of consumers, $\bar{b}_t$, is publicly observable, while individual purchase decisions are not. The firm's payoffs are increasing in the aggregate level of purchases by consumers, and decreasing in the level of investment. Average payoffs are represented as

$$r \int_0^\infty e^{-rt}(\bar{b}_t - ca_t^2)dt$$

where r is the common discount rate and $c < 1$ captures the cost of investment.

Consumers' payoffs depend on the stock quality, the firm's current investment level and their individual purchase decisions. As is standard in games of imperfect monitoring, payoffs can only depend on the firm's unobserved action through the public signal. Consumers are anonymous and their purchase decisions have a negligible impact on the aggregate purchase level. In equilibrium, they choose a purchase level that myopically optimizes their expected flow payoffs of the current stage game, represented as

$$E\left[\min\left\{b^i, 10\right\}^{1/2}[(1-\lambda)dY + \lambda X] - b^i\right]$$

Marginal utility from an additional unit of product is decreasing in the current purchase level, with a saturation point at 10, and is increasing in current investment and stock quality. The parameter λ captures the importance of current investment relative to stock investment.

This product choice game can be viewed as a stochastic game with current product quality X as the state variable and the change in product quality dX as the transition function, which depends on the investment of the firm. I am interested in characterizing equilibrium payoffs and actions in a Markov perfect public equilibrium.

The firm is subject to binding moral hazard in that it would like to commit to a higher level of investment in order to entice consumers to choose a higher purchase level. However, in the absence of such a commitment device, the firm is tempted to deviate to lower investment. This example seeks to characterize when intertemporal incentives, particularly incentives created by the dependence of future feasible payoffs on current investment through persistent quality, can provide the firm with endogenous incentives to choose a positive level of investment. Note that in the absence of intertemporal incentives, the firm always chooses an investment level of $a = 0$.

In a Markov perfect equilibrium, the continuation value can be expressed as an ordinary differential equation that depends on the stock quality. Let $U(X)$ represent the continuation value of the firm when $X_t = X$. Then, given equilibrium action profile $\left(a, \bar{b}\right)$

$$U(X) = \bar{b} - ca^2 + \frac{1}{r}\left[\theta\left(a - X\right)U'(X) + \frac{1}{2}\sigma^2 U''(X)\right]$$

describes the relationship between U and its first and second derivatives. The continuation value can be expressed as the sum of the payoff that the firm earns today, $\bar{b} - ca^2$, and the expected change in the continuation value, weighted by the discount rate. The expected change in the continuation value has two components. First, the drift of quality determines whether quality is increasing or decreasing in expectation. Given that the firm's payoffs are increasing in quality ($U' > 0$), positive quality drift increases the expected change in the continuation value, while negative quality drift decreases this expected change. Second, the volatility of quality determines how the concavity of the continuation value relates to its expected change. If the value of quality is concave ($U'' < 0$), then volatility of quality hurts the firm. The firm is more sensitive to negative quality shocks than positive quality shocks, and has a higher continuation value at the expected quality relative to the expected continuation value of quality; in simple terms, the firm is "risk averse" in quality. Positive and negative shocks are equally likely with Brownian noise; thus, volatility has a net negative impact on the continuation value. If the value of quality is convex ($U'' > 0$), then volatility of quality helps the firm: the firm benefits more from positive quality shocks than it is hurt by negative quality shocks. The continuation value is graphed in Figure 2.1.

The firm faces a trade-off when choosing its investment level: the cost of investment is borne in the current period, but yields a benefit in future periods through higher expected purchase levels by consumers. The impact of investment on future payoffs is captured by the slope of the continuation value, $U'(X)$, which measures the sensitivity of the continuation value to changes in stock quality. In equilibrium, investment in chosen to equate the marginal cost of investment with its future expected benefit:

$$a(X_t) = \min\left\{\frac{\theta}{2cr}U'(X_t), \overline{a}\right\}.$$

Marginal cost is captured by $2c$,while the marginal future benefit depends on the ratio of persistence to the discount rate. When θ is high, current investment will have a larger immediate impact on future quality, and the firm is willing to choose higher investment. Likewise, when the firm becomes more patient, it cares more about the impact investment today continues to have in future periods, and is willing to choose higher investment. It is interesting to note the trade-off between persistence and the discount rate. When investment decays at the same rate as the firm discounts future payoffs, these two parameters cancel. Thus, only the ratio of persistence to the discount rate is relevant for determining investment; as such, doubling θ has the same impact as halving the discount rate. Investment also depends on the sensitivity of the continuation value to changes in quality; when the continuation value is more sensitive to changes (captured by a steeper slope), the firm chooses a higher level of investment. As θ approaches 0, stock quality is almost entirely determined by its initial level and the intertemporal link between investment and payoffs is very small.

The boundary conditions that characterize the solution to $U(X)$ dictate that the slope of the continuation value converges to 0 as the stock quality approaches positive and negative infinity. Thus, the firm has the strongest incentive to invest at intermediate quality levels - a "reputation building" phase. When quality is very high, the firm's continuation value is less sensitive to changes in quality and the firm has a weaker incentive to invest. In effect, the firm is "riding" its good reputation for quality. The incentive to invest is also weak when quality is very low, and a firm may wait out a very bad reputation shock before beginning to rebuild its reputation - "reputation recovery". For interior values of X, the slope of the continuation value is positive, and thus the intertemporal incentives created by persistent actions allows the firm to choose a positive level of investment level.

In equilibrium, consumers myopically optimize flow payoffs by choosing a purchase level such that the marginal utility of an additional unit of product is zero:

$$b^i(a(X), X) =$$

$$\begin{array}{ll} 0 & \text{if } (1-\lambda)a(X) + \lambda X \leq 0 \\ \frac{1}{4}\left[(1-\lambda)a(X) + \lambda X\right]^2 & \text{if } (1-\lambda)a(X) + \lambda X \in \left[0, 2\sqrt{10}\right] \\ 10 & \text{if } (1-\lambda)a(X) + \lambda X > 2\sqrt{10} \end{array}$$

I show that there is a unique Perfect Public Equilibrium, which is Markovian in X_t; as such, $a(X_t)$ and $b^i(X_t)$, are uniquely determined by X_t, and are also continuous.

Note that $(a(X_t), b^i(X_t))$ is uniquely specified by and continuous in (X_t). Figures 2.2 and 2.3 graph equilibrium actions for the firm and consumers, respectively.

In this model, reputation effects are present in the long-run. Product quality is cyclical, with periods of high quality characterized by lower investment and negative drift, and periods of intermediate quality, where the firm chooses high investment and builds up its product quality. Very negative shocks can lead to periods where the firm chooses low investment and waits for its product quality to recover. Figure 2.4 illustrates the cycles of product quality across time. This contrasts with models in which reputations come from behavioral types: as Cripps et al. (2007) and Faingold and Sannikov (2011) show, reputation effects are temporary insofar as consumers eventually learn the firm's type, and so asymptotically, a firm's incentives to build reputation disappear. Additionally, conditional on the firm being strategic, reputation in these types models has negative drift.

Lastly, I compare the firm's payoffs in the stochastic game with action persistence to the benchmark without action persistence. The static Nash payoff depends on the value of stock quality. Let

$$v(X) = \min\left\{10, \frac{1}{4}\lambda^2 \max\{0, X\}^2\right\}$$

represent the static Nash payoff of the firm when the stock quality is at level X. This payoff is increasing in the stock quality. In the absence of investment persistence (this corresponds to $\theta = 0$), the unique equilibrium of the stochastic game is to play the static Nash equilibrium each period, which yields an expected continuation value at time t of

$$V(X_t) = r\int_t^{\infty} e^{-rs}E_t\left[v(X_s)\right] ds$$

Note that this expected continuation value may be above or below the static Nash equilibrium payoff of the current stage game, $v(X_t)$, depending on whether X_t is increasing or decreasing in expectation.

The firm achieves higher equilibrium payoffs when its actions are persistent, i.e. $U(X_t) \geq V(X_t)$ for all X_t. There are two complementary channels by which action persistence enhances the firm's payoffs. First, the firm chooses an investment level that equates the marginal cost of investment today with the marginal future benefit. Thus, in order for the firm to be willing to choose a positive level of investment, the future benefit of doing so must exceed the future benefit of choosing zero investment and must also exceed the current cost of this level of investment. Second, the link with future payoffs allows the firm to commit to a positive level of investment in the current period, which increases the equilibrium purchase level of consumers in the current period.

2.2.2 Policy Targeting

Elected officials and governing bodies often play a role in formulating and implementing policy targets. For example, the Federal Reserve targets interest rates, a board of directors sets growth and return targets for its company, and the housing authority targets home ownership rates. Achieving such targets requires costly effort on behalf of officials, and moral hazard issues arise because the preferences of the officials are not aligned with the population they serve. This example explores when a governing body can be provided with incentives to undertake a costly action in order to implement a target policy when the current level of the policy depends on the history of actions undertaken by the governing body.

Consider a setting where constituents elect a governing body to implement a policy target. The current policy takes on value $X_t \in [0, 2]$, and a policy target of $X_t = 1$ is optimal for constituents. In the absence of intervention, the policy drifts towards its natural level d. Each instant, the governing body chooses an action $a_t \in [-1, 1]$, where a negative action decreases the policy variable and a positive action increases the policy variable, in expectation. The policy evolves over time according to the stochastic process

$$dX_t = X_t(2 - X_t)\left[a_t dt + \theta(d - X_t)dt + dZ_t\right]$$

Constituents also choose an action b_t^i each period, which represent their campaign contributions or support for the governing body. Constituents pledge higher support to the governing body when the policy is closer to their optimal target and when the governing body is exerting higher effort to achieve this target. I model the reduced form of the aggregate best response of constituents as

$$\bar{b}\left(a_t, X_t\right) = 1 + \lambda a_t^2 - (1 - X_t)^2$$

in which λ captures the value that constituents place on the governing body's effort to achieve the policy target.

The governing body has no direct preference over the policy target; its payoffs are increasing in the support it receives from the constituents, and decreasing in the effort level it exerts.

$$g(a, \bar{b}_t, X_t) = \bar{b}_t - ca_t^2$$

The unique Nash equilibrium of the static game is for the governing body to set $a = 0$ i.e. not intervene in reaching the desired policy, and for the constituents to support the governing body based on the difference between the desired and current policy level, $b = 1 - (1 - X)^2$. Given the current policy level is X, this yields a stage game Nash equilibrium payoff of

$$v(X) = 1 - (1 - X_t)^2$$

for the governing body. This payoff is concave in the state variable, and therefore the highest PPE payoff in the stochastic game occurs at the value of the state variable that maximizes the stage game Nash equilibrium payoff, $X = 1$, which yields a stage game payoff of $v(1) = 1$. The highest PPE payoff is strictly less than the highest static game Nash equilibrium payoff. Figure 4 plots the PPE payoff of the governing body, as a function of the current policy level. This payoff is increasing in the policy level for levels below the optimal target, and decreasing in the policy level for levels above the optimal target.

The characterization of a unique Markovian equilibrium can be used to determine the equilibrium effort level of the governing body. Let $U(X)$ represent the continuation value as a function of the policy level in such an equilibrium, which is plotted in figure 2.5

The optimal effort choice of the governing body depends on the slope of the continuation value, the sensitivity of the change in the policy level to the effort level, and the cost of effort.

$$a_t(X) = \frac{X_t(2 - X_t)}{2rc} U'(X_t)$$

When the current policy level is very far from its optimal target, the effort of the governing body has a smaller impact on the policy level, and the governing body has a lower incentive to undertake costly effort. When the policy level is close to the optimal target, the continuation value approaches its maximum, and the slope of the continuation value approaches zero. Thus, the governing body also has a lower incentive to undertake costly effort when the policy is close to its target. Figure 2.6 plots the equilibrium effort choice of the governing body as a function of the policy level. As illustrated in the figure, the governing body exerts the highest effort when the policy variable is an intermediate distance from the optimal target. Figure 2.7 shows the equilibrium constituent support, which is highest when the policy level is closest to its optimal target.

2.3 Model

I study a stochastic game of imperfect monitoring between a single long run player and a continuum of small, anonymous short-run players. I refer to the long run player as the agency and the small, anonymous players $I = [0, 1]$ as members of the collective, with each individual indexed by i. Time $t \in [0, \infty)$ is continuous.

The Stage Game: At each instant t, the agency and collective members simultaneously choose actions a_t from A and b_t^i from B, respectively, where A and B are compact sets of a Euclidean space. Individual actions privately observed. Rather, the aggregate distribution of the collective's action, $\bar{b}_t \in \Delta B$ and a public signal of the agency's action, dY_t, are publicly observed. The public signal evolves according to the stochastic differential equation

$$dY_t = \mu_Y(a_t, \bar{b}_t)dt + \sigma_Y dZ_t^Y$$

where $\left(Z_t^Y\right)_{t\geq 0}$ is a Brownian motion, $\mu_Y : A \times B \rightarrow R$ is the drift and $\sigma_Y \in R$ is the volatility. Assume μ_Y is a Lipschitz continuous function. The drift term provides a signal of the agency's action and can also depend on the aggregate action of the collective, but is independent of the individual actions of the collective to preserve anonymity. The volatility is independent of players' actions.

The Stochastic Game: The stage game varies across time through its dependence on a state variable $(X_t)_{t\geq 0}$,, which takes on values in the state space $\Xi \subset R$ and evolves stochastically as a function of the current state and players' actions. The path of the state variable is publicly observable. As the state variable is not intended to provide any additional signal of players' actions, its evolution depends on actions solely through the available public information The transition of the state variable is governed by the stochastic differential equation:

$$dX_t = f_1(\bar{b}_t, X_t)\mu_Y(a_t, \bar{b}_t)dt + f_2(\bar{b}_t, X_t)dt + f_1(\bar{b}_t, X_t)\sigma_Y dZ_t^Y + \sigma_X(X_t)dZ_t^X$$

where $f_1 : B \times \Xi \rightarrow R$, $f_2 : B \times \Xi \rightarrow R$ and $\sigma_X^2 : \Xi \rightarrow R$ are Lipschitz continuous functions, and $\left(Z_t^X\right)_{t\geq 0}$ is a Brownian motion which is assumed to be orthogonal to $\left(Z_t^Y\right)_{t\geq 0}$. The drift of the state variable has two components: the first component, $f_1(\bar{b}_t, X_t)\mu_Y(a_t, \bar{b}_t)$, specifies how the agency's action influences the transition of the state, while the second component, $f_2(\bar{b}_t, X_t)$, is independent of the firm's action and allows the model to capture other channels that influence the transition of the state variable. The volatility of the state variable depends on the volatility of the public signal, $f_1(\bar{b}_t, X_t)\sigma_Y dZ_t^Y$, as well as a volatility term that is independent of the public signal, $\sigma_X(X_t)dZ_t^X$. Note that the same function multiplies the drift and volatility of the public signal; this ensures that no additional information about the agency's action is revealed by the evolution of the state variable. Let $\{F_t\}_{t\geq 0}$ represent the filtration generated by the public information, $(Y_t, X_t)_{t\geq 0}$.[3]

I assume that the volatility of the state variable is positive at all interior points of the state space. This ensures that the future path of the state variable is always stochastic. Brownian noise can take on any value in R, and as such, this assumption means that any future path of the state variable, $(X_s)_{s>t}$ can be reached from the current state $X_t \in \Xi$. This assumption is analogous to a strong form of irreducibility, since any state $X_s \in \Xi$ can be reached from the current state X_t at all times $s > t$.

Assumption 1. *For any compact proper subset $I \subset \Xi$, there exists a c such that*

$$\sigma_I = \inf_{\bar{b}\in B, X\in I} \left[f_1(\bar{b}, X)^2 \sigma_Y^2 + \sigma_X^2(X) \right] > c$$

Note that this assumption does not preclude the possibility that the state variable evolves independently of the public signal, which corresponds to $f_1 = 0$.

Define a state X as an *absorbing state* if the drift and volatility of the transition function are both zero. The following definition formalizes the conditions that characterize an absorbing state.

Definition 1. *$X \in \Xi$ is an absorbing state if there exists an action profile $\bar{b} \in B$ such that $f_1(\bar{b}, X) = 0$, $f_2(\bar{b}, X) = 0$ and $\sigma_X(X) = 0$.*

Remark 1. *The assumption that the volatility of the state variable is positive at all interior points of the state space precludes the existence of interior absorbing points. Given that Brownian motion is continuous, this is without loss of generality. To see why, suppose that $\Xi = [\underline{X}, \overline{X}]$ and there is an interior absorbing point X^*, and the*

[3] The state space may or may not be bounded. It is bounded if (i) there exists an upper bound $\overline{X}$ at which the volatility is zero and the drift is weakly negative, i.e. $f_1(\bar{b}, \overline{X}) = 0$; $\sigma_X(\overline{X}) = 0$, and $f_2(\bar{b}, \overline{X}) \leq 0$ for all $\bar{b} \in B$; and (ii) there exists a lower bound $\underline{X} < \overline{X}$ such that the volatility is zero and the drift is weakly positive, i.e. $f_1(\bar{b}, \underline{X}) = 0$; $\sigma_X(\underline{X}) = 0$ and $f_2(\bar{b}, \underline{X}) \geq 0$ for all $\bar{b} \in B$.

initial state is $X_0 < X^*$. *Then states* $X > X^*$ *are never reached under any strategy profile, and the game can be redefined on the state space* $\Xi = [\underline{X}, X^*]$.

Payoffs: The state variable determines the set of feasible payoffs in a given instant. Given an action profile $(a, \bar{b})$ and a state X, the agency receives an expected flow payoff of $g(a, \bar{b}, X)$. The agency seeks to maximize its expected normalized discounted payoff,

$$r \int_0^\infty e^{-rt} g(a_t, \bar{b}_t, X_t) dt$$

where r is the discount rate. Assume g is Lipschitz continuous and bounded for all $a \in A$, $\bar{b} \in \Delta B$ and $X \in \Xi$. The dependence of payoffs on the state variable creates a form of action persistence for the firm, since the state variable is a function of prior actions.

Collective members' have identical preferences, and each member seeks to maximize its expected flow payoff at time t,

$$h(a_t, b_t^i, \bar{b}_t, X_t)$$

which is a continuous function. Ex post payoffs can only depend on a_t through the public signal, dY_t, as is standard in games of imperfect monitoring.

Thus, in the stochastic game, at each instant t, given the current state X_t, players choose actions, and then nature stochastically determines payoffs, the public signal and next state as a function of the current state and action profile. The game defined here includes several subclasses of games, including a game where the state variable evolves independently of the agency's action ($f_1(\bar{b}, X) = 0$), the state variable evolves deterministically given the public signal ($\sigma_X(X) = 0$), or the agency's payoffs only depend on the state indirectly through the actions of the collective ($g(a, \bar{b}, X) = g(a, \bar{b})$).

Strategies: A public strategy for the agency is a stochastic process $(a_t)_{t \geq 0}$ with values $a_t \in A$ and progressively measurable with respect to $\{F_t\}_{t \geq 0}$. Likewise, a public strategy for a member of the collective is an action $b_t^i \in B$ progressively measurable with respect to $\{F_t\}_{t\geq 0}$.

2.3.1 Equilibrium Structure

Perfect Public Equilibria: I restrict attention to pure strategy perfect public equilibria (PPE). A public strategy profile is a PPE if after any public history and for all t, no player wants to deviate given the strategy profile of its opponents.

In any PPE, collective members choose b_t^i to myopically optimize expected flow payoffs each instant.[4] Let $\mathcal{B} : A \times \Delta B \times \Xi \rightrightarrows B$ represent the best response correspondence that maps an action profile and a state to the set of collective member actions that maximize payoffs in the current stage game, and $\overline{\mathcal{B}} : A \times \Xi \rightrightarrows \Delta B$ represent the aggregate best response function. In many applications, it will be sufficient to specify the aggregate best response function as a reduced form for the collective's behavior.

Define the agency's continuation value as the expected discounted payoff at time t, given the public information contained in $\{F_t\}_{t\geq 0}$ and strategy profile $S = (a_t, b_t^i)_{t\geq 0}$:

$$W_t(S) := E_t \left[r \int_t^\infty e^{-r(s-t)} g(a_s, \overline{b}_s, X_s) ds \right]$$

The agency's action at time t can impact its continuation value through two channels: (1) future equilibrium play and (2) the set of future feasible flow payoffs. It is well known that the public signal can be used to punish or reward the agency in future periods by allowing continuation play to depend on the realization of the public signal. A stochastic game adds a second link between current play and future payoffs: the agency's action affects the evolution of the state variable, which in turn determines

[4] The individual actions of a collective member, b_t^i, has a negligible impact on the aggregate action $\overline{b}_t$ (and therefore X_t) and is not observable by the agency. Therefore, the model could also allow for long-run small, anonymous players.

the set of future feasible stage payoffs. Each channel provides a potential source of intertemporal incentives.

This paper applies recursive techniques for continuous time games with imperfect monitoring to characterize the evolution of the continuation value and the agency's incentive constraint in a PPE. Fix an initial value for the state variable, X_0.

Lemma 3. *A public strategy profile $S = (a_t, b_t^i)_{t\geq 0}$ is a PPE with continuation values $(W_t)_{t\geq 0}$ if and only if for some $\{F_t\} - measurable$ process $(\beta_t)_{t\geq 0}$ in $\mathcal{L}$*

1. *$(W_t)_{t\geq 0}$ is a bounded process and satisfies:*

$$\begin{aligned} dW_t(S) &= r\left(W_t(S) - g(a_t, \overline{b}_t, X_t)\right) dt \\ &\quad + r\beta_{1t}\left[dY_t - \mu_Y(a_t, \overline{b}_t)dt\right] \\ &\quad + r\beta_{2t}\sigma_X(X_t)dZ_t^X \end{aligned}$$

   given $(\beta_t)_{t\geq 0}$

2. *Strategies $(a_t, b_t^i)_{t\geq 0}$ are sequentially rational given $(\beta_t)_{t\geq 0}$. For all t, (a_t, b_t^i) satisfy:*

$$\begin{aligned} a_t &\in \arg\max \; g(a', \overline{b}_t, X_t) + \beta_{1t}\mu_Y(a', \overline{b}_t) \\ b_t^i &\in \mathcal{B}\left(a_t, X_t\right) \end{aligned}$$

The continuation value of the agency is a stochastic process that is measurable with respect to public information, $\{F_t\}_{t\geq 0}$. Two components govern the motion of the continuation value, a drift term that captures the difference between the current continuation value and the current flow payoff. This is the expected change in the continuation value. A volatility term β_{1t} determines the sensitivity of the continuation value to the public signal: the agency's future payoffs are more sensitive to good or bad signal realizations when the volatility of the continuation value is larger. A second volatility term β_{2t} captures the sensitivity of the continuation value to the stochastic element of the state variable that is independent of the public signal.

The condition for sequential rationality depends on the process $(\beta_t)_{t\geq 0}$, which specifies how the continuation value changes with respect to the public information. Today's action impacts future payoffs through the drift of the public signal, $\mu_Y(a, \bar{b})$, and the sensitivity of the continuation value to the public signal, β_1, while it impacts current payoffs through the flow payoff of the agency, $g(a, \bar{b}, X)$. A strategy for the agency is sequentially rational if it maximizes the sum of flow payoffs today and the expected impact of today's action on future payoffs. This condition is analogous to the one-shot deviation principle in discrete time.

A key feature of this characterization is the linearity of the continuation value and incentive constraint with respect to the Brownian information. Brownian information can only be used linearly to provide effective incentives in continuous time (Sannikov and Skrzypacz, 2010). Therefore, the agency's incentive constraint takes a very tractable linear form, in which the process $(\beta_t)_{t\geq 0}$ captures all potential channels through which the agency's current action may impact future payoffs, including coordination of equilibrium play and the set of future feasible payoffs that depend on the state variable.

Remark 2. *The key aspect of this model that allows for this tractable characterization of the agency's incentive constraint is the assumption that the volatility of the state variable is always positive (except at the boundary of the state space), which ensures that any future path of states can be reached from the current state. This assumption, coupled with the linear incentive structure of Brownian information, ensures the condition for sequential rationality takes the form in Lemma 3. To see this, consider a deviation from a_t to $\widetilde{a}_t$ at time t. This deviation impacts future payoffs by inducing a different probability measure over the future path of the state variable, $(X_s)_{s>t}$, but doesn't affect the set of feasible sample paths. Given that all paths of the state variable are feasible under a_t and $\widetilde{a}_t$, the continuation value under both strategies is a* *non-degenerate expectation with respect to the future path of the state variable. Thus, the change in the continuation value when the agency deviates from* a_t *to* $\widetilde{a}_t$ *depends solely on the different measures* a_t *and* $\widetilde{a}_t$ *induce over future sample paths, and, given the requirement that Brownian information is used linearly, this change is linear with respect to the difference in the drift of the public signal,* $\mu_Y\left(\widetilde{a}_t, \overline{b}_t\right) - \mu_Y\left(a_t, \overline{b}_t\right)$.

Remark 3. *It is of interest to note that it is precisely this linear structure with respect to the Brownian information, coupled with the inability to transfer continuation payoffs between players, that precludes the effective provision of intertemporal incentives in a standard repeated game between a long-run and short-run player. The short-run player acts myopically, so it is not possible to tangentially transfer continuation values between players. Using Brownian information linearly, but non-tangentially, results in the continuation value escaping the boundary of the payoff set with positive probability, and Brownian information cannot be used effectively in a non-linear manner. This paper will illustrate that a stochastic game permits the provision of intertemporal incentives by introducing the possibility of linearly using Brownian information for some values of the state variable.*

The sequential rationality condition can be used to specify an auxiliary stage game parameterized by the state variable and the process linking current play to the continuation value. Let $S^*(X, \beta_1) = \left\{\left(a, \overline{b}\right)\right\}$ represent the correspondence of static Nash equilibrium action profiles in this auxiliary game, defined as:

Definition 2. *Define* $S^*(X, \beta_1) = \Xi \times R \rightrightarrows A \times \Delta B$ *as the correspondence that describes the Nash equilibrium of the static game parameterized by* $(X, \beta_1) \in \Xi \times R$*:*

$$S^*(X, \beta) = \left\{ \begin{array}{c} a \in \arg\max_{a'} g(a', \overline{b}, X) + \beta_1 \mu_Y(a', \overline{b}) \\ \overline{b} \in \overline{\mathcal{B}}(a, X) \end{array} \right\}$$

In any PPE strategy profile $(a_t, \overline{b}_t)_{t \geq 0}$ of the stochastic game, given some processes $(X_t)_{t>0}$ and $(\beta_{1t})_{t>0}$, the action profile at each instant must be a static Nash equilibrium of the auxiliary game i.e. $(a_t, \bar{b}_t) \in S^*(X_t, \beta_{1t})$ for all t. I assume that this auxiliary stage game has a unique static Nash equilibrium with an atomic distribution over small players' actions. While this assumption is somewhat restrictive, it still allows for a broad class of games, including those discussed in the previous examples.

Assumption 2. *Assume $S^*(X, \beta)$ is non-empty and single-valued for all $(X, \beta) \in \Xi \times R$, Lipschitz continuous on any subset of $\Xi \times R$, and the small players choose identical actions $b^i = \bar{b}$.*

Note that $S^*(X, 0)$ corresponds to the Nash equilibrium of the stage game in the current model when the state variable is equal to X.

Static Equilibria Payoffs: The feasible payoffs of the current stage game depend on the state variable, as do stage game Nash equilibrium payoffs. The presence of myopic players imposes restrictions on the payoffs that can be achieved by the long-run player, given that the myopic players must play a static best response.

Define $v : \Xi \to R$ as the payoff to the agency in the Nash equilibrium of the stage game, parameterized by the state variable, where $v(X) := g(S^*(X, 0), X)$. The assumption that the Nash equilibrium correspondence of the stage game is Lipschitz continuous, non-empty and single-valued guarantees $v(X)$ is a Lipschitz continuous function. When the state space is bounded, $\Xi = [\underline{X}, \overline{X}]$, $v(X)$ has a well-defined limit as it approaches the highest and lowest state. If the state space is unbounded, an additional assumption is necessary to guarantee that $v(X)$ has well-defined limits.

Assumption 3. *If the state space is unbounded, $\Xi = R$, then there exists a δ such that for $|X| > \delta$, $v(X)$ is monotonic in X.*

This assumption ensures that $v(X)$ doesn't oscillate as it approaches infinity, a technical assumption that is necessary for the equilibrium uniqueness result. Represent the highest and lowest stage Nash equilibrium payoffs across all states as:

$$\overline{v}^* = \sup_{X \in \Xi} v(X)$$
$$\underline{v}^* = \inf_{X \in \Xi} v(X)$$

These values are well-defined given that g is bounded.

This subsection illustrates the model and definitions introduced above.

Example 1. ***Pricing Quality:*** *Consider a setting where a firm invests in developing a product, and consumers choose the price they are willing to pay to purchase this product. Each instant, a firm chooses an investment level $a_t \in [0, 1]$ and consumers choose a price $b^i \in \left[0, \overline{B}\right]$. A public signal provides information about the firm's instantaneous investment through the process*

$$dY_t = a_t dt + dZ_t^Y$$

The state variable is product quality, which is a function of past investments and takes on values in the bounded support $\Xi = [0, \overline{X}]$. The change in product quality is governed by the process

$$dX_t = X_t(\overline{X} - X_t)\left(a_t dt + dZ_t^Y\right) - X_t dt$$

which is increasing in the firm's investment, and decreasing in the current product quality. Note this corresponds to $\mu_Y = a_t$, $\sigma_Y = 1$, $f_1 = X_t(\overline{X} - X_t)$, $f_2 = -X_t$ and $\sigma_X^2 = 0$. At the upper bound of product quality, investment no longer impacts product quality and the process has negative drift. At the lower bound, investment also no longer impacts product quality, and the process has zero drift. As such, $X = 0$ is an absorbing state but $X = \overline{X}$ is not.

