Information and the Bandit: Breakdown Learning in the Lab*

Johannes C. Hoelzemann† Nicolas Klein‡

March 22, 2025

Abstract

We experimentally investigate a game of strategic experimentation in which information arrives through fully revealing, publicly observable, *breakdowns*. As predicted by theory, we find that players experiment significantly less, and payoffs are lower, when actions are hidden. We view this as evidence that behavior is systematically affected by the informational environment and consistent with strategic free-riding.

JEL Classification: C73, C92, D83, O32

Keywords: Dynamic Public-Good Problem, Strategic Experimentation, Exponential Bandits, Learning, Dynamic Games, Laboratory Experiments

---

*A previous version of this paper was circulated under the title "Information and the Bandit: The Good, the bad and the Ugly." 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). We thank Ben Greiner, Stanton Hudja, Sven Rady, and Karl Schlag for helpful comments.

†Universität Wien: www.johanneshoelzemann.com & jc.hoelzemann@gmail.com.

‡Université de Montréal: www.nicolasklein.com & kleinnic@yahoo.com.

1 Introduction

The question of feedback provision in organizations is an important concern in both management and economics (see, e.g., Mihm and Schlapp 2019, Halac, Kartik, and Liu 2017, as well as the references therein). In this paper, we provide experimental evidence showing that the effect of feedback depends on the nature of the news process. To isolate the informational dimension, we examine a setting with purely informational externalities, deviating from the contest framework employed by, e.g., Mihm and Schlapp (2019) and Halac, Kartik, and Liu (2017). In line with theoretical predictions, we find that private information hurts effort provision with an uncertain technology if news arrives in the form of lumpy breakdowns.

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 *breakthrough*, 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 *breakdowns*, i.e., 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 with good news, we are, to the best of our knowledge, the first to experimentally investigate a bad-news strategic-experimentation setting. As predicted behavior contrasts sharply with that in breakthrough environments, we believe our investigation to be filling an important gap in the literature.

The scant attention given to settings with breakdowns 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. Empirically, we indeed find that both experimentation and payoffs are higher with observable actions, as predicted by theory. 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. Further, participants use the risky option more frequently over time, reflecting growing optimism consistent with Bayesian updating.

In summary, the paper makes two main contributions. First, we present evidence that, in a *bad-news* setting, behavior is systematically affected by the informational environment. We find that both experimentation and payoffs are *higher* with observable actions. Second, behavior is consistent with strategic free-riding, as information is a public good and participants produce inefficiently little of it. Participants experiment, on average, too little even when the efficient solution is an equilibrium.

2 The Environment

We have endeavored to come up with the simplest possible environment in which theory would predict lower welfare with unobservable actions. 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, simultaneous-move game, calibrating the parameters in such a way that the game features the strategic effects we are interested in. 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.

Specifically, there are two risk-neutral players, and the game is played over three periods $t = 1, 2, 3$. In each period, players make a simultaneous choice. At the end of the period, outcomes are revealed; we vary whether a player's choice is observable. If the safe arm is used, the payoff will be 0 for certain in that period. Using the risky arm entails a *benefit* of $s = 2857$. 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 leads to a breakdown, imposing a cost of 20000, with a probability of $\lambda = 0.25$ in any period it is used. Conditionally on the risky arm's type, the draws are i.i.d. between players and across periods; there are thus no payoff externalities between the players, as only the player whose arm incurs the breakdown bears its cost. 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 $p_0 = 0.676$. 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 20000 from the risky arm or not.

More formally, the solution concept is that of perfect Bayesian equilibrium.[1] Sequential rationality is verified by moving backwards in time. 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 normalize the cost of a breakdown to 1.[2] We write $p_i(t) \equiv p_n$ for player i's Bayesian belief in period $t \in \{1, 2, 3\}$, if $n = \sum_{z=1}^{t-1}(k_{i,z} + \hat{k}_{-i,z})$, 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 suffering 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*

$$s - p_1\lambda + (1 - p_1 + p_1(1 - \lambda)^2)s - p_1\lambda(1 - \lambda)^2 \geq (1 - p_1 + p_1(1 - \lambda))s - p_1\lambda(1 - \lambda)$$

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

which holds for our parameters. Thus, $k_{i,2} = 1$ is a best response to $k_{-i,2} = 1$.

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

[1] 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.

