Information and the Bandit: The Good, the Bad and the Ugly[*]

Johannes C. Hoelzemann[†] Nicolas Klein[‡]

April 5, 2024

Abstract

We study a game of strategic experimentation in which information arrives through fully revealing, publicly observable, breakdowns. In line with our theoretical predictions, we find that players experiment significantly less and payoffs are lower when actions are hidden. We run a robustness test where we study a game of strategic experimentation in which information arrives through fully revealing, publicly observable, breakthroughs. In the case of breakthroughs, both experimentation and payoffs are higher with hidden actions—even when theory does not predict any difference in equilibrium. We view this as evidence that behavior is systematically affected by the informational environment. Moreover, behavior is consistent with strategic free-riding, as information is a public good and players produce inefficiently little of it.

JEL Classification: C73, C92, D83, O32
Keywords: Dynamic Public-Good Problem, Strategic Experimentation, Exponential Bandits, Learning, Dynamic Games, Laboratory Experiments

[*] Financial support by the Université Paris-Panthéon-Assas and the University of Vienna is gratefully acknowledged. The experimental part of this study was approved by the University of Vienna's VCEE (Vienna Center for Experimental Economics).

[†] UniVie: www.johanneshoelzemann.com & jc.hoelzemann@gmail.com.

[‡] UdeM: www.nicolasklein.com & kleinnic@yahoo.com.

1 Introduction

Games of pure informational externalities have received a lot of attention in the literature (see, e.g., Bolton and Harris 1999, Keller, Rady, and Cripps 2005 or Hörner, Klein, and Rady 2022). In these games, the information produced by a given player benefits other players as well—information production is a public good, and players tend to produce inefficiently little of it in equilibrium. Following Keller, Rady, and Cripps (2005), most papers in this literature have focused on so-called good-news environments, where discontinuous events bring good news; the absence of news consequently leads to a continuous deterioration in beliefs. In many real-world applications, however, discontinuous news events are in the form of bad news; think of severe side effects stemming from a medical drug, or the catastrophic malfunctioning of some technology, for instance. Theoretically, it is well understood (see, e.g., Keller and Rady 2015, or Wagner and Klein 2022) that the mechanisms underlying the bad-news strategic-learning models differ sharply from those under good news. While Hoelzemann and Klein (2021) has experimentally investigated strategic experimentation under good news, and Hoelzemann, Manso, Nagaraj, and Tranchero (2024) investigates the role of players' information in a strategic setting, we are, to the best of our knowledge, the first to experimentally investigate a bad-news strategic-experimentation setting.

The scant attention given to bad-news settings is surprising because of their economic importance: Bad-news learning processes naturally occur upon the introduction of a new technology that holds out hopes of cost savings but entails risks. Such risky technologies include new drugs and medical devices, and innovative processes such as hydraulic fracturing for oil production. Some technologies that are socially undesirable, perhaps because they impose negative externalities on other sectors, also fit in this broad class. Consider financial fraud or tax evasion when agents have incomplete information about the effectiveness of the detection technology. In all these cases, there also exist significant barriers to the flow of information, making unobservable actions a good starting point for the analysis. For example, the decision to evade taxes is private, but getting caught is typically a public event.

In this paper, we are investigating in particular the role of the observability of actions in a bad-news game of strategic experimentation with bandits. These are games of purely informational externalities, where players have an incentive to free-ride on the information produced by the other players. In a continuous-time, infinite-horizon, setting, it is theoretically known that, in a conclusive bad-news model, private information tends to be bad for welfare (Bonatti and Hörner 2017). This is because, in the absence of conclusive news, observing a player's shirking in information production makes the other player(s) more pessimistic than they would be on the equilibrium path if the conclusive bad news fails to materialize. Therefore, with conclusive bad news, players will be less prone to slack off in information production if their actions are observable, because, after observing a deviation, the other player(s) will be warier about the risky option than they would be absent a deviation. Because the only externality in the game is the positive informational externality between players, leading to a tendency toward under-production of information in equilibrium, we should expect that making deviations unobservable ought to dampen welfare in a conclusive bad-news environment.

The main goal of this investigation is to test whether this qualitative prediction of the theory is borne out by actual behavior in a controlled laboratory environment. In order to do so, we have endeavored to come up with the simplest possible environment in which theory would predict the qualitative effect just described to arise. For the effect to arise, we need at least three periods. This is because, in the last period, a player does not care what their opponent will do, as they have no future use for the information learned in this period. So, only in the first period do players want to alter their opponent's future behavior for strategic considerations. We therefore construct a three-period, two-player, game, calibrating the parameters in such a way that the game features the strategic effects we are interested in. We have constructed our game in a particularly stark way so that it has the feature that the efficient solution is an equilibrium *if and only if* actions are observable. The efficient solution has both players using the risky option in all periods (absent a breakdown); the unique equilibrium with unobservable actions has both players never using the risky option, while either always or never playing risky are the two equilibria with observable actions. Empirically, both experimentation and payoffs are higher with observable actions. Further, participants use the risky option more frequently over time, reflecting growing optimism.

To understand whether the differences in behavior between the informational settings depend on whether they are predicted by (perfect Bayesian) equilibrium in the particular game, or whether they are a more general feature of behavior, we study a three-period game in the good-news setting. To do so, we chose simple numerical values for the parameters that additionally have the property that there is no difference in equilibrium predictions depending on whether actions are observable or not. Information now arrives through fully revealing, publicly observable, breakthroughs instead of breakdowns. In contrast to our bad-news game, it is known that, in a continuous-time, infinite-horizon, setting, private information is good for welfare in a conclusive good-news game (Bonatti and Hörner 2011). This is because, in the absence of conclusive news, observing a player's shirking in information production makes the other player(s) more *optimistic* than they should be on the equilibrium path. Therefore, the other player(s) will tend to pick up the slack in information production after an observable deviation to shirking. Therefore, with conclusive good news in continuous time, players will be more prone to slacking off in information production if their actions are observable. Because the only externality in the game is the positive informational externality between players, leading to a tendency toward under-production of information in equilibrium, making deviations unobservable improves welfare in a continuous-time, infinite-horizon, conclusive good-news environment. Thus, in contrast to naïve intuition, less observability, and hence less information, can be welfare-improving in a game of purely informational externalities. We thus want to test if, in our good-news game, participants will free-ride more, and therefore achieve lower average payoffs, when actions are observable, even though this is not an equilibrium feature of our three-period game. Empirically, participants experiment indeed significantly less when actions are observable. Moreover, their payoffs are lower. Over time, participants decrease their use of the risky option, reflecting growing pessimism.

In summary, the paper makes two main contributions. First, we present evidence that behavior is systematically affected by the informational environment. In the *bad-news* environment, we find that both experimentation and payoffs are *higher* with observable actions. By contrast, in the *good-news* environment, participants experiment significantly *less*, and their payoffs are *lower*, when actions are observable, even though, in our game, there is no difference in equilibrium predictions depending on whether actions are observable or not.

Second, behavior is consistent with strategic free-riding, as information is a public good and participants produce inefficiently little of it. In the *bad-news* environment, participants experiment, on average, too little even when the efficient solution is an equilibrium. For the *good-news* environment, we design a three-period game such that, under either informational assumption, equilibrium always features underproduction of information. Participants' behavior is indeed characterized by too little experimentation, especially when actions are observable.

The rest of the paper is organized as follows. Section 2 explains our environment and design. Section 3 sets out our experimental implementation and presents our main findings. Section 4 introduces the good-news game and presents its analysis. This is followed by an econometric robustness test, which is presented in Section 5. Section 6 explains our theoretical frameworks in detail. Section 7 offers a discussion on economic significance, reviews some additional related literature and concludes with some thoughts on free-riding on information.

2 The *Bad-News* Environment

In this section, we provide a brief description of our theoretical framework to build intuition and to guide our experimental design, identification strategy, and econometric analysis. A complete formal analysis of the game will be provided in Section 6.

2.1 The Design

The game is played over three periods $t = 1, 2, 3$. If the safe arm is used, the payoff will be 0 for certain in that period. Using the risky arm entails a *benefit* of $s = 2,857$ (Experimental \$). The risky arm is either *good* or *bad*, its type remaining constant over the three periods of the game. If it is good, its use never imposes a cost. If it is bad, it imposes a cost of 20,000 with a probability of $\lambda = 1/4$ in any period it is used; conditionally on the risky arm's type, the draws are i.i.d. between players and across periods. Players do not initially know if the risky arm is good or bad; they know that Nature (or the computer) makes the risky arm bad with a probability of $p_0 = 0.676392$. After a breakdown is observed, the risky arm is known to be bad with probability 1. In the absence of a breakdown and n successful tries of the risky arm, Bayes' rule implies that an observer knowing this information should hold the belief $p_n = \frac{p_0(1-\lambda)^n}{p_0(1-\lambda)^n+1-p_0}$ that the risky arm is bad; i.e., observing that the risky arm has been used without a breakdown makes players increasingly *optimistic* about the quality of the risky arm. Thus, the updated posterior belief either jumps to 1 in case of a breakdown, or declines with the number of unsuccessful tries n. Arm types are i.i.d. across games. One player's risky arm is good *if and only if* the other one's is as well. In the treatment with *observable* actions, a player observes all of the other player's previous actions as well as the outcomes of these actions. In the treatment with *unobservable* actions, a player observes only if the other player has suffered a breakdown of 20,000 from the risky arm or not.

Our (admittedly somewhat buckled) numerical values allow us to make stark theoretical predictions. One computes that the solution maximizing the sum of the players' payoffs has both players playing risky in the first period, and continuing to play risky in the subsequent periods unless and until a breakdown occurs. Clearly, in equilibrium, once a player knows the risky arm to be bad because they have observed a breakdown, they will use the safe arm in all subsequent periods, as is efficient. Furthermore, one verifies by backward induction that, with unobservable actions, the only equilibrium is for both players always to play safe.[1] With observable actions, however, while always playing safe remains an equilibrium, the efficient solution is an equilibrium as well. This latter equilibrium is sustained by the threat of the other player switching to always playing safe if one player unilaterally deviates to playing safe in the first period, and thus requires that actions be observable.

Implications for Behavior Consequently, our behavioral hypotheses are as follows:

- We observe efficient behavior more often with observable than with unobservable actions.
- Participants use the risky arm *more* when actions are observable.
- Participants' payoffs are *higher* when actions are observable.
- Updating of beliefs: Conditionally on no breakdown having occurred, participants use the risky arm *more* in later periods.