The firm earns the price the consumers are willing to pay for the product, and *pays a cost of* $c(a)$ *for an investment level of* a*, with* $c(0) = 0, c' > 0$ *and* $c'' > 0$*. Its payoff function is*

$$g(a, \overline{b}, X) = \overline{b} - c(a)$$

which is independent of product quality.

Consumers receive an instantaneous value of $X + a$ *from purchasing a unit of the product. Their flow payoff is the difference between the purchase price and the value of the product,*

$$h(a, b^i, \overline{b}, X) = -\left(b^i - X - a\right)^2$$

In equilibrium, consumers myopically optimize flow payoffs, and thus pay a price equal to their expected utility from purchasing the product, which is increasing in the stock quality and investment of the firm.[5] *The aggregate consumer best response function takes the form*

$$\overline{b}\,(a, X) = X + a$$

In the static game, given a product quality of X*, the unique Nash equilibrium is for the firm to choose an investment level of* $a^* = 0$ *and the consumers to pay a price of* $\overline{b}^*(0, X) = X$ *for the good. This yields a stage game Nash equilibrium payoff of* $v(X) = X$ *for the firm, a maximum stage NE payoff of* $\overline{v}^* = \overline{X}$ *at* $X = \overline{X}$ *and a minimum stage NE payoff of* $\underline{v}^* = 0$ *at* $X = 0$*.*

In the stochastic game, the firm also considers the impact that current investment has on future product quality. Using the condition for sequential rationality specified in Lemma 3, the firm chooses an investment level to maximize

$$a \in \arg\max_{a'} X + a - c(a') + \beta a'$$

[5] While it may seem unusual that the consumer receives negative utility when they pay a price lower than the value of the product, this setting can be interpreted as the reduced for a monopolistic market in which the firm captures all of the surplus from the product quality. Such a setting would yield the same aggregate best response function, which is the only relevant aspect of consumer behavior for equilibrium analysis.

which yields an equilibrium action

$$a^*(X,\beta) = (c')^{-1}(\beta)$$

Equilibrium investment is strictly positive in the stochastic game when $\beta > 0$. Thus, persistent investment allows the firm to overcome the binding moral hazard present in the static game and earn a higher price for its product.

Note that

$$S^*(X,\beta) = \left((c')^{-1}(\beta), X + (c')^{-1}(\beta)\right)$$

which is non-empty, single-valued, unique and Lipschitz continuous for each (X,β).

2.4 Equilibrium Analysis

This section presents the main results of the paper, and proceeds as follows. First, I construct a Markovian equilibrium in the state variable, which simultaneously establishes the existence of at least one Markovian equilibria and characterizes equilibrium behavior and payoffs in such an equilibrium. Next, I establish conditions for a Markovian equilibrium to be the unique equilibrium in the class of all Perfect Public Equilibria. Following is a brief discussion on the role action persistence plays in using Brownian information to create effective intertemporal incentives. An application of the existence and uniqueness results to a stochastic game *without* action persistence shows that the agency acts myopically in the unique Perfect Public Equilibria of this setting. Finally, I use the equilibrium characterization to describe several interesting properties relating the agency's equilibrium payoffs to the structure of the underlying stage game.

2.4.1 Existence of Markov Perfect Equilibria

The first main result of the paper establishes the existence of a Markovian equilibrium in the state variable. The existence proof is constructive, and as such, characterizes the explicit form of equilibrium continuation values and actions in Markovian equilibria. This result applies to a general setting in which:

- The state space may be bounded or unbounded.
- The transition function governing the law of motion of the state variable is stochastic and depends on the agency's action through a public signal, as well as the aggregate action of the collective and the current value of the state.
- There may or may not be absorbing states at the endpoints of the state space.

Theorem 5. *Suppose Assumptions 1 and 2 hold. Then given an initial state X_0 and action profile $(a, \bar{b}) = S^*(X, U'(X)f_1(\bar{b}, X))$, any bounded solution $U(X)$ to the second order differential equation:*

$$U''(X) = \frac{2r\left[U(X) - g(a, \bar{b}, X)\right]}{f_1(\bar{b}, X)^2\sigma_Y^2 + \sigma_X^2(X)} - \frac{2\left[f_1(\bar{b}, X)\mu_Y(a, \bar{b}) + f_2(\bar{b}, X)\right]}{f_1(\bar{b}, X)^2\sigma_Y^2 + \sigma_X^2(X)}U'(X)$$

referred to as the optimality equation, characterizes a Markovian equilibrium in the state variable $(X_t)_{t\geq 0}$ with

1. *Equilibrium payoffs $U(X_0)$*
2. *Continuation values $(W_t)_{t\geq 0} = (U(X_t))_{t\geq 0}$*
3. *Equilibrium actions $(a_t, \bar{b}_t)_{t\geq 0}$ uniquely specified by*

$$S^*(X, U'(X)f_1(\bar{b}, X)) = \left\{ \begin{array}{c} a = \arg\max_{a'} rg(a', \bar{b}, X) + U'(X)f_1(\bar{b}, X)\mu(a', \bar{b}) \\ \bar{b} = \overline{\mathcal{B}}(a, X) \end{array} \right\}$$

The optimality equation has at least one solution $U \in C^2(R)$ that lies in the range of feasible payoffs for the agency $U(X) \in [\underline{g}, \overline{g}]$ for all states $X \in \Xi$. Thus, there exists at least one Markovian equilibrium.

Theorem 5 shows that the stochastic game has at least one Markovian equilibrium. Continuation values in this equilibrium are represented by a second order ordinary differential equation. Rearranging the optimality equation as:

$$\begin{aligned} U(X) &= g(a,\bar{b},X) + \frac{1}{r}\left[f_1(\bar{b},X)\mu_Y(a,\bar{b}) + f_2(\bar{b},X)\right]U'(X) \\ &\quad + \frac{1}{2r}U''(X)\left[f_1(\bar{b},X)^2\sigma_Y^2 + \sigma_X^2(X)\right] \end{aligned}$$

lends insight into the relationship between the continuation value and the transition of state variable. The continuation value is equal to the sum of the flow payoff today, $g(a, \bar{b}, X)$, and the expected change in the continuation value, weighted by the discount rate. The second term captures how the continuation value changes with respect to the drift of the state variable. For example, if the state variable has positive drift ($f_1(\bar{b}, X)\mu_Y(a, \bar{b}) + f_2(\bar{b}, X) > 0$), and the continuation value is increasing in the state variable ($U' > 0$), then this increases the expected change in the continuation value. The third term captures how the continuation value changes with respect to the volatility of the state variable. If U is concave ($U'' < 0$), it is more sensitive to negative shocks than positive shocks. Positive and negative shocks are equally likely, and therefore, the continuation value is decreasing in the volatility of the state variable. If U is linear ($U'' = 0$), then the continuation value is equally sensitive to positive and negative shocks, and the volatility of the state variable does not impact the continuation value.

Now consider a value of the state variable that yields a local maximum $U(X^*)$ (note this implies $U' = 0$). Since the continuation value is at a local maximum, it must be decreasing as X moves away from X^* in either direction. This is captured by the fact that $U''(X) < 0$. Larger volatility of the state variable or a more concave function lead to a larger expected decrease in the continuation value.

I now outline the intuition behind the proof of Theorem 5. The first step in proving this existence is to show that if a Markovian equilibrium exists, then continuation values must be characterized by the solution to the optimality equation. In a Markovian equilibrium, continuation values take the form $W_t = U(X_t)$ for some function U. Using Ito's formula to differentiate $U(X_t)$ with respect to X_t yields an expression for the law of motion of the continuation value in any Markovian equilibrium $dW_t = dU(X_t)$, as a function of the law of motion for the state variable:

$$\begin{aligned} dU(X_t) &= U'(X_t)\left[f_1(\bar{b}_t, X_t)\mu_Y(a_t, \bar{b}_t) + f_2(\bar{b}_t, X_t)\right] dt \\ &\quad + \frac{1}{2}U''(X_t)\left[f_1(\bar{b}_t, X_t)^2\sigma_Y^2 + \sigma_X^2(X_t)\right] dt \\ &\quad + U'(X_t)\left[f_1(\bar{b}_t, X_t)\sigma_Y dZ_t^Y + \sigma_X(X_t) dZ_t^X\right] \end{aligned}$$

In order for this to be an equilibrium, continuation values must also follow the law of motion specified in Lemma 3, with drift

$$r\left(U(X_t) - g(a_t, \bar{b}_t, X_t)\right) dt$$

Matching the drifts of these two laws of motion yields the optimality equation, a second order ordinary differential equation that specifies continuation payoffs as a function of the state variable.

The next step in the existence proof is to show that this ODE has at least one solution that lies in the range of feasible payoffs for the agency. The technical condition to guarantee the existence of a solution is that the second derivative of U is bounded with respect the first derivative of U on any bounded interval of the state space. The denominator of the optimality equation depends on the volatility of the state variable. Thus, the assumption that the volatility of the state variable is positive on any open interval of the state space (Assumption 1) is crucial to ensure this condition is satisfied. The numerator of the optimality equation depends on the drift of the state variable, and the agency's flow payoff. Lipschitz continuity of these functions ensures that they are bounded on any bounded interval of the state space. These conditions are sufficient to guarantee the optimality equation has at least one bounded solution that lies in the range of feasible payoffs for the agency.

The final step of the existence proof is to construct a Markovian equilibrium that satisfies the conditions of a PPE established in Lemma 3. The incentive constraint for the agency is constructed by matching the volatility of the laws of motion for the continuation value established in Lemma 3 with the volatility of the law of motion for the continuation value as a function of the state variable, $dU(X_t)$. Lemma 3 established that the volatility of the continuation value must be

$$r\beta_{1t}\sigma_Y dZ_t^Y + r\beta_{2t}\sigma_X(X_t)dZ_t^X$$

in any PPE. Thus, in a Markovian equilibrium

$$\begin{aligned} r\beta_{1t}\sigma_Y &= U'(X_t)f_1(\bar{b}_t, X_t)\sigma_Y \\ r\beta_{2t}\sigma_X(X_t) &= U'(X_t)\sigma_X(X_t) \end{aligned}$$

This characterizes the process $(\beta_t)_{t\geq 0}$ governing incentives, and as such, the incentive constraint for the agency. This incentive constraint takes an intuitive form. The impact of the current action on future payoffs is captured by the impact the current action has on the state variable, $f_1(\bar{b}, X)\mu(a', \bar{b})$, as well as the slope of the continuation value, $U'(X_t)$,which captures how the continuation value changes with respect to the state variable.

Theorem 5 also establishes that *each* solution U to the optimality equation characterizes a single Markovian equilibrium. This is a direct consequence of the assumption that the Nash equilibrium correspondence of the auxiliary stage game $S^*(X, \beta)$ is single-valued, Assumption 2, which guarantees that U uniquely determines equilibrium actions. Note that if there are multiple solutions to the optimality equation, then each solution characterizes a single Markovian equilibrium. The formal proof of Theorem 5 is presented in the Appendix.

Markovian equilibria have an intuitive appeal in stochastic games. Advantages of Markovian equilibria include their simplicity and their dependence on payoff relevant variables to specify incentives. Theorem 5 yields a tractable expression that can be used to construct equilibrium behavior and payoffs in a Markovian equilibrium. The continuation value of the agency is specified by the solution to a second order differential equation defined over the state space. The agency's incentives are governed by the slope of this solution, which determines how the continuation value changes with the state variable. As such, this result provides a tool to analyze equilibrium behavior in a broad range of applied settings. Once functional forms are specified for the agency's payoffs and the transition function of the state variable, it is straightforward to use Theorem 5 to characterize the optimality equation and incentive constraint for the agency, as a function of the state variable. This constructs a Markovian equilibrium. Numerical methods for ordinary differential equations can then be used to estimate a solution to the optimality equation and explicitly calculate equilibrium payoffs and actions. These calculations yield empirically testable predictions about equilibrium behavior. Note that numerical methods are used for estimation in an equilibrium that has been characterized analytically, and not to simulate an approximate equilibrium. This is an important distinction.

Example to illustrate Theorem 5

The following example illustrates how to use Theorem 5 to construct equilibrium behavior.

Example 2. *Consider the persistent investment model presented in Section 2.2.1. The state variable evolves according to:*

$$dX_t = \theta(a_t - X_t)\,dt + \sigma dZ_t$$

which corresponds to $f_1 = 1$*,* $\mu_Y = \theta a$*,* $f_2 = -\theta X, \sigma_Y = \sigma$ *and* $\sigma_X = 0$*, and the firm's flow payoff is:*

$$g(a, \overline{b}, X) = \overline{b} - ca^2$$

Using Theorem 5 to characterize the optimality equation yields

$$U''(X) = \frac{2r}{\sigma^2}\left(U(X) - \overline{b}^* + c\,(a^*)^2\right) - \frac{2\theta}{\sigma^2}\,(a^* - X)\,U'(X)$$

The sequential rationality condition for the firm is

$$a = \arg\max_{a'} \overline{b} - ca^2 + U'(X)\theta a'$$

In equilibrium, the firm chooses action

$$a^*(X, U'(X)) = \min\left\{\frac{\sigma\theta}{2cr}U'(X_t), \overline{a}\right\}$$

This constructs equilibrium behavior and payoffs as a function of the current product quality X *and the solution to the optimality equation,* U*. Numerical methods can now be used to estimate a solution* U *to the optimality equation. Calibrating the model with a set of parameters will then fully determine equilibrium actions and payoffs as a function of the current product quality. As discussed in 2.2.1, the empirical predictions* *of this application are:*

1. *The firm's continuation value is increasing in the current product quality*
2. *The firm's incentives to invest in quality are highest when the current product quality is at an intermediate level. As such, the firm goes through phases of "reputation building", during which the firm chooses high investment levels and product quality increases in expectation, and "reputation riding", during which the firm chooses low investment levels and reaps the benefits of having a high product quality.*

The firm's equilibrium payoff captures the future value of owning a product of a given quality level, and as such, can be interpreted as the asset value of the firm.

2.4.2 Uniqueness of Markovian Equilibrium

The second main result of the paper establishes conditions under which there is a unique Markovian equilibrium, which is also the *unique equilibrium* in the class of *all Perfect Public Equilibria.* The first step of this result is to establish when the optimality equation has a unique bounded solution. Recall that each solution to the optimality equation characterized in Theorem 5 characterizes a single Markovian equilibrium. Thus, when the optimality equation has a unique solution, there is a unique Markovian equilibrium. The second step of the result is to prove that there are no non-Markovian PPE, and as such, this unique Markovian equilibrium is the unique PPE.

The optimality equation will have a unique solution when its solution satisfies certain boundary conditions as the state variable approaches its upper and lower bound (in the case of an unbounded state space, as the state variable converges to positive or negative infinity). The boundary conditions for the optimality equation depend on the rate at which the drift and volatility of the state variable converge as the state variable approaches its upper and lower bound. As such, the key condition that ensures a unique solution to the optimality equation is an assumption on the limiting behavior of the drift and volatility of the state variable.

Assumption 4. *1. If the state space is bounded, $\Xi = [\underline{X}, \overline{X}]$, then as X approaches its upper and lower bound $\{\underline{X}, \overline{X}\}$, the functions governing the transition of the state variable satisfy the following limiting behavior:*

(a) The drift of the state variable converges to zero at a linear rate, or faster: $f_2(\bar{b}, X)$ and $f_1(\bar{b}, X)$ are $O(X^* - X)$ as $X \to X^* \in \{\underline{X}, \overline{X}\}$.

(b) The volatility of the state variable converges to zero at a linear rate, or faster: $1/f_1(\bar{b}, X)\sigma_Y(\bar{b}) + \sigma_X(X)$ is $O(1/(X^* - X))$ as $X \to X^* \in \{\underline{X}, \overline{X}\}$.

2. If the state space is unbounded, $\Xi = R$, then as X approaches positive and negative infinity, the functions governing the transition of the state variable satisfy the following limiting behavior:

(a) The drift of the state variable grows linearly, or slower: $f_2(\bar{b}, X)$ and $f_1(\bar{b}, X)$ are $O(X)$ as $X \to \{-\infty, \infty\}$.

(b) The volatility of the state variable is bounded: $f_1(\bar{b}, X)\sigma_Y(\bar{b}) + \sigma_X(X)$ is $O(1)$ as $X \to \{-\infty, \infty\}$.

When the support is bounded, this assumption requires that the upper and lower bounds of the state space are absorbing points. The drift and volatility of the state variable must converge to zero at a linear rate, or faster, as the state variable approaches its boundary. When the support is unbounded, these assumptions require that the drift of the state variable grows at a linear rate, or slower, as the magnitude of the state becomes arbitrarily large, and that the volatility of the state variable is uniformly bounded. The role this assumption plays in establishing equilibrium uniqueness is discussed following the presentation of the result.

Remark 4. *When the endpoints of the state space are absorbing points, whether the state variable actually converges to one of its absorbing points with positive probability will depend on the relationship between the drift and the volatility as the state variable approaches its boundary points. It is possible that the state variable converges to an absorbing point with probability zero.*

The following theorem establishes the uniqueness of a Markovian equilibrium in the class of all Perfect Public Equilibria.

Theorem 6. *Suppose Assumptions 1, 2, 3 and 4 hold. Then, for each initial value of the state variable $X_0 \in \Xi$, there exists a unique perfect public equilibrium, which is Markovian, with continuation values characterized by the unique bounded solution U of the optimality equation, yielding equilibrium payoff $U(X_0)$.*

1. *When the state space is bounded, $\Xi = [\underline{X}, \overline{X}]$, then the solution satisfies the following boundary conditions:*

$$\begin{aligned} \lim_{X \to \overline{X}} U(X) &= v(\overline{X}) \textit{ and } \lim_{X \to \underline{X}} U(X) = v\left(\underline{X}\right) \\ \lim_{X \to \overline{X}} (\overline{X} - X) U'(X) &= \lim_{X \to \underline{X}} (\underline{X} - X) U'(X) = 0 \end{aligned}$$

2. *When the state space is unbounded, $\Xi = \mathcal{R}$, then the solution satisfies the following boundary conditions:*

$$\begin{aligned} \lim_{X \to \infty} U(X) &= v_{\infty} \textit{ and } \lim_{X \to -\infty} U(X) = v_{-\infty} \\ \lim_{X \to \infty} X U'(X) &= \lim_{X \to -\infty} X U'(X) = 0 \end{aligned}$$

I briefly relate the boundary conditions characterized in Theorem 6 to equilibrium behavior and payoffs, and then outline the intuition behind the uniqueness result. These boundary conditions have several implications for equilibrium play. Recall the incentive constraint for the agency from Theorem 5. The link between the agency's action and future payoffs is proportional to the slope of the continuation value and the drift component of the state variable that depends on the public signal, $U'(X)f_1(\bar{b}, X)$. The assumption on the growth rate of $f_1(\bar{b}, X)$ ensures that $U'(X)f_1(\bar{b}, X)$ converges to zero at the boundary points (in the unbounded case, as the state variable approaches positive or negative infinity). When this is the case, the agency's incentive constraint is reduced to the myopic optimization of its instantaneous flow payoff at the boundary points. Thus, at the upper and lower bound of the state space (in the limit for an unbounded state space), the agency plays a static Nash equilibrium action. Additionally, continuation payoffs are equal to the Nash equilibrium payoff of the static game at the boundary points.

I next provide a sketch of the proof for the existence of a unique PPE, which is Markovian. The first step in proving this result is establishing that the optimality equation has a unique solution. This is done so in two parts: (i) showing that any solution to the optimality equation must satisfy the same boundary conditions, and (ii) showing that it is not possible for two different solutions to the optimality equation to satisfy the same boundary conditions.

I discuss the boundary conditions for an unbounded state space; the case of a bounded state space is analogous. Suppose U is a bounded solution to the optimality equation. The U, and its first and second derivative, must satisfy the following set of boundary conditions. U will have a well-defined limit at the boundary points when the static Nash equilibrium payoff function has a well-defined limit, which is guaranteed given the Lipschitz continuity of the Nash equilibrium correspondence and the agency's payoff function. The boundedness of U coupled with the assumption that the static Nash equilibrium payoff function is monotonic for large X ensures that the first derivative of U converges to zero, and does so faster than $1/X$.[6] This establishes the

[6]The monotonicity assumption on the static Nash equilibrium payoff function, $v(X)$, (Assumption 3) plays a key role in ensuring the limit of the first derivative exists. It is possible for a bounded function to converge to a finite limit, but have a derivative that oscillates. This assumption guarantees that U is monotonic for large X, and prevents U' from oscillating. A similar assumption is not necessary in

boundary condition on U' presented in Theorem 6,

$$\lim_{X\to\infty} XU'(X) = \lim_{X\to-\infty} XU'(X) = 0$$

The boundedness of U also ensures that the second derivative, U'', doesn't converge to a constant.[7] The optimality equation in Theorem 5 specifies the relationship between U and its first and second derivative. This relationship, coupled with Assumption 4 is used to establish the boundary condition for U. From the optimality equation,

$$\begin{aligned}\left(f_1(\bar{b},X)^2\sigma_Y^2 + \sigma_X^2(X)\right)U''(X) &= 2r\left[U(X) - g(a,\bar{b},X)\right] \\ &\quad -2\left[f_1(\bar{b},X)\mu_Y(a,\bar{b}) + f_2(\bar{b},X)\right]U'(X)\end{aligned}$$

Consider the limit of the optimality equation as the state variable approaches positive infinity. Under the assumption that the drift of the state variable has linear growth (f_1 and f_2), the second term on the right hand side converges to zero, and the flow payoff $g(a,\bar{b},X)$ converges to v_∞. (Recall that the agency plays a myopic best response at the boundaries, which yields a flow payoff equal to the static Nash equilibrium payoff v_∞). Then when the volatility of the state variable, $f_1(\bar{b},X)^2\sigma_Y^2 + \sigma_X^2(X)$, is bounded, as is assumed in Assumption 4, U must also converge to v_∞ to prevent the U'' from converging to a constant. This establishes the boundary condition on U presented in Theorem 6,

$$\lim_{X\to\infty} U(X) = v_\infty \text{ and } \lim_{X\to-\infty} U(X) = v_{-\infty}$$

Given Assumption 4, any solution to the optimality equation must satisfy these boundary conditions. Showing that it is not possible for two different solutions U_1 and U_2 to both satisfy these boundary conditions concludes the proof that the optimality equation has a unique solution. This establishes the existence of a unique Markovian

the bounded state space case, as the Lipschitz continuity of v is sufficient to ensure the limit of U' is well-defined.

[7]The boundedness of U ensures that U'' either converges to zero, or oscillates around zero.

equilibrium.

The second step in proving the existence of a unique PPE is showing that there are no non-Markovian PPE, and as such, this unique Markovian equilibrium is the unique PPE. The intuition behind this result, and its relationship with the continuous time literature, is discussed in depth following an example to illustrate Theorem 6.

Example to illustrate Theorem 6

I next illustrate that the persistent investment example satisfies the assumptions for Theorem 6 and has a unique PPE, which is Markovian.

Example 3. *In the persistent investment model, the state variable evolves according to:*

$$dX_t = \theta \left(a_t - X_t\right) dt + \sigma dZ_t$$

The drift of the state variable is $\theta \left(a_t - X_t\right)$, which grows linearly as $|X|$ approaches infinity. The volatility of the state variable is σ, which is bounded uniformly with respect to X, since it is constant. This example satisfies Assumption 4. As characterized in Section 2.2.1, the unique stage game Nash equilibrium is for the firm to choose zero investment, and consumers to choose a purchase level of $\overline{b} = 3$ when $X > 3/\lambda$. This ensures the monotonicity assumption of the static Nash payoffs, Assumption 3, is satisfied. Section 2.4.1 established that this example satisfies the other required assumptions for Theorem 6. Therefore, this example has a unique Markovian equilibrium that satisfies the following boundary conditions:

$$\begin{aligned} \lim_{X\to\infty} U(X) &= 3 \text{ and } \lim_{X\to-\infty} U(X) = 0 \\ \lim_{X\to\infty} XU'(X) &= \lim_{X\to-\infty} XU'(X) = 0 \end{aligned}$$

In this equilibrium, the firm's action converges to the static Nash best response of zero investment as the product quality becomes large, and the firm receives an equilibrium *payoff of* 3*, the highest feasible payoff for the firm.*

Intertemporal Incentives in Stochastic Games

The fact that the unique PPE is Markovian yields an important insight on the role action persistence plays in generating intertemporal incentives. In a stochastic game of imperfect monitoring, intertemporal incentives can be generated through two potential channels: (1) conditioning future equilibrium play on the public signal and (2) the effect of the current action on the set of future feasible payoffs via the state variable. Equilibrium play in a Markovian equilibrium is completely specified by the current value of the state variable, and the public signal is ignored. As such, the sole source of intertemporal incentives in a Markovian equilibrium is from the impact that the current action has on the set of future feasible payoffs. When this equilibrium is unique, it precludes the existence of any equilibria that use the public signal to generate intertemporal incentives via continuation play. *As such, the ability to generate effective intertemporal incentives in a stochastic game of imperfect monitoring stems entirely from the effect of the current action on the set of future feasible payoffs via the state variable.*

This insight relates to equilibrium degeneracy results from the continuous time repeated games literature, which show that it is not possible to provide effective intertemporal incentives in an imperfect monitoring game between a long-run and short-run player. In a standard repeated game, conditioning future equilibrium play on the public signal is the only potential channel for generating intertemporal incentives, and, as is the case in the stochastic game, this is not an effective channel for incentive provision. Thus, the introduction of action persistence creates an essential avenue for intertemporal incentive provision, and the ability to create effective intertemporal incentives in the stochastic game is entirely due to this additional this channel.

I next comment on the features of a stochastic game that allow Brownian information to be used to effectively provide intertemporal incentives. First consider the intuition behind why coordinating future equilibrium play is not an effective channel for incentive provision in a game between a short-run and long-run player. As discussed following lemma 3, Brownian information must be used linearly. Given this restriction, consider what happens when the long-run player's continuation value is at its upper bound. Using Brownian information linearly in a direction that is non-tangential to the boundary of the equilibrium payoff set will result in the continuation value escaping its upper bound with positive probability, a contradiction. Using Brownian information linearly in a tangential manner is precluded by the presence of myopic short-run players. Thus, it is not possible to linearly use Brownian information to structure incentives at the long-run player's highest continuation payoff, and both the long-run player and short-run player will play a myopic best response at this point. But this is precisely the definition of a static Nash equilibrium, and therefore long-run player's highest continuation payoff is bounded above by the highest static Nash equilibrium payoff.

Now consider the introduction of action persistence. The firm's incentive constraint and the evolution of the continuation value is still linear with respect to the Brownian information, as captured by the process $(\beta_t)_{t\geq 0}$ that governs incentives and the volatility of the continuation value. However, it is possible to characterize this process in a manner that depends on the state variable, and ensures the continuation value has zero volatility at states that yield the highest continuation payoff. This prevents the continuation value from escaping its upper bound with positive probability. Note this also implies that the long-run player must be playing a myopic best response at the state that yields the highest continuation payoff. However, for other states, it is possible to structure incentives such that the firm plays a non-myopic action.

Recall the characterization of $(\beta_t)_{t\geq 0}$ in Theorem 5, where β_t is the volatility of the agency's continuation value and also governs the agency's incentive constraint:

$$\beta_{1t} = U'(X_t) f_1(\bar{b}_t, X_t)$$

Suppose an interior state yields the highest continuation value. Then the slope of the continuation value is zero at this point, $U'(X_t) = 0$, which ensures the volatility of the continuation value is zero at its upper bound. Suppose that a boundary of the state space yields the highest continuation value. Then the boundary conditions characterized in Theorem 6 ensure that β_{1t} converges to zero at this boundary, and therefore the continuation value has zero volatility at its upper bound. However, for states that do not yield the highest or lowest continuation value, it is possible for $|\beta_{1t}| > 0$, which allows intertemporal to be structured in a manner such that the agency plays a nonmyopic action.

2.4.3 Stochastic Games without Action Persistence

Suppose that the state variable evolves independently of the public signal. This removes the link between the agency's action and evolution of the state variable, creating a stochastic game without action persistence. This section establishes that the unique PPE in this setting is one in which the firm acts myopically and plays the static Nash equilibrium of the current stage game.

Given an initial state X_0, define the average discounted payoff from playing the static Nash equilibrium action profile in each state as:

$$V_{NE}(X_0) = E\left[r\int_0^{\infty} e^{-rt}v(X_t)dt\right]$$

and the expected continuation payoff from playing a static Nash equilibrium action profile as:

$$W_{NE}(X_t) = E_t\left[r\int_t^{\infty} e^{-rs}v(X_s)dt\right]$$

where these expectations are taken with respect to the state variable, given that the state evolves according to the measure generated by the static Nash equilibrium action profile $S^*(X,0) = (a(X,0), \bar{b}(X,0))_{X\in\Xi}$. This expression defines the stochastic game payoff that the agency will earn if it myopically optimizes flow payoffs each instant. It is important to note that repeated play of the static Nash action profile is not necessarily an equilibrium strategy profile of the stochastic game; this is a general property of stochastic games.

The following Corollary establishes that in a stochastic game without action persistence, repeated play of the static Nash action profile is an equilibrium strategy profile; in fact, it is the unique equilibrium strategy profile, and yields the firm an equilibrium payoff of $V_{NE}(X_0)$. The corollary also directly characterizes $V_{NE}(X_0)$ as the solution to a *non-stochastic* second order differential equation (recall that $V_{NE}(X_0)$ is an expectation with respect to the state variable).