[2] For our parameters, this implies that $s = 2857/20000 = 0.14285$. Recall furthermore that $\lambda = 0.25$, and $p_0 = 0.676$.

if and only if

$$s - p_1\lambda + (1 - p_1\lambda)s - p_1\lambda(1-\lambda) \geq 0$$
$$\iff p_1 \leq \frac{s}{\lambda}\frac{2}{2-(\lambda - s)},$$

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*

$$0 \geq s - p_1\lambda + (1 - p_1\lambda)s - p_1\lambda(1-\lambda)$$

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

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 $k_{i,1} + k_{-i,1} = 1$ that 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*

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

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

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*

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

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

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 $k_{-i,1} = 0$ with $k_{i,2} = k_{-i,2} = k_{i,3} = k_{-i,3} = 0$. 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, $s - p_1\lambda + \big(1 - p_1 + p_1(1-\lambda)\big)s - p_1(1-\lambda)\lambda < 0$. 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.*

Implications for Behavior Consequently, we hypothesize that action observability matters. 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. Participants earned on average approximately €10.57 from one randomly selected game and all payments were in cash and in Euros (with a conversion rate of 1000 = €1). The instructions and experimental interface are reproduced in the Online Appendix.

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.[3] 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. As illustrated in Figure 1, in the treatment with observable actions, participants could see their opponent's as well as their own past action choices and payoffs. In the treatment with unobservable actions, participants could see only if, and when, their opponent has suffered a breakdown so far as well as their own past action choices and payoffs.[4,5]

4 Findings

This section is dedicated to examining the implications for behavior as detailed in Section 2. For each of the 25 games, we implemented two treatments, with observable and unobservable actions respectively, within two-player groups, comprising 52 groups in total. 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, $\sum_{t=1}^{3} \frac{k_{1,t}+k_{2,t}}{2}$ (wh-

[3]Details are available upon request.

[4]Screenshots can be found in the Online Appendix.

[5]We chose not to elicit risk preferences due to the small stakes involved and prior studies not detecting any statistically significant effect or impact on behavior in similar bandit environments with good news (Hoelzemann and Klein 2021, Hoelzemann, Manso, Nagaraj, and Tranchero 2024).

Left: The *Observable Actions* treatment with public information; right: The *Unobservable Actions* treatment with private information.

Figure 1: Experimental Implementation

ere $k_{i,t} = 1$ ($k_{i,t} = 0$) if player i played risky (safe) in period t), 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 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 the Online Appendix. Our results show that the number and order of games previously played by participants does not significantly influence their behavior.

4.1 Experimentation and Payoffs

As outlined in Section 2, we anticipate that average experimentation intensities and group payoffs will be higher in the treatment where monitoring by others is possible. We measure experimentation intensity for each player 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.

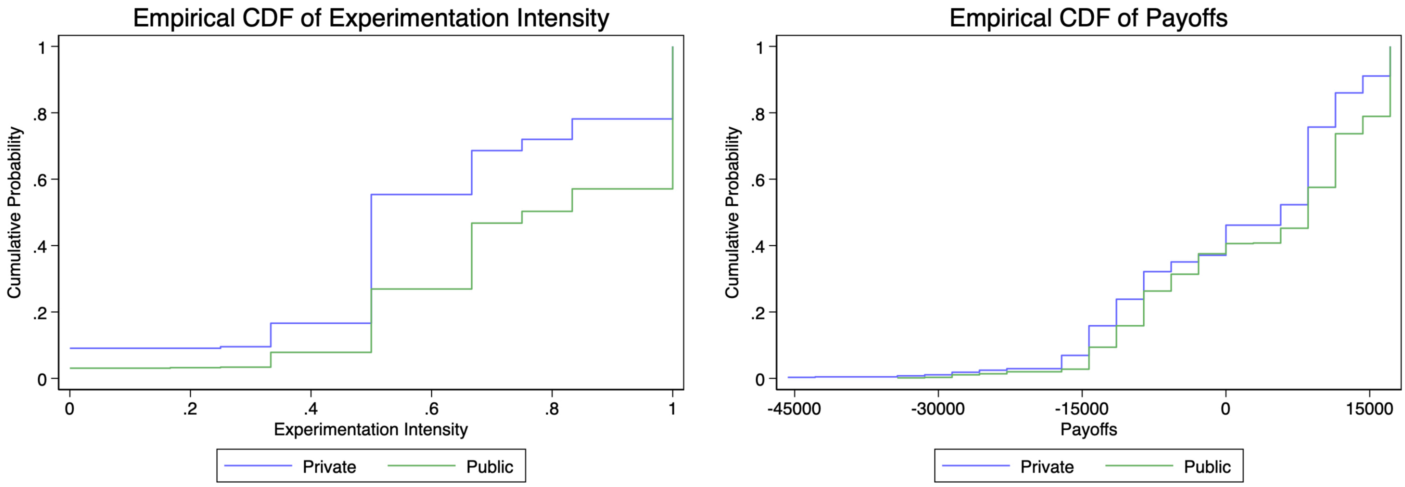
Table I: OLS Estimations

	Experimentation Intensity	Payoffs
Intercept	0.603***	1159.982***
	(0.045)	(255.138)
Public	0.156***	2813.060***
	(0.062)	(446.359)
N	1300	1300
R-squared	0.076	0.014

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 across the different treatments.