3 The Experiment

3.1 Organization

We conducted all experiments in the months of July to November 2023 at the University of Vienna. Participants were recruited from the Vienna Center for Experimental Economics (VCEE) subject pool using ORSEE (Greiner 2015). No one participated in more than one session. During the experiments, participants could contact an experimenter anytime for assistance. After reading the instructions, participants had to correctly answer several comprehension questions before starting the main part of the experiment. The experiment was programmed in oTree (Chen, Schonger, and Wickens 2016). We recruited 104 participants and all payments were made in cash. The average participant earned approximately €10.57 from one randomly selected game and all payments were in Euros. The instructions and experimental interface are reproduced in the online appendix.

[1] See Section 6 for details.

3.2 Implementation

In order to increase the computational efficiency of the implementation and to increase control, we had simulated all the relevant parameters ahead of time. As all our stochastic processes are Bernoulli processes, simulating their realizations ahead of time is equivalent to simulating them as the game progresses. These included separate processes for the quality of the risky arm and the timing of breakdowns on the risky arm in case it was bad.[2] We generated 25 different sets of realizations of the random parameters controlling the quality of the risky arm and the arrivals of the bad risky arm. These corresponded to 25 different games that each of our participants played. To make our findings more easily comparable, we have kept the same realizations for both observable and hidden actions. Participants were randomly assigned to groups of two players and randomly rematched within a matching group of six to eight participants after each game. Each participant was randomly assigned either to the treatment with observable or hidden actions, and played the 25 games in random order. To ensure a balanced data-collection process, we replicated any order of the 25 games that was used for a matching group in the treatment with observable actions for a matching group in the treatment with unobservable actions. Participants could see their fellow group members' action choices and payoffs, depending on the randomly assigned treatment, on their computer screens. Figure 1 shows how information was displayed, observable actions being illustrated at the top and unobservable actions—"the ugly"—being highlighted at the bottom.

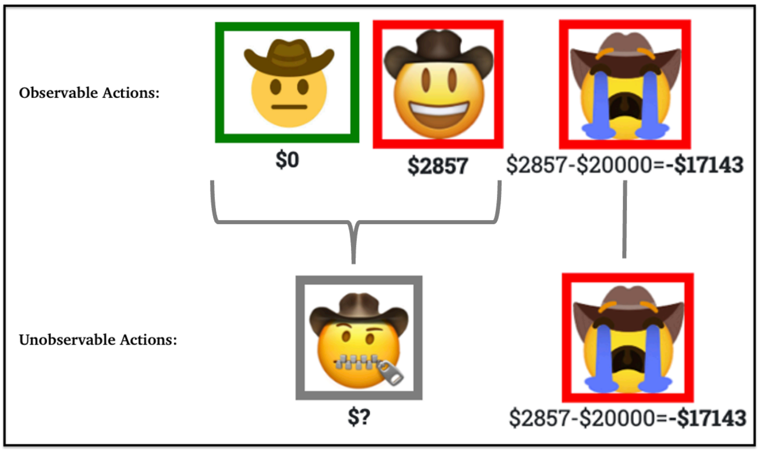


Figure 1: Bad News – Experimental Implementation

3.3 Experimental Results

This section is dedicated to examining the implications for behavior as detailed in Section 2.1. For each of the 25 games, we implemented two treatments, actions being either observable or unobservable, within two-player groups, comprising 52 groups in total.

[2] Details are available upon request.

These groups were randomly re-matched within a matching group after each game; all relevant parameters were simulated in advance.

We divide our analysis into four distinct sections. Initially, we provide summary statistics, highlighting both the average intensity of experimentation and the overall group payoffs. Following this, our primary analysis examines the aggregate experimental outcomes, with an initial focus on the distribution patterns of experimentation intensities and group payoffs. Subsequently, we assess efficiency by comparing observed behavior to the theoretical efficient solution. In addition, we study how behavior relates to our theoretical predictions, in particular the consistency with equilibrium. The final part of our analysis examines the evolution of behavior over time, specifically how participants adjust their action choices in games where no breakdowns occur. To enhance the robustness of our findings, we include a robustness test, utilizing ordinary least squares (OLS) regressions with random effects and clustering of standard errors at the matching-group level. These results are reported in Section 5. Our results show that the number and order of games previously played by participants does not significantly influence their behavior, thereby affirming the robustness of our findings across the study.

Experimentation and Payoffs

As outlined in Section 2.1, we anticipate that average experimentation intensities and group payoffs would be higher in the treatment where monitoring by others is possible. Recall that experimentation intensity for each player is measured up until the moment a first breakdown occurs to any player in the group. Table I presents the observed mean experimentation intensities and the average total payoffs, calculated using group averages across all games for both treatments.

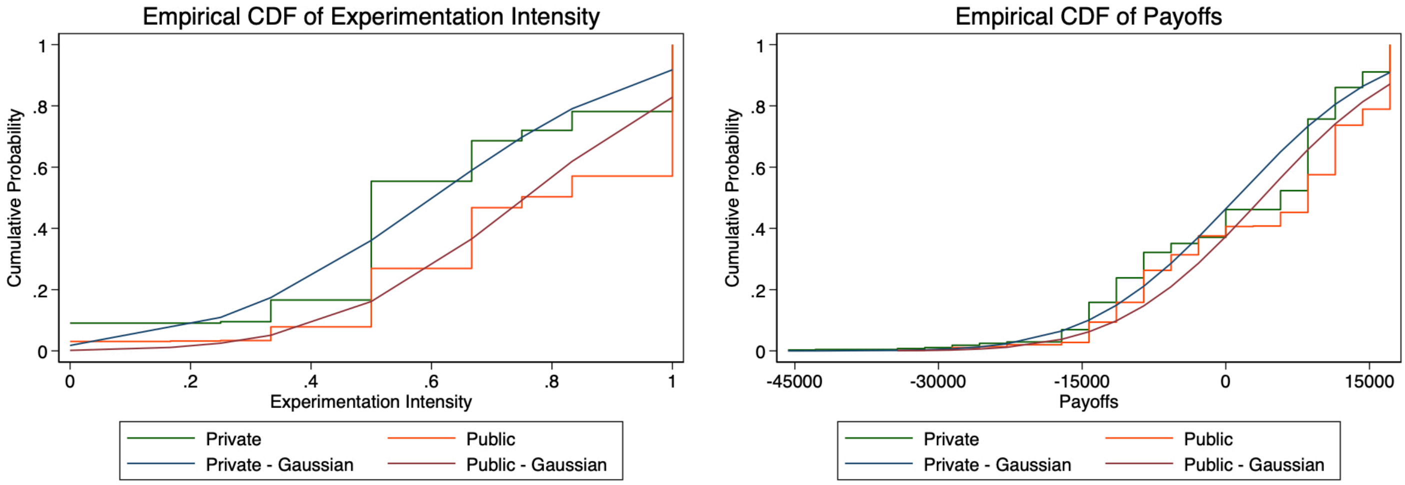
Table I: Bad News – OLS Estimations

	Experimentation Intensity	Payoffs
Intercept	0.602^{***}	1085.255^{***}
	(0.044)	(298.817)
Public	0.153^{***}	2716.357^{***}
	(0.052)	(418.066)
N	1300	1300
R-squared	0.074	0.013

For all estimations, robust standard errors are clustered at the matching-group level and shown in brackets.

We observe a pronounced positive impact of action observability on both experimentation intensity and payoffs. Specifically, participants pull the risky arm considerably more often when monitoring is feasible, resulting in markedly higher payoffs at the group level.

Extending our analysis beyond mere point estimates, Figure 2 illustrates the empirical distribution of experimentation intensities and group payoffs with the best fitting normal Gaussian model being superimposed over the sample cumulative density function across the different treatments.



The sample cumulative distribution functions for experimentation intensity and payoffs are shown, by information condition. The best fitting normal (Gaussian) model is superimposed over the sample CDF.

Figure 2: Bad News – Empirical CDFs of Experimentation Intensity and Payoffs

Both experimentation intensities and group payoffs significantly exhibit stochastic dominance in the treatment with observable actions over the treatment where actions are unobservable. This difference is statistically significant, as demonstrated by Kolmogorov-Smirnov tests, which yield p-values of 0.001.

We summarize these findings in the following:

Experimentation Participants use the risky arm *more* when actions are observable.

Payoffs Group payoffs are *higher* when actions are observable.

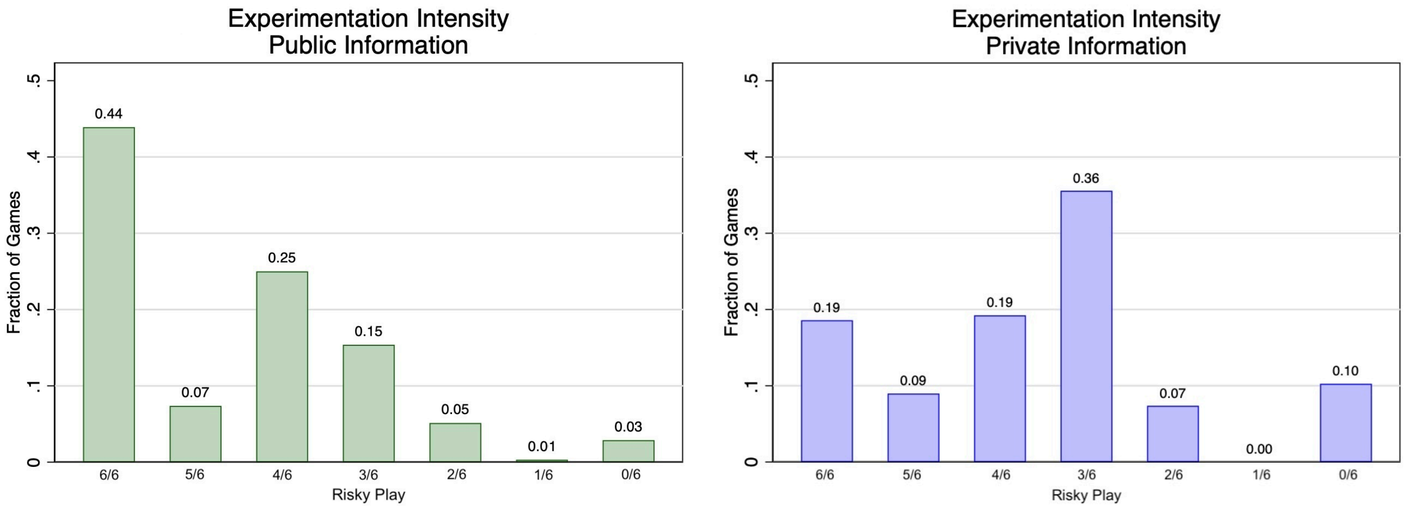
Efficiency Benchmark

To assess the efficiency of participants' behavior, our analysis concentrates on games where no breakdown was suffered during the three interaction periods. In the treatment where actions are observable, the average experimentation intensity stands at 0.762 (with a standard deviation of 0.256, N=312), significantly deviating from the theoretical efficient solution of 1. Out of 312 observations, 137 are in alignment with the efficient solution. Conversely, for games with unobservable actions, the average experimentation intensity with 0.591 and a standard deviation of 0.285 for N=312 is also significantly different from the efficient solution, and only 58 of 312 observations directly coincide with the efficient solution.