Corollary 2. *Suppose that the transition function of the state variable is independent of the public signal, $f_1 = 0$, for all X, and suppose Assumptions 1 and 2 hold. Then, given an initial state X_0, there is a unique perfect public equilibrium characterized by the unique bounded $U(X)$ solution to:*

$$U''(X) = \frac{2r\left[U(X) - g(a,\bar{b},X) - f_2(\bar{b},X)U'(X)\right]}{\sigma_X^2(X)}$$

This solution characterizes a Markovian equilibrium with:

1. *Equilibrium payoff $U(X_0)$*
2. *Continuation values $(W_t)_{t\geq 0} = (U(X_t))_{t\geq 0}$*
3. *Equilibrium actions $(a_t,\bar{b}_t)_{t\geq 0}$ uniquely specified by the static Nash equilibrium action profile, for each X:*

$$S^*(X,0) = \left\{ \begin{array}{c} a = \arg\max_{a'} g(a',\bar{b},X) \\ \bar{b} = \overline{\mathcal{B}}(a,X) \end{array} \right\}$$

Additionally, this equilibrium payoff and the continuation values correspond *to the expected payoff from playing the static Nash equilibrium action profile,* $U(X_0) = V_{NE}(X_0)$ *and* $(U(X_t))_{t\geq 0} = (W_{NE}(X_t))_{t\geq 0}$.

This result is a direct application of Theorem 5. The equilibrium characterization in Theorem 5 is used to characterize the optimality equation and incentive constraint for the firm when $f_1 = 0$. In the absence of a link between the agency's action and the state variable, the firm plays a static best response each instant.

It is interesting to note that the uniqueness result in a game without action persistence stems directly from the existence characterization in Theorem 5, and does not require the additional assumptions (Assumptions 3 and 4) necessary for Theorem 6. To see why, note that the incentive constraint is independent of the solution U to the optimality equation. As such, any solution to the optimality equation yields the same equilibrium action profile - namely, the action profile in which the agency plays a static best response each instant. Thus, all Markovian equilibrium action profiles induce the same measure over the path of the state variable. When a static best response is played each period, the continuation value in any Markovian equilibrium evolves according to the expected payoff from playing the static Nash equilibrium action profile in each state:

$$\begin{aligned} W_t &= E_t\left[r\int_t^{\infty} e^{-rs} v(X_s)ds\right] \\ &= W_{NE}(X_t) \end{aligned}$$

where the expectation is taken with respect to the measure over the state variable. Given that this measure over the state variable is the same in any Markovian equilibrium, any solution to the optimality equation must yield the same continuation values $(U(X_t))_{t\geq 0} = (W_{NE}(X_t))_{t\geq 0}$ for all $X \in \Xi$. Therefore, this solution must be unique. The solution to the optimality equation in Corollary 2 can be used to explicitly characterize the expectation $V_{NE}(X_0)$ and $.W_{NE}(X_t)$.

The existence of a solution to the optimality equation requires that the volatility of the state variable is bounded away from zero at interior points (Assumption 1). Note that when the state variable does not depend on the public signal, this assumption requires $\sigma_X^2(X_t)$ to be bounded away from zero to ensure that the state variable evolves stochastically.

Although Assumption 4 and 3 are not required to establish the existence of a unique PPE in the stochastic game without action persistence, if they do hold, then Theorem 6 can be used to characterize continuation payoffs at the boundary points of the state space. As established in Theorem 6, as the state variable approaches its boundary points (when the state space is unbounded, positive and negative infinity), then continuation payoffs approach the static Nash equilibrium payoff at the boundary points. Thus, the expected payoff from repeated play of the static game Nash equilibrium action profile approaches the payoff of the static Nash equilibrium at the boundary point of the state space.

Corollary 3. *Suppose Assumptions 3 and 4 also hold. Then if the state space is bounded,*

$$\lim_{X\to\overline{X}} W_{NE}(X) = v(\overline{X}) \textit{ and } \lim_{X\to\underline{X}} W_{NE}(X) = v\left(\underline{X}\right)$$

and if the state space is unbounded

$$\lim_{X\to\infty} W_{NE}(X) = v_{\infty} \textit{ and } \lim_{X\to-\infty} W_{NE}(X) = v_{-\infty}$$

This result is a direct implication of Theorem 6.

The degeneracy result of this section relates to the discussion of intertemporal incentive provision in Section 2.4.2. When the state variable evolves independently of the agency's action, this removes the second channel for intertemporal incentives that links the agency's action to the set of future feasible payoffs. Therefore, the only potential channel for intertemporal incentive provision is the coordination of equilibrium play, and, as discussed in 2.4.2, it is not possible to effectively structure incentives via this channel. The agency plays myopically in the unique equilibrium of this setting.

2.4.4 Properties of Equilibrium Payoffs

This section uses Theorem 5 and 6 to describe several interesting properties relating the agency's equilibrium payoffs to the structure of the underlying stage game. The main results of this section are to provide an upper and lower bound on the PPE payoffs of the stochastic game across all states and characterize how the PPE payoff of the agency varies with the state variable

First, I examine properties of the highest and lowest PPE payoffs across all states, represented by $\overline{W}$ and $\underline{W}$, respectively. Incentives for agency to choose a non-myopic action in the current period are provided through the link between the agency's action, the transition of the state variable and future feasible payoffs. When the continuation value is at its upper or lower bound, then the continuation value must have zero volatility so as not to escape its bound. The volatility of the continuation value is proportional to β_t, which also determines the incentive constraint for the agency. If the continuation value doesn't respond to the public signal, then the agency will myopically best respond by choosing the action that maximizes current flow payoffs. Therefore, the action profile at the set of states that yield the highest and lowest PPE payoffs across all states must be a Nash equilibrium of the stage game at that state.

At its upper bound, the drift of the continuation value must be negative. Using the law of motion for the continuation value characterized above, this means that the current flow payoff must exceed the continuation value. The current flow payoff in any stage Nash equilibrium is bounded above by $\overline{v}^*$, the highest stage Nash equilibrium payoff across all states. Thus, the highest PPE payoff across all states is bounded above by the highest static Nash equilibrium payoff across all states. Similar reasoning applies to showing that the lowest PPE payoff across all states is bounded below by the lowest static Nash equilibrium payoff.

Theorem 7. $\overline{W} \leq \overline{v}^*$ *and* $\underline{W} \geq \underline{v}^*$

If there is an absorbing state X in the set of states that yields the highest stage

Nash equilibrium payoff, then it is possible to remain in this state once it is reached. Thus, the highest PPE payoff across all states arises from repeated play of this stage Nash equilibrium, and yields a continuation value of $\overline{W} = \overline{v}^*$. By construction, this continuation value occurs at the state X that yields the highest static payoff. An analogous result holds for the lowest PPE payoff.

This result has an intuitive relation to reputation models of incomplete information. Recall that the state variable in this model is the consumers' beliefs about the firm's type. Therefore, $X = 0$ and $X = 1$ are absorbing states. When $X = 0$, consumers place probability one on the firm being a normal type, and it is not possible for the firm to earn payoffs above the static Nash equilibrium payoff of the game with complete information. Provided the firm's payoffs are increasing in the belief it is a behavioral type, the lowest PPE payoff occurs at $X = 0$ and is equal to the Nash equilibrium payoff of the complete information game, $\underline{v}^*$. Conditional on the firm being a normal type, the transition function governing beliefs has negative drift, and beliefs converge to the absorbing state $X = 0$. This captures the temporary reputation phenomenon associated with reputation models of incomplete information. Once consumers learn the firm's type, it is not possible to return to a state $X > 0$. Note that although $X = 1$ is also an absorbing state, but conditional on the firm being the normal type, the state variable never converges to $X = 1$

In the current model, if either endpoint is an absorbing state, *and* the state variable converges to this endpoint, then the intertemporal incentives created by the stochastic game will be temporary. Once this absorbing state is reached, the dynamic game is reduced to a standard repeated game and the unique equilibrium involves repeated play of the static Nash equilibrium. On the other hand, if neither endpoint is an absorbing state, or if the state variable doesn't converge to its absorbing states with positive probability, then the intertemporal incentives created by the stochastic game are permanent. As noted in the above discussion, it is possible to have an absorbing state that the state variable converges to with probability zero.

The second main result on PPE payoffs relates how the continuation value of the agency changes with the state variable to how the stage Nash equilibrium of the underlying stage game varies with the state variable.

Theorem 8. *Assume 1, 2, 3 and 4. The following properties show how PPE payoffs change with the state variable:*

1. *Suppose $v(X)$ is increasing (decreasing) in X.Then $U(X)$ is also increasing (decreasing) in X. The state that yields the highest static Nash payoff also yields the highest PPE payoff; likewise, the state that yields the lowest static Nash payoff also yields the lowest PPE payoff.*

2. *Suppose $v(X)$ has a unique interior maximum X^*, and v is monotonically increasing (decreasing) for $X < X^*$ $(X > X^*)$. Then $U(X)$ has a unique interior maximum at X^*,and U is monotonically increasing (decreasing) for $X < X^*$ $(X > X^*)$. The state that yields the highest static Nash payoff also yields the highest PPE payoff, whereas and the state that yields the lowest PPE payoff is a boundary point.*

3. *Suppose $v(X)$ has a unique interior minimum X^*, and v is monotonically decreasing (increasing) for $X < X^*$ $(X > X^*)$. Then $U(X)$ has a unique interior minimum at X^*,and U is monotonically decreasing (increasing) for $X < X^*$ $(X > X^*)$. The state that yields the lowest static Nash payoff also yields the lowest PPE payoff, whereas and the state that yields the highest PPE payoff is a boundary point.*

If the stage Nash equilibrium payoff of the agency is increasing in the state variable, then the PPE payoff to the agency is also increasing in the the state variable. The state that yields the highest PPE payoff to the agency corresponds to the state that yields the highest stage Nash equilibrium payoff, and the state that yields the lowest PPE payoff to the agency corresponds to the state that yields the lowest PPE payoff.

Stochastic games differ from standard repeated games in that it is not necessarily possible to achieve an equilibrium payoff of the stochastic game that is equal to the stage Nash equilibrium payoff. Thus, for intermediate values of the state variable, the PPE payoff may lie above or below the static Nash equilibrium payoff. A symmetric result holds if the stage Nash equilibrium payoff of the agency is decreasing in the state variable.

If the stage Nash equilibrium payoff of the agency is concave in the state variable, then the PPE payoff will be increasing over the region that the stage Nash equilibrium payoff is increasing, and decreasing over the region that the stage Nash equilibrium payoff is decreasing. The maximum PPE payoff will occur at the state that yields the maximum Nash equilibrium payoff of the stage game, and this PPE payoff will lie below maximum stage Nash equilibrium payoff, $\overline{W} < \overline{v}^*$. The state that yields the lowest PPE payoff will occur at either endpoint of the state space. If the endpoint that yields the lowest stage Nash equilibrium payoff is an absorbing state, then this state also yields the lowest PPE payoff and $\underline{W} = \underline{v}^*$. Otherwise, the endpoint that yields the lowest stage Nash equilibrium payoff will depend on the transition function. A symmetric result holds if the stage Nash equilibrium payoff of the agency is convex in the state variable.

This result characterizes properties of the PPE payoff as a function of the state variable for several relevant classes of games. More generally, if the stage game Nash equilibrium payoff is not monotonic or single-peaked in the state variable, then the highest and lowest PPE payoffs of the stochastic game may not coincide with the states that yield the maximum or minimum stage game Nash equilibrium payoffs.

2.5 Conclusion

Persistence and rigidities are pervasive in economics. There are many situations in which a payoff-relevant stock variable is determined not only by actions chosen today, but also by the history of past actions. This paper shows that this realistic departure from a standard repeated game provides a new channel for intertemporal incentives. The long-run player realizes that the impact of the action it chooses today will continue to be felt tomorrow, and incorporates the future value of this action into its decision. Persistence is a particularly important source of intertemporal incentives in the class of games examined in this paper; in the absence of such persistence, the long-run player cannot earn payoffs higher than those earned by playing a myopic best response.

The main results of this paper are to establish conditions on the structure of the game that guarantee existence of Markovian equilibria, and uniqueness of a perfect public equilibria, which is Markovian. Markovian equilibria have attractive features for use in applied work. These results not only provide a theoretical justification for restricting attention to such equilibria, but also develop a tractable method to characterize equilibrium behavior and payoffs in a Markovian equilibrium. The equilibrium dynamics can be directly related to observable features of a firm, or other long-run player, and used to generate empirically testable predictions.

This paper leaves open several interesting avenues for future research. Continuous time provides a tractable framework for studying games of imperfect monitoring. Ideally, equilibria of the continuous time will be robust in the sense that nearby discrete time games will exhibit similar equilibrium properties, as the period length becomes small. Faingold (2008) establishes such a robustness property in the context of a reputation game with commitment types. Whether the current setting is robust to the period length remains an open question.

Often, multiple long-run players may compete for the support of a fixed population of small players. For instance, rival firms may strive for a larger consumer base, political parties may contend for office, or universities may vie for the brightest students. These examples describe a setting in which each long-run player takes an action that persistently affects its state variable. Analyzing a setting with multiple state variables is technically challenging; if one could reduce such a game to a setting with a single payoff-relevant state, this simplification could yield a tractable characterization of equilibrium dynamics. For example, perhaps it is only the difference between two firms' product qualities that guide consumers' purchase behavior, or the difference between the platform of two political parties that determines constituents voting behavior.

Additionally, examining other classes of stochastic games, such as games between two long-run players whose actions jointly determine a stock variable, or games with different information structures governing the state transitions, remain unexplored.

I thank David Miller, Paul Niehaus, Joel Sobel, Jeroen Swinkels, Joel Watson and especially Nageeb Ali for useful comments. I also thank participants of the UCSD theory seminar and graduate student research seminar for helpful feedback.

2.6 Appendix

2.6.1 Proof of Lemma 3

Evolution of the continuation value

Let $W_t(S)$ be the firm's continuation value at time t given $X_t = X$, where $S = (a_t, \bar{b}_t)_{t\geq 0}$, and let $V_t(S)$ be the average discounted payoff conditional on info at time t.

$$\begin{aligned} V_t(S) \quad : \quad &= E_t\left[r\int_0^{\infty} e^{-rs} g(a_s, \bar{b}_s, X_s) ds\right] \\ &= r\int_0^{t} e^{-rs} g(a_s, \bar{b}_s, X_s) ds + e^{-rt} W_t(S) \end{aligned}$$

Lemma 4. *The average discounted payoff at time t, $V_t(S)$, is a martingale.*

$$\begin{aligned}
E_t[V_{t+k}(S)] &= E_t\left[r\int_0^{t+k} e^{-rs}g(a_s,\bar{b}_s,X_s)ds + e^{-r(t+k)}W_{t+k}(S)\right] \\
&= r\int_0^t e^{-rs}g(a_s,\bar{b}_s,X_s)ds \\
&\quad +E_t[r\int_t^{t+k} e^{-rs}g(a_s,\bar{b}_s,X_s)ds \\
&\quad +e^{-r(t+k)}E_{t+k}\left[r\int_{t+k}^{\infty} e^{-r(s-(t+k))}g(a_s,\bar{b}_s,X_s)ds\right]] \\
&= r\int_0^t e^{-rs}g(a_s,\bar{b}_s,X_s)ds \\
&\quad +e^{-rt}E_t[r\int_t^{t+k} e^{-r(s-t)}g(a_s,\bar{b}_s,X_s)ds \\
&\quad +r\int_{t+k}^{\infty} e^{-r(s-t)}g(a_s,\bar{b}_s,X_s)ds] \\
&= r\int_0^t e^{-rs}g(a_s,\bar{b}_s,X_s)ds + e^{-rt}W_t(S) = V_t(S)
\end{aligned}$$

Lemma 5. *In any PPE, the continuation value evolves according to the stochastic differential equation*

$$dW_t(S) = r\left(W_t(S) - g(a_t,\bar{b}_t,X_t)\right)dt + r\beta_{1t}\left[dY_t - \mu_Y(a_t,\bar{b}_t)dt\right] + r\beta_{2t}\sigma_X(X_t)dZ_t^X$$

Take the derivative of $V_t(S)$ wrt t :

$$dV_t(S) = re^{-rt}g(a_t,\bar{b}_t,X_t)dt - re^{-rt}W_t(S)dt + e^{-rt}dW_t(S)$$

By the martingale representation theorem, there exists a progressively measurable process $(\beta_t)_{t\geq 0}$ such that V_t can be represented as:

$$dV_t(S) = re^{-rt}\left[\beta_{1t}\sigma_Y dZ_t^Y + \beta_{2t}\sigma_X(X_t)dZ_t^X\right]$$

Combining these two expressions for $dV_t(S)$ yields the law of motion for the continuation value:

$$\begin{aligned} & re^{-rt}g(a_t,\overline{b}_t,X_t)dt - re^{-rt}W_t(S)dt + e^{-rt}dW_t(S) \\ = \quad & re^{-rt}\left[\beta_{1t}\sigma_Y dZ_t^Y + \beta_{2t}\sigma_X(X_t)dZ_t^X\right] \\ \Rightarrow \quad & dW_t(S) = rW_t(S)dt - rg(a_t,\overline{b}_t,X_t)dt \\ & +r\beta_{1t}\sigma_Y dZ_t^Y + r\beta_{2t}\sigma_X(X_t)dZ_t^X \\ \Rightarrow \quad & dW_t(S) = r\left(W_t(S) - g(a_t,\overline{b}_t,X_t)\right)dt \\ & +r\beta_{1t}\left[dY_t - \mu_Y(a_t,\overline{b}_t)dt\right] + r\beta_{2t}\sigma_X(X_t)dZ_t^X \end{aligned}$$

Q.E.D.

Sequential Rationality

Lemma 6. *A strategy $(a_t)_{t\geq 0}$ is sequentially rational for the agency if, given $(\beta_t)_{t\geq 0}$, for all t:*

$$a_t \in \arg\max \; g(a',\overline{b}_t,X_t) + \beta_{1t}\mu_Y(a',\overline{b}_t)$$

Consider strategy profile $(a_t,\overline{b}_t)_{t\geq 0}$ played from period τ onwards and alternative strategy $(\widetilde{a}_t,\overline{b}_t)_{t\geq 0}$ played up to time τ. Recall that all values of X_t are possible under both strategies, but that each strategy induces a different measure over sample paths $(X_t)_{t\geq 0}$.

At time τ, the state variable is equal to X_τ. Action a_τ will induce $dY_\tau = \mu_Y(a_\tau,\overline{b}_\tau)d\tau + \sigma_Y dZ_\tau^Y$ whereas action $\widetilde{a}_\tau$ will induce $dY_\tau = \mu_Y(\widetilde{a}_\tau,\overline{b}_\tau)d\tau + \sigma_Y dZ_\tau^Y$. Let $\widetilde{V}_\tau$ be is expected average payoff conditional on info at time τ when follows $\widetilde{a}$ up to τ and a afterwards, and let W_τ be the continuation value when the firm follows strategy $(a_t)_{t\geq 0}$ starting at time τ.

$$\widetilde{V}_\tau = r\int_0^\tau e^{-rs}g(\widetilde{a}_s,\overline{b}_s,X_s)ds + e^{-r\tau}W_\tau$$

Consider changing τ so that firm plays strategy $(\widetilde{a}_t, \overline{b}_t)$ for another instant: $d\widetilde{V}_\tau$ is the change in average expected payoffs when the firm switches to $(a_t)_{t\geq 0}$ at $\tau + d\tau$ instead of τ. Note

$$\begin{aligned} dW_\tau &= r\left(W_\tau - g(a_\tau, \overline{b}_\tau, X_\tau)\right) d\tau \\ &\quad + r\beta_{1t}\left[dY_\tau - \mu_Y(a_\tau, \overline{b}_\tau)d\tau\right] \\ &\quad + r\beta_{2\tau}\sigma_X(\overline{b}_\tau, X_\tau)dZ^X_\tau \end{aligned}$$

when firm switches strategies at time τ.

$$\begin{aligned} d\widetilde{V}_\tau &= re^{-r\tau}\left[g(\widetilde{a}_\tau, \overline{b}_\tau, X_\tau) - W_\tau\right] d\tau + e^{-r\tau} dW_\tau \\ &= re^{-r\tau}\left[g(\widetilde{a}_\tau, \overline{b}_\tau, X_\tau) - g(a_\tau, \overline{b}_\tau, X_\tau)\right] d\tau \\ &\quad + re^{-r\tau}\beta_{1\tau}\left[dY_\tau - \mu_Y(a_\tau, \overline{b}_\tau)d\tau\right] \\ &\quad + re^{-r\tau}\beta_{2\tau}\sigma_X(\overline{b}_\tau, X_\tau)dZ^X_\tau \\ &= re^{-r\tau}\left\{ \begin{array}{c} \left[g(\widetilde{a}_\tau, \overline{b}_\tau, X_\tau) - g(a_\tau, \overline{b}_\tau, X_\tau)\right] d\tau \\ + re^{-r\tau} r\beta_{1\tau}\left[\mu_Y(\widetilde{a}_\tau, \overline{b}_\tau)d\tau - \mu_Y(a_\tau, \overline{b}_\tau)d\tau + \sigma_Y dZ^Y_\tau .\right] \\ + re^{-r\tau}\beta_{2\tau}\sigma_X(\overline{b}_\tau, X_\tau)dZ^X_\tau \end{array} \right\} \end{aligned}$$

There are two components to this strategy change: how it affects the immediate flow payoff and how it affects future public signals Y_t, which impacts the continuation value (captured in process β). The profile $(\widetilde{a}_t, \overline{b}_t)_{t\geq 0}$ yields the firm a payoff of:

$$\begin{aligned} \widetilde{W}_0 &= E_0\left[\widetilde{V}_\infty\right] = E_0\left[\widetilde{V}_0 + \int_0^\infty d\widetilde{V}_t\right] \\ &= W_0 + E_0\left[r\int_0^\infty e^{-rt}\left\{ \begin{array}{c} g(\widetilde{a}_t, \overline{b}_t, X_t) + \beta_{1t}\mu_Y(\widetilde{a}_t, \overline{b}_t) \\ -g(a_t, \overline{b}_t, X_t) - \beta_t\mu_Y(a_t, \overline{b}_t) \end{array} \right\} dt\right] \end{aligned}$$

If

$$g(a_t, \overline{b}_t, X_t) + \beta_{1t}\mu_Y(a_t, \overline{b}_t) \geq g(\widetilde{a}_t, \overline{b}_t, X_t) + \beta_{1t}\mu_Y(\widetilde{a}_t, \overline{b}_t)$$

holds for all $t \geq 0$, Then $W_0 \geq \widetilde{W}_0$ and deviating to $S = (\widetilde{a}_t, \overline{b}_t)$ is not a profitable deviation. This yields the condition for sequential rationality for the firm.

Q.E.D.

2.6.2 Proof of Theorem 5: Characterization of Markovian Equilibrium

Theorem 9. *Suppose Assumptions 1 and 2 hold. Then given X_0, any solution $U(X)$ to the second order differential equation,*

$$U''(X) = \frac{2r\left[U(X) - g(a, \overline{b}, X)\right]}{f_1^2(\overline{b}, X)\sigma_Y^2 + \sigma_X^2(X)} - \frac{2\left[f_1(\overline{b}, X)\mu_Y(a, \overline{b}) + f_2(\overline{b}, X)\right]U'(X)}{f_1^2(\overline{b}, X)\sigma_Y^2 + \sigma_X^2(X)}$$

referred to as the optimality equation, characterizes a unique Markovian equilibrium in the state variable $(X_t)_{t\geq 0}$ with

1. *Equilibrium payoffs $U(X_0)$*
2. *Continuation values $(W_t)_{t\geq 0} = (U(X_t))_{t\geq 0}$*
3. *Equilibrium actions $(a_t, \overline{b}_t)_{t\geq 0} = \left(a^*(X_t), \overline{b}^*(X_t)\right)_{t\geq 0}$ uniquely specified by*

$$S^*(X, U'(X)f_1(\overline{b}, X)) = \left\{ \begin{array}{c} a = \arg\max_{a'} rg(a', \overline{b}, X) + U'(X)f_1(\overline{b}, X)\mu(a', \overline{b}) \\ \overline{b} = \overline{\mathcal{B}}(a, X) \end{array} \right\}$$

The optimality equation has at least one solution $U \in C^2(R)$ that lies in the range of feasible payoffs for the agency $U(X) \in \left[\underline{g}, \overline{g}\right]$ for all $X \in \Xi$. Thus, there exists at least on e Markovian equilibrium.

Proof of form of Optimality Equation:

Lemma 7. *If a Markovian equilibrium exists, it takes the following form:*

1. *Continuation values are characterized by a solution $U(X)$ to the optimality equation:*

$$U''(X) = \frac{2r\left(U(X) - g(a,\bar{b},X)\right)}{\left[f_1(\bar{b},X)^2\sigma_Y^2 + \sigma_X^2(X)\right]} - \frac{2\left[f_1(\bar{b},X)\mu_Y(a,\bar{b}) + f_2(\bar{b},X)\right]}{\left[f_1(\bar{b},X)^2\sigma_Y^2 + \sigma_X^2(X)\right]}U'(X)$$

2. *Given solution $U(X)$, the process governing incentives for the agency is characterized by:*

$$\begin{aligned} r\beta_1 &= U'(X)f_1(\bar{b},X) \\ r\beta_2 &= U'(X) \end{aligned}$$

Look for a Markovian equilibrium in the state variable X_t. In a Markovian equilibrium, the continuation value and equilibrium actions are characterized as a function of the state variable as $W_t = U(X_t), a_t = a(X_t)$ and $\bar{b}_t = \bar{b}(X_t)$. Note that Z_t^Y and Z_t^X are orthogonal. By Ito's formula, in a Markovian equilibrium, the continuation value will evolve according to

$$\begin{aligned} dW_t &= U'(X_t)dX_t + \frac{1}{2}U''(X_t)\left[f_1(\bar{b}_t,X_t)^2\sigma_Y^2 + \sigma_X^2(X_t)\right]dt \\ &= U'(X_t)\left[f_1(\bar{b}_t,X_t)\mu_Y(a_t,\bar{b}_t) + f_2(\bar{b}_t,X_t)\right]dt \\ &\quad + \frac{1}{2}U''(X_t)\left[f_1(\bar{b}_t,X_t)^2\sigma_Y^2 + \sigma_X^2(X_t)\right]dt \\ &\quad + U'(X_t)\left[f_1(\bar{b}_t,X_t)\sigma_Y dZ_t^Y + \sigma_X(X_t)dZ_t^X\right] \end{aligned}$$

Also, given players are playing strategy $\left(a_t,\bar{b}_t\right)_{t\geq 0}$ and the state variable evolves according to the transition function dX_t, the continuation value evolves according to:

$$dW_t = r\left(W_t - g(a_t, \bar{b}_t, X_t)\right)dt + r\beta_{1t}\sigma_Y dZ_t^Y + r\beta_{2t}\sigma_X(X_t)dZ_t^X$$

We can match the drift of these two characterizations to obtain the optimality equation:

$$\begin{aligned} r\left(U(X) - g(a, \bar{b}, X)\right) &= U'(X)\left[f_1(\bar{b}, X)\mu_Y(a, \bar{b}) + f_2(\bar{b}, X)\right] \\ &\quad + \frac{1}{2}U''(X)\left[f_1(\bar{b}, X)^2\sigma_Y^2 + \sigma_X^2(X)\right] \\ \Rightarrow U''(X) &= \frac{2r\left(U(X) - g(a, \bar{b}, X)\right)}{\left[f_1(\bar{b}, X)^2\sigma_Y^2 + \sigma_X^2(X)\right]} \\ &\quad - \frac{2\left[f_1(\bar{b}, X)\mu_Y(a, \bar{b}) + f_2(\bar{b}, X)\right]}{\left[f_1(\bar{b}, X)^2\sigma_Y^2 + \sigma_X^2(X)\right]}U'(X) \end{aligned}$$

for strategy profile $(a, \bar{b}) = (a(X), \bar{b}(X))$, which is a second order non-homogenous differential equation. Matching the volatility characterizes the process governing incentives:

$$\begin{aligned} r\beta_{1t} &= U'(X_t)f_1(\bar{b}_t, X_t) \\ r\beta_{2\tau} &= U'(X_t) \end{aligned}$$

Plugging these into the constraints for sequential rationality yields

$$\begin{aligned} S^*(X, U'(X)f_1(\bar{b}_t, X_t)) &= (a, \bar{b}) \text{ s.t.} \\ a &= \arg\max \ \ rg(a', \bar{b}, X) + U'(X)f_1(\bar{b}, X)\mu(a, \bar{b}) \\ \bar{b} &= \overline{\mathcal{B}}(a, X) \end{aligned}$$

which are unique by Assumption 2.

Q.E.D.

Prove existence of bounded solution to optimality equation

Linear Growth

Lemma 8. *The optimality equation has linear growth. Suppose Assumption 1 holds. For all $M > 0$ and compact intervals $I \subset \Xi$, there exists a $K_I > 0$ such that for all $X \in I$, $\left(a, \bar{b}\right) \in A \times B$, $u \in [-M, M]$ and $u' \in R$,*

$$u'' = \frac{r\left[u - g(a, \bar{b}, X)\right] - \left[f_1(\bar{b}, X)\mu_Y(a, \bar{b}) + f_2(\bar{b}, X)\right]u'}{\left[f_1(\bar{b}, X)^2\sigma_Y^2 + \sigma_X(X)^2\right]} \leq K\left(1 + |u'|\right)$$

Follows directly from the fact that $u \in [-M, M]$, g, f_1, μ_Y and f_2 are Lipschitz continuous, the bound on $f_1(\bar{b}, X)^2\sigma_Y^2 + \sigma_X^2(X)$, and $X \in I$.

Q.E.D.

Existence for Unbounded support

Theorem 10. *The optimality equation has at least one solution $U \in C^2(R)$ that lies in the range of feasible payoffs for the agency i.e. for all $X \in R$*

$$\inf g(a, \bar{b}, X) \leq U(X) \leq \sup g(a, \bar{b}, X)$$

The existence proof uses the following theorem from Schmitt, which gives sufficient conditions for the existence of a bounded solution to a second order differential equation defined on R^3. The Theorem is reproduced below.