The sample cumulative distribution functions for experimentation intensity and payoffs are shown, by information condition.

Figure 2: Empirical CDFs of Experimentation Intensities and Payoffs

As can be seen in Figure 2, both experimentation intensities and group payoffs exhibit first-order stochastic dominance in the treatment with observable actions over the treatment where actions are unobservable. Kolmogorov-Smirnov tests for evaluating similarity of distributions 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.

That participants should use the risky arm more with observable actions is consistent with the (larger) equilibrium set, unless one expected for some reason that the equilibrium in which everyone plays safe all the time was the only equilibrium being played. In particular, note that there is a stark difference in first-round experimentation intensities, see Figure 4, an effect that cannot be accounted for by channels other than the strategic forces related to action observability.

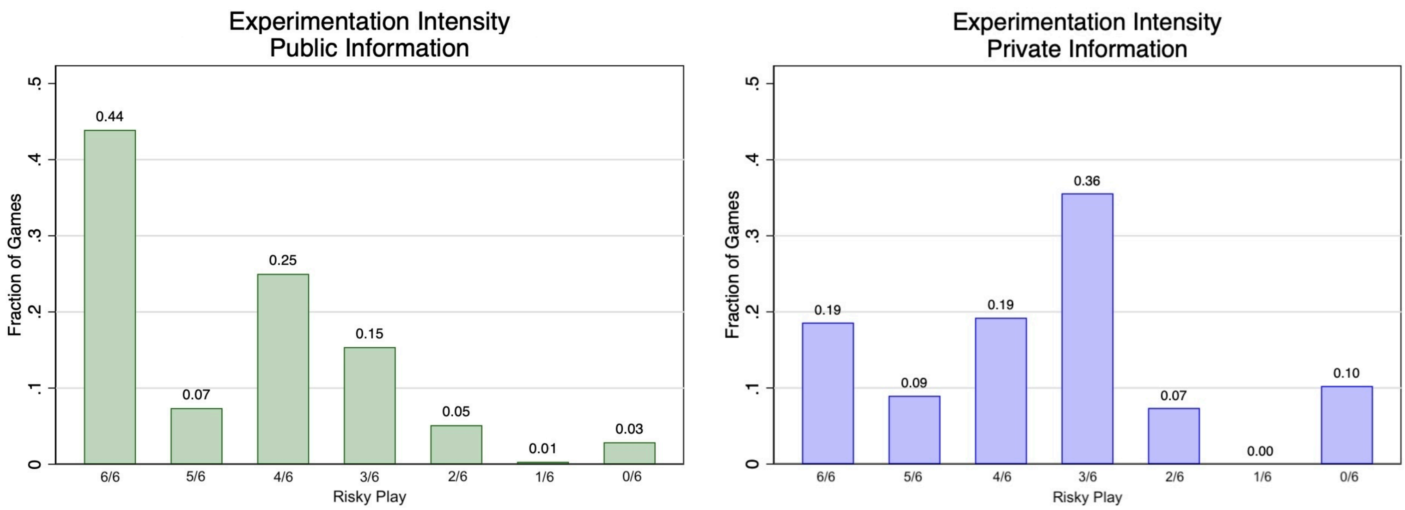
4.2 Efficiency Benchmark

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 0.253, N=312), significantly deviating from the theoretical efficient solution of 1. Out of 312 observations, 136 are in alignment with the efficient solution. Conversely, for games with unobservable actions, the average experimentation intensity with 0.592 and a standard deviation of 0.280 for N=312 is also significantly different from the efficient solution, and only 57 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.