Efficiency Participants free-ride, i.e., they use the risky arm *less* than what would be efficient with either observable or unobservable actions.

Consistency of Behavior with Equilibrium

Here too, our analysis focuses on the 312 games where no breakdown was suffered during the three interaction periods. In Figure 3, we highlight the observed risky play for each treatment separately, providing a detailed view of how behavior is associated with equilibrium play. In particular, we plot the fraction of games in which risky was played r times divided by the number of periods t multiplied by the number of players n, i.e., $\frac{r}{t \times n}$.



The experimentation intensity in games without breakdowns is shown, by information condition.

Figure 3: Bad News – Experimentation Intensity

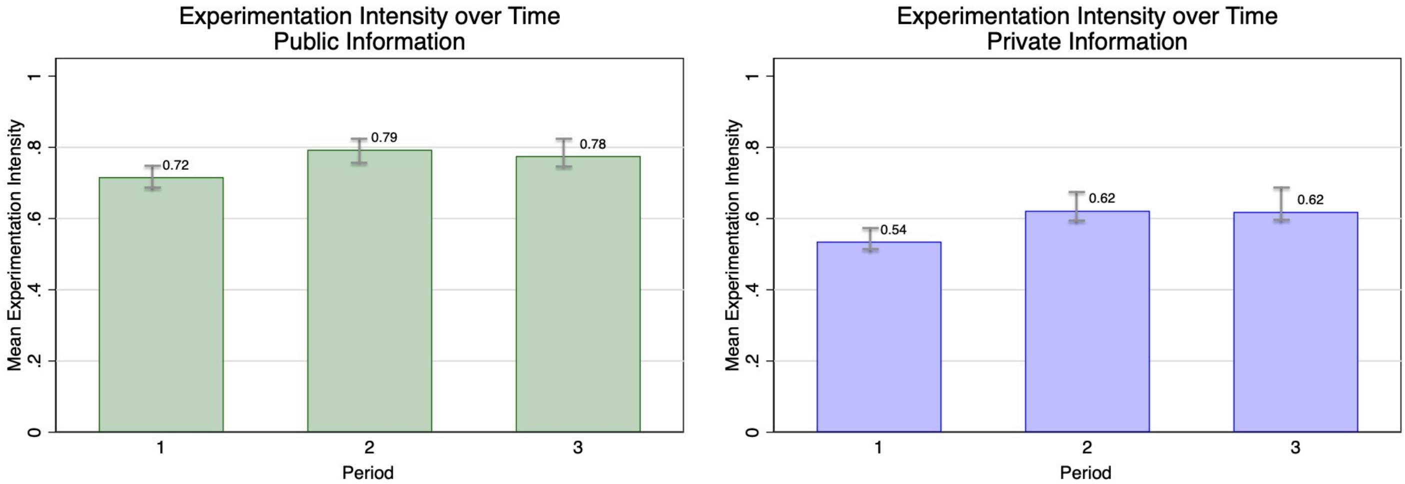
When actions are observable, 146 of 312 games are consistent with equilibrium; among these, the overwhelming majority, namely 137 games, coincide with the efficient solution. By contrast, with hidden actions, play that is consistent with the—now smaller—equilibrium set significantly decreases, with only 32 of 312 in line with the theoretical prediction. Unsurprisingly, the difference in equilibrium play by treatment is highly statistically significant with p-values of 0.001 for both a t-test and a two-sided Wilcoxon rank-sum test.

This leads to the following:

Equilibrium Behavior is more often consistent with equilibrium when actions are observable.

Dynamic Evolution of Behavior

We now shift our focus to the dynamics of observed behavior. We are particularly interested in whether, consistently with Bayesian updating, participants increase their use of the risky arm as the game progresses. Our attention remains on games where no breakdown is incurred throughout the three periods of interaction. In Figure 4, we graph the observed experimentation intensities for each period and treatment separately, providing a detailed view of how behavior evolves over the course of the game.



The average experimentation intensity over time in games without breakdowns is shown, by information condition.

Figure 4: Bad News – Experimentation Intensity over Time

While it is unrealistic to expect our participants to calculate posterior beliefs precisely using Bayes' rule, we nevertheless anticipate that, in the absence of a breakdown, participants will increasingly use the risky arm as the game progresses, reflecting growing optimism. At the beginning of a game without any breakdowns, participants are indeed significantly less likely to choose the risky arm than in later periods. To examine changes in behavior over time, we employ two-sided t-tests for parametric analysis and two-sided Wilcoxon rank-sum (Mann-Whitney) tests for non-parametric analysis, treating group averages as independent observations. Regardless of whether actions were observable, we find that the differences in experimentation intensities across time are highly statistically significant when comparing behavior in the very first period to those in either the second or the last period, with p-values in both treatments for either test being less than 0.050. In the treatment with observable actions, t-tests (two-sided Wilcoxon rank-sum tests) produce p-values of $p_{12} = 0.004$ (0.009), $p_{13} = 0.015$ (0.049), and $p_{23} = 0.759$ (0.515), where $p_{\alpha\beta}$ is the p-value from comparing periods α vs. β. With unobservable actions, by contrast, we find p-values of $p_{12} = 0.001$ (0.002), $p_{13} = 0.002$ (0.003), and $p_{23} = 0.546$ (0.987), respectively.

Additionally, we also analyze mean experimentation intensities across treatments for each period. As illustrated in Figure 2, participants engage with the risky arm more frequently when actions are observable. The differences in all periods are highly statistically significant, with p-values of 0.001 for both tests.

We summarize these results as follows:

Belief Updating Conditionally on no breakdown having occurred, participants use the risky arm *more* in later periods.

4 The *Good-News* Environment

In the previous section, we have seen that experimentation intensity and payoffs are significantly higher with public information. Our game's equilibrium set is larger, containing the efficient solution, when actions are observable; by contrast, never using the risky arm is the only equilibrium with unobservable actions. We have indeed calibrated the parameters of our game in such a way as to generate stark theoretical predictions illustrating the impact of the informational setting on behavior, which is known from continuous-time theory, in the starkest possible way. So, in a way, we have stacked the cards in our favor. In this section, we report on an experiment, where, in a sense, we do the mirror opposite: We look at a three-period game in the *good-news* setting, choosing very simple numerical values for the parameters that additionally have the property that there is *no difference* in equilibrium predictions depending on whether actions are observable or not, so as to understand whether the differences in behavior between the informational settings depend on whether they are predicted by (perfect Bayesian) equilibrium in the particular game, or whether it is a more general feature of behavior. Information now arrives through fully revealing, publicly observable, *breakthroughs* instead of *breakdowns*. Recall from our discussion in the Introduction that, in a continuous-time, infinite-horizon setting, private information is good for welfare in a conclusive good-news game (Bonatti and Hörner 2011). We are thus, in the main, interested in whether participants will free-ride more, and therefore achieve lower average payoffs, when actions are observable, even though this is not a feature of perfect Bayesian equilibrium.

In the following subsection, we provide a brief description of our theoretical framework, relegating a complete formal analysis to Section 6.

4.1 The Design

The game is played over three periods $t = 1, 2, 3$. If the safe arm is used, the payoff will be 0 for certain in that period. Using the risky arm entails a cost of $s = 25$ (Experimental \$). The risky arm is either *good* or *bad*, its type remaining constant over the three periods of the game. If it is bad, it never yields a positive payoff. If it is good, it pays out a lump sum of 100 with a probability of $\lambda = 1/2$ in any period it is used; conditionally on the risky arm's type, the draws are i.i.d. between players and across periods. Players do not initially know if the risky arm is good or bad; they know that Nature (or the computer) makes the risky arm good with a probability of $p_0 = 3/4$. After a success is observed, the risky arm is known to be good with probability 1. In the absence of a success and n unsuccessful tries of the risky arm, Bayes' rule implies that an observer knowing this information should hold the belief $p_n = \frac{p_0(1-\lambda)^n}{p_0(1-\lambda)^n+1-p_0}$ that the risky arm is good. Note that p_n is strictly decreasing in n. Thus, the updated posterior belief either jumps to 1 in case of a success, or declines with the number of unsuccessful tries n. Arm types are i.i.d. across games. One player's risky arm is good *if and only if* the other one's is as well. In the treatment with *observable* actions, a player observes all of the other player's previous actions as well as the outcomes of these actions. In the treatment with *unobservable* actions, a player observes only if the other player has received the reward of 100 from the risky arm or not.

One computes that the solution maximizing the sum of the players' payoffs has both players playing risky in the first two periods, and safe in the last (conditionally on no breakthrough having occurred). As ours is a game of purely informational externalities, we should expect players to use the risky arm too little in equilibrium, as a player will not take into account that the information they produce (at a private cost) benefits the other player as well. Clearly, in equilibrium, once a player knows the risky arm to be good because they have observed a success, they will use the risky arm in all subsequent periods, as is efficient. Furthermore, one verifies by backward induction that in equilibrium both players play risky in the first period. Then, in the second period, exactly one player will play risky, provided no breakthrough has been observed in the first period. In the third period, conditionally on no breakthrough having arrived yet, both players play safe.

Our theoretical analysis thus leads us to the following behavioral hypotheses. For one, we expect players to be *free-riding*, i.e., to use the risky arm less than what would maximize the sum of their payoffs. This is because a player has to bear the full cost of experimenting with the risky arm, while sharing the benefits of the information generated. Consequently, we should expect average payoffs to be lower than in the efficient solution. Even though this is not a feature of the equilibrium set, we furthermore hypothesize that participants will free-ride more, and therefore achieve lower average payoffs, when actions are observable. Furthermore, while it would be unrealistic to presume that our participants could compute posterior beliefs p_n precisely via Bayes' rule, we hypothesize that (in the absence of a breakthrough) players will use the risky arm less as time progresses.