Theorem 11. *Let $\alpha, \beta \in R$ be such that $\alpha \leq \beta$, $E = \{(t, u, v) \in R^3 | \alpha \leq u \leq \beta\}$ and $f : E \to R$ be continuous. Assume that α and β are such that for all $t \in R$*

$$\begin{aligned} f(t, \alpha, 0) &\leq 0 \\ f(t, \beta, 0) &\geq 0 \end{aligned}$$

Assume that for any bounded interval I, there exists a positive continuous function $\phi_I : R^+ \to R$ that satisfies

$$\int_0^\infty \frac{sds}{\phi_I(s)} = \infty$$

and for all $t \in I$, $(u,v) \in R^2$ with $\alpha \le u \le \beta$,

$$|f(t,u,v)| \le \phi_I(|v|)$$

Then the equation $u'' = f(t,u,v)$ has at least one solution on $u \in C^2(R)$ such that for all $t \in R$,

$$\alpha \le u(t) \le \beta$$

Let $\overline{g} = \sup g(a,\overline{b},X)$ and $\underline{g} = \inf g(a,\overline{b},X)$, which are well defined since g is bounded. Applying the above theorem with $\alpha = \underline{g}$ and $\beta = \overline{g}$ to $h(X,U(X),U'(X))$ yields

$$\begin{aligned} h(X,\underline{g},0) &= \frac{2r}{\left[f_1(\overline{b},X)^2\sigma_Y^2 + \sigma_X^2(X)\right]}\left(\underline{g} - g(a,\overline{b},X)\right) \le 0 \\ h(X,\overline{g},0) &= \frac{2r}{\left[f_1(\overline{b},X)^2\sigma_Y^2 + \sigma_X^2(X)\right]}\left(\overline{g} - g(a,\overline{b},X)\right) \ge 0 \end{aligned}$$

for all X. For any bounded interval I, define

$$\phi_I(v) = \frac{2r}{\sigma_I^2}\left(\overline{g} - \underline{g}\right) - \frac{2\mu_I}{\sigma_I^2}v$$

where $\sigma_I = \inf_{\overline{b}\in B, X\in I}\left[f_1(\overline{b}_t,X_t)^2\sigma_Y^2 + \sigma_X^2(X_t)\right]$ which is positive by assumption.and $\mu_I = \sup_{\overline{b}\in B, X\in I} f_1(\overline{b},X)\mu_Y(a,\overline{b}) + f_2(\overline{b},X)$, which are well-defined given f_1 , f_2 , μ_Y , σ_Y and σ_X are Lipschitz continuous and B is compact. Note

$$\int_0^\infty \frac{sds}{\phi_I(s)} = \infty$$

and for all $X \in I, (u,v) \in R^2$ with $\underline{g} \leq u \leq \overline{g}$

$$\begin{aligned} |h(X,u,v)| &= \left| \begin{array}{c} \frac{2r}{f_1(\bar{b},X)^2\sigma_Y^2+\sigma_X^2(X)}\left(u - g(a(X),\bar{b}(X),X)\right) \\ -\frac{2\left[f_1(\bar{b},X)\mu_Y(a,\bar{b})+f_2(\bar{b},X)\right]}{f_1(\bar{b},X)^2\sigma_Y^2+\sigma_X^2(X)}v \end{array} \right| \\ &\leq \phi_I(|v|) \end{aligned}$$

Additionally, $h(X_t, U(X_t), U'(X_t))$ is continuous given that $f_1, f_2, \mu_Y, \sigma_Y$ and σ_X are Lipschitz continuous and $g(a(X), \bar{b}(X), X)$ is continuous. Thus, $h(X, U(X), U'(X))$ has at least one solution on $U \in C^2(R)$ such that for all $X \in R$,

$$\underline{g} \leq U(X) \leq \overline{g}$$

Q.E.D.

Existence for Bounded support

Theorem 12. *The optimality equation has at least one solution $U \in C^2(R)$ that lies in the range of feasible payoffs for the agency i.e. for all $X \in \Xi$*

$$\inf g(a, \bar{b}, X) \leq U(X) \leq \sup g(a, \bar{b}, X)$$

The existence proof utilizes standard existence results from de Coster and Habets (2006) and an extension in Faingold and Sannikov (2011), applied to the current setting. The optimality equation is undefined at $\underline{X}$ and $\overline{X}$, since the volatility of X is zero. Therefore, an extension of standard existence results for second order ODEs is necessary. The main idea is to show that the boundary value problem has a solution U_n on $X \in \left[\underline{X} + 1/n, \overline{X} - 1/n\right]$ for every $n \in$ N, and then to show that this sequence of solutions converges pointwise to a continuously differentiable function U defined on $\left(\underline{X}, \overline{X}\right)$.

Given the boundary value problem has a solution U_n for every $n \in N$, with $\underline{X} = 0$ and $\overline{X} = 1$, Faingold and Sannikov (2011) show that when the second derivative of the ODE has quadratic growth, then a subsequence of $(U_n)_{n \geq 0}$ converges pointwise to a continuously differentiable function U defined on $(0, 1)$.

In this model, the second order derivative has linear growth, and therefore a similar argument shows existence of a continuously differentiable function U defined on $\left(\underline{X}, \overline{X}\right)$.

The existence results that are relevant for the current context are reproduced below:

Lemma 9. *Let $E = \{(t, u, v) \in \Xi \times R^2\}$ and $f : E \to R$ be continuous. Assume that for any interval $I \subset \Xi$,there exists a $K_I > 0$ such that for all $t \in I$, $(u, v) \in R^2$ with $\alpha \leq u \leq \beta$,*

$$|f\,(t, u, v)| \leq K_I\,(1 + |v|)$$

Then the equation $u'' = f\,(t, u, v)$ has at least one solution on $u \in C^2(R)$ such that for all $t \in \Xi$,

$$\alpha \leq u(t) \leq \beta$$

Consider the optimality equation $h(X, U(X), U'(X))$. Let $\overline{g} = \sup g(a, \overline{b}, X)$ and $\underline{g} = \inf g(a, \overline{b}, X)$, which are well defined since g is bounded. By 8, for any bounded interval I and $u \in \left[\underline{g}, \overline{g}\right]$, there exists a K_I such that

$$|f\,(t, u, v)| \leq K_I\,(1 + |v|)$$

Additionally, $h(X_t, U(X_t), U'(X_t))$ is continuous given that $f_1, f_2, \mu_Y, \sigma_Y$ and σ_X are Lipschitz continuous and $g(a(X), \overline{b}(X), X)$ is continuous. Let $\alpha = \underline{g}$ and $\beta = \overline{g}$. Then $h(X, U(X), U'(X))$ has at least one solution on $U \in C^2(R)$ such that for all $X \in \Xi$,

$$\underline{g} \leq U(X) \leq \overline{g}$$

Q.E.D.

Construct a Markovian equilibrium

Suppose the state variable initially starts at X_0 and U is a bounded solution to the optimality equation. The action profile satisfying $(a, \bar{b}) = S^*(X, U'(X)f_1(\bar{b}, X))$ is unique and Lipschitz continuous in X and U. Thus, given X_0, U and $(a_t, \bar{b}_t)_{t\geq 0} = \left[S^*(X_t, U'(X_t)f_1(\bar{b}_t, X_t))\right]_{t\geq 0}$, the state variable uniquely evolves according to the stochastic differential equation

$$dX_t = \left[f_1(\bar{b}_t, X_t)\mu_Y(a_t, \bar{b}_t) + f_2(\bar{b}_t, X_t)\right] dt + f_1(\bar{b}_t, X_t)\sigma_Y dZ_t^Y + \sigma_X(X_t)dZ_t^X$$

yielding unique path $(X_t)_{t\geq 0}$ given initial state X_0. Given that $U(X_t)$ is a bounded process that satisfies

$$\begin{aligned} dU(X_t) &= U'(X_t)\left[f_1(\bar{b}_t, X_t)\mu_Y(a_t, \bar{b}_t) + f_2(\bar{b}_t, X_t)\right] dt \\ &\quad + \frac{1}{2}U''(X_t)\left[f_1(\bar{b}_t, X_t)^2\sigma_Y^2 + \sigma_X^2(X_t)\right] dt \\ &\quad + U'(X_t)\left[f_1(\bar{b}_t, X_t)\sigma_Y dZ_t^Y + \sigma_X(X_t)dZ_t^X\right] \\ &= r\left(U(X_t) - g(a_t, \bar{b}_t, X_t)\right) + U'(X_t)\left[f_1(\bar{b}_t, X_t)\sigma_Y dZ_t^Y + \sigma_X(X_t)dZ_t^X\right] \end{aligned}$$

this process satisfies the conditions for the continuation value in a PPE characterized in Lemma 3. Additionally, $(a_t, \bar{b}_t)_{t\geq 0}$ satisfies the condition for sequential rationality given process $(\beta_t)_{t\geq 0} = \left(U'(X_t)f_1(\bar{b}_t, X_t), U'(X_t)\right)_{t\geq 0}$. Thus, the strategy profile $(a_t, \bar{b}_t)_{t\geq 0}$ is a PPE yielding equilibrium payoff $U(X_0)$.

Q.E.D.

2.6.3 Proof of Theorem 6: Characterize Unique Markovian Equilibrium

Theorem 13. *Suppose 4, 2 and 1 hold. Then, for each $X_0 \in \Xi$, there exists a unique perfect public equilibrium with continuation values characterized by the unique bounded solution U of the optimality equation, yielding equilibrium payoff $U(X_0)$.*

1. *If $\Xi = [\underline{X}, \overline{X}]$ then the solution satisfies the following boundary conditions:*

$$\begin{aligned} \lim_{X \to \overline{X}} U(X) &= v(\overline{X}) \text{ and } \lim_{X \to \underline{X}} U(X) = v\left(\underline{X}\right) \\ \lim_{X \to \overline{X}} (\overline{X} - X) U'(X) &= \lim_{X \to \underline{X}} (\underline{X} - X) U'(X) = 0 \end{aligned}$$

2. *If $\Xi = \mathcal{R}$ then the solution satisfies the following boundary conditions:*

$$\begin{aligned} \lim_{X \to \infty} U(X) &= v_{\infty} \text{ and } \lim_{X \to -\infty} U(X) = v_{-\infty} \\ \lim_{X \to \infty} X U'(X) &= \lim_{X \to -\infty} X U'(X) = 0 \end{aligned}$$

Boundary Conditions

Boundary Conditions for Unbounded Support

Theorem 14. *Any bounded solution U of the optimality equation satisfies the following boundary conditions*

$$\begin{aligned} &\lim_{X \to \infty} U(X) = v_{\infty} \text{ and } \lim_{X \to -\infty} U(X) = v_{-\infty} \\ &\lim_{X \to \infty} X U'(X) = \lim_{X \to -\infty} X U'(X) = 0 \\ &\lim_{X \to \infty} \left(f_1(\overline{b}, X)^2 \sigma_Y^2 + \sigma_X^2(X) \right) U''(X) = 0 \\ &\lim_{X \to -\infty} \left(f_1(\overline{b}, X)^2 \sigma_Y^2 + \sigma_X^2(X) \right) U''(X) = 0 \end{aligned}$$

The proof proceeds by a series of lemmas.

Lemma 10. *If U is a bounded solution of the optimality equation, then $\lim_{X\to\infty} U(X)$ and $\lim_{X\to-\infty} U(X)$ are well-defined.*

Proof: Assume Assumption 3. This guarantees $\lim_{X\to\infty} v(X)$ exists. U is bounded, and therefore $\lim_{X\to\infty} \sup U(X)$ and $\lim_{X\to\infty} \inf U(X)$ are well-defined. Suppose $\lim_{X\to\infty} \sup U(X) \neq \lim_{X\to\infty} \inf U(X)$. Then there exists a sequence $(X_n)_{n\in N}$ that correspond to local maxima of U, so $U'(X_n) = 0$ and $U''(X_n) \leq 0$. Given the incentives for the agency, a stage nash equilibria is played when $U'(X) = 0$, yielding flow payoff $v(X)$. From the optimality equation, this implies $v(X_n) \geq U(X_n)$. Likewise, there exists a sequence $(X_m)_{m\in N}$ that correspond to local minima of U, so $U'(X_m) = 0$ and $U''(X_m) \geq 0$. This implies $v(X_m) \leq U(X_m)$. Thus, $\lim_{X\to\infty} \sup v(X) \neq \lim_{X\to\infty} \inf v(X)$. This is a contradiction, as $\lim_{X\to\infty} v(X)$ is well-defined. Thus, $\lim_{X\to\infty} U(X)$ exists. The case of $\lim_{X\to-\infty} U(X)$ is similar. Q.E.D.

Lemma 11. *If $U(X)$ is a bounded solution of the optimality equation, then there exists a δ such that for $|X| > \delta$, $U(X)$ is monotonic.*

Proof: Assume Assumption 3. Suppose that there does not exist a δ such that for $X > \delta$, U is monotonic. Then for all $\delta > 0$, there exists a $X_n > \delta$ that corresponds to a local maxima of U, so $U'(X_n) = 0$ and $U''(X_n) \leq 0$ and there exists a $X_m > \delta$ that corresponds to a local minima of U, so $U'(X_m) = 0$ and $U''(X_m) \geq 0$, by the continuity of U. Given the incentives for the agency, a stage nash equilibria is played when $U'(X) = 0$, yielding flow payoff $v(X)$. From the optimality equation, this implies $v(X_n) \geq U(X_n)$ at the maximum and $v(X_m) \leq U(X_m)$ at the minimum. Thus, the oscillation of $v(X)$ is at least as large as the oscillation of $U(X)$. This is a contradiction, as there exists a δ such that for $X > \delta$, $v(X)$ is monotonic.The case of $-X > \delta$ is similar. Q.E.D.

Lemma 12. *Given a function $f(X)$ that is $O(X)$ as $X \to \{-\infty, \infty\}$, then any bounded solution U of the optimality equation satisfies*

1.

$$\begin{aligned}\lim_{X\to\infty}\inf f(X)U'(X) &\leq 0 \leq \lim_{X\to\infty}\sup f(X)U'(X)\\ \lim_{X\to-\infty}\inf f(X)U'(X) &\leq 0 \leq \lim_{X\to-\infty}\sup f(X)U'(X)\end{aligned}$$

2.

$$\begin{aligned}\lim_{X\to\infty}\inf f(X)^2U''(X) &\leq 0 \leq \lim_{X\to\infty}\sup f(X)^2U''(X)\\ \lim_{X\to-\infty}\inf f(X)^2U''(X) &\leq 0 \leq \lim_{X\to-\infty}\sup f(X)^2U''(X)\end{aligned}$$

Note this is trivially satisfied if $f(X)$ is $O(1)$.

1. Suppose $f(X)$ is $O(X)$ and $\lim_{X\to\infty}\inf|f(X)U'(X)| > 0$. Given $f(X)$ is $O(X)$, there exists an $M \in R$ and a $\delta_1 \in R$ such that when $X > \delta_1, |f(X)| \leq M\,|X|$ Given $\lim_{X\to\infty}\inf|f(X)U'(X)| > 0$, there exists a $\delta_2 \in R$ and an $\varepsilon > 0$ such that when $X > \delta_2$, $|f(X)U'(X)| > \varepsilon$. Take $\delta = \max\{\delta_1, \delta_2\}$. Then for $X > \delta$, $|U'(X)| > \frac{\varepsilon}{|f(X)|} \geq \frac{\varepsilon}{MX}$. Then the antiderivative of $\frac{\varepsilon}{MX}$ is $\frac{\varepsilon}{M}\ln X$ which converges to ∞ as $X \to \infty$. This violates the boundedness of U. Therefore $\lim_{X\to\infty}\inf f(X)U'(X) \leq 0$. The proof is analogous for the other cases.

2. Suppose $f(X)$ is $O(X)$ and $\lim_{X\to\infty}\inf|f(X)^2U''(X)| > 0$. Given $f(X)$ is $O(X)$, there exists an $M \in R$ and a $\delta_1 \in R$ such that when $X > \delta_1, |f(X)| \leq MX$ and therefore, $f(X)^2 \leq M^2X^2$. There also exists a $\delta_2 \in R$ and an $\varepsilon > 0$ such that when $X > \delta_2$, $|f(X)^2U''(X)| > \varepsilon$. Take $\delta = \max\{\delta_1, \delta_2\}$. Then for $X > \delta$, $|U''(X)| > \frac{\varepsilon}{f(X)^2} > \frac{\varepsilon}{M^2X^2}$. Then the antiderivative of $\frac{\varepsilon}{M^2X^2}$ is $\frac{-\varepsilon}{M^2}\ln X$ which converges to $-\infty$ as $X \to \infty$. This violates the boundedness of U. Therefore $\lim_{X\to\infty}\inf|f(X)^2U''(X)| \leq 0$. The proof is analogous for the other cases.

Q.E.D.

Lemma 13. *Given a function $f(X)$ that is $O(X)$ as $X \rightarrow \{-\infty, \infty\}$, then any bounded solution U of the optimality equation satisfies*

$$\lim_{X\rightarrow\infty} f(X)U'(X) = \lim_{X\rightarrow-\infty} f(X)U'(X) = 0$$

Assume Assumption 3. By Lemma 12,

$$\lim_{X\rightarrow\infty} \inf XU'(X) \leq 0 \leq \lim_{X\rightarrow\infty} \sup XU'(X)$$

. Suppose, without loss of generality, that $\lim_{X\rightarrow\infty} \sup XU'(X) > 0$. By Lemma 11, there exists a $\delta > 0$ such that $U(X)$ is monotonic for $X > \delta$. Then for $X > \delta$, $U'(X)$ doesn't change sign and therefore, $XU'(X)$ doesn't change sign. Therefore, if $\lim_{X\rightarrow\infty} \sup XU'(X) > 0$, then $\lim_{X\rightarrow\infty} \inf XU'(X) > 0$. This is a contradiction. Thus, $\lim_{X\rightarrow\infty} \sup XU'(X) = 0$. By similar reasoning, $\lim_{X\rightarrow\infty} \inf XU'(X) = 0$, and therefore $\lim_{X\rightarrow\infty} XU'(X) = 0$. Suppose $f(X)$ is $O(X)$. Then there exists an $M \in R$ and a $\delta_1 \in R$ such that when $X > \delta_1, |f(X)| \leq M\,|X|$. Thus, for $X > \delta_1$, $|f(X)U'(X)| \leq M\,|XU'(X)| \rightarrow 0$. The case for $\lim_{X\rightarrow-\infty} f(X)U'(X) = 0$ is analogous. Note this result also implies that

$$\lim_{X\rightarrow\infty} U'(X) = \lim_{X\rightarrow-\infty} U'(X) = 0$$

Q.E.D.

Lemma 14. *Let U be a bounded solution of the optimality equation. Then the limit of U converges to the limit of the stage Nash equilibrium payoffs as $X \rightarrow \{-\infty, \infty\}$*

$$\begin{aligned} \lim_{X\rightarrow\infty} U(X) &= v_{\infty} \\ \lim_{X\rightarrow-\infty} U(X) &= v_{-\infty} \end{aligned}$$

Proof: Assume Assumption 2, 3, and 4. By 10, $\lim_{X\to\infty} U(X) = U_\infty$ exists. Suppose $U_\infty < v_\infty$, where v_∞ is the limit of the stage game Nash equilibrium payoff at positive infinity. The function f_1 is $O(X)$, and therefore by Lemma 13, $\lim_{X\to\infty} U'(X) f_1(\bar{b}, X) = 0$ and $S^*(X, U'(X) f_1(\bar{b}, X)) \to \left(a^N_\infty, \bar{b}^N_\infty\right)$ which is the stage Nash equilibrium as $X \to \infty$. Thus, $\lim_{X\to\infty} g(a(X), \bar{b}(X), X) = v_\infty$.

By Lemma 13 and the assumption that $(f_1\mu_Y + f_2)$ is $O(X)$

$$\lim_{X\to\infty} \left[f_1(\bar{b}, X)\mu_Y(a, \bar{b}) + f_2(\bar{b}, X)\right] U'(X) = 0$$

By the assumption that $f_1\sigma_Y + \sigma_X$ is $O(1)$, there exists an $M > 0$ and a δ such that when $X > \delta_1$, then $f_1(\bar{b}, X)^2\sigma_Y^2 + \sigma_X^2(X) \le M$.

Plugging the above conditions in to the optimality equation yields

$$\begin{aligned}\limsup_{X\to\infty} U''(X) &= \limsup_{X\to\infty} [\frac{2r\left(U(X) - g(a, \bar{b}, X)\right)}{f_1(\bar{b}, X)^2\sigma_Y^2 + \sigma_X^2(X)} - \\ &\quad \frac{2\left[f_1(\bar{b}, X)\mu_Y(a, \bar{b}) + f_2(\bar{b}, X)\right]}{f_1(\bar{b}, X)^2\sigma_Y^2 + \sigma_X^2(X)} U'(X)] \\ &\le \frac{2r\left(U_\infty - v_\infty\right)}{M} < 0\end{aligned}$$

which violates Lemma 12, and therefore U is unbounded. This is a contradiction. Thus, $U_\infty = v_\infty$. The proof for the other cases is analogous.

Q.E.D.

Lemma 15. *Any bounded solution U of the optimality equation satisfies*

$$\begin{aligned}\lim_{X\to\infty} \left|\left(f_1(\bar{b}, X)^2\sigma_Y^2 + \sigma_X^2(X)\right) U''(X)\right| &= 0 \\ \lim_{X\to-\infty} \left|\left(f_1(\bar{b}, X)^2\sigma_Y^2 + \sigma_X^2(X)\right) U''(X)\right| &= 0\end{aligned}$$

Note this also implies $U''(X) \to 0$.

Applying the squeeze theorem to the optimality equation yields

$$
\begin{aligned}
& \lim_{X\to\infty} \left|\left(f_1(\overline{b},X)^2\sigma_Y^2 + \sigma_X^2(X)\right)U''(X)\right| \\
= & \lim_{X\to\infty} \left|2r\left(U(X) - g(a,\overline{b},X)\right) - 2\left[f_1(\overline{b},X)\mu_Y(a,\overline{b}) + f_2(\overline{b},X)\right]U'(X)\right| \\
= & \ 0
\end{aligned}
$$

by applying Lemmas 13 and 14 and the assumption that $\left[f_1(\overline{b},X)\mu_Y(a,\overline{b}) + f_2(\overline{b},X)\right]$ is $O(X)$.

Q.E.D.

Boundary Conditions for Bounded Support

Theorem 15. *Any bounded solution U of the optimality equation satisfies the following boundary conditions:*

$$
\begin{aligned}
& \lim_{X\to\overline{X}} U(X) = v(\overline{X}) \text{ and } \lim_{X\to\underline{X}} U(X) = v\left(\underline{X}\right) \\
& \lim_{X\to\overline{X}} (\overline{X} - X)U'(X) = \lim_{X\to\underline{X}} (\underline{X} - X)U'(X) = 0 \\
& \lim_{X\to\overline{X}} \left(f_1(\overline{b},X)^2\sigma_Y^2 + \sigma_X^2(X)\right)U''(X) = 0 \\
& \lim_{X\to\underline{X}} \left(f_1(\overline{b},X)^2\sigma_Y^2 + \sigma_X^2(X)\right)U''(X) = 0
\end{aligned}
$$

The proof proceeds by a series of lemmas.

Lemma 16. *Any bounded solution U of the optimality equation has bounded variation.*

Suppose not. Then there exists a sequence $(X_n)_{n\in N}$ that correspond to local maxima of U, so $U'(X_n) = 0$ and $U''(X_n) \leq 0$. Given the incentives for the agency, a stage nash equilibria is played when $U'(X) = 0$, yielding flow payoff $v(X)$. From the optimality equation, this implies $v(X_n) \geq U(X_n)$. Likewise, there exists a sequence $(X_m)_{m\in N}$ that correspond to local minima of U, so $U'(X_m) = 0$ and $U''(X_m) \geq 0$.

This implies $v(X_m) \leq U(X_m)$. Thus, v also has unbounded variation. This is a contradiction, since v is Lipschitz continuous.

Q.E.D.

Lemma 17. *Given a function $f(X)$ that is $O(a-X)$ as $X \to a \in \left\{\underline{X}, \overline{X}\right\}$, then any bounded solution U of the optimality equation satisfies*

1.

$$\begin{aligned}
\lim_{X\to\overline{X}} \inf f(X)U'(X) &\leq 0 \leq \lim_{X\to\overline{X}} \sup f(X)U'(X) \\
\lim_{X\to\underline{X}} \inf f(X)U'(X) &\leq 0 \leq \lim_{X\to\underline{X}} \sup f(X)U'(X)
\end{aligned}$$

2.

$$\begin{aligned}
\lim_{X\to\overline{X}} \inf f(X)^2U''(X) &\leq 0 \leq \lim_{X\to\overline{X}} \sup f(X)^2U''(X) \\
\lim_{X\to\underline{X}} \inf f(X)^2U''(X) &\leq 0 \leq \lim_{X\to\underline{X}} \sup f(X)^2U''(X)
\end{aligned}$$

Note this is trivially satisfied if $f(X)$ is $O(1)$.

1. Suppose $f(X)$ is $O(\overline{X}-X)$ as $X \to \overline{X}$ and $\lim_{X\to\overline{X}} \inf |f(X)U'(X)| > 0$. There exists an $M \in R$ and a $\delta_1 > 0$ such that when $\left|\overline{X}-X\right| < \delta_1, |f(X)| \leq M\left|\overline{X}-X\right|$. Given $\lim_{X\to\overline{X}} \inf |f(X)U'(X)| > 0$, there exists a $\delta_2 \in R$ and an $\varepsilon > 0$ such that when $\left|\overline{X}-X\right| < \delta_2$, $|f(X)U'(X)| > \varepsilon$. Take $\delta = \min\{\delta_1, \delta_2\}$. Then for $\left|\overline{X}-X\right| < \delta$, $|U'(X)| > \frac{\varepsilon}{|f(X)|} \geq \frac{\varepsilon}{M\left|\overline{X}-X\right|}$. Then the antiderivative of $\frac{\varepsilon}{M\left|\overline{X}-X\right|}$ is $\frac{\varepsilon}{M}\ln\left|\overline{X}-X\right|$ which diverges to $-\infty$ as $X \to \overline{X}$. This violates the boundedness of U. Therefore $\lim_{X\to\overline{X}} \inf f(X)U'(X) \leq 0$. The proof is analogous for the other cases.

2. Suppose $f(X)$ is $O(\overline{X}-X)$ and $\lim_{X\to\infty} \inf |f(X)^2U''(X)| > 0$. There exists an $M \in R$ and a $\delta_1 > 0$ such that when $\left|\overline{X}-X\right| < \delta_1, |f(X)| \leq M\left|\overline{X}-X\right|$

and therefore, $f(X)^2 \leq M^2\left(\overline{X}-X\right)^2$. There also exists a $\delta_2 \in R$ and an $\varepsilon > 0$ such that when $\left|\overline{X}-X\right| < \delta_2$, $|f(X)^2U''(X)| > \varepsilon$. Take $\delta = \min\{\delta_1, \delta_2\}$. Then for $\left|\overline{X}-X\right| < \delta$, $|U''(X)| > \frac{\varepsilon}{f(X)^2} > \frac{\varepsilon}{M^2\left(\overline{X}-X\right)^2}$. Then the second antiderivative of $\frac{\varepsilon}{M^2\left(\overline{X}-X\right)^2}$ is $\frac{-\varepsilon}{M^2}\ln\left(\overline{X}-X\right)$ which converges to ∞ as $X \to \overline{X}$. This violates the boundedness of U. Therefore $\lim_{X\to\infty}\inf|f(X)^2U''(X)| \leq 0$. The proof is analogous for the other cases.

Q.E.D.

Lemma 18. *Given a differentiable function $f(X)$ that is $O(X^*-X)$ as $X \to \{\underline{X}, \overline{X}\}$, then any bounded solution U of the optimality equation satisfies*

$$\lim_{X\to\overline{X}} f(X)U'(X) = \lim_{X\to\underline{X}} f(X)U'(X) = 0$$

By Lemma 17, $\lim_{X\to\overline{X}}\inf(\overline{X}-X)U'(X) \leq 0 \leq \lim_{X\to\overline{X}}\sup(\overline{X}-X)U'(X)$. Suppose, without loss of generality, that $\lim_{X\to\overline{X}}\sup(\overline{X}-X)U'(X) > 0$. Then there exist constants k and K such that $(\overline{X}-X)U'$ crosses the interval (k, K) infinitely many times as X approaches $\overline{X}$. Additionally, there exists an $L > 0$ such that

$$\begin{aligned}
|U''(X)| &= \left|\frac{2r\left[U(X)-g(a,\overline{b},X)\right]-2\left[f_1(\overline{b},X)\mu_Y(a,\overline{b})+f_2(\overline{b},X)\right]U'(X)}{f_1(\overline{b},X)^2\sigma_Y^2+\sigma_X^2(X)}\right| \\
&\leq \left|\frac{L_1-L_2\left(\overline{X}-X\right)U'(X)}{\left(\overline{X}-X\right)^2}\right| \\
&\leq \left|\frac{L_1-L_2k}{\left(\overline{X}-X\right)^2}\right| = \frac{L}{\left(\overline{X}-X\right)^2}
\end{aligned}$$

Therefore,

$$\begin{aligned}\left|\left[(\overline{X}-X)U'(X)\right]'\right| &\leq |U'(X)| + \left|(\overline{X}-X)U''(X)\right| \\ &= \left(1+\left|(\overline{X}-X)\frac{U''(X)}{U'(X)}\right|\right)|U'(X)| \\ &\leq \left(1+\frac{L}{k}\right)|U'(X)|\end{aligned}$$

where the first line follows from differentiating $\left(\overline{X}-X\right)U'(X)$ and the subadditivity of the absolute value function, the next line follows from rearranging terms, the third line follows from the bound on $|U''(X)|$ and $(\overline{X}-X)U'(X) \in (k,K)$. Then

$$|U'(X)| \geq \frac{\left|\left[\left(\overline{X}-X\right)U'(X)\right]'\right|}{\left(1+\frac{L}{k}\right)}$$

Therefore, the total variation of U is at least $\frac{K-k}{\left(1+\frac{L}{k}\right)}$ on the interval $(\overline{X}-X)U'(X) \in (k,K)$, which implies that U has unbounded variation near $\overline{X}$. This is a contradiction. Thus, $\lim_{X\to\overline{X}}\sup(\overline{X}-X)U'(X)=0$. Likewise, $\lim_{X\to\overline{X}}\inf(\overline{X}-X)U'(X)=0$, and therefore $\lim_{X\to\overline{X}}(\overline{X}-X)U'(X)=0$. Then for any function $f(X)$ that is $O(\overline{X}-X)$, $|f(X)U'(X)| \leq M_1\left|\left(\overline{X}-X\right)U'(X)\right| \to 0$, and therefore $\lim_{X\to\overline{X}} f(X)U'(X)=0$

Q.E.D.