4.3 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: Experimentation Intensities

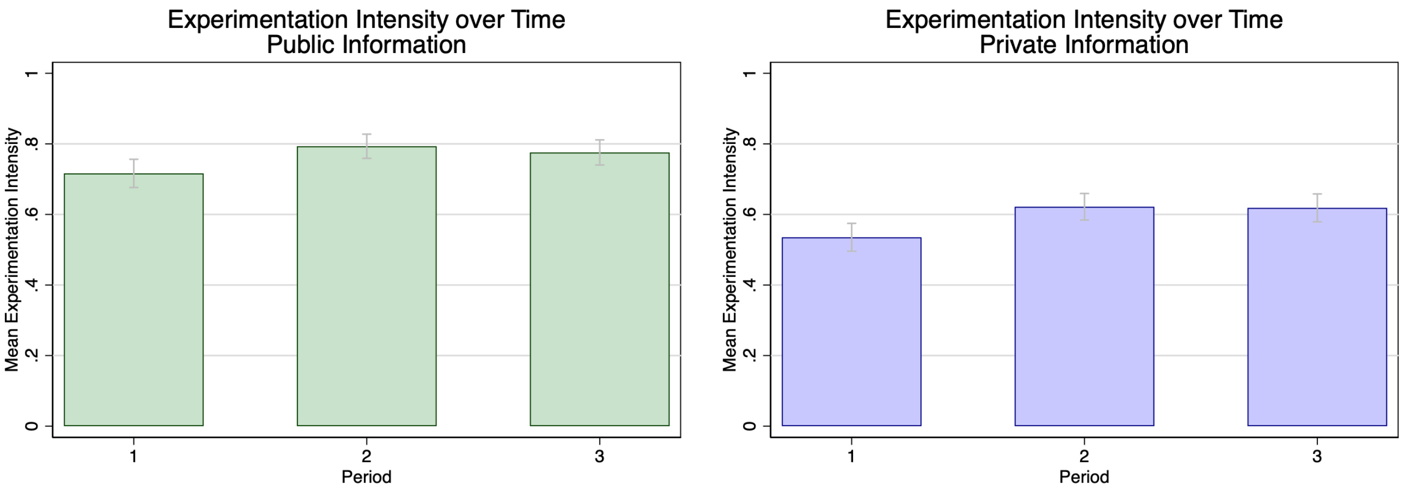
When actions are observable, 144 of 312 games are consistent with equilibrium; among these, the overwhelming majority, namely 136 games, coincide with the efficient solution. By contrast, with hidden actions, play that is consistent with the—now smaller—equilibrium set significantly decreases, with only 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.

4.4 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: Experimentation Intensities 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. In the treatment with observable actions, t-tests (two-sided Wilcoxon rank-sum tests) produce p-values of $p_{12} = 0.012$ (0.024), $p_{13} = 0.089$ (0.121), and $p_{23} = 0.411$ (0.480), 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), $p_{13} = 0.008$ (0.007), and $p_{23} = 0.643$ (0.751), respectively.

Additionally, we also analyze mean experimentation intensities across treatments for each period. 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.

5 Related Literature

Strategic Experimentation 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), and Hörner, Klein, and Rady (2022) study various bandit problems in which different players may choose different arms. 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), while

Bonatti and Hörner (2011) studies the case of observable outcomes with unobservable actions. In particular, Keller and Rady (2015) study the bad-news setting under public information, while, closest to our setting, Bonatti and Hörner (2017) study a bad-news setting where actions are not observed but outcomes are.

Strategic Experimentation Experiments in Good-News Environments Our paper is embedded in an emerging literature that studies behavior in experiments of experimentation with breakthrough-learning in groups; that is, multi-player settings with strategic links across players where decision-makers are *not* acting in isolation *and* informational externalities exist. We are aware of only four other experimental investigations of good-news strategic-experimentation problems 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. Von Essen, Huysentruyt, and Miettinen (2020) implement a treasure-hunt game in the laboratory, finding that the information externality can induce an encouragement effect. 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 genetics 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.

By contrast, this paper offers the first experimental investigation of *breakdown*-learning in a strategic setting. Observed behavior is markedly different from that documented in strategic environments with good news. Moreover, observed behavior is consistent with strategic free-riding; thus documenting that free-riding on information because of strategic concerns also exists in bad-news learning environments, while differing in its form of manifestation.

References

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.

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

Halac, M., N. Kartik, and Q. Liu (2017): "Contests for Experimentation," *Journal of Political Economy*, 125(5), 1523–1569.

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," *NBER Working Paper #32401*.

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.

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.

Mihm, J., and J. Schlapp (2019): "Sourcing Innovation: On Feedback in Contests," *Management Science*, 65(2), 559–576.

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

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

Von Essen, E., M. Huysentruyt, and T. Miettinen (2020): "Exploration in Teams and the Encouragement Effect: Theory and Experimental Evidence," *Management Science*, 66(12), 5861–5885.

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