Implications for Behavior Therefore, our behavioral hypotheses are as follows:

- Participants use the risky arm *less* than what would be efficient.
- Participants use the risky arm *less* when actions are observable (although not an equilibrium feature).
- Participants' payoffs are *lower* when actions are observable (although not an equilibrium feature).
- Updating of beliefs: Conditionally on no breakthrough having been achieved, participants use the risky arm *less* in later periods.

4.2 Experiment Details

We collected data from another 110 participants who were also recruited from the Vienna Center for Experimental Economics (VCEE) subject pool using ORSEE (Greiner 2015). The average participant earned approximately €15.01 from one randomly selected game. All payments were made in Euros and in cash. As before, we had simulated all the relevant parameters ahead of time as all our stochastic processes are Bernoulli processes. These included separate processes for the quality of the risky arm and the timing of breakthroughs on the risky arm in case it was good. We generated

25 different sets of realizations of the random parameters controlling the quality of the risky arm and the arrivals of the good risky arm. These corresponded to 25 different games that each of our participants played. As before, in order to make our findings more easily comparable, we have kept the same realizations for both observable and hidden actions. Participants were randomly assigned to groups of two players and randomly rematched within a matching group of six to eight participants after each game. Each participant was randomly assigned either to the treatment with observable or hidden actions, and played the 25 games in random order. We again ensured a balanced data-collection process by replicating any order of the 25 games that was used for a matching group in the treatment with observable actions for a matching group in the treatment with unobservable actions. Participants could see their fellow group members' action choices and payoffs, depending on the randomly assigned treatment, on their computer screens. In our experimental implementation, we attempted to keep the *good-news* environment as close as possible to our *bad-news* environment. In the *good-news* environment, the experimental implementation was similar with the difference that we implemented a "party emoji" GIF for breakthroughs instead of the "crying emoji" GIF in the *bad-news* environment. Similarly, we displayed a "sad emoji" in the *good-news* environment instead of the "happy emoji" in the *bad-news* environment when a participant had pulled unsuccessfully the risky arm.[3] Figure 5 shows how information was displayed, observable actions being illustrated at the top and unobservable actions—"the ugly"—being highlighted at the bottom.

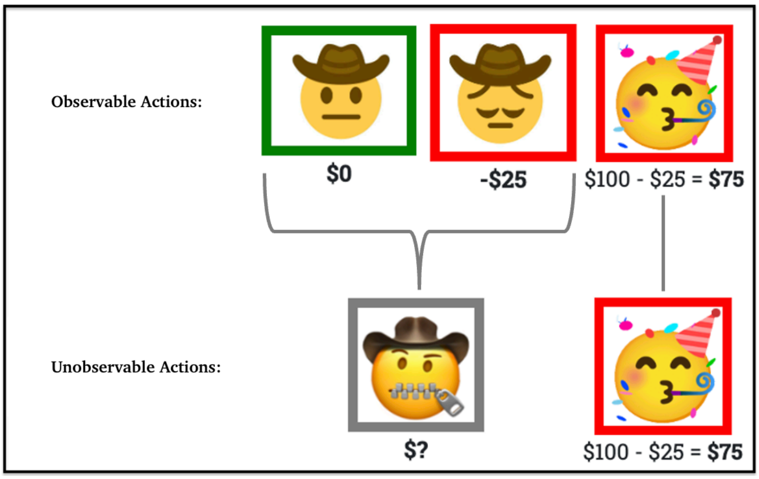


Figure 5: Good News – Experimental Implementation

4.3 Experimental Results

This section is devoted to testing our behavioral hypotheses and the implications for behavior outlined in Section 4.1. For each of the 25 games, we conducted two treat-

[3] The experimental instructions and interfaces can be found in the online appendix as well. The dynamic interface can be accessed online upon request.

ments (observable and unobservable actions with two-player groups), with 55 groups in total that were randomly re-matched within a matching group after each game.

As before, to maintain consistency and ensure comparability, we break the analysis into three sections. We start with presenting the summary statistics for both average experimentation intensity and group payoffs. Then, we begin our main analysis by presenting the aggregate experimental results focusing first on the distribution of the experimentation intensity and group payoffs. Next, we focus on efficiency and study how behavior relates to the efficient solution. Lastly, we delve into behavior over time, in particular participants' updating of beliefs in games where no breakthrough has occurred. In Section 5, we complement our analysis of this section with a robustness test by reporting results from OLS regressions with random effects and clustering of standard errors at the matching-group level. We find no effect of the number and order of games previously played on participants' behavior, and results reported throughout the paper remain robust.

Experimentation and Payoffs

As we have argued in Section 4.1, we might expect average experimentation intensities as well as (group) payoffs to be lower in the treatment when monitoring is feasible. Recall that the experimentation intensity is calculated for each player until the time of a first breakthrough by any player in a group or the end of the game, whichever arrives first.

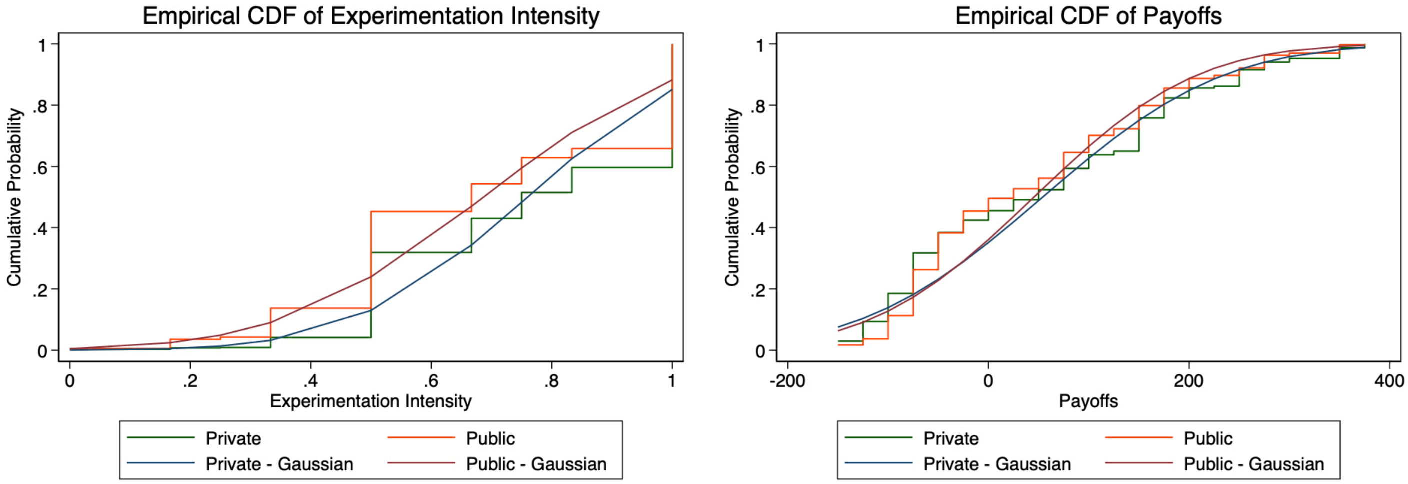
Table II: Good News – OLS Estimations

	Experimentation Intensity	Payoffs
Intercept	0.764^{***}	54.037^{***}
	(0.018)	(3.960)
Public	-0.078^{**}	-8.573^{*}
	(0.028)	(4.580)
N	1375	1375
R-squared	0.024	0.001

For all estimations, robust standard errors are clustered at the matching-group level and shown in brackets.

Table II lists the observed mean experimentation intensities and average sum of payoffs, using group averages across games for the two treatments. Thus, we find a negative effect of the observability of actions on both experimentation intensity and payoffs. This is the first piece of evidence that participants tend to shirk more when it comes to the production of information when actions are observable. Participants use the risky arm significantly less when monitoring is possible, leading to significantly lower payoffs at the group level.

Moving beyond point estimates, Figure 6 plots the empirical distribution of experimentation intensities and group payoffs with the best fitting normal Gaussian model superimposed over the sample cumulative density function by treatment.



The ample cumulative distribution functions for experimentation intensity and payoffs are shown, by information condition. The best fitting normal (Gaussian) model is superimposed over the sample CDF.

Figure 6: Good News – Empirical CDFs of Experimentation Intensity and Payoffs

While the effect of monitoring is less nuanced when it comes to payoffs compared to experimentation intensities, both measures of interest are significantly higher in stochastic dominance in the treatment with unobservable actions than in the treatment with observable actions: Kolmogorov Smirnov tests produce p-values of 0.001.

We summarize our findings in the following:

Experimentation Participants use the risky arm *less* when actions are observable.

Payoffs Group payoffs are *lower* when actions are unobservable.

Efficiency Benchmark

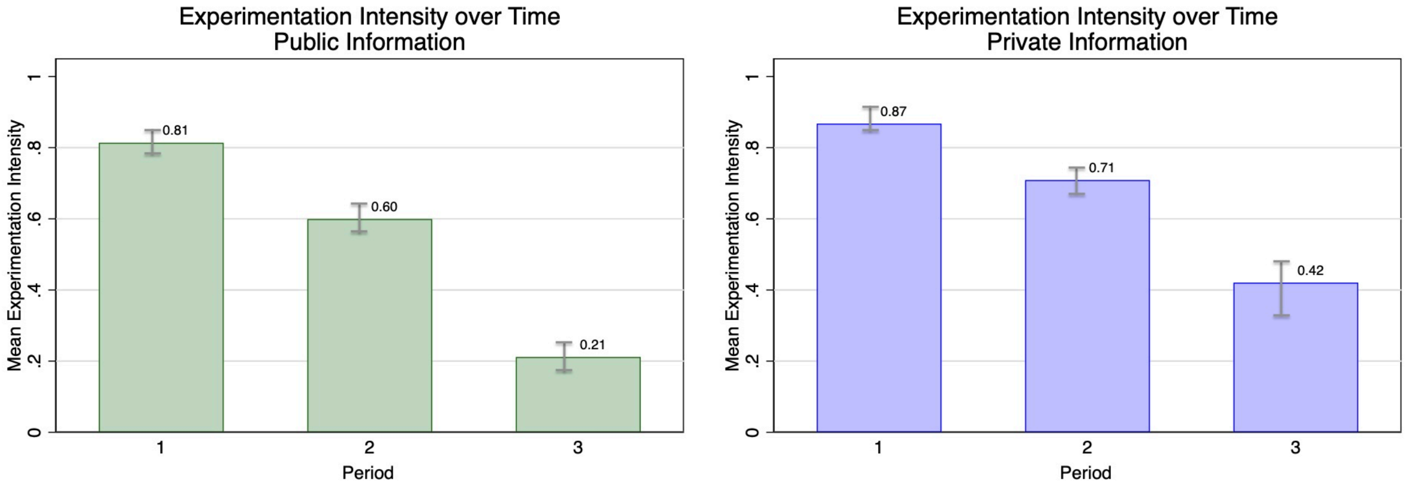
To investigate whether participants behaved efficiently, we focus on games in which no breakthrough occurs over the three periods of interaction. For observable actions, the average experimentation intensity with 0.542 (with a standard deviation of 0.199, N=196) is significantly different from the planner's solution of $2/3$. 25 of 196 observations are compatible with the efficient solution. For unobservable actions, by contrast, the average experimentation intensity coincides with the efficient solution (with a standard deviation of 0.184, N=184), while only 22 out of 189 observations coincide with the efficient solution.