Lemma 19. *Let U be a bounded solution of the optimality equation. Then the limit of U converges to the limit of the stage Nash equilibrium payoffs as $X \to \left\{\underline{X},\overline{X}\right\}$*

$$\begin{aligned}\lim_{X\to\overline{X}} U(X) &= v(\overline{X}) \\ \lim_{X\to\underline{X}} U(X) &= v\left(\underline{X}\right)\end{aligned}$$

Suppose not. By 16, $\lim_{X\to\overline{X}} U(X) = U\left(\overline{X}\right)$ exists. Suppose $U\left(\overline{X}\right) < v(\overline{X})$, where $v(\overline{X})$ is the limit of the stage game Nash equilibrium payoff at $\overline{X}$. The function f_1 is $O(\overline{X}-X)$, and therefore by Lemma 18, $\lim_{X\to\overline{X}} U'(X) f_1(\bar{b},X) = 0$ and $S^*(X, U'(X) f_1(\bar{b},X)) \to \left(a^N_{\overline{X}}, \bar{b}^N_{\overline{X}}\right)$ which is the stage Nash equilibrium as $X \to \overline{X}$. Thus, $\lim_{X\to\overline{X}} g(a(X), \bar{b}(X), X) = v(\overline{X})$.

By Lemma 18 and the assumption that $f_1\mu_Y + f_2$ is $O(\overline{X}-X)$

$$\lim_{X\to\infty} \left(f_1(\bar{b},X)\mu_Y(a,\bar{b}) + f_2(\bar{b},X)\right) U'(X) = 0$$

By the assumption that $1/\left(f_1\sigma_Y + \sigma_X\right)$ is $O(1/(\overline{X}-X))$, there exists an $M>0$ and a δ such that when $\left|\overline{X}-X\right| < \delta_1$, then $1/\left(f_1\sigma_Y + \sigma_X\right) \leq M/(\overline{X}-X)$

Plugging the above conditions in to the optimality equation yields

$$\begin{aligned}
\limsup_{X\to\overline{X}} U''(X) &= \limsup_{X\to\overline{X}} \frac{2\left[f_1(\bar{b},X)\mu_Y(a,\bar{b}) + f_2(\bar{b},X)\right]}{f_1(\bar{b},X)^2\sigma_Y^2 + \sigma_X^2(X)} U'(X) \\
&\leq \frac{2r\left(U\left(\overline{X}\right) - v(\overline{X})\right)}{M(\overline{X}-X)^2} \\
&< 0
\end{aligned}$$

which violates Lemma 17, and therefore U is unbounded. This is a contradiction. Thus, $U\left(\overline{X}\right) = v(\overline{X})$. The proof for the other cases is analogous.

Q.E.D.

Lemma 20. *Any bounded solution U of the optimality equation satisfies*

$$\begin{aligned}
\lim_{X\to\overline{X}} \left|\left(f_1(\bar{b},X)^2\sigma_Y^2 + \sigma_X^2(X)\right) U''(X)\right| &= 0 \\
\lim_{X\to\underline{X}} \left|\left(f_1(\bar{b},X)^2\sigma_Y^2 + \sigma_X^2(X)\right) U''(X)\right| &= 0
\end{aligned}$$

Applying the squeeze theorem to the optimality equation yields

$$\begin{aligned}
& \lim_{X \to \overline{X}} \left| \left(f_1(\overline{b}, X)^2 \sigma_Y^2 + \sigma_X^2(X) \right) U''(X) \right| \\
= & \lim_{X \to \overline{X}} \left| 2r \left(U(X) - g(a, \overline{b}, X) \right) - 2 \left[f_1(\overline{b}, X) \mu_Y(a, \overline{b}) + f_2(\overline{b}, X) \right] U'(X) \right| \\
= & \; 0
\end{aligned}$$

by applying Lemmas 18 and 19.

Q.E.D.

Uniqueness of Solution to Optimality Equation

This proof builds on a result from Faingold and Sannikov (2011). They prove that the optimality equation characterizing a Markovian equilibrium in a repeated game of incomplete information over the type of the long-run player has a unique solution. The key element of this proof is that all solutions have the same boundary conditions when beliefs place probability 1 on the long-run player being a normal or behavioral type. This result also applies to the optimality equation characterized in this paper, given that all solutions have the same boundary conditions. An extension of this result is necessary for the case of an unbounded state space. The proof proceeds by two lemmas.

The first lemma follows directly from Lemma C.7 in Faingold and Sannikov (2011).

Lemma 21. *If two bounded solutions of the optimality equation, U and V, satisfy $U(X_0) \leq V(X_0)$ and $U'(X_0) \leq V'(X_0)$, with at least one strict inequality, then $U(X) \leq V(X)$ and $U'(X) \leq V'(X)$ for all $X > X_0$. Similarly if $U(X_0) \leq V(X_0)$ and $U'(X_0) \geq V'(X_0)$, with at least one strict inequality, then $U(X) < V(X)$ and $U'(X) > V'(X)$ for all $X < X_0$.*

The proof is analogous to the proof in Faingold and Sannikov (2011), defining

$$X_1 = \inf \left\{ X \in [X_0, \overline{X}) : U'(X) \geq V'(X) \right\}$$

for $\Xi = [\underline{X}, \overline{X}]$, and

$$X_1 = \inf \{ X \in [X_0, \infty) : U'(X) \geq V'(X) \}$$

for $\Xi = R$.

Q.E.D.

Lemma 22. *There exists a unique solution U to the optimality equation.*

Suppose U and V are both bounded solutions to the optimality equation, and assume $U(X) - V(X) > 0$ for some $X \in \Xi$.

First consider $\Xi = R$. Given that $\lim_{X \to \infty} U(X) = \lim_{X \to \infty} V(X) = v_\infty$, for all $\varepsilon > 0$, there exists a δ such that for $X \geq \delta$, $|U(X) - v_\infty| < \varepsilon/2$ and $|V(X) - v_\infty| < \varepsilon/2$. Then for $X \geq \delta$, $|U(X) - V(X)| < \varepsilon$.

Take an interval $X \in [X_1, X_2]$, and suppose $U(X) > V(X)$ for some $X \in [X_1, X_2]$. Let X^* be the point where $U - V$ is maximized, which is well-defined given U and V are continuous functions on a compact interval. Suppose the maximum occurs at an interior point $X^* \in (X_1, X_2)$. Then $U'(X^*) = V'(X^*)$. By Lemma 21, $U'(X) \geq V'(X)$ for all $X > X^*$, and this difference is strictly increasing, a contradiction. Suppose the maximum occurs at an endpoint, $X^* = X_2$, and let $U(X_2) - V(X_2) = M > 0$. Then it must be the case that $U'(X_2) \geq V'(X_2)$. By Lemma 21, $U'(X) \geq V'(X)$ for all $X > X_2$, and this difference is strictly increasing for $X > X_2$. But then for $\varepsilon < M$, there does not exists a δ such that $|U(X) - V(X)| < \varepsilon$ when $X > \delta$. This violates the boundary condition. The argument is analogous if the maximum occurs at $X^* = X_1$. Thus, it is not possible to have $U(X) > V(X)$.

The proof for$\Xi = [\underline{X}, \overline{X}]$ is similar, using $[X_1, X_2] = [\underline{X}, \overline{X}]$, and the fact that the boundary conditions at $[\underline{X}, \overline{X}]$ ensure the point where $U - V$ is maximized is an interior point.

Q.E.D.

Uniqueness of Markovian Equilibrium in class of PPE

Let X_0 be the initial state, and let U be a bounded solution to the optimality equation. Suppose there is a PPE $(a_t, \bar{b}_t)_{t\geq 0}$ that yields an equilibrium payoff $W_0 > U(X_o)$. The continuation value in this equilibrium must evolve according to

$$\begin{aligned} dW_t(S) &= r\left(W_t(S) - g(a_t, \bar{b}_t, X_t)\right)dt \\ &+ r\beta_{1t}\left[dY_t - \mu_Y(a_t, \bar{b}_t)dt\right] \\ &+ r\beta_{2t}\sigma_X(X_t)dZ_t^X \end{aligned}$$

for some process $(\beta_t)_{t\geq 0}$ and by sequential rationality, $(a_t, \bar{b}_t) = S^*(X_t, \beta_t)$ for all t. The process $U(X_t)$ evolves according to

$$\begin{aligned} dU(X_t) &= U'(X)\left[f_1(\bar{b}, X)\mu_Y(a, \bar{b}) + f_2(\bar{b}, X)\right]dt \\ &+ \frac{1}{2}U''(X)\left[f_1(\bar{b}, X)^2\sigma_Y^2 + \sigma_X^2(X)\right]dt \\ &+ U'(X_t)\left[f_1(\bar{b}_t, X_t)\sigma_Y dZ_t^Y + \sigma_X(X_t)dZ_t^X\right] \end{aligned}$$

Define a process $D_t = W_t - U(X_t)$ with initial condition $D_0 = W_0 - U(X_o) > 0$. Then dD_t evolves with drift

$$\begin{aligned} & r\left[W_t - g(a_t, \bar{b}_t, X_t)\right] - U'(X_t)\left[f_1(\overline{b_t}, X_t)\mu_Y(a_t, \bar{b}_t) + f_2(\bar{b}_t, X_t)\right] \\ & -\frac{1}{2}U''(X_t)\left[f_1(\bar{b}_t, X_t)^2\sigma_Y^2 + \sigma_X^2(X_t)\right] \\ = \ & rD_t + r\left[U(X_t) - g(a_t, \bar{b}_t, X_t)\right] - U'(X_t)\left[f_1(\overline{b_t}, X_t)\mu_Y(a_t, \bar{b}_t) + f_2(\bar{b}_t, X_t)\right] \\ & -\frac{1}{2}U''(X_t)\left[f_1(\bar{b}_t, X_t)^2\sigma_Y^2 + \sigma_X^2(X_t)\right] \\ = \ & rD_t + d(a_t, \bar{b}_t, X_t) \end{aligned}$$

and volatility

$$f(\bar{b}, X, \beta) = \begin{matrix} r\beta_{1t}\sigma_Y \\ r\beta_{2t}\sigma_X(X_t) \end{matrix} - \begin{matrix} U'(X_t)f_1(\bar{b}_t, X_t)\sigma_Y \\ U'(X_t)\sigma_X(X_t) \end{matrix}$$

Lemma 23. *For every $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $\left(a, \bar{b}, X, \beta\right)$ satisfying the condition for sequential rationality*

$$\begin{aligned} a & \in \arg\max rg(a', \bar{b}, X) + r\beta_1\mu_Y(a, \bar{b}) \\ \bar{b} & \in \overline{\mathcal{B}}(a, X) \end{aligned}$$

either $d(a_t, \bar{b}_t, X_t) > -\varepsilon$ or $\left|f(\bar{b}, X, \beta)\right| > \delta$.

Proof:

Suppose the state space is unbounded, $\Xi = R$.

Step 1: Show that if $\left|f(\bar{b}, X, \beta)\right| = 0$, then $d(a, \bar{b}, X) = 0$. (i.e. when the volatility of D_t is 0, the Markovian action profile is used in both equilibria)

Let $\left|f(\bar{b}, X, \beta_1)\right| = 0$. Then $r\beta_1 = U'(X)f_1(\bar{b}, X)$ and for each X, there is a unique action $\left(a, \bar{b}\right)$ profile that satisfies

$$\begin{aligned} a &\in \arg\max rg(a', \overline{b}, X) + U'(X) f_1(\overline{b}, X)\mu_Y(a, \overline{b}) \\ \overline{b} &\in \overline{\mathcal{B}}(a, X) \end{aligned}$$

This action profile corresponds to the actions played in a Markovian equilibrium, and therefore $d(a_t, \overline{b}_t, X_t) = 0$, by the optimality equation.

Step 2: Fix ε. Show if $d(a_t, \overline{b}_t, X_t) \leq -\varepsilon$, then there exists a $\delta > 0$ such that $\left|f(\overline{b}, X, \beta_1)\right| > \delta$ for all $\left(a, \overline{b}, X, \beta\right)$ such that the sequential rationality condition is satisfied.

Step 2a: Show there exists an $M > 0$ such that this is true for

$$\left(a, \overline{b}, X, \beta\right) \in \{A \times B \times \Xi \times R : |\beta| > M\ \}$$

$U'(X) f_1(\overline{b}, X)$ and $U'(X)$ are bounded, so there exists an $M > 0$ and $m > 0$ such that $\left|f(\overline{b}, X, \beta)\right| > m$ for all $|\beta| > M$.

Step 2b: Show that there exists an X^* such that this is true for

$$\left(a, \overline{b}, X, \beta\right) \in \{A \times B \times \Xi \times R : |\beta| \leq M \text{ and } |X| > X^*\}$$

Show if $r\left(v_\infty - g_\infty(a, \overline{b})\right) \leq -\gamma$ then there exists a $\eta_2 > 0$ such that $|\beta| > \eta_2$. Let $\lim_{X\to\infty} g(a, \overline{b}, X) := g_\infty(a, \overline{b})$ be the limit flow payoff for actions $(a, \overline{b})$, and $\lim_{X\to\infty} \overline{\mathcal{B}}(a, X) := \overline{\mathcal{B}}_\infty(a)$. These limits *exist*, since g and $\overline{\mathcal{B}}(a)$ are Lipschitz continuous and bounded. Consider the set $\Phi = \left(a, \overline{b}, \beta\right)$ that satisfies

$$\begin{aligned} a &\in \arg\max rg_\infty(a', \overline{b}) + r\beta_1\mu_Y(a, \overline{b}) \\ \overline{b} &\in \overline{\mathcal{B}}_\infty(a) \end{aligned}$$

with $|\beta| < M$. Suppose $r\left(v_\infty - g_\infty(a, \overline{b})\right) \leq -\gamma$. Then there exists a $\eta_2 > 0$ such that $|\beta| > \eta_2$. Thus, $\lim_{X\to\infty} \left|f(\bar{b}, X, \beta)\right| = r\ |\beta| > r\eta_2$.

Note $\lim_{X\to\infty} d(a, \bar{b}, X) = r\left[g(a, \bar{b}) - g(a^N, \bar{b}^N)\right]$. Then there exists an X_1 such that for $X > X_1$, $\left|d(a, \bar{b}, X) - r\left[g(a, \bar{b}) - g(a^N, \bar{b}^N)\right]\right| < \varepsilon/2$. Consider the set $\Phi = \left(a, \bar{b}, X, \beta\right)$ that satisfy the condition for sequential rationality, with $|\beta| \leq M$ and $|X| \geq X_1$, and $d(a, \bar{b}, X) \leq -\varepsilon$. For $X > X_1$, $\left|d(a, \bar{b}, X) - r\left[g(a, \bar{b}) - g(a^N, \bar{b}^N)\right]\right| < \varepsilon/2$. Then $r\left(v_\infty - g_\infty(a, \bar{b})\right) \leq -\varepsilon/2$. Then there exists a η_2 such that $|\beta| > \eta_2$. Thus, $\lim_{X\to\infty} \left|f(\bar{b}, X, \beta)\right| = r\ |\beta| > r\eta_2$. Then there exists an X_2 such that for $X > X_2$, $\left|f(\bar{b}, X, \beta) - r\beta\right| < r\eta_2/2$. Then $\left|f(\bar{b}, X, \beta)\right| > r\eta_2/2 := \delta_2$. Take $X^* = \max\{X_1, X_2\}$. Then on the set

$$\left(a, \bar{b}, X, \beta\right) \in \{A \times B \times \Xi \times R : |\beta| \leq M \text{ and } |X| > X^*\}$$

if $d(a_t, \bar{b}_t, X_t) \leq -\varepsilon$ then $\left|f(\bar{b}, X, \beta)\right| > \delta_2$.

Step 2c: Show this is true for

$$\left(a, \bar{b}, X, \beta\right) \in \{A \times B \times \Xi \times R : |\beta| \leq M \text{ and } |X| \leq X^*\}$$

Consider the set $\Phi = \left(a, \bar{b}, X, \beta\right)$ that satisfy the condition for sequential rationality, with $|\beta| \leq M$ and $|X| \leq X^*$, and $d(a, \bar{b}, X) \leq -\varepsilon$. The function d is continuous and $\{A \times B \times \Xi \times R : |\beta| \leq M \text{ and } |X| \leq X^*\}$ is compact, so Φ is compact. f is also continuous, and therefore achieves a minimum on Φ.This minimum $\eta_1 > 0$ since $d(a_t, \bar{b}_t, X_t) < -\varepsilon$. Thus, $\left|f(\bar{b}, X, \beta)\right| > \eta_1$ for all $\left(a, \bar{b}, X, \beta\right) \in \Phi$.

Take $\delta = \min\{\eta_1, \delta_2, m\}$. Then when $d(a, \bar{b}, X) \leq -\varepsilon$, $\left|f(\bar{b}, X, \beta)\right| > \delta$.

The proof for a bounded state space is analogous, omitting step 2b.

Q.E.D.

This lemma implies that whenever the drift of D_t is less than $rD_t - \varepsilon$, the volatility is greater than δ. Take $\varepsilon = rD_0/4$ and suppose $D_t \geq D_0/2$. Then whenever the drift is less than $rD_t - \varepsilon > rD_0/2 - rD_0/4 = rD_0/4 > 0$, there exists a δ

such that $\left|f(\bar{b}, X, \beta)\right| > \delta$. Thus, whenever $D_t \geq D_0/2 > 0$, it has either positive drift or positive volatility, and grows arbitrarily large with positive probability. This is a contradiction, since D_t is the difference of two bounded processes. Thus, it cannot be that $D_0 > 0$. Likewise, it is not possible to have $D_0 < 0$. Thus, in any PPE with continuation values $(W_t)_{t\geq 0}$, it must be the case that $W_t = U(X_t)$ for all t. Therefore, it must be that $\left|f(\bar{b}, X, \beta)\right| = 0$, and actions are uniquely specified by $S^*(X, U'(X)f_1(\bar{b}_t, X_t))$.

Q.E.D.

2.6.4 Proofs: Equilibrium Payoffs

Theorem 16. *The highest PPE payoff across all states is bounded above by the highest static Nash equilibrium payoff across states, $\overline{W} \leq \overline{v}^*$ and the lowest PPE payoff across all states is bounded below by the lowest static Nash equilibrium payoff across states $\underline{W} \geq \underline{v}^*$.*

Let $\overline{W} = \sup_{\Xi} U(X)$ be the highest PPE payoff across all states. Suppose $\overline{W} = U(X)$ occurs at an interior point. Then $U'(X) = 0$ and $U''(X) \leq 0$. From the optimality equation,

$$U''(X) = \frac{2r\left[\overline{W} - v(X)\right]}{f_1(\bar{b}, X)^2\sigma_Y^2 + \sigma_X^2(X)} \leq 0$$

and therefore $\overline{W} \leq v(X) \leq \overline{v}^*$. Suppose the state space is bounded and $\overline{W} = U(X)$ occurs at an endpoint. Suppose, without loss of generality, that $\overline{W}$ occurs at $\overline{X}$. Then by the boundary conditions, $\overline{W} = v(\overline{X}) \leq \overline{v}^*$. Suppose the state space is unbounded and there is no interior maximum with $U(X) = \overline{W}$. Then $U(X)$ must converge to $\overline{W}$ at either ∞ or $-\infty$. Suppose $\lim_{X\to\infty} U(X) = \overline{W}$. Then $\overline{W} = v_\infty \leq \overline{v}^*$. The proof for $\underline{W} \geq \underline{v}^*$ is analogous.

Q.E.D.

Theorem 17. *Assume 1, 2, 3 and 4. The following properties show how PPE payoffs change with the state variable:*

1. *Suppose $v(X)$ is increasing (decreasing) in X.Then $U(X)$ is also increasing (decreasing) in X. The state that yields the highest static Nash payoff also yields the highest PPE payoff; likewise, the state that yields the lowest static Nash payoff also yields the lowest PPE payoff.*

2. *Suppose $v(X)$ has a unique interior maximum X^*, and v is monotonically increasing (decreasing) for $X < X^*$ $(X > X^*)$. Then $U(X)$ has a unique interior maximum at X^*,and U is monotonically increasing (decreasing) for $X < X^*$ $(X > X^*)$. The state that yields the highest static Nash payoff also yields the highest PPE payoff, whereas and the state that yields the lowest PPE payoff is a boundary point.*

3. *Suppose $v(X)$ has a unique interior minimum X^*, and v is monotonically decreasing (increasing) for $X < X^*$ $(X > X^*)$. Then $U(X)$ has a unique interior minimum at X^*,and U is monotonically decreasing (increasing) for $X < X^*$ $(X > X^*)$. The state that yields the lowest static Nash payoff also yields the lowest PPE payoff, whereas and the state that yields the highest PPE payoff is a boundary point.*

1. Suppose $v(X)$ is increasing in X, but $U(X)$ is not increasing in X. Thus, $U'(X) < 0$ for some $X \in \Xi$. Let $(X_1, X_2) \subset \Xi$ be a maximal subinterval such that $U'(X) < 0$ for all $X \in (X_1, X_2)$. Note $\lim_{X\to-\infty} U(X) = \underline{v}^* \leq \lim_{X\to\infty} U(X) = \overline{v}^*$ since v is increasing in X, so $U'(X)$ is not strictly decreasing on Ξ. Without loss of generality, assume $U(X)$ is increasing on $(-\infty, X_1)$. Then X_1 is an interior local maximum with $U'(X_1) = 0$ and $U''(X_1) \leq 0$. Then

by the optimality equation,

$$U''(X_1) = \frac{2r}{\left[f_1(\bar{b}, X_1)^2\sigma_Y^2 + \sigma_X^2(\bar{b}, X_1)\right]}\left(U(X_1) - v(X_1)\right) \leq 0$$

which implies $U(X_1) \leq v(X_1)$. Then

$$\lim_{X \to X_2} U(X_2) < U(X_1) \leq v(X_1) \leq \overline{v}^* = \lim_{X \to \infty} U(X)$$

Thus, since $U(X_2) < U(X_1)$ by definition, it must be that $X_2 < \infty$ and X_2 is a local minimum with $U'(X_2) = 0$ and $U''(X_2) \leq 0$. Then by the optimality equation,

$$\begin{aligned} U''(X_2) &= \frac{2r}{\left[f_1(\bar{b}, X_2)^2\sigma_Y^2 + \sigma_X^2(\bar{b}, X_2)\right]}\left(U(X_2) - v(X_2)\right) \\ &\leq \frac{2r}{\left[f_1(\bar{b}, X_2)^2\sigma_Y^2 + \sigma_X^2(\bar{b}, X_2)\right]}\left(U(X_2) - v(X_1)\right) < 0 \end{aligned}$$

which implies X_2 is a local maximum. This is a contradiction. The proof for $U(X)$ decreasing is analogous.

2. If $v(X)$ has a unique interior maximum $\widehat{X}$ such that $v'(X) > 0$ for $X < \widehat{X}$ and $v'(X) < 0$ for $X > \widehat{X}$. Assume $U'(X) < 0$ for some $X < \widehat{X}$. Let $(X_1, X_2) \subset (-\infty, \widehat{X})$ be a maximal subinterval such that $U'(X) < 0$ for all $X \in (X_1, X_2)$. First suppose $X_1 > \infty$ and $X_2 < \widehat{X}$. Then X_1 is a local maximum with $U'(X_1) = 0$ and $U''(X_1) \leq 0$, and by the optimality equation, $U(X_1) \leq v(X_1)$. Also, X_2 is a local minimum with $U'(X_2) = 0$ and $U''(X_2) \geq 0$, and by the optimality equation, $U(X_2) \geq v(X_2)$. This implies:

$$U(X_1) \leq v(X_1) \leq v(X_2) \leq U(X_2)$$

which is a contradiction. Thus, (X_1, X_2) must include a boundary point of $(-\infty, \widehat{X})$. Next suppose U is decreasing over $(-\infty, X_2)$ and $X_2 < \widehat{X}$. Thus, X_2 is a local maximum. Given that $\lim_{X\to-\infty} U(X) = \lim_{X\to-\infty} v(X)$, this implies:

$$\lim_{X\to-\infty} U(X) = \lim_{X\to-\infty} v(X) \leq v(X_2) \leq U(X_2)$$

which is a contradiction. Thus, it can only be that $(X_1, X_2) = (-\infty, \widehat{X})$. Likewise, if $U'(X) > 0$ for some $X > \widehat{X}$, then it must be the case that any maximal subinterval (X_1, X_2) on which $U'(X) > 0$ is $(X_1, X_2) = (\widehat{X}, \infty)$.

Suppose $U'(X) < 0$ on $(-\infty, \widehat{X})$ and $U'(X) > 0$ on $(X_1, X_2) = (\widehat{X}, \infty)$. Then $\widehat{X}$ is a global minimum, which implies:

$$U(X) > U(\widehat{X}) \geq v(\widehat{X}) = \overline{v}^*$$

But this is a contradiction, since $U(X) \leq \overline{v}^*$ for all X.

3. The proof is analogous to part 2.

Q.E.D.