This leads us to state the following:

Efficiency Participants use the risky arm *less* than what would be efficient only with observable actions. By contrast, when actions are unobservable the average experimentation intensity coincides with the efficient solution.

Dynamic Evolution of Behavior

We now turn to the dynamics of observed behavior. In particular, we are curious to see whether participants use the risky arm less as the game progresses. Here again, we focus on games in which no breakthrough occurs over the three periods of interaction. In Figure 7, we plot the observed experimentation intensity period-by-period and for each treatment separately.



The average experimentation intensity over time in games without breakthroughs are shown, by information condition.

Figure 7: Good News – Experimentation Intensity over Time

Obviously, we do not expect that participants compute posterior beliefs p_n precisely via Bayes' rule; however, we would expect that, in the absence of a breakthrough, participants will use the risky arm less as time progresses as they gradually grow pessimistic over time.

To test for differences over time parametrically, we apply two-sided t-tests and to test for differences non-parametrically, we apply two-sided Wilcoxon rank-sum (Mann-Whitney) tests, using group averages as independent observations. Irrespective of the observability of actions, the differences in experimentation intensities are highly statistically significant. The corresponding p-values in both cases are 0.001. At the outset of a given game without any breakthrough, participants use the risky arm significantly more often compared to later periods, the largest drop of experimentation being observed in the very last period of the game.

In addition, we also compare mean experimentation intensities across treatments period-by-period. As can be seen in Figure 7, participants use, on average, the risky arm more frequently in the treatment with unobservable actions. Differences are highly statistically significant in the last two periods with p-values of 0.001 for either test. In the very first period, differences are significant at the 5%-level, the t-test (Wilcoxon rank-sum test) produces a p-value of 0.022 (0.029).

We summarize these results as follows:

Belief Updating Conditionally on no breakthrough having been achieved, participants use the risky arm *less* in later periods.

5 Econometric Robustness Tests

As a further robustness test and to complement our previous analyses and key elements discussed so far in Sections 2, 3, and 4, we run ordinary least-square regressions with random effects controlling for learning effects. In particular, we regressed experimentation intensity and individual payoffs on the treatment dummy *Public*, which is 0 for the private information treatment and 1 for the public information treatment. Recall that participants played the 25 games in random order and any order of these games that was used for participants in the public information sessions was replicated for participants in the private information sessions. In order to verify that participants treated the games they successively played as independent games rather than as parts of a larger super-game, we define a weighted learning function $\{g_o\} = \{1/o\}$ where o ($o \in \{1, \cdots, 25\}$) corresponds to the random order in which each participant was exposed to each game. All regressions control for trends over time using this weighted learning function. The results do not qualitatively change when we replace the learning function with a linear version such that $\{g_o\} = \{o\}$. Further, the results do not qualitatively change either when we include controls for age, gender, field of study as well as attempts needed to correctly answer the quiz questions at the start of the experiment. To account for the fact that behavior within matching groups is not independent, we treat each matching group as our units of statistically independent observations and cluster standard errors by matching group. Table III lists the results from this analysis where Panel A shows the results for the *bad-news* environment and Panel B displays the results for the *good-news* environment.

In the *bad-news* environment, we find a strong positive effect of public information on experimentation intensity across all games, games with and without breakdowns, and payoffs. By contrast, in the *good-news* environment, we find a strong negative effect of public information on experimentation intensity across all games, games with and without breakthroughs, and payoffs.

Table III: OLS Estimations with Random Effects of Experimentation Intensity and Payoffs.

	Experimentation Intensity			*Individual*
	All	*No Breakdown or Breakthrough*	*Until Breakdown or Breakthrough*	*Payoffs*
Panel A: The *Bad-News* Environment				
Intercept	0.595^{***}	0.584^{***}	0.612^{***}	406.154
	(0.046)	(0.050)	(0.045)	(285.930)
Public	0.158^{***}	0.170^{**}	0.139^{**}	1368.876^{***}
	(0.062)	(0.069)	(0.059)	(216.667)
Learning	0.033	0.045	0.015	793.306
	(0.031)	(0.030)	(0.058)	(1545.339)
σ_ϵ	0.291	0.269	0.306	8583.367
σ_u	0.216	0.216	0.218	0
N	2600	1248	1352	2600
(Between) R-squared	0.113	0.124	0.082	0.187
Panel B: The *Good-News* Environment				
Intercept	0.764^{***}	0.662^{***}	0.798^{***}	27.643^{***}
	(0.017)	(0.025)	(0.018)	(2.154)
Public	-0.076^{***}	-0.124^{***}	-0.059^{*}	-4.286^{*}
	(0.028)	(0.034)	(0.031)	(2.291)
Learning	0.027	0.020	0.033	-4.090
	(0.037)	(0.027)	(0.056)	(10.781)
σ_ϵ	0.306	0.222	0.316	79.642
σ_u	0.135	0.162	0.134	0
N	2750	770	1980	2750
(Between) R-squared	0.063	0.107	0.037	0.043

For all estimations, robust standard errors are clustered at the session level and shown in brackets. ***Significant at the 1 percent level; **Significant at the 5 percent level; *Significant at the 10 percent level

6 Theoretical Analysis

In Sections 2.1 and 4.1, we provided intuitive explanations for our identification strategy. In this section, we elaborate and present a formal analysis.

6.1 The *Bad-News* Environment

The solution concept is that of perfect Bayesian equilibrium.[4] Sequential rationality is verified by backward induction. After a breakdown has been publicly observed, the risky arm is known to be bad, so that playing safe is the dominant action. Subsequently, we thus verify sequential rationality conditionally on no breakdown having been observed. We apply backward induction to this purpose, and normalize the cost of a breakdown to 1.[5] We write $p_i(t = \tau) \equiv p_n$ for player i's Bayesian belief in period $\tau \in \{1, 2, 3\}$, if $n = \sum_{z=1}^{\tau-1}(k_{i,z} + \hat{k}_{-i,z})$, where we write $k_{q,z} = 1$ ($k_{q,z} = 0$) if player $q \in \{i, -i\}$ has used the risky (safe) arm in period z without provoking a breakdown, and $\hat{k}_{-i,z}$ denotes the action that player i thinks that player $-i$ has taken in period z. In the case of observable actions, $\hat{k}_{-i,z} \equiv k_{-i,z}$; in the case of unobservable actions, $\hat{k}_{-i,z}$ is pinned down by player i's expectations, which are correct in equilibrium.

In the last period $t = 3$, players face a myopic decision problem, where playing risky is a best response *if and only if* $p(t = 3)\lambda \leq s$. For our parameters, $p_n\lambda < s$ *if and only if* $n \geq 2$.

Indeed, $p_2\lambda < s$ implies that, after a history of two tries without a breakdown, playing risky becomes a dominant action. This pins down play in all equilibrium candidates in which $k_{i,1} + k_{-i,1} = 2$.

Now, let us assume that $k_{i,1}+k_{-i,1} = 1$. Suppose that $k_{-i,2} = 1$ (so that $k_{i,3}+k_{-i,3} = 2$). In this case, $k_{i,2} = 1$ is a best response *if and only if*

$$s - p_1\lambda + \left(1 - p_1 + p_1(1-\lambda)^2\right)s - p_1\lambda(1-\lambda)^2 \overset{?}{\geq} \left(1 - p_1 + p_1(1-\lambda)\right)s - p_1\lambda(1-\lambda)$$

$$\iff p_1 \overset{!}{\leq} \frac{s}{\lambda}\frac{1}{1-(1-\lambda)(\lambda - s)},$$

which holds for our parameters. Thus, $k_{i,2} + k_{-i,2} = 2$ is incompatible with equilibrium, whether actions be observable or unobservable.

Next, let us assume that $k_{-i,2} = 0$. If actions are observable, $k_{i,2} = 1$ will induce $k_{i,3} + k_{-i,3} = 2$, and $k_{i,2} = 0$ will induce $k_{i,3} = k_{-i,3} = 0$. Thus, $k_{i,2} = 1$ is a best response *if and only if*

$$s - p_1\lambda + (1 - p_1\lambda)s - p_1\lambda(1-\lambda) \overset{?}{\geq} 0$$

$$\iff p_1 \overset{?}{\leq} \frac{s}{\lambda}\frac{2}{2-(\lambda - s)},$$

which is not satisfied for our parameters. Thus, $k_{i,2} = 0$ is the unique best response to $k_{-i,2} = 0$ if actions are observable, inducing $k_{i,3} = k_{-i,3} = 0$. Now suppose that actions are unobservable. We have already shown that $k_{i,2} = 1$ (inducing $k_{i,3} = k_{-i,3} = 1$) cannot happen on the equilibrium path. Now, $k_{i,2} = k_{-i,2} = k_{i,3} = k_{-i,3} = 0$ can be part of an equilibrium *if and only if*

$$0 \overset{?}{\geq} s - p_1\lambda + (1 - p_1\lambda)s - p_1\lambda(1-\lambda)$$

[4] Subgame perfection has no bite as an equilibrium refinement, because the game starts with an initial move of Nature, which determines the quality of the risky arm; the game therefore admits of no proper subgames.

[5] For our parameters, this implies that $s = 0.142857$. Recall furthermore that $\lambda = 1/4$, and $p_0 = 0.676392$.

$$\iff p_1 \geq^! \frac{s}{\lambda} \frac{2}{2 - (\lambda - s)},$$

which is satisfied for our parameters, as we have seen.

Thus, in conclusion, after a history such that $k_{i,1} + k_{-i,1} = 1$, both $k_{i,2} = k_{-i,2} = k_{i,3} = k_{-i,3} = 0$ and $k_{i,2} = k_{-i,2} = k_{i,3} = k_{-i,3} = 1$ are compatible with equilibrium, whether actions are observable or unobservable. It is this non-uniqueness of equilibrium play after histories $k_{i,1} + k_{-i,1} = 1$, which stands in contrast to our good-news game, and which will lead to different first-period equilibrium predictions, depending on whether actions are observable or unobservable, as we shall see below.

Let us turn to histories such that $k_{i,1} = k_{-i,1} = 0$. Clearly, $k_{-i,2} = k_{i,3} = k_{-i,3} = 0$, while $k_{i,2} = 1$ is not compatible with equilibrium, because $s - p_0\lambda < 0$ implies that i has an incentive to deviate to $k_{i,2} = 0$. Can $k_{i,2} = k_{-i,2} = k_{i,3} = k_{-i,3} = 1$ occur in equilibrium? If actions are observable (unobservable), a deviation by player i in $t = 2$ leads to a path of play of $k_{i,2} = k_{i,3} = k_{-i,3} = 0$, with $k_{-i,2} = 1$ ($k_{i,2} = k_{i,3} = 0$, with $k_{-i,2} = k_{-i,3} = 1$), giving the deviator i a payoff of 0 in both cases; such a deviation is therefore profitable *if and only if*

$$0 >^? s - p_0 + (1 - p_0 + p_0(1 - \lambda)^2)s - p_0(1 - \lambda)^2\lambda$$

$$\iff p_0 >^! \frac{s}{\lambda} \frac{2}{1 + s(2 - \lambda) + (1 - \lambda)^2},$$

which holds for our parameters. The only equilibrium candidate remaining is thus $k_{i,2} = k_{-i,2} = k_{i,3} = k_{-i,3} = 0$; this clearly is compatible with equilibrium as $s - p_0\lambda < 0$.

Let us now move to the first period $t = 1$, and assume that actions are observable. By our previous analysis, there are four equilibrium candidates: (1.) the utilitarian optimum, namely $k_{i,1} = k_{-i,1} = k_{i,2} = k_{-i,2} = k_{i,3} = k_{-i,3} = 1$, (2.) $k_{i,1} = k_{i,2} = k_{-i,2} = k_{i,3} = k_{-i,3} = 1$ and $k_{-i,1} = 0$, (3.) $k_{i,1} = k_{-i,1} = k_{i,2} = k_{-i,2} = k_{i,3} = k_{-i,3} = 0$, and (4.) $k_{-i,1} = k_{i,2} = k_{-i,2} = k_{i,3} = k_{-i,3} = 0$, while $k_{i,1} = 1$. Candidate (4.) can be ruled out right away, as $s - p_0\lambda < 0$, so that player i has an incentive to deviate in the first period, whether actions are observable or unobservable.

Let us turn to candidate (1.). With observable actions, a unilateral deviation by i in the first period can be "punished" with the continuation equilibrium $k_{i,2} = k_{-i,2} = k_{i,3} = k_{-i,3} = 0$, making the deviation unprofitable; indeed, in the absence of a deviation, players get the utilitarian optimum, which is strictly greater than 0, whereas, by deviating, i receives 0. For unobservable actions, however, this "punishment equilibrium" is not available, and (1.) is an equilibrium *if and only if*

$$s - p_0\lambda - 2p_0\lambda(1 - \lambda)s + p_0\lambda(1 - \lambda)(s + \lambda) + p_0\lambda(1 - \lambda)^3(\lambda - s) \geq^? 0$$

$$\iff p_0\lambda\left[1 - (-\lambda)(\lambda - s)\left(1 + (1 - \lambda)^2\right)\right] \leq^? s,$$

which is violated for our parameters. Therefore, the utilitarian optimum (1.) is an equilibrium *if and only if* actions are observable.

Next, let us analyze candidate (2.). If actions are observable (unobservable), a first-period deviation by player i who is supposed to play risky in that period leads to play of $k_{i,1} = k_{-i,1} = k_{i,2} = k_{-i,2} = k_{i,3} = k_{-i,3} = 0$ ($k_{i,1} = k_{-i,1} = k_{i,2} = k_{i,3} = 0$, with $k_{-i,2} = k_{-i,3} = 1$). In either case, this deviation is profitable *if and only if*

$$0 >^? s - p_0\lambda + (1 - p_0\lambda)s - p_0(1-\lambda)\lambda + \left(1 - p_0 + p_0(1-\lambda)^3\right)s - p_0(1-\lambda)^3\lambda,$$

which holds for our parameters. Thus, candidate (2.) is eliminated, whether actions be observable or unobservable.

Finally, let us turn to candidate (3.). If actions are observable, a first-period deviation leads to play of either $k_{i,1} = 1$, with $k_{-i,1} = k_{i,2} = k_{-i,2} = k_{i,3} = k_{-i,3} = 0$ or $k_{-i,1} = 0$ $k_{i,2} = k_{-i,2} = k_{i,3} = k_{-i,3} = 0$. In the latter case, the deviation is unprofitable as $s - p_0\lambda < 0$; in the former, it is unprofitable by the same argument as in the previous paragraph. With unobservable actions, deviating in only one period is unprofitable as $s - p_0\lambda < 0$. A deviation to $k_{i,1} = k_{i,2} = k_{i,3} = 1$ is profitable *if and only if*

$$0 <^? s - p_0\lambda + \left(1 - p_0 + p_0(1-\lambda)\right)\left[s - p_1\lambda + \left(1 - p_1 + p_1(1-\lambda)\right)s - p_1(1-\lambda)\lambda\right].$$

Yet, $s - p_0\lambda < 0$, and, as we have shown above, $s - p_1\lambda + \left(1 - p_1 + p_1(1-\lambda)\right)s - p_1(1-\lambda)\lambda < 0$. We thus conclude that candidate (3.) is an equilibrium, whether actions be observable or unobservable.

We summarize our findings in the following

Proposition 1. *If actions are unobservable in our* bad-news *game, players uniquely always play safe in perfect Bayesian equilibrium. This remains an equilibrium with observable actions. With observable actions, the utilitarian optimum (in which players play risky until a breakdown arrives) is an additional equilibrium, which is supported by the threat of always playing safe in case of a deviation.*

6.2 The *Good-News* Environment

We now turn to our *good-news* game. Sequential rationality is verified by backward induction. In the subsequent calculations, we normalize the value of a breakthrough success to 1. Clearly, playing risky is the dominant action after a breakthrough success has been observed. Thus, in the following, assume that no success has been observed yet. We shall write $p_i(t = \tau) \equiv p_n$ for player i's Bayesian belief in period $\tau \in \{1, 2, 3\}$, if $n = \sum_{z=1}^{\tau-1}(k_{i,z} + \hat{k}_{-i,z})$, where we write $k_{q,z} = 1$ ($k_{q,z} = 0$) if player $q \in \{i, -i\}$ has—unsuccessfully—used the risky (safe) arm in period z, and $\hat{k}_{-i,z}$ denotes the action that player i thinks that player $-i$ has taken in period z. In the case of observable actions, $\hat{k}_{-i,z} \equiv k_{-i,z}$; in the case of unobservable actions, $\hat{k}_{-i,z}$ is pinned down by player i's expectations, which are correct in equilibrium.

In the last period $t = 3$, players face a myopic decision problem, where playing risky is a best response *if and only if* $p(t = 3)\lambda \geq s$. For our parameters, $p_n\lambda \geq s$ *if and only if* $n \leq 1$.

Moving to the penultimate period $t = 2$, our previous step implies that, after a history in which both have unsuccessfully played risky in the first period, both players will play safe in the last period in the absence of a success, since $p(t = 3) \leq p(t =$

2) $= p_2$. Therefore, after both players have played risky in the first period, it is a best response for player i to play risky in $t = 2$ *if and only if*

$$-k_{i,2}s + p(t=2)\lambda\big[k_{i,2}(1+\lambda-s) + \hat{k}_{-i,2}(1-\lambda k_{i,2})(\lambda-s)\big] \geq 0,$$

which is equivalent to

$$p(t=2) \geq \frac{s}{\lambda}\frac{1}{1+(1-\lambda\hat{k}_{-i,2})(\lambda-s)}.$$

For our parameters,[6]

$$\frac{s}{\lambda}\frac{1}{1+(1-\lambda)(\lambda-s)} > p_2 > \frac{s}{\lambda}\frac{1}{1+\lambda-s}.$$

The first of these inequalities implies that the utilitarian optimum, which requires both players to experiment in both periods $t = 1, 2$, is not an equilibrium. Both inequalities together imply that, if $k_{i,1} + k_{-i,1} = 2$, safe and risky are mutually best responses in period $t = 2$.

It remains to analyze best responses after such histories that $k_{i,1} + k_{-i,1} < 2$. If $k_{i,1} + k_{-i,1} = 1$, the previous analysis implies that equilibrium continuation play will be one of the following: $k_{i,2} + k_{-i,2} = 2$ or $k_{i,2} + k_{-i,2} = 1$ (both followed by $k_{i,3} + k_{-i,3} = 0$ in the absence of a success). So suppose that $k_{i,2} = 1$ and $k_{-i,2} = 0$. As, for our parameters, $p_1 > \frac{s}{\lambda} > \frac{s}{\lambda}\frac{1}{1+(1-\lambda)(\lambda-s)}$, $-i$ prefers to deviate to $k_{-i,2} = 1$. Since this deviation does not affect continuation play with observable actions, it remains a profitable deviation, whether actions be observable or not. Let us thus turn to the possibility of $k_{i,2}+k_{-i,2} = 0$, which would be followed by $k_{i,3} + k_{-i,3} = 2$. Suppose first that actions are observable. In this case, either player i prefers to bring his experimentation forward in time, as $p_1 > \frac{s}{\lambda}$, i.e., one verifies that i prefers $k_{i,2} = 1$ and $k_{-i,2} = k_{i,3} + k_{-i,3} = 0$. If actions are unobservable, the same deviation leads to $k_{i,2} = 1$ leads to $k_{-i,2} = k_{i,3} = 0$, yet $k_{-i,3} = 1$. Since $-i$'s action choice in the last period does not impact i's payoff, the deviation remains profitable even if actions are unobservable. In contrast, $k_{i,2}+k_{-i,2} = 2$ (followed by $k_{i,2} + k_{-i,2} = 0$) is compatible with equilibrium as $p_1 > \frac{s}{\lambda} > \frac{s}{\lambda}\frac{1}{1+(1-\lambda)(\lambda-s)}$. Thus, in summary, after a history such that $k_{i,1} + k_{-i,1} = 1$, equilibrium uniquely calls for $k_{i,2} + k_{-i,2} = 2$, followed by $k_{i,3} + k_{-i,3} = 0$, irrespectively of whether actions are observable or unobservable.