2.7 Figures

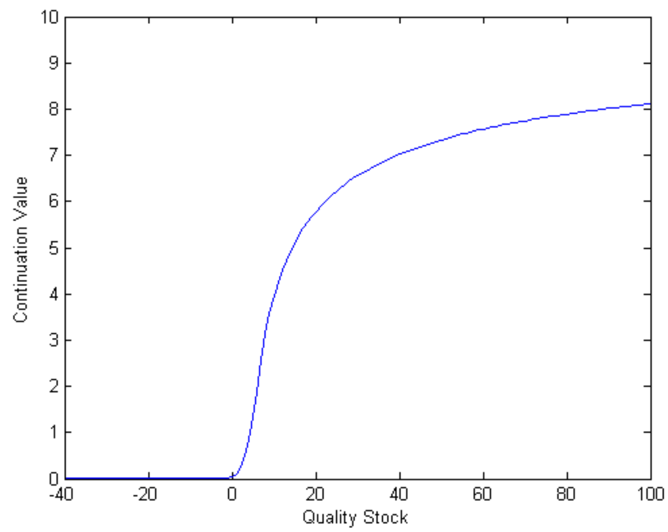

Figure 2.1: Equilibrium Payoffs

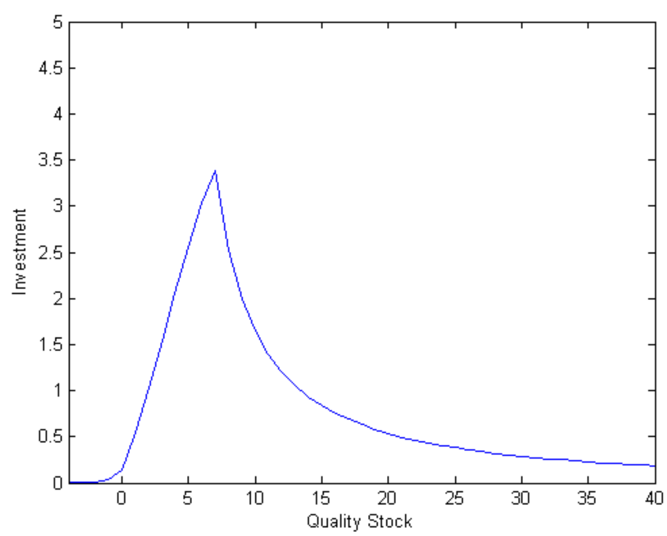

Figure 2.2: Firm Equilibrium Behavior

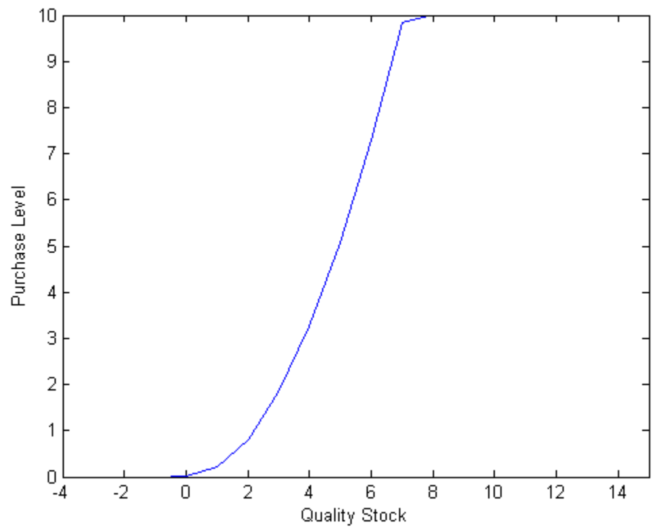

Figure 2.3: Consumer Equilibrium Behavior

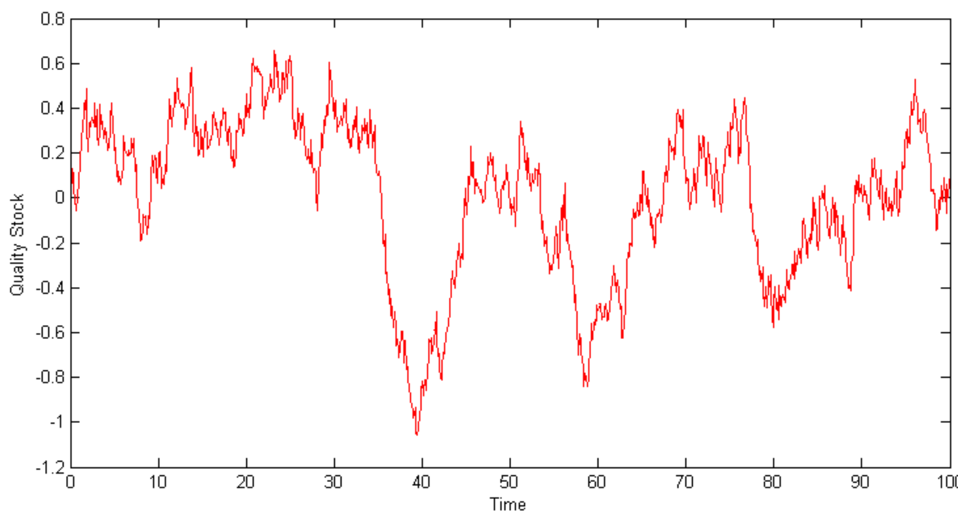

Figure 2.4: Product Quality Cycles

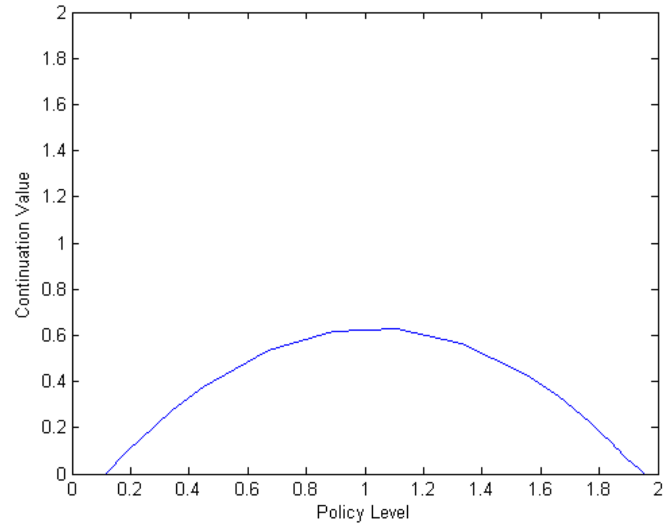

Figure 2.5: Equilibrium Payoffs

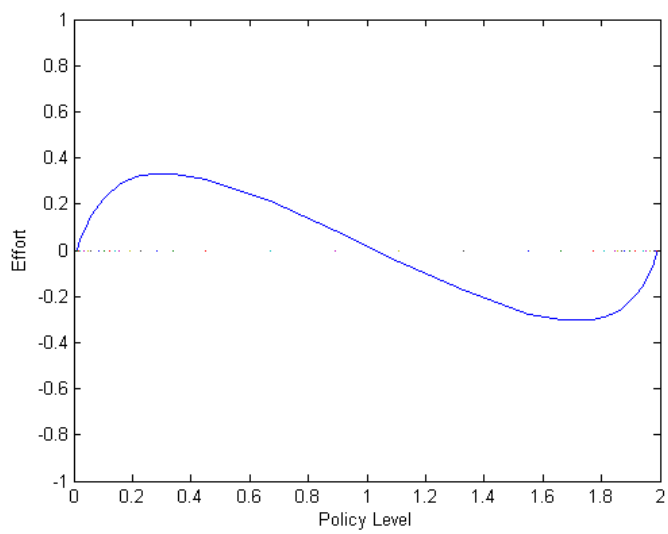

Figure 2.6: Government Equilibrium Behavior

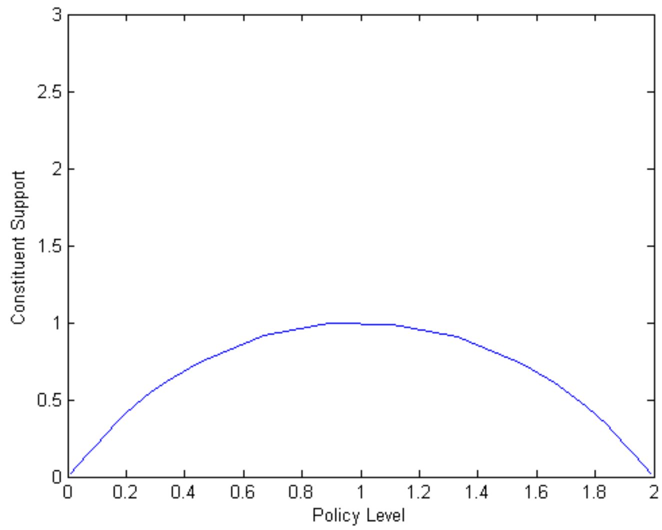

Figure 2.7: Constituent Equilibrium Behavior

Chapter 3

Incentives in Crowdsourcing

3.1 Introduction

A clothing manufacturer wants to create an online database tagging each individual attribute of its products; a public website must review and remove objectionable postings; a secretarial service needs to transcribe the recorded minutes from a company meeting. Traditionally, these tasks were performed by conventional in-house employees or outsourced workers tied to the firm via a long-term contract. Crowdsourcing offers a dynamic and flexible alternative to complete these tasks.

Crowdsourcing is the process of delegating work to an undefined group of people (a crowd) through an open call. Workers answering the call are not presumed to have any relationship with the crowdsourcing firm either preceding or following the short-lived job. Coined officially in 2006, the term crowdsourcing has become a portmanteau for many types of related sourcing protocols spanning intrinsic to extrinsic motivations. We focus on the problem posed to firms facing exclusively financially-motivated workers or, equivalently, offering work lacking non-pecuniary benefits.

The environment we consider is plagued by unobservable effort. The costly exertion of effort cannot be contractually specified and, since there is no expectation of future employment with the firm, traditional reputation mechanisms have no bite. The firm must guard against shirking, but how? The key insight here is that the firm hires multiple workers and conditions wage payment on a comparison of the output they produce. Workers are paid only when their output matches. This mechanism is wasteful compared to the observable effort benchmark: not only must additional workers be hired, but each must be paid a wage greater than their cost of effort.

We study the problem from a principal-agent perspective in which a firm offers agents a contract to induce workers to exert costly effort. A contract consists of an assignment of tasks to be completed, a protocol specifying how many workers are to be employed, a verification technology to review the agent's output, and a payment schedule. Our problem differs from standard principal-agent settings in that the firm cannot draw probabilistic inferences about an agent's effort on the basis of any observed data.[1]

Our first result is that the firm must hire a second worker with nontrivial probability. Although virtual monitoring is feasible when there is no limit on the size of potential jobs, the unboundedly large wage payments required to enforce effort with virtual monitoring preclude its optimality. Next, the firm must decide how to organize the workers. The firm has the option of informing a worker that his job is to verify the output of another worker. Doing so reduces the informed worker's incentive to shirk since, effectively, the uninformed worker is serving as his monitor. An uninformed agent then knows his output may not be checked and the wage required to prevent him from shirking is higher. We show that the firm optimally treats workers symmetrically instead of creating a hierarchy in which there are informed and uninformed agents.

As the number of tasks assigned to an agent increases, the firm reduces its monitoring probability. This leads to fewer workers being hired but each is paid a higher wage. For a given job size, the effect of decreasing the monitoring probability on the wage bill is non-monotonic. Along the optimal path, however, the reduction in the

[1]For example, the seller of a house can form beliefs about how hard his listing agent worked from the quantity and quality of received offers.

expected number of workers employed outweighs the increased per-worker wage and the firm's wage bill decreases monotonically (along with the monitoring probability) as jobs grow in size.

We consider two environmental conditions. The analysis begins with a firm hiring all workers simultaneously at the outset. Next we suppose the firm may instead move sequentially and hire additional workers conditional on the output of existing workers. Here a surprising result obtains. Suppose a particular state of the world is more likely than another. Then a worker who shirks hopes to match his fellow worker's output by following the prior. The firm responds by monitoring this output *less*: if the firm were to hire an additional worker with a higher probability when the output confirms the prior, the second agent's incentive to work evaporates since his posterior belief over the first agent's output is even more pronounced. We treat the two settings – hiring all workers at once or sequentially – separately since they impose quite different burdens upon the firm.

The paper proceeds as follows. After summarizing the related literature, we begin with a brief overview of the crowdsourcing marketplace. Next, we introduce our model, characterize the feasible set of contracts and develop optimal implementation plans. The model we present captures the salient features of the crowdsourcing environment and enables us to best communicate our main insights. We then extend our model to more general settings and examine the robustness of our findings. These theoretical results are followed by a discussion about implementing the plans in the current marketplace and some avenues for future research. The final section concludes.

3.1.1 Related Literature

Our model is similar to Rahman (2012). Rahman (2012) characterizes feasible contracts, while we are interested in optimal contracts. Rahman (2012) paper establishes when ”virtual monitoring” is feasible (in the context of our model, hiring a second agent with probability epsilon). While virtual monitoring is always feasible in our model, we establish that it is generally not optimal since it requires much larger wage payments. Gershkov and Szentes (2009) characterize what information is optimal to hide from agents in a common values setting with costly signals. Similar to the contract we consider with uninformed or informed monitoring positions, Gershkov and Szentes (2009) find it optimal to keep an agent's position unknown.

This paper also relates more broadly to the auditing literature. Townsend (1979) and Gale and Hellwig (1983) seek to characterize the optimal mechanism in a model with costly state verification. These models assume that one agent privately observes the state of the world, while it is costly for the second agent (an auditor) to observe the state.

3.2 Marketplace Overview

Of the dozens of work exchanges where firms (known as *requesters*) can hire workers (*providers*), Amazon's Mechanical Turk (AMT) is the most prominent. Created in-house in 2005 to find duplicates among the company's product webpages, the service rapidly expanded and by 2007 comprised a pool of more than 100,000 workers in over 100 countries completing various types of tasks (*New York Times*, artificial intelligence article). Quickly, the platform was found to be useful for tasks like transcribing podcasts, rating and tagging images, and writing/rewriting sentences. Wage payments typically range from one cent to $10 per task.

The current paid crowdsourcing market has grown considerably since AMT's founding. A sampling of 10 work exchanges that publish statistics tallies well over 2 million registered workers and gross payments nearing a billion dollars (Frei, 2009). The revenues of vendors connecting firms with workers are estimated at 500 million dollars per year (Frei, 2009).

Though the marketplace is rapidly evolving, we have recent data on the com-position of the crowdsourcing workforce from a paid survey conducted on AMT in February 2010.[2] Workers report 68 different countries of origin with the United States being most prevalent at 45% followed by India at 34%. Young workers are overrepresented, even when compared to the general population of Internet users. (The skew towards youth is even more pronounced for workers from India.) Self-reported education levels are also greater than those of the general populations.

Most workers spend a day or less working on AMT, completing 20-100 jobs and earning \$20 or less per week. There exists a high-end of the income distribution with workers earning more than \$1,000 per month. More than 20% of Indian workers report AMT as their primary source of income (10% of American workers), with an additional 35% using AMT as a secondary source of income (60% for American workers). Correspondingly, women are roughly twice as common as men among American workers while the reverse holds for the Indian subpopulation. Two-third of American workers reported incomes below \$60,000 (compared to 45% for the US Internet population) and more than 55% of Indian workers claim incomes below \$10,000.[3] Finally, the primary motivation for working on AMT is to earn cash while spending free time fruitfully (60% of American workers and 70% of Indian workers). Almost all of the remaining workers participate because the tasks are fun while very few workers use AMT to simply kill time.

Field experiments carried out on AMT suggest workers respond to economic incentives in a predictable fashion. Rational choice theory is borne out by findings that workers complete more tasks as the wage increases or the task difficulty decreases (Mason and Watts, 2009). The same experiments also test whether subjects reveal an expectation of being paid regardless of their performance. Explicitly informing workers that the accuracy of their responses is being measured and used to determine whether payment for their work would be provided has no discernible effect on either

[2] The data we report here are from a posting on March 9, 2010 from the blog "Behind Enemy Lines" (http://behind-the-enemy-lines.blogspot.com/2010/03/new-demographics-of-mechanical-turk.html).

[3] Less than a third of workers are unemployed or hold only a part-time job.

the quality or quantity of output: participants appear to treat their pay as necessarily performance dependent (Mason and Watts, 2009). One main departure from the rational choice model is that many workers resemble threshold earners (Horton and Chilton, 2010). These workers set an earnings target and appear otherwise unresponsive to financial incentives.

3.3 Model

A firm with access to an infinite stream of independent, identically structured *tasks* wants to hire agents to complete the work. For example, as new images are uploaded to a social networking or auction site each day the site needs to moderate the content of the images. Or, an online retailer wants to categorize attributes of its products. Individual tasks are bundled together to form a job. We first describe the stage game for a task and then model the employment contract for a job.

Each task is characterized by an unknown *state* of the world that the firm seeks to discover. A state can be thought of as whether an image is objectionable, a product bears a certain attribute, or a sentence is translated faithfully. The realized state for task j is denoted as $\omega_j \in \Omega$. The state for each task is independently and identically distributed. All agents share a common *prior* belief over Ω. To simplify exposition we consider a binary state space. This restriction is without loss of generality assuming the firm can divide non-binary tasks into binary constituent components. Let $\Omega = \{y, n\}$ and denote the prior belief the state reflects an innocuous image as $\pi = Pr(\omega_j = n) \in (1/2, 1)$.

A worker i hired to complete task j chooses whether to exert low or high *effort*, $e^i_j \in \{0, 1\}$. High effort perfectly reveals the task's state to the worker and costs c, while low effort yields no information about the state and is costless. Workers' effort choices are unobservable by the firm. After choosing an effort level and potentially learning the state, worker i sends a *message* $m^i_j \in \mathcal{M}$ to the firm about the state for task j. We will often refer to a worker's message(s) as his output. As messages acquire meaning only in equilibrium, the message space corresponds to the state space: $\mathcal{M} = \Omega$.

The firm must decide on a labor contract for each task. An *implementation plan*, P_j, for task j is an algorithm specifying a probability distribution over the number of workers to hire for each history of the game. We distinguish two types of implementation plans. A *simultaneous implementation plan* selects (perhaps randomly) how many workers to hire at the beginning of the task, whereas a *sequential implementation plan* specifies the probability of hiring an additional worker conditional on the current history.[4] More formally, let $H^k \equiv \Pi^k \mathcal{M}$ represent the space of message profiles when k agents have completed task j. Define $H \equiv \bigcup_{k=0}^{\infty} H^k$ as the set of all possible message profiles for task j with h as an element of H.[5] Let $F_X = \{f_X\}$ represent the set of all probability measures over discrete support X, with representative element f_X.[6] In the simultaneous case, $X = \mathbb{N}$ and for sequential implementation plans $X = \{0, 1\}$. When acting simultaneously, the firm chooses a probability measure specifying the distribution over the number of workers to hire at the beginning of the task. When acting sequentially, the firm chooses a sequence of probability measures specifying the probability of hiring an additional worker for a given realized history h. A simultaneous implementation plan is then represented as $P_j : H_0 \to F_{\mathbb{N}}$ and a sequential plan as $P_j : H \to F_{\{0,1\}}$.

An additional piece of notation will prove helpful. Denote as $\Delta(q)$ the probability measure over $\mathbb{N}$ that places probability $1 - q$ on $N = 1$, q on $N = 2$, and zero

---

[4]As the subsequent analysis shows, a simultaneous implementation plan can be viewed as an optimal sequential implementation plan subject to additional restrictions. We treat the two types of implementation plans separately to better connect our theoretical findings with the actual crowdsourcing marketplace. Implementing a sequential plan is informationally more intensive and operationally more complicated than implementing a simultaneous plan.

[5]We represent the history for a task before any messages have been received as H_0.

[6]The probability measure assigning all mass to x is denoted $\delta(x)$.

everywhere else.[7] Thus,

$$\Delta(q) = f_{\mathbb{N}}(N) = \begin{cases} N = 1, & \text{with probability } 1-q; \\ N = 2, & \text{with probability } q; \\ N \neq \{1,2\}, & \text{with probability } 0. \end{cases}$$

Let N_j represent the realized number of workers hired to complete task j and $m_j = (m_j^1, ..., m_j^{N_j})$ the realized profile of messages. Upon observing this message profile, the firm takes an *action* $A_j : \{m_j\} \to \mathcal{A}$. As the firm wants to match the state, let $\mathcal{A} = \Omega$.

Multiple tasks are bundled together as a *job*. We assume technological constraints prevent the firm from assigning arbitrarily many tasks to a worker. Each job is limited to $J \leq \overline{J}$ tasks.

The firm forms a *contract* to offer to prospective workers consisting of an implementation plan for each task in the job and a schedule of transfers. The *transfer* for worker i, represented as $t^i : m \to R$, is conditional on the message profile for all tasks in the job, $m = (m_1, ..., m_J)$. For clarity of exposition, we normalize the transfer to consider a per-task wage payment of $w^i(m) = t^i(m)/J$.

Given a contract $< J; \{P_j\}_{j=1}^J; w^i(m) >$, we can now specify payoffs. A worker's utility is independent of the state: his payoffs depend solely on any transfers received from the firm and the cost of his exerted effort. The utility of worker i on a given job is represented as:

$$u^i(m, e^i) = J \cdot w^i(m) - c \cdot \sum_{j=1}^{J} e_j^i$$

where $e^i = (e_1^i, ..., e_J^i)$ represents the effort choices of the worker. Note that worker i's payoffs depend on the entire profile of messages from all workers.

[7] With this notation $\Delta(0) = \delta(1)$ the Dirac measure on $N = 1$.

The firm seeks to choose an action that matches the realized state of the world. It receives a payoff of 1 if successful and 0 otherwise. This specification of payoffs is inappropriate for many of the situations we consider. It is more reasonable for the firm's payoffs for matching the state to differ depending on the state: removing objectionable content likely yields larger payoffs for the firm than allowing permissible content (and likewise for failing to remove content). The specification we consider here has no substantive impact on our results since it only affects the firm's individual rationality constraint. We allow firm payoffs for success or failure to differ by state in 3.5. When the firm chooses $A_j = n$ it permits the image while $A_j = y$ means the firm disallows the posting.

The firm must pay each agent the specified transfer for their employment. Its payoffs per task are represented as:

$$u_j^F = 1_{A_j=\omega_j} - \sum_{i=1}^{N_j} w^i(m).$$

Firm payoffs for a job are simply the sum of payoffs from the tasks making up the job:

$$u^F(A, \overline{\omega}) = \sum_{j=1}^{J} u_j^F = \sum_{j=1}^{J} \left(1_{A_j=\omega_j} - \sum_{i=1}^{N_j} w^i(m)\right)$$

where $A = (A_1, ..., A_J)$ and $\overline{\omega} = (\omega_1, ..., \omega_J)$.

Before proceeding we must discuss a few points about the regulatory environment. We assume limited liability on the part of workers so that a firm selecting the wrong action on the advice of a worker cannot seek restitution. We also rule out the possibility of fraudulent behavior of the firm by adopting the notion of a strong intermediary or work exchange. The firm submits its contract as an algorithm to an intermediary, like AMT, to execute. The intermediary prevents the firm from seeing the workers' output, rejecting it so that no workers are compensated, and then using the output to inform its action choice anyway.[8] The job, implementation plan and payment schedule are executed by the intermediary. Worker messages are submitted to the intermediary, who reveals them to the firm only if workers are compensated according to the payment schedule.[9]

3.4 Analysis

We begin by establishing the contractible effort benchmark. When effort is contractible, the firm simply hires one worker and pays him the cost of his effort. The state of the world is fully revealed and the firm captures the entirety of the surplus from the transaction. Per-task payoffs are $u_j^F = 1 - c$ for the firm and $u_j^i = 0$ for the worker. The firm optimally bundles as many tasks as possible to form a job.

Theorem 18 (Contractible Effort Benchmark)**.** *When effort is contractible, the firm optimally offers contract* $< \overline{J}; \{\Delta(0)\}_{j=1}^{\overline{J}}; w(m) = c \ \forall m >$ *with the maximum possible job size, hiring one worker for each task and compensating him exactly for his effort. Worker payoffs are* $u^i = 0$ *while firm payoffs are* $u^F = \sum_{j=1}^{\overline{J}}(1-c) = \overline{J}(1-c)$.

Even when effort is contractible, the firm has the option to select its action blindly without hiring any worker to uncover and reveal the state. The firm prefers the contract from Theorem 18 when the expected payoffs from hiring the worker are greater than those from guessing the state: $1 - c \geq \pi$. Thus, with contractible effort the firm will hire a worker whenever $c \leq 1 - \pi$.

The contractible effort benchmark is unattainable when effort is noncontractible since an isolated worker's message reveals no information about his effort choice. As

[8] This is not an academic concern. Consider the situation of a firm soliciting architectural renderings for a proposed new building. Before deciding which – if any – submission to use as a blueprint, the firm observes all submissions. We are assuming the firm cannot reject all submissions as unsatisfactory and then base their new building off of one of the submissions anyway.

[9] Optimal implementation when the firm can behave fraudulently is the subject of ongoing companion research.

the firm cannot ascertain the worker's effort, the worker will shirk if presented with this contract. The firm must hire additional workers with positive probability in order to induce effort. The additional workers duplicate tasks assigned to the first worker. The firm then compares the output across all agents and generates incentives for high effort by conditioning payment upon matching output.

The firm must hire more than one worker, at least sometimes, in any high-effort equilibrium. We begin with two lemmas. The first establishes that the firm will never assign more than two workers to each task; the second shows the firm requires agents' output to match on all jointly assigned tasks.

Lemma 24 (No more than two workers in equilibrium)**.** *No more than two workers are assigned to any task in equilibrium.*

Proof. In a high-effort equilibrium, each agent is exerting effort and learning the state on every task he's assigned. Since there is positive probability that an agent's output is compared to that of another worker, the agents' message choices will report the realized state. Thus, the firm is learning the true state for each task. Assigning more than two agents to a task does not affect worker incentives to report truthfully and serves only to increase the expected wage bill. Q.E.D.

Lemma 25 (Output must match on all applicable tasks)**.** *In equilibrium the firm requires agents' output to match on all tasks assigned to multiple agents.*

Proof. Suppose agent k is exerting effort on all tasks he is assigned and that the firm requires i's output to match k's output on fraction $\gamma < 1$ of the L tasks they are jointly assigned. Without loss of generality we may assume the firm requires i to match k's output on an integer number of tasks: suppose γL is a non-integer. Then the firm is effectively requiring i to match k on some integer l of the L tasks jointly assigned, where l is the smallest integer greater than γL. Agent i will not exert effort on all tasks since exerting effort on $J - (L - l)$ tasks results in the same probability of being paid while reducing i's costs of working. Q.E.D.

Our theoretical framework distinguishes between simultaneous and sequential implementation plans. This distinction is not a priori artificial. As the subsequent analysis shows, an optimal simultaneous plan can be viewed as the optimal sequential plan subject to additional restrictions; however, this relationship does not hold out of equilibrium. In addition to being theoretically more accurate, we believe treating the cases distinctly is valuable since the practical considerations for implementing a sequential plan are more burdensome than those for implementing a simultaneous plan. We begin by considering simultaneous implementation plans.

3.4.1 Simultaneous Implementation Plans

Consider a J-task contract with a simultaneous implementation plan where two agents are hired for a task with probability $q \in (0,1]$, and one agent is hired with probability $1-q$. That is, consider the simultaneous implementation plan $\Delta(q)$. We first focus on the case in which the firm treats agents symmetrically instead of offering each agent a different contract.

The contract specifies a transfer of $w \cdot J$ to a worker if his message choice matches the other agent's message on all tasks assigned to two agents. Otherwise, he receives no transfer. If the per-task wage w is set high enough, exerting effort is incentive compatible for the worker. The following theorem characterizes the set of feasible symmetric contracts that implement high effort in equilibrium for J tasks using monitoring probability q.

Theorem 19 (Feasible Symmetric Simultaneous Implementation)**.** *Let $J \in \mathbb{N}$ and $q \in (0,1]$ be given. The contract*

$$< J; \{\Delta(q)\}_{j=1}^{J}; w(m) = \frac{c}{1-\left[\frac{1-q+2q\pi}{1+q}\right]^{J}} \textit{ if } m_j^1 = m_j^2 \forall j \textit{ s.t. } N_j = 2, w(m) = 0 \textit{ otherwise} >$$

induces a high-effort equilibrium for $c \leq \frac{1-\pi}{1+q}\left(1-\left[\frac{1-q+2q\pi}{1+q}\right]^{J}\right)$ *and yields expected per-task payoffs for the firm of*

$$u_j^F = 1-(1+q)w.$$

Proof. Suppose agent k is exerting effort on all tasks he's assigned. Let ξ be the number of tasks agent i exerts effort on. Agent i will exert effort on J tasks if

$$\begin{aligned} Pr(paid|\xi &= J)Jw - Jc \geq Pr(paid|\xi = n)Jw - nc \\ w &\geq \left(\frac{J-n}{J}\right)\frac{c}{Pr(paid|\xi = J) - Pr(paid|\xi = n)} \end{aligned}$$

$\forall n = 0, 1, ..., J-1.$

Agent i is paid with certainty if he exerts effort on all tasks: $Pr(paid|\xi = J) = 1$. When considering shirking, what matters for incentives is the probability the agent's output is checked. Since output is necessarily unchecked when a task is only assigned to one agent, a worker needs to calculate the probability the firm is employing an additional worker on any task he's assigned. While the probability that a second agent is hired is q, this is not the probability that a worker believes he is in a two-worker pool on any given task. Let m represent the number of workers who are assigned to the task. Then

$$\begin{aligned} Pr(m &= 1) = \frac{1-q}{1+q} \\ Pr(m &= 2) = \frac{2q}{1+q} \end{aligned}$$

and the probability a worker matches on an individual task when he shirks is

$$Pr(m=1) + Pr(m=2)\pi = \frac{1-q}{1+q} + \left(\frac{2q}{1+q}\right)\pi.$$

The probability an agent expects to be paid for a job when exerting effort on n of the tasks in the job is

$$Pr(paid|\xi = n) = \left[\frac{1-q+2q\pi}{1+q}\right]^{J-n}.$$

For any n, the offered wage must dissuade the worker from exerting effort on only n tasks. This means the wage must satisfy

$$w \geq \frac{J-n}{J}\frac{c}{1-\left[\frac{1-q+2q\pi}{1+q}\right]^{J-n}}.$$

Letting $v = \frac{1-q+2q\pi}{1+q}$ and $x = J - n$, this can be rewritten as $w \geq \frac{c}{J}\frac{x}{1-v^x}$. The right-hand side is increasing in x since the sign of $\frac{\partial}{\partial x}(\frac{x}{1-v^x})$ is given by $1 - v^x + v^x \ln(v^x)$, which is positive. Thus, the wage must be large enough to dissuade shirking on all tasks. So a high-effort equilibrium for given J and q requires $w \geq \frac{c}{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}$.

The firm selects the lowest wage able to induce the agent to exert effort:

$$w(q, J) = \frac{c}{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}$$

The firm is willing to offer this contract if it outperforms its expected payoffs from hiring no workers and selecting its action blindly. Thus, the firm will employ workers if

$$1-(1+q)\left(\frac{c}{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}\right) \geq \pi,$$

which is equivalent to $c \leq \frac{1-\pi}{1+q}\left(1-\left[\frac{1-q+2q\pi}{1+q}\right]^J\right)$. Q.E.D.

The requirement $c \leq \frac{1-\pi}{1+q}\left(1-\left[\frac{1-q+2q\pi}{1+q}\right]^J\right)$ implies there exists effort costs for which the firm would undertake an employment contract were effort contractible but not when effort is non-contractible. In particular, whenever

$$c \in [1-\pi, \frac{1-\pi}{1+q}\left(1-\left[\frac{1-q+2q\pi}{1+q}\right]^J\right))$$

the firm will not employ any workers despite it being efficient to do so were shirking not a concern.

The comparative statics on the wage required to induce effort, collected in co:statics, are immediate. The incentives to shirk are reduced as the job size, J, and monitoring probability, q, increase. As agents are required to complete more tasks per job or are monitored more frequently, it becomes more difficult for them to meet the criteria for payment when they shirk. Thus, the wage required to induce high effort is decreasing in the number of tasks bundled into a job and in the monitoring probability. Changes toward more pronounced priors have the opposite effect. As the prior becomes more skewed towards one state, agents are more likely to be paid when shirking and a higher wage is required to induce effort. Indeed, as $\pi \to 1$ the required wage to induce effort goes to infinity for each J.

Corollary 4 (Comparative Statics)**.** *As the monitoring probability q increases, the job size J increases, or the prior π decreases, a lower wage is required to induce an agent to exert effort. Moreover, as q increases, J increases or π decreases, the firm is able to induce agents with higher costs of effort to exert effort.*

Proof. Immediate from consideration of $w(q,J) = \frac{c}{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}$ and $c \leq \frac{1-\pi}{1+q}\left(1-\left[\frac{1-q+2q\pi}{1+q}\right]^J\right)$. Q.E.D.

As stated in the proof of 19, the monitoring probability q has a subtle effect upon a worker's probability of being paid when shirking. Consider a given task. A worker's probability of being paid when shirking depends on whether the task is assigned to a second agent as well (which happens with probability q). The worker is necessarily paid if he is the only worker assigned the task. For a worker asked to complete a task when the monitoring probability is q, the probability a worker is completing a task assigned to two workers is not q but, rather, $\frac{2q}{1+q}$.[10] What matters for incentives is the probability of being in a two-worker pool and not, simply, the probability a task is assigned to two workers. This will loom prominently in subsequent results.

Corollary 5 (Wage is convex in the monitoring probability)**.** *The wage $w(q, J)$ required to induce high-effort is convex in the monitoring probability q.*

Proof. With $w(q, J) = \frac{c}{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}$, $\frac{\partial}{\partial q}(w(q, J)) < 0$ and $\frac{\partial^2}{\partial q^2}(w(q, J)) > 0$. Q.E.D.

Theorem 19 shows that there exists an incentive compatible transfer scheme such that high effort is an equilibrium for any monitoring probability and job size. Rahman (2012) studies virtual monitoring contracts where a second agent is hired with arbitrarily small probability. According to Theorem 19, such contracts are feasible. However, the required wage to implement a virtual monitoring apparatus tends to infinity.

[Virtual monitoring requires unbounded wage payments] For job size $J < \infty$ virtual monitoring is feasible only with unboundedly large wage payments.

Proof. For fixed job size J, the wage required to induce effort is given by $w(q, J) = \frac{c}{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}$ for monitoring probability q. As $q \to 0$, $w(q, J) \to \infty$. Q.E.D.

We now seek to characterize the optimal implementation plan for the firm for a given job size. Lemmas 24 and 25 already constrain the determination of such a plan. Lemma 24 establishes that in a high-effort equilibrium the firm need only monitor a given worker with at most one additional worker per task. Requiring a worker to

[10] A concrete example helps clarify the distinction. Consider a job of $J = 10$ tasks with monitoring probability $q = .30$. In total, 13 workers are expected to be hired. On average, seven of the ten tasks will be assigned to only one agent. So 7 of the 13 worker-tasks have only one employee while 6 worker-tasks have two employees. The probability of being in a two-worker pool is $\frac{2q}{1+q} = \frac{6}{13} > \frac{3}{10} = q$.

match the output of more than one other worker has no bearing on the probability of the worker being paid when shirking and only increases the expected number of workers paid. Lemma 25 shows that the firm provides wage payment for a worker only if his message matches his partner's on all tasks assigned to two agents.

A higher monitoring probability results in a lower required wage to induce high effort, but it also leads to the firm paying more workers in expectation. Since the wage required to induce high-effort is convex in the monitoring probability. At low monitoring probabilities, the wage is very sensitive to the monitoring probability and a small increase in monitoring lowers the required wage significantly. At high monitoring probabilities, the required wage is not as sensitive. On the other hand, the expected number of agents paid by the firm always changes linearly with the monitoring probability. The next result shows that when a job consists of a single task, the firm should hire a second agent with probability one. For jobs with more than one task, the two effects trade-off and an interior monitoring probability is better for the firm.

[Probabilistic monitoring is optimal] The firm increases her payoff by making jobs as large possible ($J = \overline{J}$) and choosing $q^* \in (0, 1)$ for $\overline{J} > 1$ and $q^* = 1$ for $\overline{J} = 1$.

Proof. It is obvious that firm payoffs are increasing in J since wages are decreasing as the job size grows larger.

As for the monitoring probability, the firm's expected payoff per task is:

$$E[u_j^F] = 1 - (1+q)\left(\frac{c}{1 - \left[\frac{1-q+2q\pi}{1+q}\right]^J}\right).$$

The best contract for a given job size J maximizes this payoff with respect to

q. Denoting $v = \frac{1-q+2q\pi}{1+q}$, the first-order condition stipulates

$$c \cdot \frac{\big[-2J(\pi-1)v^J + v^J + (2\pi-1)q(v^J-1) - 1\big]}{[(2\pi-1)q+1][v^J-1]^2} = 0,$$

which requires

$$-2J(\pi-1)v^J + v^J + (2\pi-1)q(v^J-1) - 1 = 0. \tag{1}$$

Noting that $v = 1$ when $q = 0$ and $v = \pi$ when $q = 1$, it is clear that $q = 1$ satisfies this condition when $J = 1$. For $J > 1$, when $q = 0$ the left-hand side of the expression is greater than zero. It is less than zero when $q = 1$. To see this, note that the left-hand side reduces to $-2J\pi^{J+1} + 2J\pi^J + 2\pi^{J+1} - 2\pi$. So optimality requires $-J\pi^J + J\pi^{J-1} + \pi^J = 1$. The sum $-J\pi^J + J\pi^{J-1} + \pi^J$ is strictly increasing in π, equals unity at $\pi = 1$, and is less than unity for $\pi < 1$. Thus, the left-hand side of e:FOC is less than zero when $q = 1$. Since e:FOC is continuous in q, it is satisfied with equality at some interior q^* by application of the intermediate value theorem. Q.E.D.

The firm's payoffs are increasing in the job size J so the firm should set $J = \overline{J}$. Given $\overline{J}$, the firm trades off the benefits and costs of higher monitoring probabilities. For an idiosyncratic task – where a job can consist of only that task – the optimal monitoring probability is to necessarily hire two agents. Again, as Theorem 19 shows, interior monitoring probabilities – including virtual monitoring – are feasible here, but Theorem 3.4.1 shows they are not optimal.

When $\overline{J} > 1$, an interior monitoring probability is superior. As we now show, the firm's optimal monitoring probability is decreasing in $\overline{J}$. Further, the lemma establishes that if the firm is able to bundle arbitrarily large numbers of tasks together into a job, the monitoring probability approaches zero.

Lemma 26 (Optimal monitoring probability is decreasing in job size)**.** *The optimal monitoring probability q^* is decreasing in the job size J. Moreover, as the maximum* *job size $\overline{J}$ increases without bound, the optimal monitoring probability goes to zero.*

Proof. Denote $v = \frac{1-q+2q\pi}{1+q}$. From the proof of 3.4.1, the optimal monitoring probability q must satisfy

$$-2J(\pi-1)v^J + v^J + (2\pi-1)q(v^J-1) = 1,$$

which can be rearranged to yield

$$\frac{Jv^J}{1-v^J} - \frac{1-q+2q\pi}{2(1-\pi)} = 0. \tag{3.1}$$

The partial derivative of the left-hand side of e:FOC2 with respect to J is

$$\frac{v^J(-v^J + J\ln(v) + 1)}{(v^J-1)^2}$$

The sign of this derivative is governed by $1 - v^J + J\ln(v) \equiv 1 - x + \ln(x)$ where $x = v^J$. The sum $1 - x + \ln(x)$ is increasing over $x \in (0, 1]$, so it is increasing in v^J. Since $\lim_{x\to 1} 1 - x + \ln(x) = 0$, $1 - v^J + \ln(v^J)$ is negative. Thus, the derivative of the FOC with respect to J is negative. Since the derivative of the FOC with respect to q is negative, the optimal monitoring probability is decreasing in the job size.

Since $q^*(J)$ is decreasing and is bounded below, $\hat{q} = \lim_{J\to\infty} q^*(J)$ exists. To see that the limiting monitoring probability is zero, consider fixed $\pi < 1$ and suppose $\hat{q} > 0$. Define $v^*(J) = \frac{1-q^*(J)+2\pi q^*(J)}{1+q^*(J)}$ and since $\hat{q}$ exists, $\hat{v} = \lim_{J\to\infty} \frac{1-q^*(J)+2\pi q^*(J)}{1+q^*(J)}$ exists.

The first-order condition $-2J(\pi-1)v^*(J)^J + v^*(J)^J + (2\pi-1)q^*(J)(v^*(J)^J - 1) = 1$ must hold at each J, which means $\lim_{J\to\infty} Jv^*(J)^J$ must converge to a finite value, implying $\lim_{J\to\infty} v^*(J)^J = 0$.

Then, for J sufficiently large, $2J(1-\pi)v^*(J)^J - (2\pi-1)q^*(J) \approx 1$ and so $\lim_{J\to\infty} Jv^*(J)^J = \lim_{J\to\infty} \frac{1-q^*(J)+2\pi q^*(J)}{2(1-\pi)} = \lim_{J\to\infty} \frac{1+q^*(J)}{2(1-\pi)} v^*(J)$ and beginequation* J$\lim_{J\to\infty} v^*(J)^{J-1} = \lim_{J\to\infty} \frac{1+q^*(J)}{2(1-\pi)}$

Let $\varepsilon > 0$ be given. Since $v^*(J)$ is convergent, for J large $(J-1)v^*(J-1)^{J-1} > (J-1)v^*(J)^{J-1} - \frac{\varepsilon}{2}$. Likewise, since $Jv^*(J)^J$ is convergent, it is Cauchy convergent and there exists J sufficiently large so that $(J-1)v^*(J-1)^{J-1} - Jv^*(J)^J < \frac{\varepsilon}{2}$. Thus

$$\begin{aligned}
\frac{\varepsilon}{2} &> (J-1)v^*(J-1)^{J-1} - Jv^*(J)^J \\
&> (J-1)v^*(J)^{J-1} - \frac{\varepsilon}{2} - JV^*(J)^J \\
&= \frac{1+q^*(J)}{2(1-\pi)} - v^*(J)^{J-1} - Jv^*(J)^J - \frac{\varepsilon}{2} \\
&= \frac{1+q^*(J)}{2(1-\pi)} - v^*(J)^{J-1} - \frac{1+q^*(J)}{2(1-\pi)}v^*(J) - \frac{\varepsilon}{2} \\
&= \frac{1+q^*(J)}{2(1-\pi)}[1 - v^*(J)] - v^*(J)^{J-1} - \frac{\varepsilon}{2}.
\end{aligned}$$

So for J sufficiently large, $\varepsilon > \frac{1+q^*(J)}{2(1-\pi)}[1 - v^*(J)] - v^*(J)^{J-1}$. But since $v^*(J)^J \to 0$, this means $1 = \lim_{J\to\infty} v^*(J) = \lim_{J\to\infty} \frac{1-q^*(J)+2\pi q^*(J)}{1+q^*(J)}$. Recalling that $\pi < 1$ was given, this requires $q^*(J) \to 0$. This contradicts the maintained hypothesis that $\hat{q} > 0$ and the limiting monitoring probability is zero. Q.E.D.

Theorem 3.4.1 indicates that an interior monitoring probability is optimal for $J > 1$. The intuition here can be seen by considering the per-task wage bill as a function of the monitoring probability for a given job size. co:statics establishes that the wage required to induce effort is decreasing in the monitoring probability. An increase in the monitoring probability, however, increases the expected number of agents being paid. Decreasing the monitoring probability, on the other hand, pays fewer agents in expectation but each is compensated with a higher wage. 3.6 shows that for low monitoring probabilities the reduction in the wage associated with an increase in the monitoring probability outweighs the expected effect of paying more agents; as the monitoring probability increases the effect of the decreased wage is eventually outweighed by the greater expected number of agents being paid. For any $J > 1$, the per-task wage bill is U-shaped in the monitoring probability, at first decreasing and then increasing. At an optimum, the effects trade-off non-trivially and an interior monitoring probability is optimal.

Lemma 26 establishes that the optimal monitoring probability is decreasing in the job size. As is the case outside of the optimal monitoring probability, the effect on the expected wage bill is unclear. In addition to showing that the shape of the wage bill as a function of the monitoring probability for a given job size is non-monotonic, 3.6 also shows that the per-task wage bill at the optimum decreases monotonically in the job size. The firm clearly prefers to make jobs as large as possible.

Rahman (2012) studies virtual monitoring in a principal-agent framework. Although such an implementation plan is always feasible (with unboundedly large wage payments), as can be seen clearly in 3.4, it is optimal only in the limit of bundling arbitrarily many tasks into a job. Rahman (2012) can then be thought of as either studying one particular contract in the space of feasible contracts or the optimal implementation plan when jobs are arbitrarily large. The reason virtual monitoring is optimal for arbitrarily large job sizes is that as $\overline{J} \to \infty$, the wage required to induce high effort approaches the cost of effort c. Compared to the contractible effort benchmark, the per-task efficiency loss from unobservable effort decreases as jobs grow larger. However, it is only when jobs are unboundedly large that virtual monitoring is optimal and unobservable effort leads to no efficiency loss.

[Per-task efficiency loss compared to contractible effort benchmark goes to zero as job size increases] The per-task efficiency loss from non-contractible effort goes to zero as the maximum job size grows unboundedly large.

Proof. For any $q \neq 0$, $\frac{1-q+2\pi q}{1+q} < 1$ and $\left[\frac{1-q+2\pi q}{1+q}\right]^J \to 0$ as $J \to \infty$. The result follows since $w(q^*(J), J) \to c$ and $(1-c) - (1-(1+q^*(J))w(q*(J), J)) \to 0$. Q.E.D.

3.4.2 Hierarchy

Thus far, the firm has treated each agent it hires similarly (what we call symmetrically). This need not be the case. On a task in which two agents are hired, the firm could inform one of the agents that two workers have been assigned to the task. (The firm cannot inform both agents since, then, a worker would realize the absence of such notice implies they are the only worker assigned to the task and incentives for high effort would be impossible to generate.) The worker who is informed that two agents have been hired knows his probability of being paid when shirking is lower than in the symmetric case and a lower wage is required to induce him to exert effort. Conversely, the uninformed worker knows the probability he is the only worker hired for a task is greater than in the symmetric case and a higher wage is required for inducement. We now investigate whether creating such a hierarchy of informed and uninformed agents is good for the firm.

Suppose two agents are hired for a task with probability $q \in (0, 1]$, and one agent is hired with probability $1 - q$. The agents are offered transfers $w^1 \neq w^2$ and each must complete a job consisting of J tasks. The agents will be paid $w^i \cdot J$ if their output matches the other agent's output on all tasks assigned to two agents, otherwise the agent receives a transfer of zero. The following theorem characterizes the set of feasible contracts that implement high effort in equilibrium for a given number of tasks J and monitoring probability q when agents are presented with different information.

Theorem 20 (Feasible Asymmetric Simultaneous Implementation)**.** *Let* $J \in \mathbb{N}$ *and* $q \in (0, 1]$ *be given. When the principal offers contract*

$$J, \{\Delta(q)\}_{j=1}^{J}, (w^1, w^2) = \left(\frac{c}{1 - [1 - q + q\pi]^J}, \frac{c}{1 - \pi^J} \right) \textit{ if } m_j^1 = m_j^2 \forall\, j \textit{ s.t. } N_j = 2$$

there exists a high effort equilibrium for $c \leq (1 - \pi) \left[\frac{1}{1 - (1 - q + q\pi)^J} + \frac{q}{1 - \pi^J} \right]^{-1}$ *which* *yields expected per-task payoffs*

$$u_j^F = 1 - w^1 - qw^2$$

for the firm.

Proof. Now we need to consider separate incentive constraints for each worker. Denote by $i = 2$ the worker who is informed that he's in a two-agent pool; in our description above this is the second agent hired. Let $i = 1$ be the other worker who is uninformed about whether he is part of a one-worker or two-worker pool. First consider worker 1 and suppose worker 2 is exerting effort on all tasks that he's assigned. Let ξ^1 be the number of tasks that agent 1 exerts effort on. As before, agent 1 is paid with certainty if he exerts effort on all tasks:

$$P(paid|\xi^i = 1) = 1$$

If agent 1 shirks on some tasks, then his probability of being paid depends on the probability he is monitored on this task. Let m represent the number of workers who are assigned to the task with the current worker. Agent 1 believes his work will be checked with probability q:

$$\begin{aligned} Pr(m &= 1) = 1 - q \\ Pr(m &= 2) = q. \end{aligned}$$

The probability agent 1 is paid on a job when he exerts effort on n tasks is:

$$Pr(paid|\xi^i = n) = [1 - q + q\pi]^{J-n}$$

and, once again, the most profitable deviation for a worker is to exert effort on no tasks.

In an equilibrium with agents exerting effort, the principal chooses the lowest price that satisfies the worker's incentive constraint, and so the firm sets the wage so that the worker is just indifferent between exerting effort on all tasks and exerting effort on no tasks:

$$w^1 = \frac{c}{1 - [1 - q + q\pi]^J}.$$

By a similar analysis, the incentive compatible wage for worker 2 is:

$$w^2 = \frac{c}{1 - \pi^J}.$$

The firm prefers this implementation to guessing the state blindly if $1 - w^1 - qw^2 \geq \pi$. Q.E.D.

As before, the wages required to induce a high-effort equilibrium are increasing in the prior and decreasing in both the job size and the monitoring probability. The following corollary of Lemma (3) allows us to fully characterize the optimal contracts for the firm for a given job size.

Theorem 21 (Symmetric implementation outperforms asymmetric)**.** *Treating agents symmetrically is better for the firm than the asymmetric implementation in which some workers are informed and some are uninformed.*

Proof. First consider the case in which agents are treated symmetrically. If a second agent is hired for a task with probability q, each worker believes he is in a pool with another worker with probability $\frac{2q}{1+q}$. The optimal wage given J and q is $w = \frac{c}{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}$.

With monitoring probability q, in expectation $1+q$ agents are hired for a given task. Since with probability $1-q$ a second agent is not hired and with probability q a second agent is hired, a worker hired for a given task believes he is the first agent (in the parlance of asymmetric implementation) with probability $\frac{q}{1+q} + \frac{1-q}{1+q} = \frac{1}{1+q}$ and the second agent with probability $\frac{q}{1+q}$. (Of the fraction $\frac{2q}{1+q}$ of workers toiling in a two-worker pool, half can be treated as the first agent and half as the second agent hired.)

When evaluating a decision to exert effort or shirk, a worker must calculate the probability he obtains payment even if he shirks. This probability depends on whether the agent's output is going to be compared to that of another agent. The probability an agent's output is checked is given by

$$\begin{aligned} Pr(work\ checked) &= Pr(work\ checked|2nd\ agent)Pr(2nd\ agent) \\ &\quad +Pr(work\ checked|1st\ agent)Pr(1st\ agent) \\ &= 1 \cdot \frac{q}{1+q} + q \cdot \frac{1}{1+q} \\ &= \frac{2q}{1+q} \end{aligned}$$

Now consider the asymmetric situation in which the second agent is informed that he is the second agent on a given task. This agent knows his output is necessarily compared to that of another agent, and so $Pr(work\ checked) = 1$ and $w^2 = \frac{c}{1-\pi^J}$. The agent who is informed that he is the first agent hired on a given task has $Pr(work\ checked) = q$ and $w^1 = \frac{c}{1-[1-q+q\pi]^J}$.

Whether in the symmetric or asymmetric case, an agent's wage depends on the probability he is able to obtain payment when shirking: the general form of an agent's wage is $w = \frac{c}{Pr(paid|effort)-Pr(paid|shirking)}$. In a high-effort equilibrium, the probability an agent is paid when exerting effort is unity. We may rewrite the probability an agent expects payment when shirking as

$$\begin{aligned} Pr(paid|shirking) &= Pr(work\ checked)\pi + Pr(work\ unchecked) \\ &= Pr(work\ checked)\pi + 1 - Pr(work\ checked) \\ &= 1 - (1-\pi)Pr(work\ checked) \end{aligned}$$

If we denote $x = Pr(work\ checked)$, we may then rewrite the general form of an agent's wage as $w = \frac{c}{1-Pr(paid|shirking)} = \frac{c}{1-[1-(1-\pi)x]^J}$. The wage is convex in the probability an agent is paid when shirking: $\frac{\partial}{\partial x}(w) < 0$ and $\frac{\partial^2}{\partial x^2}(w) > 0$.

All that remains to be shown is that the probability an agent's work is checked in the asymmetric case is a mean-preserving spread of the probability an agent's work is checked in the symmetric case. We have already established that the probability an agent believes his output will be checked is $\frac{2q}{1+q}$ in the symmetric case. In the asymmetric case, fraction $\frac{1}{1+q}$ agents are first-agents and $\frac{q}{1+q}$ are second agents. First agents have their output checked with probability q while second agents necessarily have the output checked. Thus, the overall probability an agent's work is checked in the asymmetric case is also $\frac{2q}{1+q}$. Since the wage is convex in this probability, the asymmetric case results in a higher wage bill for the firm. Q.E.D.

We have now established the optimal contracts for the firm to implement a high-effort equilibrium. We state this as 22.

Theorem 22 (Optimal Simultaneous Implementation)**.** *When J tasks are available to bundle into a job, the firm should offer each of two agents the following contract:* $< \overline{J}; \{\Delta(q^*(\overline{J}))\}_{j=1}^{\overline{J}}; w(m) = \frac{c}{1-\left[\frac{1-q^*(\overline{J})+2q^*(\overline{J})\pi}{1+q^*(\overline{J})}\right]^{\overline{J}}}$ *if* $m_j^1 = m_j^2\ \forall j$ *s.t.* $N_j = 2, w(m) = 0$ *otherwise* $>$.

Proof. Follows from previous Theorems. Q.E.D.

3.4.3 Sequential Implementation Plan

Section 3.4.1 considered a firm choosing an implementation plan at the outset of the job. In particular, the firm does not change its behavior in light of the workers' output. This specification is appropriate for settings in which the firm is either unable or unwilling to observe worker output in real-time on a task-by-task basis. For many of the applications of crowdsourcing this is accurate. One of the attractions of the medium is that workers are able to complete tasks at their own discretion, free of the bounds of normal business hours or turn-around times. But this need not be the case, especially if firms are able delay compensating a worker upon observing his output.[11]

Now suppose the firm can condition the probability of hiring a second worker on the first worker's action. If $m_j^1 = y$, then hire a second worker with probability q_y and if $m_j^1 = n$, then hire a second worker with probability q_n. As we have already shown, hiring at most two workers is sufficient to induce a high-effort equilibrium. Thus, we can ignore consideration of sequential implementation plans calling for more than two workers. This means that the simultaneous plans we considered in 3.4.1 can be implemented as a sequential plan subject to the additional restriction that $q_n = q_y$. It is therefore immediate that the optimal sequential implementation plan is (weakly) superior to the optimal simultaneous implementation plan. This superiority, however, comes at the cost of more burdensome demands upon the firm implementing the plan.

We begin by establishing the space of feasible sequential implementation plans.

Theorem 23 (Feasible Sequential Implementation). *There exist a continuum of equilibria parameterized by (q_n, q_y) such that all workers exert effort. In equilibrium, workers are paid wage*

$$\begin{aligned} w_1(q_n, q_y) &= \frac{c}{\min\{q_n(1-\pi), q_y\pi\}} \\ w_2(q_n, q_y) &= \frac{c\left[\pi q_n + (1-\pi)q_y\right]}{\min\{q_y(1-\pi), q_n\pi\}} \end{aligned}$$

contingent upon matching actions when two workers are hired.

Proof. See Appendix. Q.E.D.

The prior favors action n, so a shirking agent 1 will choose $m^1 = n$. The firm must monitor action n more frequently to reduce the incentive to shirk for agent

[11]Then, in this case, the firm could utilize a sequential implementation plan by having one agent complete an entire job and using the agent's output to carry out the specified monitoring probabilities. All workers would then be compensated after the completion of all work, not just their own.

1. However, this comes at the cost of increasing the incentive to shirk for agent 2. Conditional on being hired, agent 2 knows $m^1 = n$ is very likely and therefore must be paid a high wage w^2 to incentively effort. In fact, in the optimal contract, the firm prefers to monitor message $m^1 = y$ more frequently. This contract reduces the incentives to shirk for agent 2 at the expense of increasing incentives to shirk for agent 1. The firm maximizes its payoffs by giving the monitor the least incentive to shirk, and paying the monitor a lower wage than the initial worker.

Theorem 24 (Optimal Sequential Implementation)**.** *Consider the class of contracts parameterized by sequential monitoring probabilities* (q_y, q_n)*. The optimal equilibria in this class is*

$$(q_n^*, q_y^*) = (q_n^*, 1)$$

where q_n^* *solves*

$$2\pi^2 q_n^3 + 2\pi(1-\pi)q_n^2 = 1$$

Proof. See Appendix. Q.E.D.

By the argument preceding the Theorem statement of 23 we immediately have the optimal implementation plan among all simultaneous and sequential plans.

Theorem 25 (Optimal Implementation)**.** *This sequential contract is optimal relative to any simultaneous contract with* $q = q_y = q_n$

3.5 Extensions

3.5.1 More General Firm Payoffs

In Section 3.4, the firm's payoff was 1 for matching the state and 0 for failing to match the state. Allowing the firm's payoffs for matching (and failing to match) the state to depend on the state changes little of the results of Section 3.4. Suppose the firm suffers greater disutility from permitting an objectionable image to be posted on its site compared to disallowing the upload of a harmless image. Let $\alpha_\omega \geq 0$ be the firm's payoff from matching the state when the state is ω and $\beta_\omega \leq 0$ be the firm's payoff from failing to match state ω. So α_n (β_n) corresponds to permitting (prohibiting) harmless content to be shared and α_y (β_y) represents removing (failing to remove) an objectionable item.

Firm optimization remains as before. The sole change to the results of Section 3.4 is to the firm's individual rationality constraint used to establish the space of feasible contracts. The final step in the proofs of Theorems 19 and 23 is to show the firm prefers the proposed contract to guessing the state blindly. With payoffs α_ω and β_ω, the firm's expected payoff from selecting $A_j = y$ ($A_j = n$) without hiring any workers is $(1-\pi)\alpha_y + \pi\beta_n$ ($\pi\alpha_n + (1-\pi)\beta_y$). The proposed contract must exceed both of these values, so the firm's simultaneous implementation individual rationality constraint becomes

$$\pi\alpha_n+(1-\pi)\alpha_y-(1+q)\left(\frac{c}{1-\left[\frac{1-q+2q\pi}{1+q}\right]^J}\right)\geq \max\{(1-\pi)\alpha_y+\pi\beta_n;\pi\alpha_n+(1-\pi)\beta_y\},$$

which is equivalent to $c \leq \frac{\min\{\pi(\alpha_n-\beta_n);(1-\pi)(\alpha_y-\beta_y)\}}{(1+q)}\left(1-\left[\frac{1-q+2q\pi}{1+q}\right]^J\right)$.

When $\min\{\pi(\alpha_n-\beta_n);(1-\pi)(\alpha_y-\beta_y)\} > 1-\pi$, the feasible space is larger than in 3.4. co:statics shows that as the prior flattens (π decreases), the firm is able to induce agents with higher costs of effort to exert effort. This remains the case when an uninformed firm follows the prior so that the default option is to allow questionable postings to be uploaded. Since the optimal monitoring probability did not depend on the payoffs from successfully matching the state, all other results remain unchanged.

3.5.2 More General State Spaces

Section 3.4 considers a binary state space $\Omega = \{y, n\}$ to make the exposition cleaner. First, for the applications most relevant to the analysis of our model, it is reasonable to presume tasks can be subdivided into binary choices: instead of asking workers to reveal whether a product is a "white, long-sleeve blouse," the firm can instead structure the tasks so that workers separately categorize whether the item is "white," "long-sleeve," and a "blouse."

Even without the possibility of such a subdivision process, the results presented in Section 3.4 continue to apply with more general state spaces. Let $\Omega = \{\omega^1, \omega^2, ...\}$ and redefine $\pi = \max_k Pr(\omega_j = \omega^k)$. From the firm's perspective, only the most likely state (in terms of having the highest prior probability) matters for inducing an agent to exert effort, and so all of the results in Section 3.4 carry-through unchanged. This would also remain the case with an uncountably infinite state space as long as there is a discrete jump in probability at some $\omega \in \Omega$. (The analysis is uninteresting when $Pr(\omega_j = \omega) = 0$ for all ω since a worker's probability of obtaining payment when shirking is zero.)

3.5.3 Imperfect Signals

The assumption of perfect signals limits the scope of the present paper. The main purpose of assuming perfect signals, however, is to separate inducing effort from learning as the firm's motivation for hiring multiple workers: when signals are noisy, there is a learning justification for the firm hiring multiple workers even when effort is contractible. In other words, with noisy signals the underlying signal structure induces an additional justification for hiring multiple agents, and this justification holds regardless of whether effort is contractible. With perfect signals, there is a wedge between the contractible effort and non-contractible effort settings and the learning justification for hiring multiple agents arises only when effort is non-contractible and the equilibrium specifies shirking.

3.6 Conclusion

Crowdsourcing is a new and non-traditional means of completing work. It offers the promise of a fast, flexible and scalable workforce available at a moment's notice. But it also presents some unique difficulties. Workers have no relationship with the firm for which they are working, so conventional reputation mechanisms have no bite. Effort is non-contractible, so the costly exertion of it must be achieved through incentives. We show that despite these pitfalls, effort can be induced and valuable information communicated. The mechanisms we design to achieve this rely upon hiring multiple workers to complete the same tasks and making payment conditional upon workers producing the same output. We consider two types of implementation technologies, studying both feasible and optimal contracts within each technology. The mechanisms we put forth are simple enough to be easily implemented by crowdsourcing firms yet they offer tangible improvements upon those currently used in practice.

I especially thank Troy Kravitz for his role as coauthor on this chapter. I also thank Nageeb Ali, David Miller, Joel Sobel, Joel Watson and participants of the UCSD theory seminar and the 2011 North American Summer Meeting of the Econometric Society.