We are now ready to move up to the initial period $t = 1$. By our preceding analysis, there are three types of candidate equilibria, depending on the experimentation intensity in the first period: (1.) $k_{i,1}+k_{-i,1} = 2$, followed by $k_{i,2}+k_{-i,2} = 1$, and $k_{i,3}+k_{-i,3} = 0$; (2.) $k_{i,1} + k_{-i,1} = 1$, followed by $k_{i,2} + k_{-i,2} = 2$, and $k_{i,3} + k_{-i,3} = 0$; (3.) $k_{i,1} + k_{-i,1} = 0$, followed by $k_{i,2}+k_{-i,2} = 2$, and $k_{i,3}+k_{-i,3} = 0$. One computes that, in candidate equilibrium (3.), a player wants to deviate to playing risky in the first period, thereby inducing play according to candidate (2.), *if and only if*

$$-s + p_0\lambda\big(1+2(\lambda-s)\big) - (1-p_0\lambda)s + p_0\lambda(1-\lambda)\big[1+\lambda-s+(1-\lambda)(\lambda-s)\big] \stackrel{?}{>}$$

[6] The normalization of the breakthrough value to 1 implies $s = 1/4$. Recall that $\lambda = 1/2$ and $p_0 = 3/4$.

$$-s + p_0\lambda\left[1 + \lambda - s + (1-\lambda)(\lambda - s)\right]$$
$$\iff -s + p_0\lambda\left[1 + (\lambda - s)(1-\lambda)^2\right] >^! 0,$$

which is verified for our parameters. Therefore, candidate (3.) cannot be an equilibrium, whether actions are observable or unobservable.

To check whether candidate (2.) is an equilibrium, assume that player i is supposed to play $k_{i,1} = 0$. As $p_0 > \frac{s}{\lambda}$, we have to deter a deviation to $k_{i,1} = 1$. Such a deviation leads to a play of $k_{i,1} = k_{-i,1} = 1$, $k_{i,2} = 1$, $k_{-i,2} = 0$, and $k_{i,3} = k_{-i,3} = 0$ (if actions are observable), or $k_{i,1} = k_{-i,1} = 1$, $k_{i,2} = 0$, $k_{-i,2} = 1$, and $k_{i,3} = k_{-i,3} = 0$ (this latter path being possible whether actions are observable or unobservable). If continuation play is given by the former option, the deviation is unprofitable *if and only if*

$$2p_0\lambda(\lambda - s) - (1 - p_0\lambda)s + p_0(1-\lambda)\lambda\left[1 + \lambda - s + (1-\lambda)(\lambda - s)\right] \geq^?$$
$$-s + p_0\lambda\left[1 + 2(\lambda - s) + 2(1-\lambda)(\lambda - s)\right] - \left(1 - p_0 + p_0(1-\lambda)^2\right)s + p_0\lambda(1-\lambda)^2(1 + \lambda - s)$$
$$\iff p_0 \leq^? \frac{s}{\lambda},$$

which is *not* the case for our parameters; thus, the deviation is profitable. Now, if continuation play is given by the latter option, the deviation is unprofitable *if and only if*

$$2p_0\lambda(\lambda - s) - (1 - p_0\lambda)s + p_0(1-\lambda)\lambda\left[1 + \lambda - s + (1-\lambda)(\lambda - s)\right] \geq^?$$
$$-s + p_0\lambda\left[1 + 2(\lambda - s) + 2(1-\lambda)(\lambda - s)\right] + p_0\lambda(1-\lambda)^2(\lambda - s)$$
$$\iff \lambda \geq^? 2,$$

which is, of course, violated, λ being a probability. We therefore conclude that candidate (2.) is not an equilibrium either, whether actions be observable or unobservable.

It remains to show that candidate (1.) is indeed an equilibrium. To do so, first let actions be observable. The latter (former) calculations from the previous paragraph show that the player who does not play risky (who plays risky) in the second period does not want to deviate. We can thus conclude that candidate (1.) is indeed an equilibrium for observable actions. Next, assume that actions are unobservable. The same argument as above shows that a deviation by the player who plays safe in the second period is unprofitable. A deviation by the player who is supposed to play risky in the second period leads to the path of play $k_{i,1} = 0$, $k_{-i,1} = 1$, $k_{i,2} = 1$, $k_{-i,2} = 0$, $k_{i,3} = k_{-i,3} = 0$; such a deviation is therefore unprofitable *if and only if*

$$-s + p_0\lambda\left[1 + 2(\lambda - s) + 2(1-\lambda)(\lambda - s)\right] - \left[1 - p_0 + p_0(1-\lambda)^2\right]s + p_0(1-\lambda)^2\lambda(1 + \lambda - s) \geq^?$$
$$2p_0\lambda(\lambda - s) - (1 - p_0\lambda)s + p_0\lambda(1-\lambda)(1 + \lambda - s)$$
$$\iff -s + p_0\lambda\left[1 + (1-\lambda)^2(\lambda - s)\right] \geq^! 0,$$

which holds for our parameters. We can thus conclude that candidate (1.) is an equilibrium for unobservable actions as well.

We summarize our findings in the following

Proposition 2. *In perfect Bayesian equilibrium in our* good-news *game, both players use the risky arm in the first period. Conditionally on no breakthrough having been observed, exactly one player plays risky in the second period, while they both play safe in the last period. There is no difference in the equilibrium prediction whether actions are observable or unobservable.*

7 Discussion and Final Thoughts

More on Economic Significance

Teams and Partnerships The economic landscape is increasingly shaped by teamwork, moving from individual efforts to collective action, as seen in research and business practices. This shift emphasizes collaboration, but it also introduces challenges like managing joint ventures and maintaining mutual trust amidst the complexities of shared responsibilities. Success depends on discerning genuine efforts from free-riding, with failure to do so potentially leading to skepticism, decreased participation, and even dissolution of the team. Extensive management literature, including works by Luo (2002) and Madhok (2006), explores these dynamics, noting especially that larger teams face greater risks of opportunistic behavior. In our setting, the team-produced good is the information, which benefits all the players.

Public Goods This paper presents a theoretical framework and an experiment illustrating how teams navigate uncertainties surrounding outcomes, and explores the impact of the observability of actions on the prevailing free-riding incentives. The framework's applicability extends to both intra-firm settings, such as research teams, and inter-firm collaborations, such as R&D joint ventures and alliances, where the public good produced is useful information. Collaborative research is widely acknowledged for its benefits, and R&D joint ventures are encouraged under both US and EU competition laws and funding programs. Nonetheless, firms considering investment in such projects must contend with the challenges associated with contributing to a public good.

***Good-News* Environment** In our *good-news* environment, we study a game of strategic experimentation in which information arrives through public *breakthroughs*. This setting mimics real-world scenarios where new but risky technologies are introduced, such as novel medical treatments, innovative manufacturing processes, or resource exploration. Understanding the trade-offs in public information production is crucial, especially in the context of innovation and social learning. Innovators often take on the initial risks and costs of experimenting with new ideas, thereby generating beneficial informational externalities for the broader community. This dynamic underscores the significant role that pioneers play in fostering progress and knowledge dissemination, highlighting the importance of their contributions to the collective understanding and advancement in various fields. In all these cases, the benefits of shared information from experimentation are evident. Examples include fishing locations being observable by others, consumers researching to find the best products, farmers choosing between traditional and genetically modified crops, and graduate students deciding on their research fields. These examples highlight how shared information influences decision-making across various contexts.

Recent Study and Gender Pay Gap Recently, Bardhi, Guo, and Strulovici (2023) investigate whether workers from social groups with similar productivity levels achieve comparable lifetime earnings, focusing on the impact of early-career discrimination. They find that in environments where failures are emphasized—that is, *breakdowns*—such discrimination leads to significant lifetime earnings gaps among equally productive groups. Conversely, in environments focusing on successes—i.e., *breakthroughs*—early discrimination tends to self-correct, ensuring comparable earnings. This outcome remains consistent across varying labor market sizes, wage flexibility, learning outcomes, productivity investments, and even with employers' misjudged beliefs. Importantly, their theoretical findings are consistent with the persisting gender pay gap among surgeons documented by Lo Sasso, Richards, Chou, and Gerber (2011) and Sarsons (2019).

Related Literature

This paper is related to several strands of literature, which we will discuss in turn.

Strategic Experimentation First, our theoretical framework can be viewed as a model of experimentation. Until fairly recently, the literature focussed on the trade-off of an individual decision maker who acts in isolation. Bolton and Harris (1999) and Keller, Rady, and Cripps (2005) have extended the individual decision problem to a multi-player framework. Since then this literature is steadily growing. For example, Klein and Rady (2011), Klein (2013), Keller and Rady (2015), and Hörner, Klein, and Rady (2022) study various bandit problems in which different players may choose different arms.[7] While free-riding is a central element in these studies as well, players' actions are observable. Several studies analyzed experimentation in teams where the outcome of each player's action is unobservable while their actions are observable (Rosenberg, Solan, and Vieille 2007, Murto and Välimäki 2011, Hopenhayn and Squintani 2011). Closest to our paper are Bonatti and Hörner (2011) and Bonatti and Hörner (2017), who study settings where actions are not observed, but outcomes are.

Free-Riding in Groups Second, our setting is related to an old literature on free-riding in groups that emerged with Olson Jr (1971) and Alchian and Demsetz (1972), and was further explored by Holmstrom (1982), Legros and Matthews (1993) and Winter (2004). Our framework relates to this literature on free-riding in that it studies the timing of free-riding in teams that are working on a project whose outcome is uncertain. For example, it can be viewed as a dynamic version of moral hazard in teams with uncertain outcome.

[7] For a boundedly rational approach to multi-armed bandits and imitation learning, see Schlag (1996, 1998, and 1999).

Dynamic Contributions to Public Goods Third, our paper ties into the literature on dynamic contributions to public goods, starting with Admati and Perry (1991), Fershtman and Nitzan (1991), Marx and Matthews (2000), Lockwood and Thomas (2002) and Compte and Jehiel (2004). Relevant to this study, especially for our *good-news* environment, is Fershtman and Nitzan (1991), who analyze equilibria in a setting with complete information and find that observability worsens free-riding. In a laboratory setting, Battaglini, Nunnari, and Palfrey (2016) test Battaglini, Nunnari, and Palfrey (2014) by investigating a game of dynamic contributions to a durable public good where the stock of the public good builds up over time. In contrast to our setting, only conventional payoff externalities—and not informational externalities—are studied.