3.7 Appendix

Firm chooses (q_L, q_R) to minimize:

$$\min_{q_L, q_R} \frac{1}{\min\left\{q_L\left(1-\pi\right), q_R\pi\right\}} + \frac{\left(\pi q_L + \left(1-\pi\right)q_R\right)^2}{\min\left\{\pi q_L, \left(1-\pi\right)q_R\right\}}$$

1. Suppose $q_R < \left(\frac{1-\pi}{\pi}\right) q_L$ ($\Rightarrow q_R < q_L$).

Best shirking action is $a_1 = R$ and $a_2 = L$

$$\begin{aligned} w_2 &= \frac{c\left(\pi q_L + (1-\pi)\, q_R\right)}{(1-\pi)\, q_R} \\ w_1 &= \frac{c}{q_R \pi} \end{aligned}$$

Firm:

$$\begin{aligned} & \min_{q_L, q_R} \frac{1}{q_R \pi} + \frac{\left(\pi q_L + (1-\pi)\, q_R\right)^2}{(1-\pi)\, q_R} \\ s.t. q_R &\in \left[0, \left(\frac{1-\pi}{\pi}\right) q_L\right] \text{ and } q_L \in [0,1] \\ \text{or } q_R &\in \left[0, \left(\frac{1-\pi}{\pi}\right)\right] \text{ and } q_L \in \left[\left(\frac{\pi}{1-\pi}\right) q_R, 1\right] \end{aligned}$$

Fix q_R. Optimal $q_L(q_R)$ is to minimize q_L, since q_L only appears in numerator. Therefore, to satisfy constraint,

$$q_L^*(q_R) = \left(\frac{\pi}{1-\pi}\right) q_R$$

and agent 1 is indifferent between L and R when shirking. Now find q_R^* :

$$\begin{aligned} & \min_{q_R} \frac{1}{q_R \pi} + \frac{\left(\pi^2 + (1-\pi)^2\right)^2}{(1-\pi)^3} q_R \\ FOC &: \frac{1}{\pi q_R^2} = \frac{\left(\pi^2 + (1-\pi)^2\right)^2}{(1-\pi)^3} \\ q_R^* &= \sqrt{\frac{(1-\pi)^3}{\pi\left(\pi^2 + (1-\pi)^2\right)^2}} \leq 1 \text{ (MATLAB)} \end{aligned}$$

2. Suppose $q_R \in \left[\left(\frac{1-\pi}{\pi}\right) q_L, \left(\frac{\pi}{1-\pi}\right) q_L\right]$

Best shirking action is $a_1 = L$ and $a_2 = L$

$$
\begin{aligned}
w_1 &= \frac{c}{q_L(1-\pi)} \\
w_2 &= \frac{c(\pi q_L + (1-\pi)q_R)}{(1-\pi)q_R}
\end{aligned}
$$

Firm:

$$
\begin{aligned}
& \min_{q_L, q_R} \frac{1}{q_L(1-\pi)} + \frac{(\pi q_L + (1-\pi)q_R)^2}{(1-\pi)q_R} \\
s.t. q_R \in & \left[\left(\frac{1-\pi}{\pi}\right)q_L, \min\left\{\left(\frac{\pi}{1-\pi}\right)q_L, 1\right\}\right] \text{ and } q_L \in [0,1]
\end{aligned}
$$

Boundary conditions:

- never need to worry about $q_L = 0$ or $q_R = 0$ since pushes term to infinity
- $q_L = 1, q_R = \left(\frac{1-\pi}{\pi}\right)$ (x)
- $q_L = 1, q_R = 1$ (x)
- $q_L = \left(\frac{1-\pi}{\pi}\right), q_R = 1$ (x)

Fix q_R and find $q_L^*(q_R)$

$$
\begin{aligned}
FOC &: \frac{1}{q_L^2(1-\pi)} = \frac{2\pi^2 q_L + 2\pi(1-\pi)q_R}{(1-\pi)q_R} \\
q_R &= 2\pi^2 q_L^3 + 2\pi(1-\pi)q_R q_L^2 \\
q_R &= \frac{2\pi^2 q_L^3}{[1 - 2\pi(1-\pi)q_L^2]} \\
q_L^*(q_R) &= \text{no closed form}
\end{aligned}
$$

Now find q_R^*

****

Fix q_L and find $q_R^*(q_L)$

$$\begin{aligned} FOC &: 2(1-\pi)^2(\pi q_L + (1-\pi) q_R) q_R = (\pi q_L + (1-\pi) q_R)^2 (1-\pi) \\ \frac{d}{dq_R} &< 0 \text{ for } q_R < \frac{\pi}{1-\pi} q_L \text{ so want to make } q_R \text{ as big as possible} \\ q_R^*(q_L) &= \min\left\{\left(\frac{\pi}{1-\pi}\right) q_L, 1\right\} \end{aligned}$$

Now find q_L^*

$$\begin{aligned} &\min_{q_R} \frac{1}{q_L(1-\pi)} + 4\pi q_L \\ FOC &: \frac{1}{q_L^2(1-\pi)} = 4\pi \\ q_L^* &= \frac{1}{2\sqrt{\pi(1-\pi)}} \geq 1 \end{aligned}$$

Boundary condition binds:

$$\begin{aligned} q_R^*(q_L) &= 1 \\ q_L^* &= \left(\frac{1-\pi}{\pi}\right) \end{aligned}$$

Firm profit:

$$\frac{\pi}{(1-\pi)^2} + 4(1-\pi)$$

Reoptimize for q_L^* at $q_R^* = 1$

$$\begin{aligned}
& \min_{q_L} \frac{1}{q_L(1-\pi)} + \frac{(\pi q_L + (1-\pi))^2}{(1-\pi)} \\
FOC \quad &: \quad \frac{1}{q_L^2(1-\pi)} = \frac{2\pi^2 q_L + 2\pi(1-\pi)}{(1-\pi)} \\
2\pi^2 q_L^3 + 2\pi(1-\pi)q_L^2 \quad &= \quad 1
\end{aligned}$$

Note $q_L = \left(\frac{1-\pi}{\pi}\right)$ does not solve this equation and neither does $q_L = 1$ (unless $\pi = 1/2$).

Reoptimize for q_R^* at $q_L^* = 1$

$$\begin{aligned}
& \min_{q_R} \frac{(\pi + (1-\pi)\, q_R)^2}{(1-\pi)\, q_R} \\
FOC : 2(1-\pi)^2 q_R (\pi + (1-\pi)\, q_R) \quad &= \quad (1-\pi)(\pi + (1-\pi)\, q_R)^2 \\
q_R^*(q_L^* \quad &= \quad 1) = \frac{\pi}{1-\pi} > q_L^*
\end{aligned}$$

3. Suppose $q_R > \left(\frac{\pi}{1-\pi}\right) q_L$

   Best shirking action is $a_1 = L$ and $a_2 = R$

$$\begin{aligned}
w_1 \quad &= \quad \frac{c}{q_L(1-\pi)} \\
w_2 \quad &= \quad \frac{c(\pi q_L + (1-\pi)\, q_R)}{\pi q_L}
\end{aligned}$$

Firm:

$$\begin{aligned}
& \min_{q_L, q_R} \frac{1}{q_L(1-\pi)} + \frac{(\pi q_L + (1-\pi) q_R)^2}{\pi q_L} \\
s.t. q_R & \in \left[\left(\frac{\pi}{1-\pi}\right) q_L, 1\right] \text{ and } q_L \in \left[0, \left(\frac{1-\pi}{\pi}\right)\right] \\
\text{or } q_R & \in [0,1] \text{ and } q_L \in \left[\left(\frac{1-\pi}{\pi}\right) q_R, 1\right]
\end{aligned}$$

Fix q_L. Optimal $q_R(q_L)$ is to minimize q_R, since q_R only appears in numerator. Therefore, to satisfy constraint,

$$q_R^*(q_L) = \left(\frac{\pi}{1-\pi}\right) q_L$$

and agent 1 is indifferent between L and R when shirking. Now find q_L^* :

$$\begin{aligned}
& \min_{q_L} \frac{1}{q_L(1-\pi)} + 4\pi q_L \\
FOC &: \frac{1}{(1-\pi) q_L^2} = 4\pi \\
q_L^* &= \frac{1}{2\sqrt{\pi(1-\pi)}} \geq 1
\end{aligned}$$

Boundary condition binds

$$\begin{aligned}
q_R^* &= 1 \\
q_L^* &= \left(\frac{1-\pi}{\pi}\right)
\end{aligned}$$

3.8 Figures

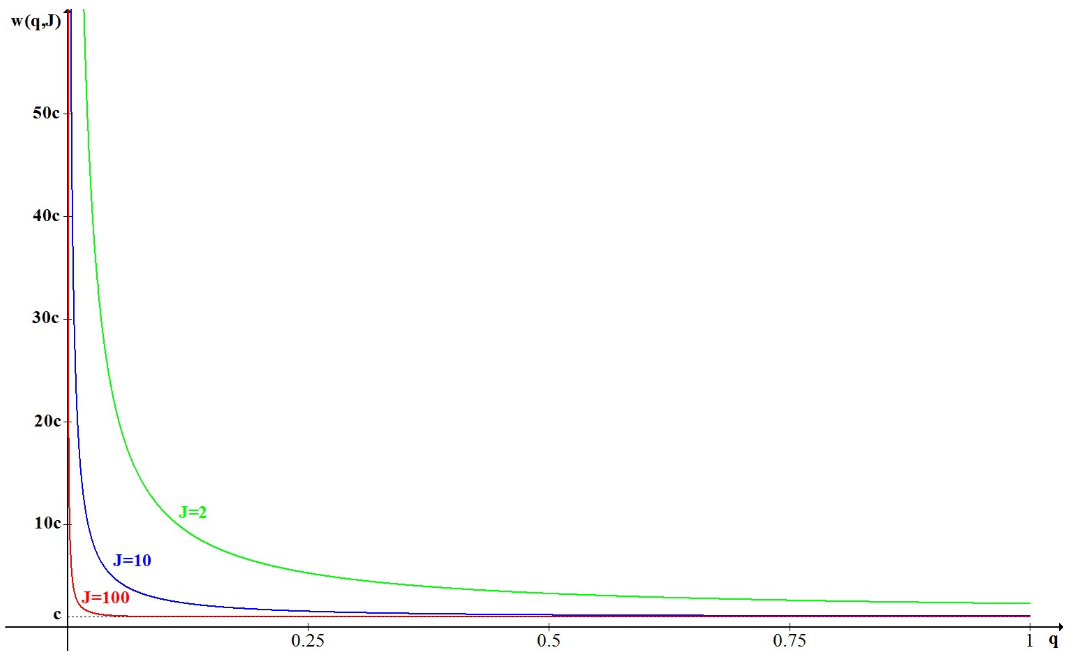

Figure 3.1: Feasible wages as a function of the monitoring probability

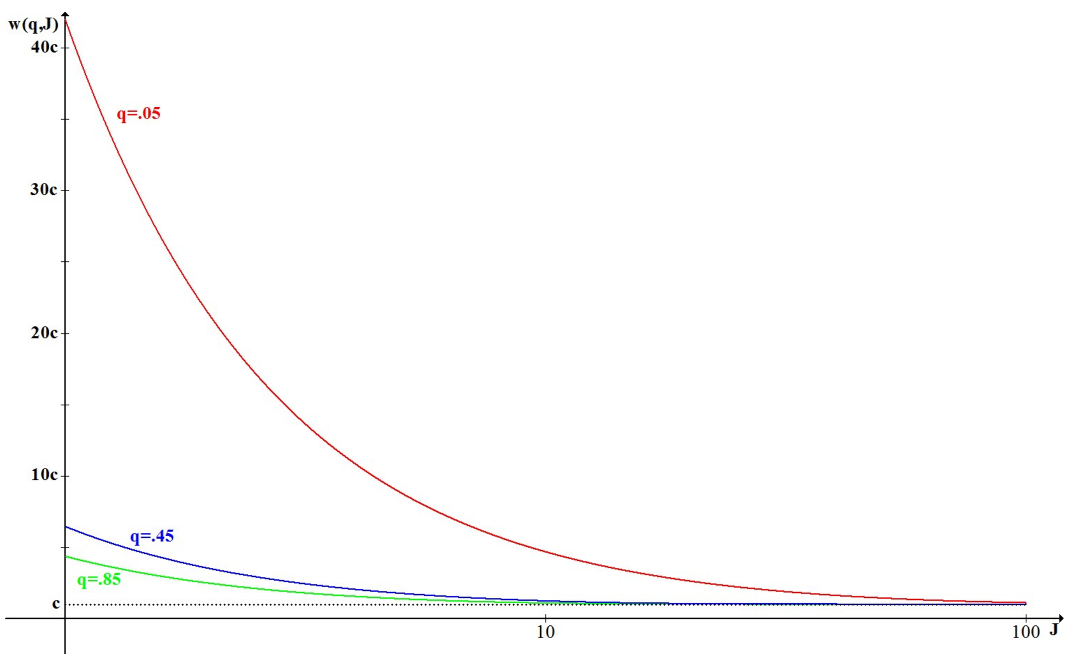

Figure 3.2: Feasible wages as a function of the job size

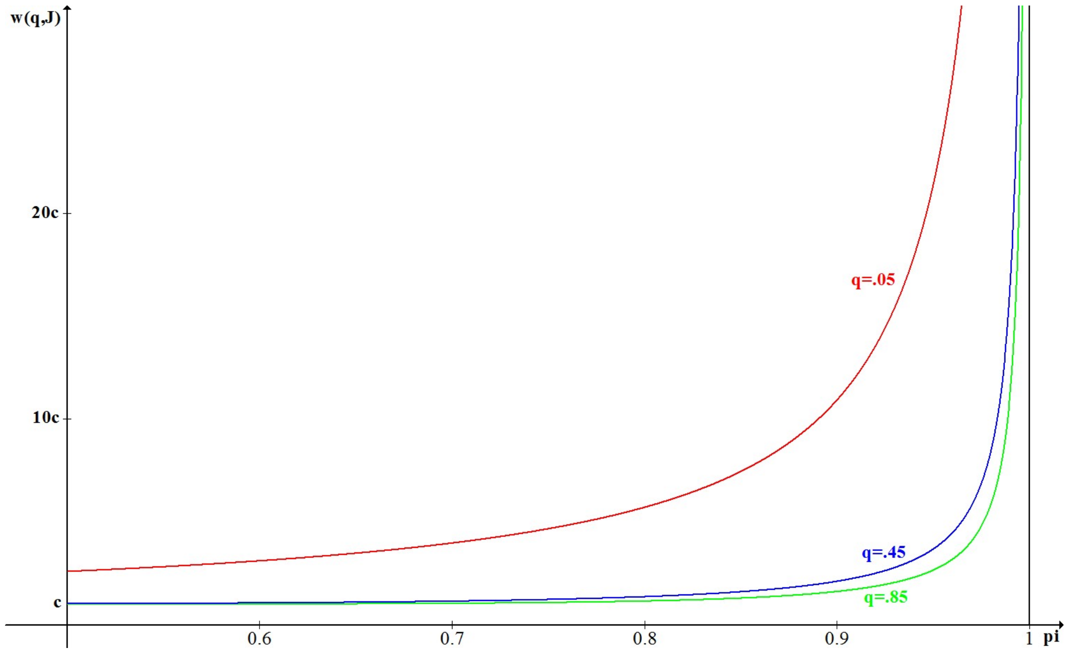

Figure 3.3: Feasible wages as a function of the prior

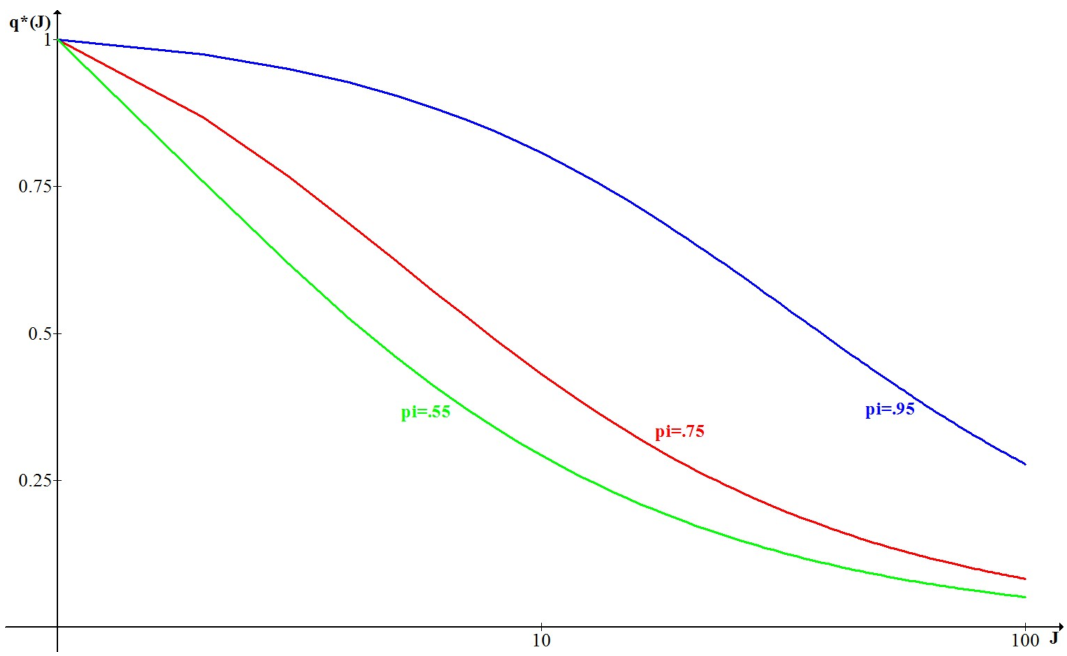

Figure 3.4: Optimal monitoring probability as a function of the job size

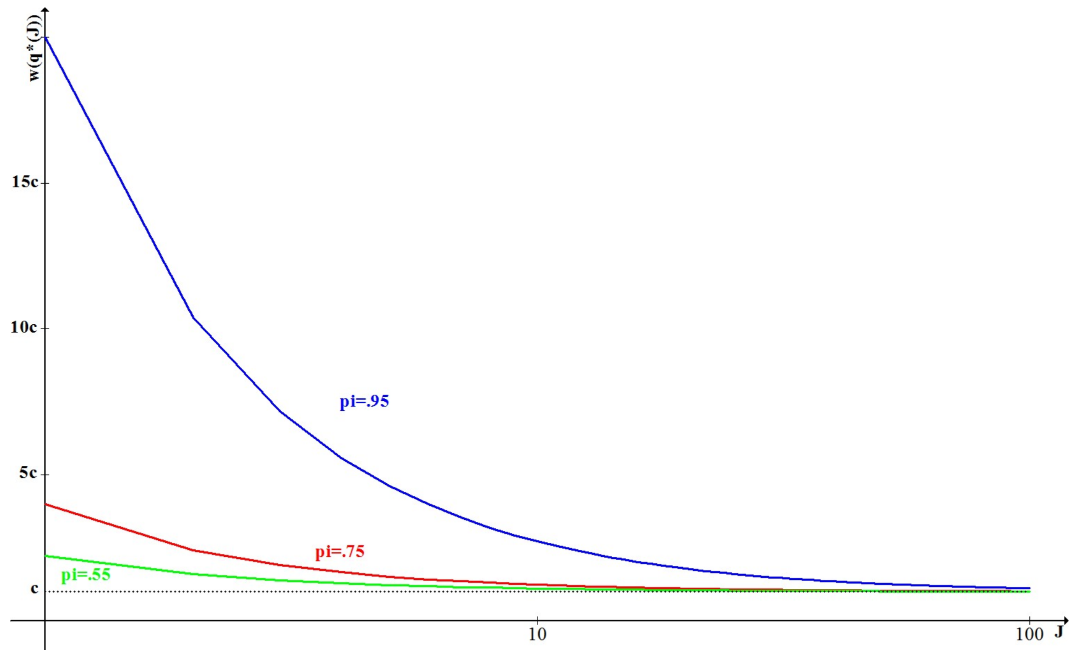

Figure 3.5: Optimal wage as a function of the job size

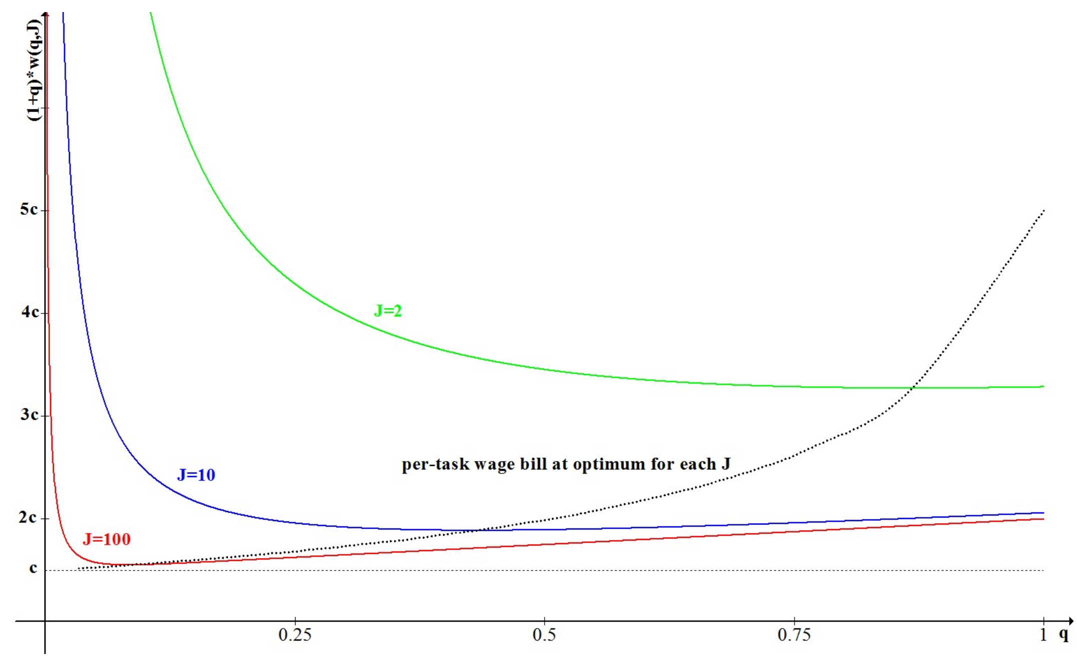

Figure 3.6: Per-task wage bill as a function of the monitoring probability

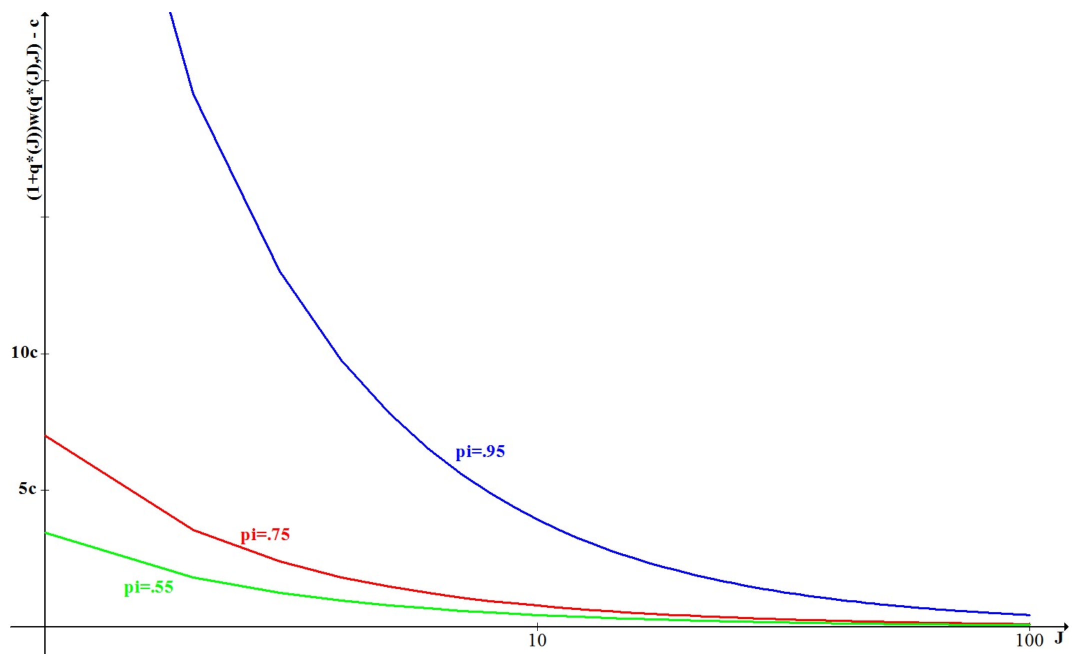

Figure 3.7: Per-task efficiency loss as a function of the job size

References

ACEMOGLU, D., M. DAHLEH, I. LOBEL, AND A. E. OZDAGLAR (2010): "Bayesian Learning in Social Networks," *SSRN eLibrary*.

BANERJEE, A. V. (1992): "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics*, 107, 797–817.

BIKHCHANDANI, S., D. HIRSHLEIFER, AND I. WELCH (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *The Journal of Political Economy*, 100, 992–1026.

BOARD, S. AND M. MEYER-TER-VEHN (2011): "Reputation for Quality," Tech. rep., UCLA.

CAMERER, C. F., T.-H. HO, AND J.-K. CHONG (2004): "A Cognitive Hierarchy Model of Games," *The Quarterly Journal of Economics*, 119, 861–898.

COSTA-GOMES, M. A., V. P. CRAWFORD, AND N. IRIBERRI (2009): "Comparing Models of Strategic Thinking in Van Huyck, Battalio, and Beil's Coordination Games," *Journal of the European Economic Association*, 7, 365–376.

CRIPPS, M. W., G. J. MAILATH, AND L. SAMUELSON (2007): "Disappearing private reputations in long-run relationships," *Journal of Economic Theory*, 134, 287 – 316.

DEMARZO, P. M., D. VAYANOS, AND J. H. ZWIEBEL (2001): "Persuasion Bias, Social Influence, and Uni-Dimensional Opinions," *SSRN eLibrary*.

EYSTER, E. AND M. RABIN (2005): "Cursed Equilibrium," *Econometrica*, 73, 1623–1672.

——— (2009): "Rational and Naive Herding," .

FAINGOLD, E. (2008): "Building a Reputation under Frequent Decisions," Tech. rep.

FAINGOLD, E. AND Y. SANNIKOV (2011): "Reputation in Continuous-Time Games," *Econometrica*, 79, 773–876.

FREI, B. (2009): "Paid Crowdsourcing: Current State and Progress Toward Mainstream Business Use," Tech. rep.

FUDENBERG, D. AND D. LEVINE (2007): "Continuous Time Limits of Repeated Games with Imperfect Public Monitoring," *Review of Economic Dynamics*, 10, 173–192.

FUDENBERG, D. AND D. K. LEVINE (1992): "Maintaining a Reputation When Strategies Are Imperfectly Observed," *Review of Economic Studies*, 59, 561–79.

——— (1994): "Efficiency and Observability with Long-Run and Short-Run Players," *Journal of Economic Theory*, 62, 103–135.

GALE, D. AND M. HELLWIG (1983): "Incentive Compatible Debt Contracts I: The One-Period Problem (Revised version now published in Review of Economic Studies (1985).)," Tech. rep.

GERSHKOV, A. AND B. SZENTES (2009): "Optimal voting schemes with costly information acquisition," *Journal of Economic Theory*, 144, 36–68.

GOEREE, J. K., T. R. PALFREY, B. W. ROGERS, AND R. D. MCKELVEY (2007): "Self-Correcting Information Cascades," *Review of Economic Studies*, 74, 733–762.

GUARINO, A. AND P. JEHIEL (2009): "Social Leanring with Course Inference," WEF Working Papers 0050, ESRC World Economy and Finance Research Programme, Birkbeck, University of London.

HARTMAN, P. AND A. WINTNER (1941): "On the Law of the Iterated Logarithm," *American Journal of Mathematics*, 63, 169–176.

HORTON, J. J. AND L. B. CHILTON (2010): "The Labor Economics of Paid Crowdsourcing," in *Proceedings of the 11th ACM Conference on Electronic Commerce*, New York, NY: ACM, EC '10, 209–218.

KLEIN, B. AND K. B. LEFFLER (1981): "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy*, 89, 615–41.

KOESSLER, F., A. ZIEGELMEYER, J. BRACHT, AND E. WINTER (2008): "Fragility of Information Cascades: An Experimental Study using Elicited Beliefs," Jena Economic Research Papers in Economics 2008-094, Friedrich-Schiller-University Jena, Max-Planck-Institute of Economics.

KREPS, D. M., P. MILGROM, J. ROBERTS, AND R. WILSON (1982): "Rational cooperation in the finitely repeated prisoners' dilemma," *Journal of Economic Theory*, 27, 245–252.

KUBLER, D. AND G. WEIZSACKER (2004): "Limited Depth of Reasoning and Failure of Cascade Formation in the Laboratory," *Review of Economic Studies*, 71, 425–441.

MAILATH, G. J. AND L. SAMUELSON (2001): "Who Wants a Good Reputation?" *Review of Economic Studies*, 68, 415–41.

MASON, W. A. AND D. J. WATTS (2009): "Financial Incentives and the 'Performance of Crowds'," *Knowledge Discovery and Data Mining - Human Computation (KDD-HCOMP)*, 11, 100–108.

RAHMAN, D. (2012): "But Who Will Monitor the Monitor?" *forthcoming, American Economic Review*.

SANNIKOV, Y. (2007): "Games with Imperfectly Observable Actions in Continuous Time," *Econometrica*, 75, 1285–1329.

SANNIKOV, Y. AND A. SKRZYPACZ (2007): "Impossibility of Collusion under Imperfect Monitoring with Flexible Production," *American Economic Review*, 97, 1794–1823.

——— (2010): "The Role of Information in Repeated Games With Frequent Actions," *Econometrica*, 78, 847–882.

SHEU, S. S. (1974): "Some Iterated Logarithm Results for Sums of Independent Two-dimensional Random Variables," *The Annals of Probability*, 2, 1139–1151.

SMITH, L. AND P. SORENSEN (2000): "Pathological Outcomes of Observational Learning," *Econometrica*, 68, 371–398.

TOWNSEND, R. M. (1979): "Optimal contracts and competitive markets with costly state verification," *Journal of Economic Theory*, 21, 265–293.