Experimentation Experiments Lastly, our paper is embedded in an emerging literature that studies behavior in experiments of experimentation both individually and in groups. We are aware of only four other experimental investigations of a strategic-experimentation problem with bandits. Hoelzemann and Klein (2021) experimentally implement a dynamic public-good problem where information about agents' common state of the world is dynamically evolving. Observed behavior is consistent with free-riding because of strategic concerns, and participants adopt non-cut-off behavior and frequent switches of action. Boyce, Bruner, and McKee (2016) study a setting with ambiguity concerning the type of the risky arm to test strategic free-riding in a two-player, two-period, game. Players are asymmetric in their costs in that one player was known to have lower opportunity costs for playing risky than the other, so that it was clear which player ought to play the free-rider in the first period. Hudja (2019) experimentally implements Strulovici (2010)'s collective experimentation model. An individual experimentation problem is compared to a collective experimentation problem where groups of three players face a majority-vote. Hoelzemann, Manso, Nagaraj, and Tranchero (2024) study an environment where players must explore across different options with varying but uncertain payoffs. While informative signals, interpreted as data, can typically reduce uncertainty and improve welfare, in their setting it can instead decrease individual and group payoffs. When data highlights sufficiently attractive but dominated options, it can crowd-out exploration and thus lower payoffs as compared to when no data is provided. Importantly, empirical evidence from the field of genetic research provides a real-world confirmation of their framework and shows that data on genetic targets of medium promise can significantly increase the delay of valuable discoveries.

Other papers that carry out experimental tests of bandit problems consider single-agent problems, where participants act in isolation. Banks, Olson, and Porter (1997) experimentally implement a bandit setting with simple Bernoulli payout distributions, and test whether participants value information gained through experimentation. Meyer and Shi (1995) and Gans, Knox, and Croson (2007) study choice patterns that are consistent with a list of simple decision rules. Meyer and Shi (1995) test decision-making under ambiguity and use experimental data to generate hypotheses about participants' possible heuristics. Gans, Knox, and Croson (2007) consider several simple discrete-choice models in a two-armed bandit set-up. Anderson (2001, 2012) uses arms with payout distributions in his experiments and finds that participants experiment suboptimally, and are willing to pay more for receiving perfect information than theory would predict. Hudja and Woods (2022) studies individual behavior in multi-armed bandit problems and implement four bandit problems that vary based on the horizon and number of bandit arms. They find that most participants are best fit by either a simple probabilistic 'win-stay lose-shift' strategy or standard reinforcement learning. Hudja, Woods, and Gately (2023) investigates behavior in settings with forced experimentation where participants are randomly blocked from implementing a specific option. In contrast to this study, there are no strategic links across players.

Free-Riding on Information

Understanding the trade-offs in generating public information is a first-order topic, particularly as innovation and social learning often originate from pioneers. These individuals take on the initial costs of exploring new methods, thereby generating benefits for a wider audience through informational spillovers. This dynamic is evident in various domains, including R&D, resource discovery, and drug trials, where the efforts of a few can inform and benefit many. R&D, in particular, is universally acknowledged as a pivotal driver of economic expansion, as highlighted by Romer (1990) and Grossman and Helpman (1993). The productivity and innovative capacity of an economy are heavily reliant on the continuous accumulation of R&D knowledge and the broader base of existing knowledge (Griliches 1988 and Coe and Helpman 1995). In all these environments we have numerous instances where the communal sharing of information from experiments is commonplace. In this paper, we provide the first experimental examination of the impact of observable and hidden actions on strategic experimentation, outlining the conditions and the environments under which the observability of actions leads players to free-ride rather than engage in socially beneficial exploration.

References

ADMATI, A. R., AND M. PERRY (1991): "Joint Projects without Commitment," *The Review of Economic Studies*, 58(2), 259–276.

ALCHIAN, A. A., AND H. DEMSETZ (1972): "Production, Information Costs, and Economic Organization," *American Economic Review*, 62(5), 777–795.

ANDERSON, C. M. (2001): "Behavioral Models of Strategies in Multi-Armed Bandit Problems," *PhD Thesis, California Institute of Technology*.

——— (2012): "Ambiguity Aversion in Multi-Armed Bandit Problems," *Theory and Decision*, 72(1), 15–33.

BANKS, J., M. OLSON, AND D. PORTER (1997): "An Experimental Analysis of the Bandit Problem," *Economic Theory*, 10(1), 55–77.

BARDHI, A., Y. GUO, AND B. STRULOVICI (2023): "Early-Career Discrimination: Spiraling or Self-Correcting?," *Working Paper*.

BATTAGLINI, M., S. NUNNARI, AND T. R. PALFREY (2014): "Dynamic Free Riding with Irreversible Investments," *American Economic Review*, 104(9), 2858–71.

——— (2016): "The Dynamic Free Rider Problem: A Laboratory Study," *American Economic Journal: Microeconomics*, 8(4), 268–308.

BOLTON, P., AND C. HARRIS (1999): "Strategic Experimentation," *Econometrica*, 67(2), 349–374.

BONATTI, A., AND J. HÖRNER (2011): "Collaborating," *American Economic Review*, 101(2), 632–63.

——— (2017): "Learning to Disagree in a Game of Experimentation," *Journal of Economic Theory*, 169, 234–269.

BOYCE, J. R., D. M. BRUNER, AND M. MCKEE (2016): "Strategic Experimentation in the Lab," *Managerial and Decision Economics*, 37(6), 375–391.

CHEN, D. L., M. SCHONGER, AND C. WICKENS (2016): "oTree – An Open-Source Platform for Laboratory, Online, and Field Experiments," *Journal of Behavioral and Experimental Finance*, 9, 88–97.

COE, D. T., AND E. HELPMAN (1995): "International R&D Spillovers," *European Economic Review*, 39(5), 859–887.

COMPTE, O., AND P. JEHIEL (2004): "Gradualism in Bargaining and Contribution Games," *The Review of Economic Studies*, 71(4), 975–1000.

FERSHTMAN, C., AND S. NITZAN (1991): "Dynamic Voluntary Provision of Public Goods," *European Economic Review*, 35(5), 1057–1067.

Gans, N., G. Knox, and R. Croson (2007): "Simple Models of Discrete Choice and Their Performance in Bandit Experiments," *Manufacturing & Service Operations Management*, 9(4), 383–408.

Greiner, B. (2015): "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE," *Journal of the Economic Science Association*, 1(1), 114–125.

Griliches, Z. (1988): "Productivity Puzzles and R&D: Another Nonexplanation," *Journal of Economic Perspectives*, 2(4), 9–21.

Grossman, G. M., and E. Helpman (1993): *Innovation and Growth in the Global Economy*. MIT Press.

Hoelzemann, J., and N. Klein (2021): "Bandits in the Lab," *Quantitative Economics*, 12(3), 1021–1051.

Hoelzemann, J., G. Manso, A. Nagaraj, and M. Tranchero (2024): "The Streetlight Effect in Data-Driven Exploration," *Working Paper*.

Holmstrom, B. (1982): "Moral Hazard in Teams," *The Bell Journal of Economics*, pp. 324–340.

Hopenhayn, H. A., and F. Squintani (2011): "Preemption Games with Private Information," *The Review of Economic Studies*, 78(2), 667–692.

Hörner, J., N. Klein, and S. Rady (2022): "Overcoming Free-Riding in Bandit Games," *The Review of Economic Studies*, 89(4), 1948–1992.

Hudja, S. (2019): "Voting for Experimentation: A Continuous Time Analysis," *Working Paper*.

Hudja, S., and D. Woods (2022): "An Exploratory Analysis of the Multi-Armed Bandit Problem," *Working Paper*.

Hudja, S., D. Woods, and B. Gately (2023): "Forced Experimentation," *Working Paper*.

Keller, G., and S. Rady (2015): "Breakdowns," *Theoretical Economics*, 10(1), 175–202.

Keller, G., S. Rady, and M. Cripps (2005): "Strategic Experimentation with Exponential Bandits," *Econometrica*, 73(1), 39–68.

Klein, N. (2013): "Strategic Learning in Teams," *Games and Economic Behavior*, 82, 636–657.

Klein, N., and S. Rady (2011): "Negatively Correlated Bandits," *The Review of Economic Studies*, 78(2), 693–732.

Legros, P., and S. A. Matthews (1993): "Efficient and Nearly-Efficient Partnerships," *The Review of Economic Studies*, 60(3), 599–611.

Lo Sasso, A. T., M. R. Richards, C.-F. Chou, and S. E. Gerber (2011): "The $16,819 Pay Gap for Newly Trained Physicians: The Unexplained Trend of Men Earning More than Women," *Health Affairs*, 30(2), 193–201.

Lockwood, B., and J. P. Thomas (2002): "Gradualism and Irreversibility," *The Review of Economic Studies*, 69(2), 339–356.

Luo, Y. (2002): "Contract, Cooperation, and Performance in International Joint Ventures," *Strategic Management Journal*, 23(10), 903–919.

Madhok, A. (2006): "Revisiting Multinational Firms' Tolerance for Joint Ventures: A Trust-Based Approach," *Journal of international Business studies*, 37, 30–43.

Marx, L. M., and S. A. Matthews (2000): "Dynamic Voluntary Contribution to a Public Project," *The Review of Economic Studies*, 67(2), 327–358.

Meyer, R. J., and Y. Shi (1995): "Sequential Choice under Ambiguity: Intuitive Solutions to the Armed-Bandit Problem," *Management Science*, 41(5), 817–834.

Murto, P., and J. Välimäki (2011): "Learning and Information Aggregation in an Exit Game," *The Review of Economic Studies*, 78(4), 1426–1461.

Olson Jr, M. (1971): *The Logic of Collective Action: Public Goods and the Theory of Groups*, vol. 124. Harvard University Press.

Romer, P. M. (1990): "Endogenous Technological Change," *Journal of Political Economy*, 98(5, Part 2), S71–S102.

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

Sarsons, H. (2019): "Interpreting Signals in the Labor Market: Evidence from Medical Referrals," *Working Paper*.

Schlag, K. H. (1996): "'Imitate Best' vs. 'Imitate Best Average'," *Working Paper*.

——— (1998): "Why Imitate, and If so, How?: A Boundedly Rational Approach to Multi-Armed Bandits," *Journal of Economic Theory*, 78(1), 130–156.

——— (1999): "Which One Should I Imitate?," *Journal of Mathematical Economics*, 31(4), 493–522.

Strulovici, B. (2010): "Learning while Voting: Determinants of Collective eExperimentation," *Econometrica*, 78(3), 933–971.

Wagner, P. A., and N. Klein (2022): "Strategic Investment and Learning with Private Information," *Journal of Economic Theory*, 204, 105523.

Winter, E. (2004): "Incentives and Discrimination," *American Economic Review*, 94(3), 764–773.